PNGases: A diverse family of enzymes related by function rather than catalytic mechanism

Jana Filitcheva

2010

PNGases:
A diverse family of enzymes related by function rather than catalytic mechanism

A thesis presented in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

in the

Institute of Molecular BioSciences
Massey University
Palmerston North, New Zealand

Jana Filitcheva

2010

ABSTRACT

Peptide:*N*-glycanases (PNGases, EC 3.5.1.52) release *N*-linked glycan moieties from glycoproteins and glycopeptides. They catalyse the cleavage of the amide bond between the proximal *N*-acetylglucosamine and the asparagine side chain of the polypeptide, resulting in the conversion of the asparagine residue to aspartic acid and the concomitant release of the intact glycan and free ammonia. PNGases, especially PNGase F, are valuable tools for the removal of glycan moieties from glycoproteins for subsequent analyses of the released glycan and/or protein.

In the first part of this work, a classification for PNGases has been proposed, dividing these enzymes into three types based on their primary amino acid sequence, and also on their subcellular localisation, phylogenetic distribution (to date) and physiological function (if known). It appears that the three PNGase-types developed by convergent evolution. Gene expression studies for one putative type I (*Deinococcus radiodurans*) and two putative type II (*Aspergillus niger*, *Streptomyces avermitilis*) PNGases showed that these proteins were expressed in their native organisms. Recombinant expression of these proteins and the putative PNGase from *Sulfolobus solfataricus* yielded soluble protein for the *S. avermitilis* and *D. radiodurans* proteins and PNGase activity could be shown once for the latter enzyme.

In the second part of this work, site-specific mutants of PNGase F, the only characterised type I PNGase to date, were generated, expressed and characterised using enzyme kinetic methods. From the kinetic results obtained here, a catalytic mechanism can be proposed for PNGase F. In this mechanism a bound water molecule acts as the nucleophile after being activated by the abstraction of a proton by a conserved glutamate residue. The carbonyl carbon of the scissile bond is primed for the nucleophilic attack by another conserved residue, Arg248, probably by the donation of a proton.

A 1.57 Å crystal structure of the recombinant wildtype PNGase F that has three glycerol molecules non-covalently bound in the active site is also presented. This crystallographic analysis shows that the recombinant protein has a structure identical to that of the native protein, validating the basis of the kinetic studies, and showing why glycerol acts as an inhibitor of this enzyme.

ACKNOWLEDGEMENTS

Many people have greatly supported me during my time working on this project. It is impossible to mention everyone here, but I would like to thank especially the following people for their support and assistance:

My supervisor Gill Norris who has given me the opportunity to 'start over' during what was a difficult time for me. Thank you for your encouragement, advice and guidance in good times and in times of frustration and despair.

My *co*-supervisor Mark Patchett, who has always been approachable and never short of excellent advice if things were not going to plan.

All my colleagues and friends in the X-lab and the 'Neighbour-lab' past and present, especially Alice Clark, Meekyung Ahn, Jan Richter, Judith Stepper, Simon Oakley, Matthew Bennett and Greg Sawyer. A big thank-you to Trevor Loo for his assistance with all sorts of equipment and general good advice.

Bryan Anderson and Geoff Jameson for their assistance with X-ray crystallography.

Everyone in the Institute of Molecular BioSciences, who has supported me during my time here.

Finally, all this would not have been possible without the unconditional love, support and constant encouragement of my family in particular my Mum, my Grandmas, my sister Rauna and my husband Viatcheslav. You were always able to raise my spirits even in the most difficult times. Thank you so much! Большое спасибо! Vielen, vielen Dank! Diese Arbeit ist Euch gewidmet!

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

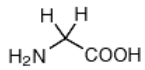
AMFR	Autocrine Motility Factor Receptor
Ani	*Aspergillus niger*
Amp	Ampicillin
BLAST	Basic Local Alignment Search Tool
bp	base pair(s)
BTP	Bis-Tris propane
BVES	Baculovirus expression system
°C	degree Celsius
CBM	Carbohydrate-binding module
Cm	Chloramphenicol
CNBr	Cyanogen bromide
CV	Column Volume(s)
Da	Dalton
DEPC	Diethyl pyrocarbonate
dn	denatured
DNA	Deoxyribonucleic acid
Dra	*Deinococcus radiodurans*
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetate
et al.	*et alteri* (and others)
ER	Endoplasmic Reticulum
ERAD	ER Associated Degradation
EtOH	Ethanol
FPLC	Fast Protein Liquid Chromatography
Fuc	Fucose
g	gram
g	g-force
Gal	Galactose
GalNAc	*N*-Acetylgalactosamine
GlcNAc	*N*-Acetylglucosamine

GOL	Glycerol
GST	Glutathione S-Transferase
h	hour(s)
h...	human...
H	α-Helix
His_6	hexahistidine-tag
HPLC	High Performance Liquid Chromatography
Hyl	Hydroxylysine
Hyp	Hydroxyproline
IMAC	Immobilised Metal Affinity Chromatography
IPTG	Isopropyl-β-D-thiogalacto-pyranoside
Kan	Kanamycin
kDa	kilodalton
L	Litre
LB	Luria-Bertani
m...	mouse...
M	Molar, Mega...
MALDI-TOF-MS	Matrix Assisted Laser Desorption/Ionization-Time Of Flight-Mass Spectrometry
MBP	Maltose Binding Protein
mg	milligram
min	minute
mL	millilitre
mM	millimolar
Man	Mannose
Mw	Molecular weight
n	native
NCBI	National Centre for Biotechnology Information
OD_{600}	Optical Density at Wavelength of 600 nanometres
OmpA	Outer Membrane Protein A
ORF	Open Reading Frame
Pa	Pascal
PAGE	Polyacrylamide-Gel-Electrophoresis

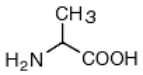
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PEG	Polyethylene glycol
pfu	plaque forming unit
PNGase	Peptide:*N*-glycanase
pI	Isoelectric Point
PUB	PNGase/Ubiquitin-associated or UBX-containing Protein Domain
rmsd	Root mean square deviation
RNase B	Ribonuclease B
rpm	Revolutions per Minute
RT	Room Temperature
***S**av*	*Streptomyces avermitilis*
SDS	Sodium-dodecylsulfate
SEC	Size Exclusion Chromatography
Sso	*Sulfolobus solfataricus*
Tc	Tetracycline
UBA	Ubiquitin-Associated Domain
UBL	Ubiquitin-Like Domain
UBX	Ubiquitin Regulatory X Domain
UV	Ultraviolet
v/v	volume per volume
w/v	weight per volume
XPCB	Xeroderma pigmentosum protein C-Binding Domain
y...	yeast...

AMINO ACIDS

Small

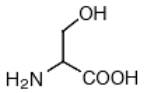


Glycine (Gly, G)
MW: 57.05

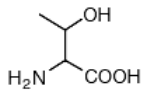


Alanine (Ala, A)
MW: 71.09

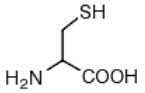
Nucleophilic



Serine (Ser, S)
MW: 87.08, pK_a ~ 16

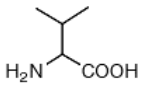


Threonine (Thr, T)
MW: 101.11, pK_a ~ 16

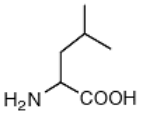


Cysteine (Cys, C)
MW: 103.15, pK_a = 8.35

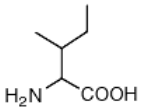
Hydrophobic



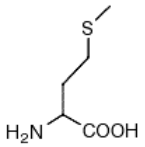
Valine (Val, V)
MW: 99.14



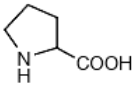
Leucine (Leu, L)
MW: 113.16



Isoleucine (Ile, I)
MW: 113.16

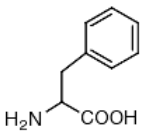


Methionine (Met, M)
MW: 131.19

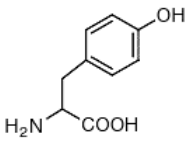


Proline (Pro, P)
MW: 97.12

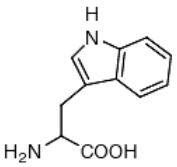
Aromatic



Phenylalanine (Phe, F)
MW: 147.18

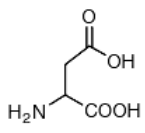


Tyrosine (Tyr, Y)
MW: 163.18

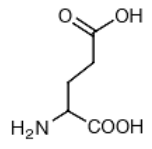


Tryptophan (Trp, W)
MW: 186.21

Acidic

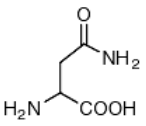


Aspartic Acid (Asp, D)
MW: 115.09, pK_a = 3.9

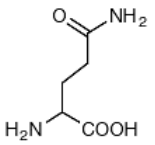


Glutamic Acid (Glu, E)
MW: 129.12, pK_a = 4.07

Amide

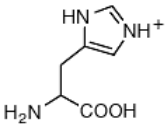


Asparagine (Asn, N)
MW: 114.11

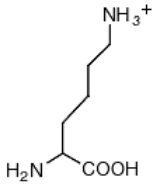


Glutamine (Gln, Q)
MW: 128.14

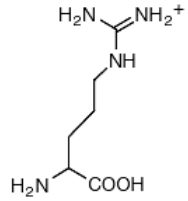
Basic



Histidine (His, H)
MW: 137.14, pK_a = 6.04



Lysine (Lys, K)
MW: 128.17, pK_a = 10.79



Arginine (Arg, R)
MW: 156.19, pK_a = 12.48

NUCLEIC ACID ABBREVIATIONS

A	Adenine
T	Thymine
C	Cytosine
G	Guanine
U	Uridine
R	G or A
Y	T or C
K	G or T
M	A or C
S	G or C
W	A or T
B	G or T or C
D	G or A or T
H	A or C or T
V	G or C or A
N	Any

STANDARD GENETIC CODE

	T			**C**			**A**			**G**		
T	TTT	F	Phe	TCT	S	Ser	TAT	Y	Tyr	TGT	C	Cys
	TTC	F	Phe	TCC	S	Ser	TAC	Y	Tyr	TGC	C	Cys
	TTA	L	Leu	TCA	S	Ser	**TAA**	*	**Stop**	**TGA**	*	**Stop**
	TTG	L	Leu	TCG	S	Ser	**TAG**	*	**Stop**	TGG	W	Trp
C	CTT	L	Leu	CCT	P	Pro	CAT	H	His	CGT	R	Arg
	CTC	L	Leu	CCC	P	Pro	CAC	H	His	CGC	R	Arg
	CTA	L	Leu	CCA	P	Pro	CAA	Q	Gln	CGA	R	Arg
	CTG	L	Leu	CCG	P	Pro	CAG	Q	Gln	CGG	R	Arg
A	ATT	I	Ile	ACT	T	Thr	AAT	N	Asn	AGT	S	Ser
	ATC	I	Ile	ACC	T	Thr	AAC	N	Asn	AGC	S	Ser
	ATA	I	Ile	ACA	T	Thr	AAA	K	Lys	AGA	R	Arg
	ATG	**M**	**Met**	ACG	T	Thr	AAG	K	Lys	AGG	R	Arg
G	GTT	V	Val	GCT	A	Ala	GAT	D	Asp	GGT	G	Gly
	GTC	V	Val	GCC	A	Ala	GAC	D	Asp	GGC	G	Gly
	GTA	V	Val	GCA	A	Ala	GAA	E	Glu	GGA	G	Gly
	GTG	V	Val	GCG	A	Ala	GAG	E	Glu	GGG	G	Gly

Chapter 1

Introduction & Literature Review

1 Introduction & Literature Review

1.1 Protein Glycosylation - Classes of Covalent Glycan-Protein-Bonds

The attachment of sugar moieties to proteins has been acknowledged as one of the most prevalent, diverse and complex co- or post-translational modifications a protein may undergo. For a long time protein glycosylation was believed to be restricted to eukaryotes, but has been described in recent years for both bacteria and archaea (Abu-Qarn *et al.*, 2008a). There are five main groups of covalent glycosidic bonds to a protein: *N*-glycosidic bonds, *O*-glycosidic bonds, *C*-mannosyl bonds, phosphoglycosyl bonds and the glycosylphosphatidylinositol (GPI) anchors (Spiro, 2002).

1.1.1 C-Mannosylation

C-Mannosylation was first described by de Beer *et al.* (1995) as a carbohydrate-protein linkage of an α-mannosyl residue to the *C*-2 of the indole ring of a tryptophan in the protein RNase U_s and subsequently in human interleukin (IL)-12 (de Beer *et al.*, 1995; Doucey *et al.*, 1999). The recognition sequon Trp-X-X-Trp has been identified in which the first tryptophan becomes mannosylated whereas the +3 tryptophan seems to play an important role in the glycosylation reaction as the transfer activity was shown to be strongly decreased (Trp→Phe) or completely abolished (Trp→Ala) after site directed mutagenesis (Doucey *et al.*, 1998; Hartmann & Hofsteenge, 2000; Krieg *et al.*, 1998). However, the *C*-mannosylated human terminal complement proteins C6, C7, C8α, C8β and C9 (Hofsteenge *et al.*, 1999), properdin (Hartmann & Hofsteenge, 2000) and thrombospondin-1 (de Peredo *et al.*, 2002; Hofsteenge *et al.*, 2001) do not possess this recognition sequon. In these proteins the thrombospondin type 1 repeats (TSR modules) having the motif (W/Y/F)XXWXX(W/C/V) contain one or more mannosylated tryptophans. A

microsomal transferase was shown to catalyse *C*-mannosylation using dolichyl-phosphate mannose as a precursor (Doucey *et al.*, 1998).

In 2008, the first *C*-mannosylated non-mammalian protein was identified in the stick insect *Carausius morosus* by Munte *et al.* (Munte *et al.*, 2008). The hypertrehalosaemic hormone Cam-HrTH-I showed a modification of residue Trp-8 by an α-mannopyranose. This protein also lacks the proposed recognition motif Trp-X-X-Trp.

The function of *C*-mannosylation, however, remains to be elucidated.

1.1.2 Phosphoglycosylation

Phosphoglycosylation defines the enzymatic attachment of a sugar to a protein through a phosphodiester bridge catalysed by phosphotransferases.

The first protein reported to contain a GlcNAc-1-PO_4-moiety linked to a serine was proteinase I, which was isolated from the slime mould *Dictyostelium discoideum* (Gustafson & Milner, 1980; Haynes, 1998). Several other proteinases, mainly cysteine proteases, have been shown to carry the GlcNAc-1-PO_4 modification (**Figure 1.1** (C)) in this organism and therefore have been grouped to form a family of such enzymes in *D. discoideum*, including cprD, cprE, cprF and cprG (Ord *et al.*, 1996; Souza *et al.*, 1995).

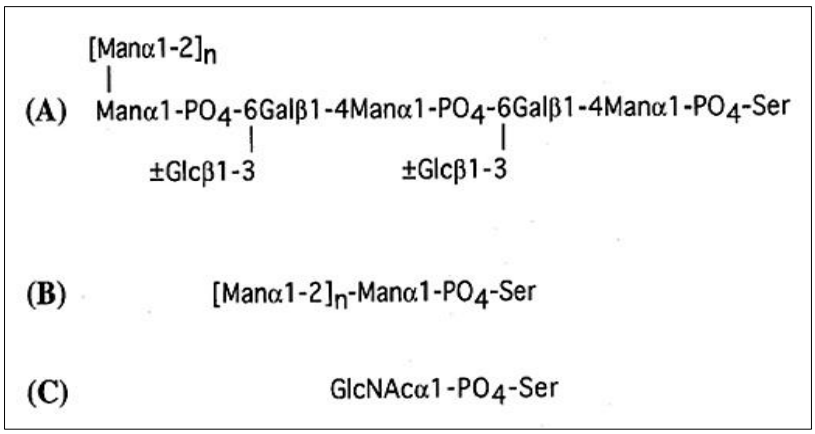


Figure 1.1: Oligosaccharide structures of three phospho-glycosylated proteins.
(A, B) Glycans from *L. mexicana* secreted acid phosphatase with n = 0–5*;* **(C)** glycan from *D. discoideum* proteinase I. Figure adapted from (Haynes, 1998)

Phosphoglycosylation was also found in several species of the protozoan parasite *Leishmania* and appears to be a major form of protein glycosylation and significant post-translational modification in these organisms (Ilg, 2000). It was identified and characterised in the secreted acid phosphatase (sAP) of *L. mexicana*, where one Man-α-1-PO_4 is linked to a serine side chain (Ilg *et al.*, 1994a). The carbohydrate moiety consists of monomeric mannose and a series of either phosphorylated and/or neutral glycans with the structures shown in **Figure 1.1 (A, B)**. Phosphoglycosylation occurs in Ser/Thr-rich repetitive domains, where the length of these repeats seems to control the phosphoglycosylation pattern (Wiese *et al.*, 1995). The fact that most phospho-glycosylated products, not only proteins, are secreted in *Leishmania ssp.* led to the speculation phosphoglycosylation may function as a secretory signal in these parasites (Haynes, 1998; Ilg *et al.*, 1994b).

Enzyme activities involved in the biosynthesis of phosphoglycoproteins were first identified in the protozoan opportunistic pathogen *Acanthamoeba castellani* and *D. discoideum* (Lang *et al.*, 1986). A UDP-GlcNAc:Ser-protein *N*-acetyl-glucosamine-1-phosphotransferase was later purified from cell membranes of *D. discoideum* (Merello *et al.*, 1995). A glycan phosphotransferase involved in synthesis of Man-α-1-PO_4-serine was later isolated from *L. mexicana* (Haynes, 1998).

As phosphoglycans in general seem to be generally absent in mammals and other vertebrates, phosphoglycan biosynthesis has been proposed as a target for the design of new drugs against Leishmaniasis, an infectious disease causing severe skin lesions, deformations of the face and other symptoms, some of which ultimately lead to death if left untreated (Ilg, 2000).

1.1.3 Glycosylphosphatidylinositol (GPI) anchoring

Glycosylphosphatidylinositol (GPI) is a complex glycophospholipid that is post-translationally attached to ~10-20% of eukaryotic membrane proteins entering the secretory pathway and serves to anchor them to the cell surface. GPI proteins are functionally diverse and include cell surface receptors, cell

adhesion molecules, cell surface hydrolases, complement regulatory proteins, the scrapie prion and protozoal coat proteins (Orlean & Menon, 2007).

The biosynthesis of GPI anchors can be regarded as being analogous to the dolichol pathway for the biosynthesis of *N*-linked glycoproteins ((Helenius & Aebi, 2002; Orlean & Menon, 2007); 1.1.5). GPI-anchors are sequentially preassembled by a series of enzymatic steps that are catalysed by enzymes located in the membrane of the endoplasmic reticulum (ER). While the first two reactions take place on the cytoplasmic side of the ER, the final steps as well as the *en-bloc*-transfer to the protein occur in the ER lumen, requiring the 'flipping' of a glycolipid intermediate through the ER membrane.

The signals within a protein's primary amino acid sequence that are required for GPI attachment are firstly a hydrophobic N-terminal signal sequence for *co*-translational translocation of the protein into the ER lumen and secondly a C-terminal GPI signal anchor sequence. This GPI signal anchor sequence typically consists of: *(i)* the ω amino acid to which the GPI will be attached, usually G, A, S, N, D, or C; *(ii)* a stretch of ~10 polar amino acids directly N-terminal to the amino acid attached (ω-10); *(iii)* the ω+2 amino acid, typically G, A, or S; *(iv)* a spacer region of moderately polar amino acids (ω+3 to at least ω+9); and *(v)* a hydrophobic sequence capable of spanning the membrane. The bond between ω and ω+1 is cleaved concomitantly with the transfer of the GPI anchor (Eisenhaber *et al.*, 1998; Orlean & Menon, 2007; Udenfriend & Kodukula, 1995).

GPI anchors are widespread in eukaryotes, they have, however, not as yet been found in bacteria although they are present in archaea (Eichler & Adams, 2005; Kobayashi *et al.*, 1997).

1.1.4 O-Glycosylation

O-Glycosylation describes the linkage formation between an amino acid side chain containing a hydroxyl group (Ser, Thr, Tyr, Hyp, Hyl) and a carbohydrate. This modification can be found in a great variety of proteins and involves a wide range of possible sugars. The GalNAc-α-Ser/Thr linkage is considered a

distinctive feature of the "mucin-type" glycoproteins, where at least nine GalNAc-transferases hierarchically catalyse the formation of the clustered Ser/Thr-linked oligosaccharides (Ten Hagen *et al.*, 2001). No conserved sequon has so far been identified for this modification, but it is generally found in clusters of Ser/Thr residues. GlcNAc-β-Ser/Thr is found in a multiplicity of eukaryotic proteins, including nuclear and cytoskeletal proteins and was the first example of a glycoprotein that is not part of the secretory pathway (Hart, 1997; Spiro, 2002). The β-linked GlcNAc residue is usually not elongated in contrast to most other peptide-linked sugars which have oligosaccharide extensions added to their core glycan.

In recent years *O*-glycosylation has been shown to occur not only in eukaryotes, but also in bacteria and archaea. It has been characterised to some extent in several bacterial pathogens including *Neisseria gonorrhoeae* and *Helicobacter pylori*, with glycosylation-defective mutants displaying reduced virulence (Abu-Qarn *et al.*, 2008a; Szymanski & Wren, 2005).

So far very little is known about the processes involved in archaeal *O*-glycosylation.

1.1.5 N-Glycosylation

In *N*-linked glycosylation a β-glycosylamine linkage between a GlcNAc and the side chain of an asparagine residue (GlcNAc-β-Asn) is formed by the *en-bloc* transfer of a preassembled oligosaccharide to a nascent polypeptide on the luminal side of the ER. First discovered in the early 1960s, it is now recognised as the most common covalent protein modification in eukaryotes. However, after its first discovery it took almost two decades for *N*-glycosylation of proteins to be observed in Archaea and even longer for the first evidence of a *N*-glycosylation machinery in Bacteria to emerge. This reflects the phylogenetic distribution of this modification, which is ubiquitous in eukaryotes, apparently wide spread amongst Archaea, but still rarely observed in Bacteria.

1.1.5.1 N-Glycosylation in Eukaryotes

N-Glycosylation of eukaryotic proteins was first observed for hen egg ovalbumin in the early 1960s (Helenius & Aebi, 2004; Johansen *et al.*, 1961). All eukaryotic cells contain, and most of them produce, *N*-glycans and have the conserved early steps of the biosynthetic process as well as some later processing reactions in common. It has been predicted that more than half of the eukaryotic proteome is glycosylated with about 90% of these glycoproteins likely to be *N*-glycosylated (Apweiler *et al.*, 1999). In 1974, Marshall postulated that the consensus sequence Asn-X-Ser/Thr was a recognition sequon for *N*-glycosylation, with 'X' being any amino acid except proline or aspartate. However, not every Asn-X-Ser/Thr sequence is necessarily glycosylated; even the same sequon in the same protein may not be always modified. The reason for this is unknown, although it has been postulated that conformational factors may play a principal role (Apweiler *et al.*, 1999). In eukaryotes *N*-glycosylation occurs in both the ER and the Golgi apparatus, with the early stages taking place in the lumen of the ER and the later maturation and differentiation of the glycan moieties being carried out by Golgi enzymes. The synthesis starts with the *co*-translational transfer of a preassembled core-oligosaccharide from a membrane-bound dolicholpyrophosphate carrier to an asparagine side chain in a nascent polypeptide. This reaction is catalysed by an oligosaccharyl-transferase, a multi-subunit protein residing in the ER membrane (Dempski & Imperiali, 2002; Silberstein & Gilmore, 1996). The core glycan is a branched oligosaccharide that is practically identical in all eukaryotes and consists of three glucose, nine mannose, and two *N*-acetylglucosamine units. For the newly folded glycoprotein to exit the ER the terminal three glucose residues and one specific terminal mannose residue are sequentially removed from the core-oligosaccharide before the transfer, a prerequisite for the properly folded glycoprotein to be released to the Golgi apparatus for further processing and maturation (Helenius & Aebi, 2001; Kornfeld & Kornfeld, 1985). Maturation involves the removal of further mannosyl-residues by special mannosidases until a trimannosyl core is formed which may then be further decorated by a range of different sugars to form the three subgroups of *N*-glycans: complex-, high-mannose- and hybrid-glycans (**Figure 1.2**). Complex glycans do not contain other mannose residues besides those in the trimannosyl core. Two of

the core α-mannosyl residues may be linked to up to five *N*-acetylglucosamine residues. Further variations arise through the presence or absence of an α-fucosyl residue at the C_3 or C_6 position of the proximal GlcNAc and the addition of a β-*N*-GlcNAc to the C_4 of the core's β-mannosyl residue (bisecting GlcNAc). In high mannose-type glycans only α-mannosyl residues are added to the core structure. Features of both, the complex type and the high mannose type are found in the hybrid glycans (Kobata, 2000).

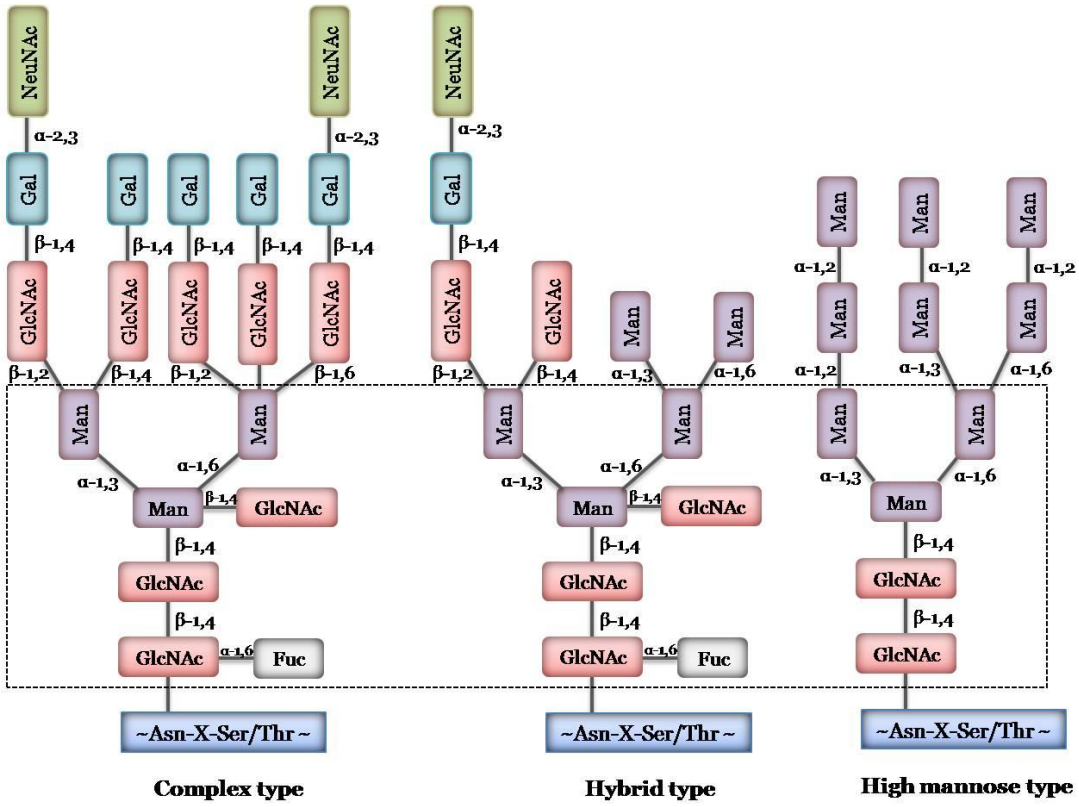


Figure 1.2: The subgroups of *N*-glycans.
Within the dotted line is the trimannosyl core structure found in all *N*-glycans. Structures outside the core are variable.

1.1.5.2 N-Glycosylation in Archaea and Bacteria

In terms of evolution, *N*-glycosylation in the ER of eukaryotes has its origin in homologous processes in the plasma membrane of prokaryotes (Bugg & Brandish, 1994; Burda & Aebi, 1999). However, the occurrence of glycoproteins in Bacteria and Archaea is a comparably recent discovery with the first non-

eukaryal *N*-glycosylated protein being found in 1976 in the envelope of the haloarchaeon *Halobacterium salinarium* (Mescher & Strominger, 1976). Since its first discovery it has been established that *N*-glycosylation in archaea is a rather frequent post-translational protein modification and much more common than in bacteria where it is considered to be a relatively rare event (Abu-Qarn *et al.*, 2008a). Given the proposed evolutionary relationship between the *N*-glycosylation mechanisms, it is not surprising to find common features in the eu- and prokaryotic processes.

In recent years several proteins involved in archaeal *N*-glycosylation have been identified mainly in the Archaea *Methanococcus voltae* and *Haloferax volcanii*, and to some extent in *Pyrococcus furiosus*. The corresponding genes, known as archaeal glycosylation (*agl*) genes, have been cloned and the proteins characterised (Abu-Qarn & Eichler, 2006; Chaban *et al.*, 2006). In *M. voltae* AglH, AglC and AglA have been identified as glycosyltransferases, each delivering one sugar residue to a dolichol carrier to form a trisaccharide found in the organism's S-layer (Chaban *et al.*, 2006; Yurist-Doutsch *et al.*, 2008). Five homologous proteins (AglD, AglE, AglF, AglG, AglI) have been described that are involved in the assembly of a pentasaccharide decorating the S-layer protein of *H. volcanii* (Abu-Qarn & Eichler, 2006; Abu-Qarn *et al.*, 2008b). Following preassembly, the lipid-linked glycans are translocated across the plasma membrane, to face the cell exterior/periplasm. In bacteria and eukaryotes the 'flippase'-protein responsible for this process has been identified, but the archaeal 'flippase' is yet to be identified and characterised (Yurist-Doutsch *et al.*, 2008). Very recently three new ORFs, *aglP*, *aglQ* and *aglR*, were located within the *agl*-gene cluster of *H. volcanii*, although their exact functions remain to be analysed (Yurist-Doutsch & Eichler, 2009). In both species, the final step, the *en-bloc* transfer of the preassembled glycan from the lipid carrier onto the protein, is catalysed by the oligosaccharyl transferase AglB (Abu-Qarn *et al.*, 2007; Chaban *et al.*, 2006). The amino acid sequon recognised in archaeal proteins is consistent with the eukaryal Asn-X-Ser/Thr motif.

The first evidence for the presence of an *N*-glycosylation system in bacteria was obtained for the Gram-negative human intestinal pathogen *Campylobacter jejuni* (Szymanski *et al.*, 1999; Young *et al.*, 2002). The genes

responsible for the biosynthesis and attachment of an *N*-linked heptasaccharide in *C. jejuni* were identified and located in the *pgl*-gene cluster consisting of 12 genes (Abu-Qarn *et al.*, 2008a; Linton *et al.*, 2005). Major evidence for the functionality of the *pgl*-gene cluster was provided by Wacker *et al.* in 2002. They transferred the *pgl*-gene cluster into *E. coli*, resulting in a modified *E. coli* strain that was indeed able to glycosylate the two *C. jejuni* proteins AcrA and PEB3 (Wacker *et al.*, 2002). Computational and functional analyses of the *pgl*-genes have identified five putative glycosyltransferases (PglA, PglC, PglH, PglI and PglJ), which are involved in the assembly of the heptasaccharide on an undecaprenyl phosphate carrier (Linton *et al.*, 2005). The gene products PglD, PglE and PglF are involved in sugar biosynthesis (Weerapana & Imperiali, 2006). The final step of the *N*-glycosylation process, the *en-bloc* transfer of the assembled glycan from the lipid carrier onto the protein, is performed by the oligosaccharyl transferase PglB (Kelly *et al.*, 2006). The sequon recognised by PglB is similar to the eukaryotic motif (Asn-X-Ser/Thr), but extended N-terminally to Asp/Glu-Z-Asn-X-Ser/Thr, where 'X' and 'Z' may be any amino acid except proline (Kowarik *et al.*, 2006b). Lectin pull down experiments led to the identification of up to 38 potentially *N*-glycosylated proteins in *C. jejuni*, which are predominantly located in the periplasm (Young *et al.*, 2002). This indicates that *N*-glycosylation in *C. jejuni* is a process specific for proteins located in the periplasm. Although it has been shown that PglB is able to glycosylate the fully folded *C. jejuni* protein AcrA *in vitro*, it is still unclear whether bacterial *N*-glycosylation is a co- or post-translational process (Kowarik *et al.*, 2006a; Weerapana & Imperiali, 2006).

1.1.5.3 Functions of N-Glycans and N-Glycoproteins

One of the most important functions of *N*-linked carbohydrates is their role in the protein folding process within the ER due to their influence on the physicochemical properties of whole domains. This explains the requirement for *N*-glycosylation to occur *co*-translationally before the folding process begins (Helenius & Aebi, 2001; Messner, 1997; Varki, 1993). Furthermore, the *N*-glycan is the ticket for entry of the newly synthesised protein into the calnexin-

calreticulin-chaperone-cycle, present in the ER of most eukaryotes (Hammond & Helenius, 1994; Nauseef *et al.*, 1995). Once properly folded, the influence of the oligosaccharide on the protein structure is usually rather limited and modification or removal of the glycan moiety has no major consequences for the overall structure, apart from influencing its physicochemical properties (i.e. stability, isoelectric point, viscosity etc.) (Imperiali & O'Connor, 1999; Olden *et al.*, 1982).

Although the influence of the oligosaccharide chains on the tertiary structure might be limited in many cases, they are often involved in the biological function of the protein and in some cases abnormalities in the *N*-glycan moiety or defects in its attachment can be lethal. For humans, some inborn glycosylation disorders have major consequences with neurological and developmental deficiencies being most common (Freeze & Westphal, 2001; Schachter, 2001; Spiro, 2002). Biological functions of the *N*-glycan moiety of glycoproteins are very varied and include transport and targeting, regulation of hormonal activity, cell-cell recognition and symbiotic communication (Varki, 1993).

In bacteria several roles for protein glycosylation have been suggested. These include maintenance of protein stability, surface recognition, resistance against proteases, cell adhesion and invasion and immune evasion (Banerjee *et al.*, 2002; Lee *et al.*, 2002; Schmidt *et al.*, 2003; Szymanski *et al.*, 2002).

In archaea, *N*-glycosylation might be involved in the organisms' ability to survive and thrive in extreme ecological niches. It has been shown that *M. volcanii* expressing no or a defective *N*-glycosylated S-layer glycoprotein has a greatly reduced ability to withstand elevated salinity (Abu-Qarn *et al.*, 2007). So far, however, it seems that *N*-glycosylation is not an essential requirement for the survival of archaea in extreme environments (Yurist-Doutsch *et al.*, 2008).

1.2 Protein Deglycosylation

As described in the previous section, the attachment of glycan moieties to proteins is essential for a wide range of biological processes and has been shown to occur in all three phylogenetic domains, Archaea, Bacteria and Eukarya. Most organisms also produce enzymes that can remove the entire carbohydrate moieties from glycoproteins, a process which is becoming increasingly recognised as being biologically significant. However, compared to the enzymes responsible for the synthesis of glycoproteins, little is known about the enzymes that catalyse the cleavage of whole, intact glycans from glycoproteins.

The term 'proximal glycanases (PROXIases)' was introduced by Suzuki *et al.* in 1994 (Suzuki *et al.*, 1994b). It describes those enzymes that are responsible for the removal of intact oligosaccharide chains from glycoconjugates (proteins or ceramides) to form a free glycan and apo-glycoconjugates (Suzuki *et al.*, 1994b). PROXIases catalyse the cleavage of the bond between the proximal monosaccharide and the core protein or the linkage between the two proximal sugar residues, and release mono- or oligo-saccharides and the apo-glycoconjugates. These enzymes have received a lot of attention for their use in glycoconjugate research, as the enzymatic removal of glycans can be performed under mild, physiological conditions leaves both cleavage products intact for further research. Chemical de-*N*-glycosylation or hydazinolysis (Takasaki *et al.*, 1982) in contrast, is unspecific, has to be carried out under much harsher reaction conditions and does not necessarily yield intact reaction products that are suitable for further analysis. Other advantages of the enzymatic deglycosylation include the absence of side reactions, which allows the oligosaccharide linkage to be identified on the basis of the substrate specificity of the PROXIase used (Maley *et al.*, 1989). Due to the interest in PROXIases as tools for structural and functional studies, almost no attention has been paid to their actual biological function in organisms (Suzuki *et al.*, 1994b).

Proximal glycanases can be divided into five subgroups:

(i) Peptide:*N*-glycanase (PNGase)
(ii) Peptide:*O*-glycanase (POGase)
(iii) Cytoplasmic β-*N*-acetylglucosaminidase (*O*-GlcNAcase)
(iv) Endo-*N*-glycanase (ENGase)

(v) Endoglycoceramidase (EGCase)

The next sections will focus on the peptide:*N*-glycanases, including a proposal for a new classification scheme based on amino acid sequence similarity and phylogenetic distribution.

1.2.1 Peptide:N-Glycanases (PNGase)

Peptide:*N*-glycanases (EC 3.5.1.52; systematic name: *N*-linked-glycopeptide-(*N*-acetyl-β-D-glucosaminyl)-L-asparagine amidohydrolase; recommended name: Peptide-N_4-(*N*-acetyl-β-D-glucosaminyl)asparagine amidase; synonyms: PNGase; Glycopeptidase, Glycoamidase, *N*-Glycanase) release *N*-linked glycan moieties from glycoproteins and glycopeptides. They catalyse the cleavage of the amide bond between the proximal *N*-acetyl-β-D-glucosamine and the asparagine side chain of the polypeptide, resulting in the conversion of the asparagine residue to aspartic acid and the concomitant release of the intact glycan and free ammonia (**Figure 1.3**). This reaction is actually a two-step reaction with amide bond hydrolysis being the first step, generating the intermediate reaction product 1-amino-*N*-acetylglucosamine and the aspartic acid-containing polypeptide. The second step is the non-enzymatic breakdown of 1-amino-*N*-acetylglucosamine into *N*-acetylglucosaminyl oligosaccharide and free ammonia (Risley & Vanetten, 1985).

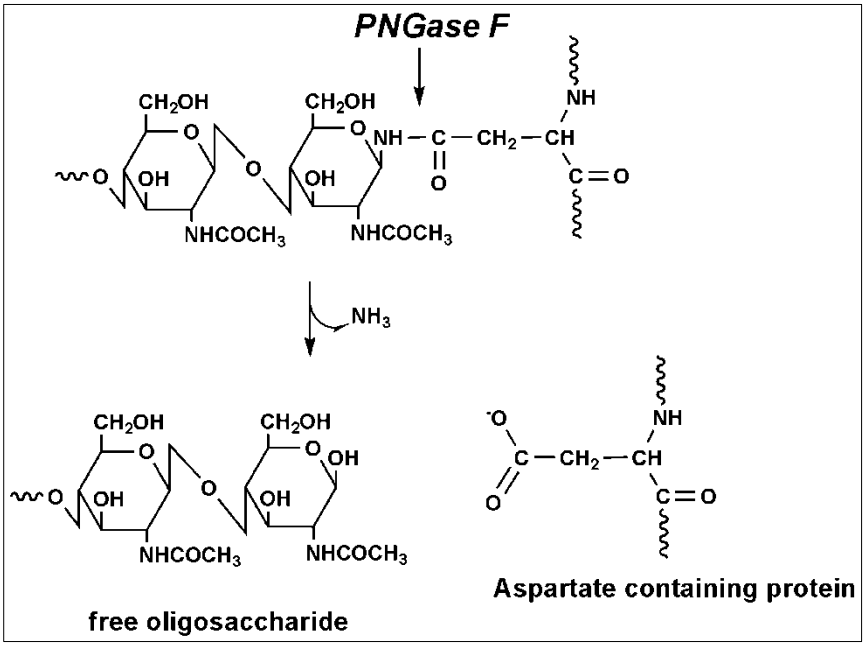


Figure 1.3: The PNGase F reaction - Cleavage of the linkage between the proximal GlcNAc and the asparagine side chain in *N*-glycoproteins.
The reaction results in the release of the oligosaccharide and the aspartic acid-containing polypeptide. Not shown is the intermediate 1-amino-*N*-acetylglucosaminyl oligosaccharide, which is spontaneously hydrolysed generating the *N*-acetylglucosaminyl oligosaccharide and ammonia.

The first PNGase described was PNGase A, found in almond emulsion by Takahashi in 1977 (Takahashi, 1977). It was partially purified from almond emulsion and shown to remove the entire *N*-glycan from stem bromelain releasing an intact oligosaccharide and a peptide lacking carbohydrate residues. This stood in contrast to the previously described endo-*N*-acetyl-β-D-glucoamidases (ENGases; (Muramats, 1971)), which cleave the linkage between the two proximal GlcNAc residues of the invariant pentasaccharide-core of *N*-linked glycans (**Figure 1.2**) producing an intact peptide with asparagine-linked GlcNAc and a glycan with one GlcNAc at the reducing end. Therefore PNGase A gave rise to the definition of a new class of amidases.

This initial discovery was followed by the identification of similar proteins in other plants and in various plant seeds (Berger *et al.*, 1995; Plummer *et al.*, 1987; Sugiyama *et al.*, 1983; Yet & Wold, 1988). A PNGase identified in 1991 in *Arabidopsis thaliana* using computational analyses is the most recent functionally characterised PNGase in plants (Diepold *et al.*, 2007; Suzuki *et al.*, 2001b). However, there is controversy regarding its actual enzymatic activity as

another research group earlier claimed it to be a *bona fide* transglutaminase, the first to be described in plants (Della Mea *et al.*, 2004).

In 1992, Lhernould *et al.* found the first fungal PNGase activity in the cultured cells from the white champignon *Silene alba*, followed by the discovery of PNGase *At* from *Aspergillus tubingensis* by Ftouhi-Paquin *et al.* in 1997 and a cytoplasmic PNGase in the budding yeast *Saccharomyces cerevisiae* by Suzuki *et al.* in 1998 (Ftouhi-Paquin *et al.*, 1997; Lhernould *et al.*, 1992; Suzuki *et al.*, 1998). In 2008, a cytoplasmic PNGase was identified and characterised in the fission yeast *Schizosaccharomyces pombe*, displaying structure and characteristics similar to its homologue in *S. cerevisiae* (Xin *et al.*, 2008).

The first, and still only, PNGase from a bacterial source was shown to be secreted from the Gram-negative soil bacterium *Flavobacterium meningosepticum* (*Chryseobacterium meningosepticum, Elizabethkingia meningoseptica*) by Plummer and colleagues in 1984 (Plummer *et al.*, 1984). This enzyme, designated PNGase F, is one of best characterised PNGases and is used extensively as a tool for the structural and functional analysis of *N*-glycans.

The discovery of peptide:*N*-glycanases in animals started with the observation of free oligosaccharide-accumulations in early embryos of Medaka fish (*Oryzias latipes*) and the subsequent identification of the PNGase being responsible for their generation. This PNGase showed its highest activity at acidic pH and therefore was thought to be located in the lysosome (Seko *et al.*, 1991). Although similar activity was not found in mammalian lysosome it led to the discovery of PNGase activity in the cytoplasm of various mammalian cells and tissues (Seko *et al.*, 1991; Suzuki *et al.*, 1993a; Suzuki *et al.*, 1993b; Suzuki *et al.*, 1994a; Suzuki *et al.*, 1994b; Suzuki *et al.*, 1994c). In contrast to the Medaka fish PNGase, all subsequently identified PNGases were found to be located in the cytoplasm and consequently were found to be most active at neutral pH. In fact, cytoplasmic PNGase is ubiquitously present in mammalian cells, indicating its involvement in essential processes. Only a few of these enzymes have, however, been purified and characterised in detail. One of the most recent functionally characterised cytoplasmic PNGases was found in 2007 in the nematode *Caenorhabditis elegans*. Interestingly, besides its deglycosylating

activity this PNGase was found to also function as a thioredoxin (Kato *et al.*, 2007; Suzuki *et al.*, 2007). An overview of some known PNGases and their sources is given in **Table 1.1**.

Table 1.1: Distribution of peptide:*N*-glycanases among the phylogenetic domains.

Domain	Source	Name	Reference
Bacteria	*Flavobacterium meningosepticum*	PNGase F	(Plummer *et al.*, 1984)
Fungi	*Silene alba* (White champignon)	PNGase Se	(Lhernould *et al.*, 1992)
	Aspergillus tubingensis	PNGase *At*	(Ftouhi-Paquin *et al.*, 1997)
	Saccharomyces cerevisiae	yPng1p	(Suzuki *et al.*, 1998)
	Schizosaccharomyces pombe	SpPNGase	(Xin *et al.*, 2008)
Plants	*Prunus amygdalus* (sweet almond)	PNGase A	(Takahashi, 1977)
	Canavalia ensiformis (Jack bean)	PNGase J	(Sugiyama *et al.*, 1983)
	Pisum sativum (Split pea)	PNGase P	(Plummer *et al.*, 1987)
	Raphanus sativus (Radish)	PNGase R	(Berger *et al.*, 1995)
	Glycine max (Soybean)	PNGase GM	(Kimura & Ohno, 1998)
	Oryza sativa (Rice)	PNGase Os	(Chang *et al.*, 2000)
	Arabidopsis thaliana	*At*PNG1	(Diepold *et al.*, 2007)
Animals	*Oryzias latipes* (Medaka fish)	PNGase M	(Seko *et al.*, 1991)
	Mouse (L-929 fibroblasts)	PNGase L-929	(Suzuki *et al.*, 1994c)
	various mammalian cell cultures	--	(Suzuki *et al.*, 1993b)
	various mouse organs	mPNGase	(Kitajima *et al.*, 1995)
	Hen oviduct	PNGase HO	(Suzuki *et al.*, 1997)
	Caenorhabditis elegans	CePNG-1	(Kato *et al.*, 2007; Suzuki *et al.*, 2007)

This short outline shows that there is an increasing number of PNGases, found in a wide range of species. Amazingly, they all were shown to catalyse the same reaction, the hydrolysis of the β-aspartylglucosaminylamine bond between the polypeptide and the attached glycan, but they are in fact quite different in several aspects, including their nucleotide and primary amino acid sequence, quaternary structure, localisation, molecular weight, pH optimum, substrate specificity and biological function. Due to these differences, Ftouhi-Paquin *et al.* (1997) predicted that PNGase F and PNGase *At* had developed along different evolutionary lines (Ftouhi-Paquin *et al.*, 1997). However, there is no clear

classification that acknowledges these differences and gives a more ordered view of this growing enzyme class. For that reason we have proposed a classification that separates the PNGases into three types based on comparison of their primary amino acid sequences. This classification, already provided here in order to bring some order to the otherwise disorganised class of PNGases, highlights the possibilities of convergent evolution and potential horizontal gene transfer, explaining the current diversity and distribution of PNGases observed (**Table 1.2**).

Table 1.2: Proposed classification of peptide:*N*-glycanases.

Type	Main characteristics	Enzyme (e.g.)	Source
I	secreted; bacterial or of bacterial origin	PNGase F	*F. meningosepticum*
II	secreted/exoplasmic; archaea, bacteria, fungi, plants	PNGase A PNGase *At*	*P. amygdalus* (sweet almond) *A. tubingensis*
III	cytoplasmic, proteasome-associated; ubiquitous in eukaryotes; not found in bacteria or archaea	yPng1p mPNGase hPng1p	*S. cerevisiae* *Mus musculus* *Homo sapiens*

Convergent evolution of enzymes describes non-homologous enzymes evolving in different organisms or biological niches to catalyse the same or at least very similar enzymatic reactions (Gherardini *et al.*, 2007). In convergent enzyme evolution one can differentiate between two situations. The first situation describes non-homologous enzymes that catalyse the same reaction using the same or a very similar mechanism, dictated by similar active site residues and geometry. Such enzymes have been named mechanistic analogues. An example of an active site-conformation that features in non-homologous enzymes is the Ser-His-Asp triad in serine proteases and the structurally different enzyme subtilisin (Gherardini *et al.*, 2007; Kraut, 1977; Matthews *et al.*, 1977). The other kind of convergent evolution leads to transformational analogues. These enzymes do not share a common structure in any way. Following this terminology, PNGases of the three types could be described as transformational analogues, as they are not homologues and, where known, do not employ a comparable mechanism to catalyse the same overall reaction.

Gherardini *et al.* (2007) describe three patterns to explain the phylogenetic distribution of transformational analogues, where two or more unrelated enzymes can either: (*i*) be uniformly distributed in different kingdoms, (*ii*) be very distinctly distributed with each form being present in a different kingdom with little or no overlap or (*iii*) be unevenly distributed, with one enzyme appearing almost everywhere and another that occupies only a small niche (Gherardini *et al.*, 2007).

Generally, and for PNGases in particular, it is intriguing how nature has found different solutions to the same problem and how apparently totally different enzymes developed to catalyse the same overall chemical reaction in very different ways.

1.2.1.1 PNGase F: The Only Example of a Bacterial PNGase

PNGase F is one of the best characterised PNGases and was for more than ten years the only PNGase for which a three-dimensional high-resolution-structure was available (Kuhn *et al.*, 1994; Norris *et al.*, 1994b). Because PNGase F was and still is a highly valued tool for studying the structure and function of *N*-linked glycoproteins, the gene has been cloned and heterologously expressed in *E. coli* by several research groups (Barsomian *et al.*, 1990; Lemp *et al.*, 1990; Loo *et al.*, 2002; Tarentino *et al.*, 1990).

PNGase F activity was observed first in preparations of another oligosaccharide chain-cleaving enzyme secreted by *F. meningosepticum*, endo-β-*N*-acetylglucosaminidase F (Endo F; (Elder & Alexander, 1982; Plummer *et al.*, 1984)), as well as in commercially available preparations of Endo F. Endo F cleaves the oligosaccharide chain between the two proximal GlcNAcs of the diacetylchitobiose moiety of high-mannose asparagine-linked glycans, but it does not cleave complex type *N*-glycans. Very weak deglycosylation activity was shown for complex biantennary oligosaccharides only at high enzyme concentrations (Tarentino *et al.*, 1985). PNGase F is most active between pH 7.5 and 9.5 with an optimum at pH 8.5. The mature enzyme consists of 314 amino

acids and its molecular weight was determined to be 34.8 kDa (Tarentino *et al.*, 1990).

PNGase F requires the α-amino and carboxyl groups of the asparagine residue to be in a peptide linkage. It was also demonstrated that it actually does act on both native and denatured glycoproteins although far higher enzyme concentrations are required for deglycosylation of native substrates (Tarentino *et al.*, 1985). Initially described as an "all-purpose enzyme to hydrolyse high-mannose, hybrid, and bi-, tri and tetra-antennary oligosaccharides" by Tarentino *et al.* (1985), it was later shown that modifications of the proximal GlcNAc residue greatly impaired PNGase F activity (Tretter *et al.*, 1991). PNGase F was not able to deglycosylate pineapple bromelain glycopeptide and horseradish peroxidase-C glycoprotein, which contain xylose linked β1-2 to β-mannose and fucose linked α1-3 to the proximal GlcNAc. After removal of the α1-3 fucose residue, PNGase F removed the glycan moiety from these substrates. However, an α1-6 fucose substituent did not block PNGase F activity. As indicated earlier by the wide range of substrates suitable for this enzyme, it was shown that the glycan structure outside the asparagine-linked dichitobiose core had no impact on its activity. As expected, the modification of the outer oligosaccharide structure by exoglycosidase treatment was shown to have minimal effect on PNGase F activity (Altmann *et al.*, 1995). The most extensive studies on substrate structure requirements were performed in 1997 by Fan & Lee (Fan & Lee, 1997). They synthesised 31 glycopeptides with different types and lengths of the carbohydrate, different lengths and sequences of the peptide, and different glycosylated amino acids. They showed that PNGase F cannot hydrolyse cellobiose and lactose glycopeptides, indicating the importance of the 2-acetoamide group of GlcNAc in naturally occurring substrates. Consistent with previous findings, they demonstrated the inability of PNGase F to act on carbohydrates linked to a single asparagine residue, and established that the minimum peptide chain length requirement was a tripeptide with the asparagine residue being the central residue. Glycopeptides containing only one GlcNAc residue were hydrolysed, albeit very slowly, by PNGase F, a finding contrary to an earlier study in which single GlcNAc hydrolysis was not observed (Chu, 1986). Furthermore, it was observed that the strict consensus sequence

for *N*-glycosylation, Asn-X-Ser/Thr, is not mandatory for PNGase F activity (Fan & Lee, 1997).

The three-dimensional structure of PNGase F was determined in 1994 independently by two research groups (Kuhn *et al.*, 1994; Norris *et al.*, 1994b) using the protein from two different *F. meningosepticum* strains (ATCC 33958 and CDC strain 3352). These were the first crystal structures obtained for a PNGase and although the crystallisation conditions were different, both groups obtained essentially identical structures. The enzyme consists of two tightly associated all-β-domains with the amino-terminal domain reaching from residues 1 to 135 and a carboxy-terminal domain comprising residues 142-314. Both domains have the same eight-stranded ('4+4') antiparallel β-jelly roll fold, where eight antiparallel β-strands are arranged as a sandwich of two four-stranded β-sheets (**Figure 1.4**).

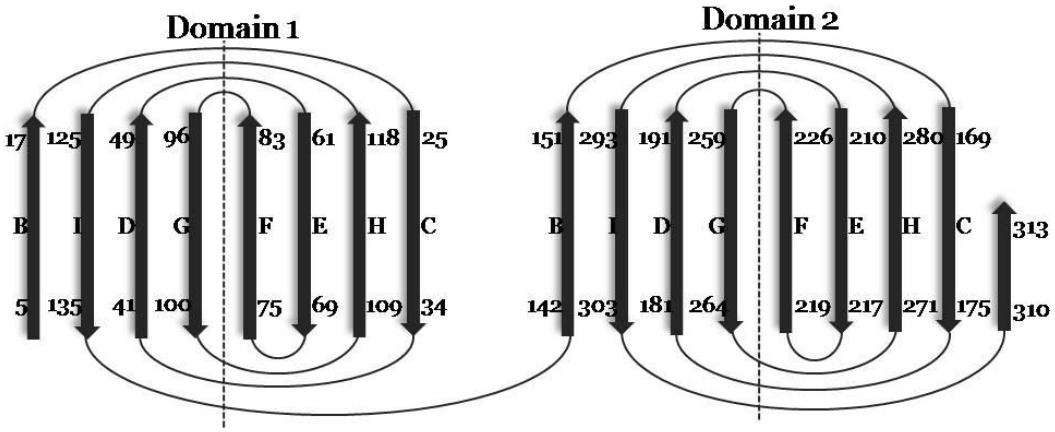


Figure 1.4: Topology of PNGase F.
The β-strands are named using the convention adopted for viral coat proteins (Rossmann *et al.*, 1983). Numbers identify the residues comprised in each strand.

This topology is often found in viral coat and capsid proteins (Rossmann *et al.*, 1983) as well as in plant and animal lectins and other carbohydrate-binding proteins. In fact, the fold most commonly found in the non-catalytic carbohydrate-binding modules (CBMs) of carbohydrate-active enzymes is the β-sandwich and amongst those the β-jelly roll fold is the most prevalent (Boraston *et al.*, 1999; Boraston *et al.*, 2004; Hashimoto, 2006). CBMs are defined as a contiguous amino acid sequence within a carbohydrate-active enzyme with a discrete fold possessing carbohydrate-binding activity (Boraston *et al.*, 1999). As the overall folds of these modules are similar, the substrate specificity usually is

determined by the location of aromatic amino acid side chains and loop structures. These two factors shape the binding site to mirror the ligands' conformations (Boraston *et al.*, 2004).

Amino acids 136-141 are located at the 'bottom' of the molecule where they form a link between the two domains. At the 'top' of the molecule, two loops from domain 2 reach across to form non-covalent interactions with domain 1, tying the two domains together. The most important of these loops (residues 227-257) links strands F and G of domain 2 and forms a double loop in which residues 227-249 form the first part. This loop extends to domain 1, then returns to domain 2 to form a wide Ω-loop between residues 231-245. The second part of the double loop is formed by residues 250-257 and is connected to the first part by a disulfide bridge formed between residues 231 and 252. The loop between residues 151-169, which links strands B and C of domain 2, also reaches across to domain 1. These loops play an important role in interdomain interactions and in forming the active site (Norris *et al.*, 1994b).

The active site and some residues essential for PNGase F activity were identified in 1995 by Kuhn *et al.* using site-directed mutagenesis and crystallographic analysis (**Table 1.3**; (Kuhn *et al.*, 1995)). Fifteen site-specific mutants were generated in different areas of the molecule and tested for catalytic activity. Enzyme activity was lost entirely in the mutant D60N and almost but not quite completely abolished in the two mutants E206Q (0.01% of wild-type activity) and E118Q (0.1% of wild-type activity). This analysis indicated that these three acidic residues, located in a cleft at the interface between the two domains at the top of the protein, are essential for activity. This cleft is formed by long loops that connect the β-strands between the individual β-sheets and is lined by the residues Trp59, Trp86, Trp120, Trp191, His193 and Trp207. The catalytically essential residues, Asp60, Glu206 and Glu118, are located at the bottom of this cleft. To confirm that the loss of enzyme activity was a result of diminished catalytic function or impaired substrate binding, and not the result of a conformational change to the protein structure, these mutant proteins were overexpressed and crystallised. The authors claim that the mutants' structures were basically identical to that of the wild-type enzyme. However, the corresponding crystallisation data have not been published nor deposited in the protein data bank (PDB). In another crystallographic approach,

aimed to determine the substrate binding site, the wildtype protein was crystallised in complex with the disaccharide *N*,*N*'-diacetylchitobiose, which had been found to partially inhibit the activity of PNGase F. The structure of the complex showed that the O_1 of the first GlcNAc residue formed a hydrogen bond with Asp60 and a water molecule (Wat346) that connected Asp60 and Glu206. The O_6 of the second GlcNAc formed a hydrogen bond with Glu118. Glu206 was shown not to be in direct contact with the substrate and was therefore predicted to play a stabilising role during catalysis (Kuhn *et al.*, 1995). **Figure 1.5** shows a schematic diagram of the intermolecular hydrogen bonding network between PNGase F and the bound diacetylchitobiose, demonstrating extensive contacts between the protein and the first GlcNAc. Contacts between the enzyme and the second GlcNAc residue were shown to be much weaker, involving only O_6 and O_7.

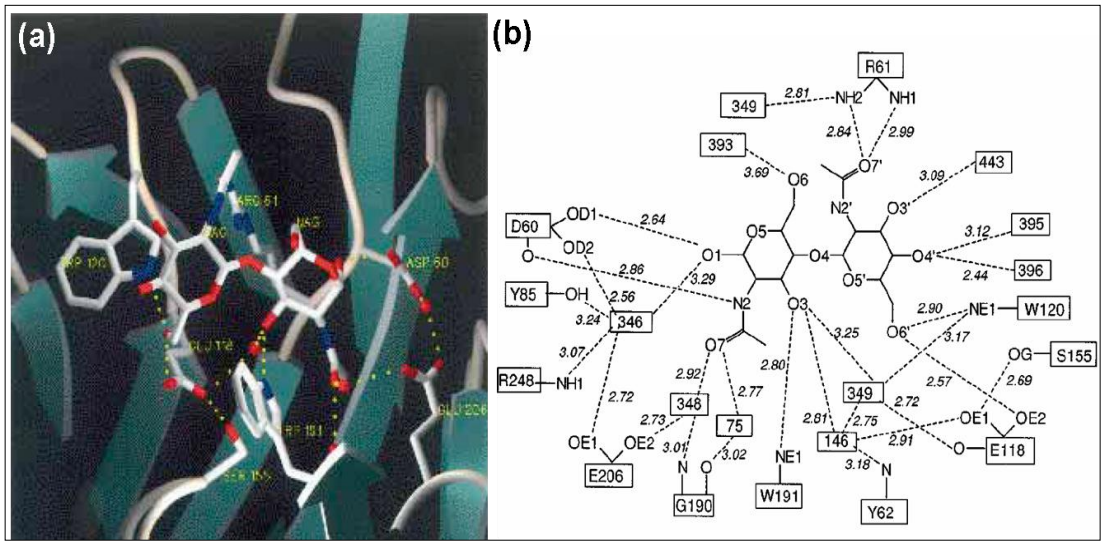


Figure 1.5: (a) Detailed and (b) schematic image of interactions of *N*-*N*'-diacetylchitobiose with PNGase F.
Hydrogen bonds between the disaccharide, water molecules and protein are indicated by dotted lines (adapted from (Kuhn *et al.*, 1995)). The proximal GlcNAc residue is on the left.

Table 1.3 provides a summary of site-specific mutants that have been generated and analysed. It also shows mutants analysed by Loo *et al.* (personal communication Dr. G.E. Norris) in addition to mutants published by Kuhn *et al.* (1995).

Table 1.3: Effects of site directed mutagenesis on PNGase F activity.
Amino acid residues proposed to be important for PNGase F activity following site directed mutagenesis experiments and their predicted functions. Results shown were obtained in two independent studies using different methods.

Amino acid residue	Mutant(s)	Relative activity [%]	Proposed function of residue
Asp60	D60N[1]	Not detectable	Catalytic mechanism
	D60E[1]	0.1	
Tyr85	Y85F[1]	> 10	?
Glu118	E118Q[1]	0.1	Substrate binding / recognition
Trp120	W120V[2]	1.8	Substrate binding / recognition
His193	H193A[2]	3	Substrate binding / recognition
Glu206	E206Q[1]	< 0.01	Stabilising
	E206D[1]	0.01	
Arg248	R248A[2]	0.1	Catalytic mechanism

Another observation from this study was the explanation for the inability of PNGase F to act on substrates with a α-1,3-fucose substitution on the proximal GlcNAc in contrast to 1,6-substitution. The orientation of the disaccharide suggests that any substituent on O_3 would not be able to fit into the space provided, whereas O_6 is fully exposed to the solvent.

The physiological function of this enzyme in the organism has not yet been investigated. As it is a secreted protein, it might serve nutritional purposes, i.e. it might deglycosylate foreign *N*-glycoproteins/-peptides available in the natural habitat, in order to make the protein/peptide more susceptible to proteolytic degradation.

Until recently, PNGase F was the only member of type I PNGases (**Table 1.2**) as there were no DNA or amino acid sequences homologous to the PNGase F sequences in public databases. All other PNGases have absolutely no sequence similarity to PNGase F. They do, however, catalyse a similar reaction using similar substrates. Last year, two sequences were published that showed amino acid similarity to PNGase F. The first sequence was found in the genome sequence of the bacterium *Deinococcus radiodurans* P1 and the second sequence was published as part of the genome project for *Danio rerio* (zebrafish) (White *et al.*, 1999). In case of the latter organism, it might be

[1] Kuhn *et al.* (1995)
[2] Loo *et al.* unpublished data

justifiable to say that the PNGase-gene was obtained *via* horizontal gene transfer from bacteria, as it is the only type I sequence so far identified in eukaryotes and *Flavobacterium* species are known zebrafish pathogens. It is clear however, that the occurrence of this type of PNGase is restricted to bacteria or is at least of bacterial origin, and therefore PNGase F-like proteins are thought to be the oldest type of existing PNGase in phylogenetic terms.

1.2.1.2 PNGases A and At: Examples of Type II PNGases

As mentioned above, PNGase A from *P. amygdalus* (sweet almond) was the first PNGase to be described in 1977 by Takahashi (Takahashi, 1977).

First studies described PNGase A as a 66.8 Da protein, which was found to be glycosylated, as it bound to ConA resin, and GlcNAc, mannose and fucose were identified as constituents of pure protein preparations. Circular dichroism spectra indicated the presence of approximately 80% α-helix content (Taga *et al.*, 1984). Plummer *et al.* estimated the molecular weight of PNGase A to be considerably higher at 79.5 kDa by HPLC and showed that it had a carbohydrate content of 27% with significant amounts of glucosamine, mannose, galactose, fucose, arabinose, xylose and glucose. The pH-optimum for the enzyme was found to be 4.5 (Plummer *et al.*, 1987). It was later discovered that PNGase A is in fact a heterodimer consisting of a 54.2 kDa subunit and a smaller 21.2 kDa subunit as determined at first by SDS-PAGE and then more accurately by MALDI-TOF-MS (Altmann *et al.*, 1998). The same study showed that PNGase A was itself an *N*-glycoprotein with 9 (±1) mol *N*-glycan/mol of protein and that these glycans were distributed over both subunits (Altmann *et al.*, 1998).

PNGase A was found to act on all three types of *N*-glycans: high-mannose, complex and hybrid. Different results were published concerning its preference for certain glycan types and its ability to cleave glycans from intact glycoproteins. While Plummer *et al.* (Plummer & Tarentino, 1981) reported a preference for complex type glycopeptides, Altmann *et al.* (Altmann *et al.*, 1995) could not detect differences between these glycan types. Tarentino & Plummer (Tarentino & Plummer, 1982) and Taga *et al.* (Taga *et al.*, 1984) found

PNGase A able to act on denatured glycoproteins, whereas Altmann *et al.* (1995) could not detect any activity for glycoproteins, even if they had been denatured. In contrast to PNGase F, PNGase A was shown to act on substrates containing an α-1,3-fucose residue on the proximal GlcNAc, a characteristic of glycoproteins from plants and insects (Fan & Lee, 1997; Faye *et al.*, 1989; Kubelka *et al.*, 1994). This PNGase is also able to cleave glycopeptides containing only one GlcNAc provided that it was covalently linked to a peptide larger than a tripeptide, although the hydrolytic rates were slower than for the corresponding diacetylchitobiose-containing peptides. The minimum peptide length was later shown to be most likely a dipeptide (Fan & Lee, 1997).

A second example of this type of PNGase was later isolated from the fungus *Aspergillus tubingensis*. PNGase *At* was discovered by Ftouhi-Paquin *et al.* in 1997 (Ftouhi-Paquin *et al.*, 1997) in a concentrated commercial extract of secretory enzymes derived from *A. tubingensis*. The PNGase *At* gene was cloned, sequenced and the amino acid sequence deduced. Comparison of the deduced amino acid sequence with the result of the Edman analysis of the N-terminus of the native mature protein confirmed the presence of a hydrophobic 21 amino acid signal sequence typical of secreted proteins. The mature protein comprises 537 amino acids with a predicted molecular weight of 59.3 kDa. Addition of seven to nine high-mannose glycans was shown to increase the molecular weight by 9 to 11 kDa, leading to an overall mass of approximately 70 kDa. Although the DNA sequence clearly showed that PNGase *At* was translated as a single polypeptide chain, the mature protein appeared as heterogeneous bands in SDS-PAGE with a molecular weight of ~ 43 kDa, suggesting that PNGase *At*, like PNGase A, consists of two glycosylated subunits. Analysis of the deglycosylated protein identified two distinct subunits, with an α-subunit of 38 kDa and a β-subunit of approximately 28 kDa, proving the heterodimeric nature of the native PNGase *At*. Cleavage of the pre-protein was shown to occur in a Ser/Thr-rich hydrophilic region of the protein between residues Thr-335 and Thr-336 (Ftouhi-Paquin *et al.*, 1997). Initially the authors suspected that a self-cleavage mechanism led to the subunit formation. Later it became obvious that an *A. tubingensis* protease must be responsible for processing of the primary translation product in the native organism, as recombinant expression of PNGase *At* in either insect cells using a baculovirus expression system or in

Aspergillus awamori did not produce the native heterodimer. Nevertheless, there appeared to be no difference in specific activity between the native and recombinant forms of PNGase *At* (Ftouhi Paquin *et al.*, 1998).

As well as its size and subunit structure, substrate specificity and pH optimum (pH 5) also indicate that PNGase *At* is more similar to PNGase A (almond) than to PNGase F (*F. meningosepticum*).

In a situation reminiscent of that for PNGase F, there appeared, until recently, to be no other sequences in accessible protein or DNA databases homologous to PNGase A or *At*. Due to the availability of an increasing amount of both genome and proteome data, more sequences homologous to PNGase A and *At* have been identified. These include sequences from plants (e.g. *A. thaliana*, *Oryza sativa*), fungi (e.g. *Neurospora crassa*, *Candida albicans*), bacteria (*Streptomyces avermitilis*) and archaea (e.g. *Sulfolobus solfataricus*). Intriguingly, no homologues have yet been identified in animals.

1.2.1.3 Cytoplasmic PNGases of Eukaryotes

This PNGase type comprises the cytoplasmic PNGases found exclusively and ubiquitously in eukaryotes. Due to the fact that *N*-glycosylation in eukaryotes is a process inherited from earlier phylogenetic groups, i.e. Bacteria and Archaea (Bugg & Brandish, 1994; Burda & Aebi, 1999), it is reasonable to suggest that de-*N*-glycosylation and the corresponding enzymes also developed later than those in Bacteria and Archaea. Therefore, these eukaryotic cytoplasmic PNGases have been chosen to represent the last group, PNGase type III.

The first PNGase in animals was discovered in fish embryos and, due to its acidic pH optimum (pH 4), was thought to be of lysosomal origin (Seko *et al.*, 1991; Seko *et al.*, 1999). After this initial discovery, PNGase activity was detected in several other eukarya, including mammalian cells (Kitajima *et al.*, 1995; Suzuki *et al.*, 1993b; Suzuki *et al.*, 1994c), birds (Suzuki *et al.*, 1997), the budding yeast *S. cerevisiae* (Suzuki *et al.*, 1998) and the fission yeast *Schizosaccharomyces pombe* (Xin *et al.*, 2008), indicating a widespread occurrence as well as an essential function for this protein in eukaryotes.

A gene encoding the cytoplasmic PNGase, PNG1, was first identified in the yeast *S. cerevisiae* where it encodes a 42.5 kDa soluble protein with no evident signal sequence. Subsequent database analyses revealed the existence of highly related genes in various eukaryotic organisms, consistent with the findings of PNGase activities in a wide range of eukaryotic cell lines and organisms (Suzuki *et al.*, 2000).

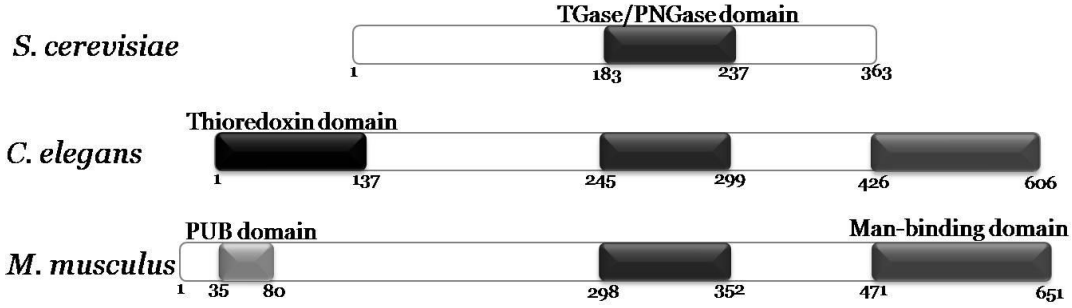


Figure 1.6: Schematic illustration of the primary structure of yeast, nematode and mouse Png1.
TGase/PNGase domain: transglutaminase domain essential for PNGase activity. Thioredoxin domain: N-terminal domain unique to *C. elegans* Png1. Man-binding domain: Mannose-binding domain. Pub domain: protein-protein interaction domain. Figure modified from (Suzuki *et al.*, 2007).

Although PNGases seem to be highly conserved in eukaryotes and carry out the same basic function, there are some differences between orthologues (**Figure 1.6**). Lower eukaryotes (i.e. *S. cerevisiae*, *S. pombe*) possess a PNGase comprising mainly the basic common 'core' sequence (residues 65-362 in *S. cerevisiae* Png1) containing the PNGase domain, present in all cytoplasmic PNGase homologues. In higher eukaryotes however, this common sequence region is extended at both, N- and C-terminus of the core-region (Suzuki *et al.*, 2000). Sequence analysis revealed the presence of a transglutaminase-motif in the 'core'-sequence common to all cytoplasmic PNGases identified so far. Therefore, they have been classified as members of the transglutaminase-like superfamily (Makarova *et al.*, 1999; Suzuki *et al.*, 2002). Transglutaminases (TGase) catalyse the formation of covalent intra- and inter-molecular linkages by cross-linking the side chains of glutamine and lysine residues of proteins. Members of this family usually possess a catalytic triad consisting of Cys, His and Asp. This potential catalytic triad is conserved in all cytoplasmic PNGase

homologues (Suzuki *et al.*, 2002). However, there are no reports on a dual TGase/PNGase function of a protein to date. In 2004 Della Mea *et al.* reported the gene encoding a putative cytoplasmic PNGase in *A. thaliana*, *At*Png1, to be the first plant transglutaminase, rather than a PNGase, despite remarkable amino acid sequence similarities to the PNGases from yeast and mouse (Della Mea *et al.*, 2004). However, in 2007 Diepold *et al.* presented very convincing results contradicting Della Meas' conclusions (Diepold *et al.*, 2007). In this study, besides other evidence, *At*PNG1 was able to rescue a PNGase-negative yeast mutant. Additionally, no decrease in transglutaminase activity could be detected in an *At*Png1-negative *A. thaliana* mutant.

Recently, a PNGase, unique amongst those functionally characterised so far, was identified in the nematode *C. elegans* (Kato *et al.*, 2007; Suzuki *et al.*, 2007). This PNGase was shown to contain an N-terminal thioredoxin-domain (**Figure 1.6**) and to in fact exhibit disulfide reductase activity *in vitro* and *in vivo*.

The first indication of the biological function of cytoplasmic PNGases was the accumulation of intermediate de-*N*-glycosylated proteins in the cytoplasm of cells in the presence of a proteasome inhibitor (Wiertz *et al.*, 1996). It is now well established that cytoplasmic PNGase participates in the 'endoplasmic reticulum-associated degradation (ERAD) pathway'. In eukaryotes, glycoproteins that are destined for the secretory pathway have to pass a stringent quality control in the ER, to ensure they assume their native conformation. This test may include interactions with chaperones and several rounds of glycosylation and deglycosylation. Proteins that fail to mature correctly in the ER are retro-translocated into the cytoplasm, where they are ubiquitinated and targeted to the proteasome for degradation (Baumeister & Pouch, 1998). Misfolded glycoproteins are deglycosylated by cytoplasmic PNGase prior to proteasomal degradation (Hirsch *et al.*, 2003). PNGase has been reported to be localised free in the cytoplasm (Suzuki *et al.*, 1998) as well as to be associated with the ER membrane (Suzuki *et al.*, 1997). As mentioned earlier, PNGases from animals have N- and C-terminal extensions to the common PNGase domain of the protein (**Figure 1.6**). It has been shown that the N-terminal region of these extensions contains a PNGase/ubiquitin-associated or UBX-containing (PUB) domain, which is thought to mediate

protein-protein-interactions (Suzuki *et al.*, 2001b). Employing yeast-two-hybrid analysis, *in vitro* GST pull-downs and *in vivo* co-localisation-experiments, recent studies identified several proteins that interact with cytoplasmic mouse PNGase and/or with each other, which lead to a model proposing the coupling of protein retro-translocation, ubiquitination, deglycosylation and degradation (Li *et al.*, 2005; Li *et al.*, 2006; Park *et al.*, 2001). **Figure 1.7** shows the latest model integrating the mPNGase with the ERAD pathway (Li *et al.*, 2006). The link between PNGase and the proteasome mediated by mHR23B was first identified in yeast, where yPNGase is linked to the proteasome *via* the N-terminal ubiquitin-like domain of the yeast-mHR23B-homologue Rad23, a protein originally identified to be required for DNA repair (Suzuki *et al.*, 2001a). In yeast, no other proteins were shown to interact with the PNGase, which stands in contrast to the system in mouse. It was recently demonstrated that the PUB domain of mPNGase is critical for the interaction between mPNGase and mp97, which in turn forms complexes with the proteins mAMFR (mouse autocrine motility factor receptor) and mY33K. The protein p97 is an AAA ATPase, which is involved in various cellular functions, including protein degradation, cell cycle, apoptosis, DNA repair and membrane vesicle fusion (Woodman, 2003). It is also thought to aid in extracting the misfolded glycoprotein from the ER to the cytoplasm and was shown to interact with Derlin-1, an integral ER-membrane protein required for the translocation of certain misfolded proteins from the ER to the cytoplasm (Lilley & Ploegh, 2004). The exact translocation mechanism is unknown, but it is thought to involve either other ER-membrane proteins or Derlin-1 oligomerisation. The mAMFR is an E3 ligase located in the ER membrane. The recruitment of ubiquitin E3 ligases by p97 to the site of the protein retro-translocation channel, coupling ubiquitination and retro-translocation, was shown earlier by Lilley & Ploegh. Y33K is a protein of unknown function with an ubiquitin-like and ubiquitin-associated domain. Li *et al.* (2006) hypothesise that the PUB domain is an evolutionary addition to the PNGase of higher organisms, such as insects and vertebrates, facilitating the assembly of a more complex ERAD system in mammals that involves multiple regulatory protein-protein-interactions.

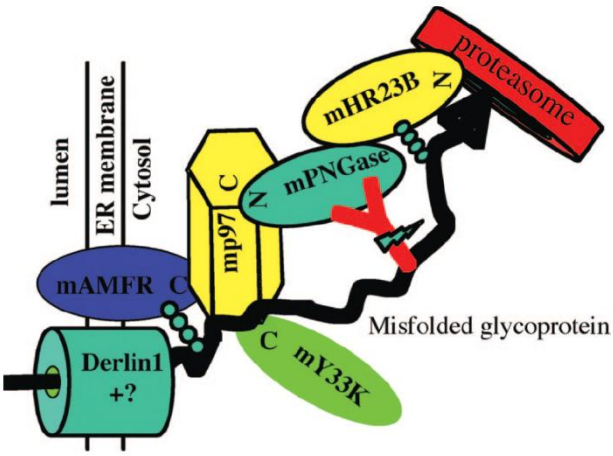


Figure 1.7: Model showing retro-translocation, ubiquitination, deglycosylation, and degradation of a glycosylated ERAD substrate.
The arrangement of the glycoprotein being degraded is hypothetical. Figure taken from Li *et al.*, 2006.

Earlier studies investigating the substrate specificity of cytoplasmic PNGase raised questions regarding the proposed involvement of cytoplasmic PNGase in the ERAD pathway as they found the enzyme unable to deglycosylate full-length glycoproteins. Only short glycopeptides appeared to be substrates for this type of PNGase (Suzuki *et al.*, 1998; Suzuki *et al.*, 2000). However, Hirsch *et al.* demonstrated that mammalian PNGase was able to distinguish between native and non-native proteins (Hirsch *et al.*, 2003; Hirsch *et al.*, 2004). They used mammalian tissue culture cells expressing only the α-chain of the T-cell receptor (TCRα), a known substrate of the ERAD pathway that fails to fold properly in absence of the other members of the TCR complex, to investigate the deglycosylation of misfolded full-length glycoproteins. TCRα fails to progress to the Golgi complex and therefore retains its high-mannose *N*-glycans. This *in vivo* experiment as well as *in vitro* experiments using recombinant yPNGase and mPNGase demonstrated the ability of this type of PNGase to deglycosylate non-native full-length proteins (Hirsch *et al.*, 2003). Similar results were obtained using RNase B, carboxypeptidase Y and ovalbumin, which were only deglycosylated when previously denatured (Hirsch *et al.*, 2004; Joshi *et al.*, 2005), indicating that the PNGase is specific for misfolded substrates. Regarding the glycan-type, it was shown that high-mannose *N*-glycans are preferred over complex-type oligosaccharides (Hirsch *et al.*, 2003), which is in agreement with the proposed involvement of PNGase activity in the ERAD

pathway, as misfolded proteins retro-translocated from the ER are not exposed to the Golgi complex glycosidases that are responsible for synthesis of complex and hybrid *N*-glycans (**Figure 1.2**).

Recently the crystal structures of the cytoplasmic PNGases from yeast and mouse were published in complex with the Xeroderma pigmentosum protein C-binding domain (XPCB) of yRad23 and mHR23B, respectively. This domain in yRad23 and mHR23B was shown to be responsible and sufficient for complex formation with XPC (Rad4 in yeast) as well as with PNGase. The Rad23-Rad4 complex in yeast and HR23-XPC complex in mammals play an important role in DNA repair (Lee *et al.*, 2005; Zhao *et al.*, 2006). As already reflected by the differences in primary structure, the structures of these two proteins are completely different to the crystal structure determined for PNGase F (**1.2.1.1**), again demonstrating the divergence of these two PNGase types and the most likely convergent evolution of their functions.

Lee *et al.* (2005) demonstrated that yPNGase is a zinc metalloenzyme and that it folds into α/β structure with an overall structure formed by three domains: a Rad23 binding domain, a core domain and a Zn^{2+}-binding domain (**Figure 1.8**). The core domain contains six central β-strands (S6-S11) supported by three α-helices (H3, H5, H6) and several loops. The Zn^{2+}-binding domain is comprised of five β-strands (S1-S5) and two helices (H7, H8). Loops that link the strands S1 and S2, and S3 and S4 provide the four thiol ligands to the Zn^{2+} ion (Lee *et al.*, 2005; Zhao *et al.*, 2006). It was shown previously that the mutation of each of the four special cysteine residues to alanine abolished PNGase activity (Katiyar *et al.*, 2002). These four cysteine residues, Cys-129, Cys-132, Cys-165 and Cys-158, were shown to coordinate the Zn^{2+} ion in two CxxC motifs in the crystal structure.

An antiparallel β-sheet, formed by strands S6-S8 and S10 and helix H3 from the core domain, are packed against helix H8 and strands S1, S4 and S5 of the Zn^{2+}-binding domain forming a deep interdomain cleft. This cleft was found to contain two binding regions, one for carbohydrate-binding and one for protein-binding. Located deep inside this cleft between the core and the Zn^{2+}-binding domain is the transglutaminase-like catalytic Cys-His-Asp triad. This structural feature provides an insight into the specificity of PNGase for denatured

glycoproteins, as it was demonstrated that native substrates would simply not fit properly into this deep cleft, whereas denatured proteins are more flexible and therefore can access the active site without constraints by the cleft walls.

The Rad23-binding domain consists of the N-terminal helix H1, the C-terminal helix H12 and helices H2 and H11. H1 and H12 were found to form two interface regions with Rad23-XPCB, both essential for Rad23 binding. In the first interface, H1 binds to a groove formed by the four helices of Rad23-XPCB, whereas in the second interface H12 interacts exclusively with the N-terminal helix of Rad23-XPCB.

In 2009, Zhao *et al.* published the crystal structure of the yeast PNGase in complex with the inhibitor $GlcNAc_2$-iodoacetamide (Zhao *et al.*, 2009). It had been demonstrated previously that $GlcNAc_2$-iodoacetamide and its derivatives covalently bind to the active site cysteine in yeast PNGase in a highly specific manner (Suzuki *et al.*, 2006). They found that residues His-218, Glu-238 and Lys-253 formed critical contacts with the two GlcNAcs of the core glycan; in fact, all interactions appear to occur between these residues and just the proximal GlcNAc. The distal GlcNAc is primarily recognised by Trp-251 *via* van der Waals contacts. Yeast PNGase prefers high-mannose type over complex type substrates. Terminally truncated high-mannose-type glycoproteins (for example, Manβ-1,4$GlcNac_2$) are not deglycosylated at all (Hirsch *et al.*, 2003). Still, it remains unclear how the enzyme recognises and interacts with the mannose moiety of its substrates. It has been suggested, that the C-terminal residues contribute to mannose binding; however, in the yeast PNGase crystal structures these residues were generally disordered (Zhao *et al.*, 2009).

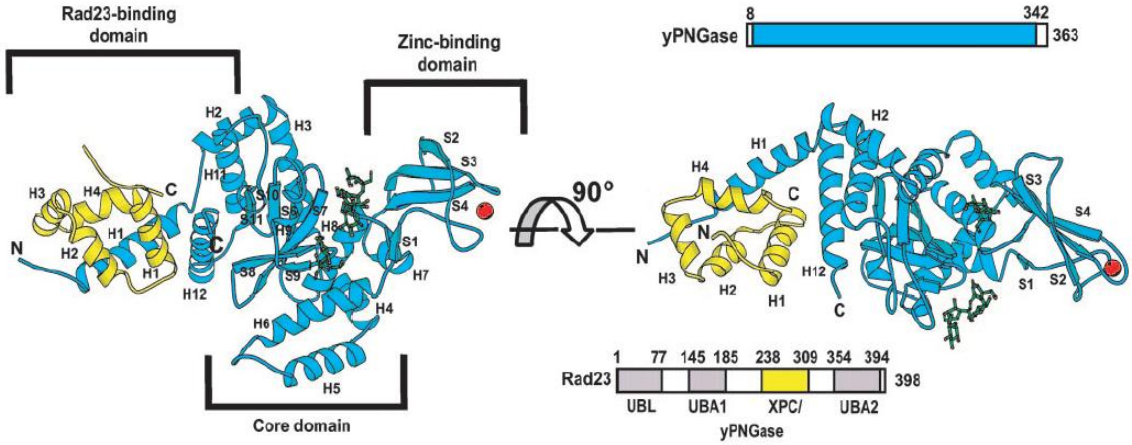


Figure 1.8: The crystal structure of the yPNGase-yRad23-complex.
yPNGase is shown in blue, yRad23 in yellow and the Zn^{2+} in red. UBL, ubiquitin-like domain; UBA, ubiquitin-associated domain. Figure was taken from (Lee *et al.*, 2005).

The overall structure of mouse PNGase was shown to be similar to yPNGase, although complex formation with the XPCB domain of mHR23B was found to be fundamentally different. In this protein H12 and not H1 has the primary XPCB interaction function. H1 is absent in mPNGase, and XPCB binding in mouse is mediated by H11 and especially H12. Zhao *et al.* suggest, based on the comparison of H12 sequences from different species, that H12 has evolved in such a way that the primary XPCB-interacting function migrated from H1 to H12 (Zhao *et al.*, 2006). In contrast to yPNGase, it could be established that the C-terminal domain of the mouse PNGase is involved in mannose recognition (Zhou *et al.*, 2006).

1.3 Aims of Thesis

In recent years significant progress has been made towards understanding cytoplasmic PNGases in different organisms. As discussed in section **1.2.1.3** this type of PNGase is ubiquitous in eukaryotic cells, where it functions as part of the proteasome complex involved in the degradation of misfolded glycoproteins. In contrast, no studies have been undertaken to further characterise the other groups of PNGases, i.e. PNGase F-like and PNGase A- and *At*-like proteins. PNGase F and PNGase A are widely used in proteomic research, biotechnology and glycobiology to liberate *N*-glycan moieties from peptides or proteins for a variety of reasons, including structure/function studies of either the protein or glycan moiety, or both, and the removal of glycans that potentially may interfere with protein crystallisation.

For a long time PNGase F appeared to be unique as no homologues could be identified in various databases. This changed with the appearance of a homologous sequence found in the bacterium *D. radiodurans* and the subsequent emergence of a few other homologues. Yet, the number of homologues is very limited.

The main aim of this thesis was to identify (using bioinformatic methods), clone, heterologously express, and characterise structurally and functionally, proteins homologous to PNGase F and PNGase A and *At*.

The second aim of this work was the further characterisation of the catalytic mechanism used by PNGase F. To achieve this, several site-specific mutants were generated and functionally characterised.

Figure 1.9 shows an overview of the aims set for this thesis.

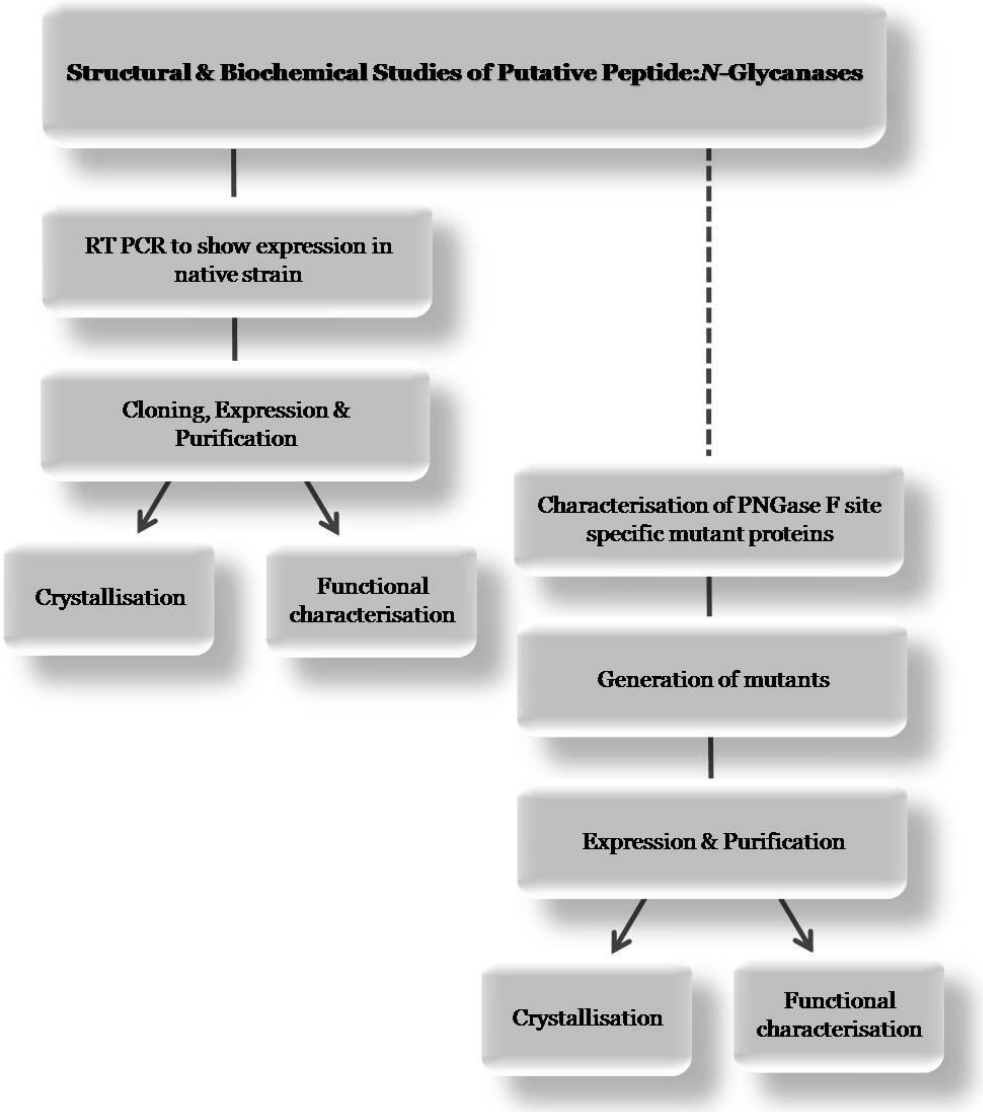


Figure 1.9: Schematic overview of the aims of this thesis.

Chapter 2

Materials & General Methods

2 Materials & General Methods

2.1 Materials, Chemicals & Kits

Acros Organics, Geel, Belgium:
Cyanogen Bromide

Ansell, Red Bank, NJ, USA:
Latex Gloves

Axygen Scientific Inc., Union City, CA, USA:
Pipet tips (10 µL, 200 µL, 1000 µL);
PCR tubes (0.2mL thin-wall flat cap);
Microtubes (Standard 1.7mL MaxyClear)

Amresco, Solon, Ohio, USA:
EDTA disodium salt dehydrate

Bioline (Aust) Pty Ltd, Alexandria, Australia:
IPTG

Eppendorf AG, Hamburg, Germany:
Perfectprep® Gel Cleanup Kit

GE Healthcare (formerly Amersham Pharmacia Biotech):
Superdex™ 75 10/300 GL;
HiTrap™ Chelating HP;
Chelating Sepharose™ Fast Flow;
Illustra™ RNAspin Mini Isolation Kit

Greiner bio-one, Frickenhausen, Germany:
PP-Test Tubes, 15 mL;
PP-Test Tubes, 50 mL;
Serological Pipettes (5 mL, 10 mL, 50 mL);
Cellstar® 6 Well Cell Culture Multiwell Plates;
Cellstar® 12 Well Cell Culture Multiwell Plates

Hellma GmbH & Co. KG, Müllheim, Germany:
Quartz SUPRASIL® precision cell, 0.1 mm path length;
Quartz SUPRASIL® precision cell holder

Integrated DNA Technologies (IDT), Coralville, USA:
Oligonucleotides

Invitrogen™, Carlsbad, USA:
LB Broth Base (Miller's LB Broth Base);
Cellfectin® Reagent;
Sf-900 II SFM (Gibco®);
Grace's Insect Cell Culture Medium, Unsupplemented (Gibco®);
Platinum® *Taq* DNA polymerase;
SuperScript™ II One-Step RT-PCR System with Platinum® *Taq* DNA Polymerase;
ChargeSwitch®-Pro Plasmid MiniPrep Kit;
PureLink® HQ Mini Plasmid Purification Kit;
Restriction Endonucleases;
DNAzol® Reagent (Gibco®)

LabServ, Biolab Ltd., Auckland, New Zealand:
Petri dishes;
Nitrile Gloves

Merck, KGaA, Darmstadt, Germany:
KOD DNA polymerase (Novagen®);
Peptone;
Acrylamide:Bis ready-to-use solution 40% (19:1)

Millipore, MA, USA:
Filter membranes;

New England Biolabs® Inc., Ipswich, USA:
Amylose resin;
Genenase™ I

Nunc™ (Part of Thermo Fisher Scientific Inc.), Roskilde, Denmark:
CryoTubes™

Oxoid LTD (Part of Thermo Fisher Scientific Inc.), Basingstoke, England:
Bacto-Agar

Phenomenex®, Torrance, USA:
Jupiter® 5 µm C18 300 Å (250 x 4.6 mm);
Jupiter® 10 µm C18 300 Å (250 x 10 mm);

Pure Science Limited, Porirua, New Zealand:
Glycine

Roche Applied Science, Roche Diagnostics N.Z., Ltd., Auckland, New Zealand:
Complete™ Mini, EDTA free Protease Inhibitor Cocktail Tablets;
High Pure Plasmid Isolation Kit;
DIG Glycan Detection Kit;
Pwo DNA polymerase;
Taq DNA polymerase;
Restriction Endonucleases;
T4 DNA Ligase;
BM Chemiluminescence Blotting Substrate (POD)

Roth GmbH + Co. KG, Karlsruhe, Germany:
Tris;
Malt extract

Sartorius AG, Göttingen, Germany:
Minisart® filter 0.2 µm;
Minisart® filter 0.8 µm;
Vivaspin™ centrifugal concentrators (0.5 mL, 2 mL, 20 mL)

Sigma-Aldrich, St. Louis, USA:
Oligonucleotides;
Ampicillin sodium salt;
Gentamycin sulfate salt;
Chloramphenicol;
Kanamycin sulfate;
D(+)-Glucose;
Imidazole (Sigma-Aldrich/Fluka);
Rubidium chloride;
EPPS;
HEPES;
EDTA disodium salt dihydrate;
Nickel (II) chloride hexahydrate;
Coomassie Brilliant Blue R 250 (Fluka analytical);
Coomassie Brilliant Blue G 250 (Fluka analytical);
RNaseB (from bovine pancreas);
DNase I;
Anti-Rabbit IgG (whole molecule)-Peroxidase antibody produced in goat;
Trypsin;

Hen egg ovalbumin;
Bovine serum albumin

United States Biochemical (USB) Corp., Cleaveland, USA:
Bromophenol Blue sodium salt;

2.2 Technical Equipment

Applied Photophysics, Surrey, UK:
Chirascan™ Circular Dichroism Spectrometers

Biometra biomedizinische Analytik GmbH, Göttingen, Germany:
TGradient Thermocycler

Bio-Rad Laboratories, Hercules, USA:
SmartSpec™ Plus Spectrophotometer;
Mini-PROTEAN® II Electrophoresis System;
Mini Trans-Blot® Electrophoretic Transfer Cell;
PowerPac™ Basic;
PowerPac™ 300;
Sub-Cell® System;
Econo-Column® Columns (empty);
UV-Trans-Illuminator, Bio Rad Gel Doc

Dionex, Sunnyvale, USA:
Dionex UltiMate® 3000 HPLC system

Eppendorf AG, Hamburg, Germany:
Eppendorf miniSpin® plus

Fujifilm Corporation, Tokyo, Japan:
Intelligent Dark Box II

GE Healthcare (formerly Amersham Pharmacia Biotech):
Äkta™ Explorer Chromatography System

Infors HT AG, Bottmingen, Switzerland:
Minifors Benchtop Bioreactor

Rigaku Americas Cooperation:
MicroMax-007 microfocus rotating anode generator
R-AXIS IV++ imaging plate area detector

SLM Aminco Instruments Incorporated, Rochester, USA:
French press

Thermo Fisher Scientific Inc., Waltham, USA:
Sorvall Centrifuge RT7;
Sorvall Evolution® RC Centrifuge;
NanoDrop™ 1000 Spectrophotometer;
Barnstead NANOpure® DIamond Life Science (UV/UF) Ultrapure Water System;
Heraeus Biofuge® fresco;
Heraeus Biofuge® pico

TTP Labtech, Royston, UK:
Mosquito® Crystallisation Robot

Waters, Milford, USA:
Micromass M@LDI mass spectrometer

2.3 Deionised water

Deionised water was obtained from a Barnstead NANOpure® DIamond Life Science (UV/UF) ultrapure water system (Thermo Fisher Scientific), containing two ion exchange and two organic filter cartridges. This system provides pure DNase-, RNase- and DNA-free water with less than 10 ppb TOC (total organic carbon) and up to 18.2 megohm/cm resistivity. It is bacteria- and particle-free due to a final 0.2 µm filter. Throughout this thesis the term 'pure water' or the short form H_2O_{pure} will be used to refer to this purified water.

2.4 Storage and Propagation of Bacterial Cultures

Bacterial strains to be stored long term were kept as glycerol stock cultures at -80°C. These were prepared by growing cultures in 5 mL of culture medium containing, if required, the appropriate antibiotics (2.6) until the OD_{600} (2.9)

reached approximately 0.8. The bacteria were pelleted for 5 minutes at 2,600 *g* and resuspended in 1 mL of fresh culture medium containing 20% (v/v) glycerol. After transfer into a 1 mL CryoTube™ (Nunc™, Thermo Fisher Scientific) the suspension was snap frozen in liquid nitrogen and stored at -80°C.

For propagation of frozen cultures, cells were streaked on appropriate agar plates containing the required antibiotics and incubated for 12-15 h at 37°C. Following purity control, single colonies were grown with shaking (200 rpm) in liquid medium at 37°C for 12-16 h.

2.5 Cultivation of Bacterial Cells

Culture media for growth of bacterial cultures were prepared as described below and sterilised at 121°C and 2×10^5 Pa for 20 minutes. If required, media were cooled to less than 50°C and antibiotics were added at appropriate concentrations (2.6) to liquid and solid media.

Liquid cultures were grown in Erlenmeyer flasks, with a flask to liquid volume ratio of 5:1. Typically, the cultures were incubated at 37°C on a rotary shaker at approximately 180 rpm (G25; New Brunswick Scientific, USA).

For preparation of solid media, 1.6% (w/v) Bacto-Agar (Oxoid) was added before autoclaving. After autoclaving approximately 25 mL of the media were dispensed into each sterile Petri dish. After solidifying of the agar, the plates were used immediately or stored at 4°C for a maximum of one month.

2.5.1 Luria Bertani (LB) Medium:

LB medium was prepared from premixed LB Broth Base powder (Miller's LB Broth Base; Invitrogen™) using 25 g/L.

Formulation per 1 L

Peptone	10.0 g
Yeast extract	5.0 g
NaCl	10.0 g

2.5.2 GYM Streptomyces Medium

GYM Streptomyces medium was used to cultivate *S. avermitilis*.

Glucose	4.0 g
Yeast extract	4.0 g
Malt extract	10.0 g
H_2O_{pure}	up to 1.0 L

Adjust pH to 7.2.

2.5.3 Oatmeal Agar (DSM Medium 425)

Oatmeal medium was used to cultivate *S. avermitilis*.

Oat flakes	10.0 g
Oatmeal	10.0 g
Agar	15.0 g
H_2O_{pure}	up to 1.0 L

Adjust pH to 7.0-7.2.

For preparation of liquid Oatmeal medium the mixture excluding the agar was cooked for approximately 5 minutes in a microwave. To remove larger particles, the mixture was filtered through Nr.1 Whatman filter paper (Whatman®) and the remaining liquid autoclaved.

2.5.4 Corynebacterium Medium

Corynebacterium medium was used to cultivate *D. radiodurans*.

Casein peptone	10.0 g
Yeast extract	5.0 g
Glucose	5.0 g
NaCl	5.0 g
H_2O_{pure}	up to 1.0 L

Adjust pH to 7.2-7.4.

2.5.5 Malt Extract Medium

Malt extract medium was used to cultivate *A. niger*.

Malt extract	30.0 g
Peptone	5.0 g
H_2O_{pure}	up to 1.0 L

2.6 Antibiotics

Stock solutions of generally used antibiotics were prepared at the concentrations given in **Table 2.1**. All stock solutions that were made up in H_2O_{pure} were sterilised by filtration using a 0.22 µm filter attached to a sterile syringe. The stocks were stored in 1 mL aliquots at -20°C and added to sterile medium that had been cooled to ~50°C.

Table 2.1: Antibiotic stock solutions and final concentration for *E. coli*.

Antibiotic	Stock concentration [mg/mL]	Final concentration [µg/mL]
Ampicillin Na-salt	100 in H_2O_{pure}	100
Tetracycline-HCl	12.5 in 70% EtOH	12.5
Kanamycin sulfate	30 in H_2O_{pure}	30 or 15
Gentamycin sulfate	10 in H_2O_{pure}	10
Chloramphenicol	34 in EtOH	34

2.7 Bacterial Strains

Bacterial strains used throughout this project for various purposes are summarised in **Table 2.2**.

Table 2.2: Bacterial strains used in this project.

Strain	Genotype	Antibiotic Resistance	Application	Source
XL1 Blue	*endA1 gyrA96*(*nalR*) *thi-1 recA1 relA1* l*ac glnV44* F'[::Tn10 *proAB*$^{+}$ *lacI*^{q} Δ(*lacZ*)M15] *hsdR17*(r_K^- m_K^+)	Tetracycline (12.5 µg/mL)	general purpose cloning strain	Stratagene
DH5α	F$^-$ *endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG* Φ80d*lacZ*ΔM15 Δ(*lacZYA-argF*)U169, *hsdR17*(rK^- mK^+), λ$^-$	none	general purpose cloning strain	Invitrogen™
One Shot® TOP10	F$^-$ *mcrA* Δ(*mrr-hsdRMS-mcrBC*) φ80*lacZ*Δ*M15* Δ*lacX74 recA1 araD139* Δ(*ara-leu*)7697 *galU galK rpsL* (Str^R) *endA1 nupG*	Streptomycin (50 µg/mL)	general purpose cloning strain	Invitrogen™
One Shot® Mach1™-T1R	F$^-$φ80(*lacZ*)ΔM15 Δ*lacX74 hsdR*($r_K^-m_K^+$) Δ*recA1398 endA1 tonA*	none	general purpose cloning strain	Invitrogen™
BL21(DE3)	F$^-$ *ompT hsdSB*(r_B^- m_B^-) *gal dcm* (DE3)	none	general purpose expression host	Novagen®
TB1	F$^-$ *ara* Δ(*lac-proAB*) [Φ80d*lac* Δ(*lacZ*)*M15*] *rpsL*(Str^R) *thi hsdR*	Streptomycin (50 µg/mL)	recommended protein expression host for pMAL™-c2x vectors	NEB®
Origami™(DE3)	Δ*ara–leu7697* Δ*lacX74* Δ*phoA Pvu*II *phoR araD139 ahpC galE galK rpsL* F'[*lac*$^+$ *lacI*^{q} *pro*] (DE3) *gor522* ::Tn10 *trxB* (Kan^R, Str^R, Tet^R)	Kanamycin (15 µg/mL) Streptomycin (50 µg/mL) Tetracycline (12.5 µg/mL)	general expression host, with enhanced disulfide bond formation in *E. coli* cytoplasm	Novagen®
Origami B™(DE3)	F$^-$ *ompT hsdSB*(r_B^- m_B^-) *gal dcm lacY1 ahpC gor522*::Tn10 *trxB* (Kan^R, Tet^R)	Kanamycin (15 µg/mL) Tetracycline (12.5 µg/mL)	general expression host, with Tuner lac permease mutation and enhanced disulfide bond formation in *E. coli* cytoplasm	Novagen®
RosettaBlue™(DE3)	*endA1 hsdR17*(r_{K12}^- m_{K12}^+) *supE44 thi-1 recA1 gyr96 relA1 lac* (DE3) F'[*proA*$^+$*B*$^+$ *lacI*^{q}Δ*Z M15* ::Tn10] pRARE (Cam^R, Tet^R)	Chloramphenicol (34 µg/mL) Tetracycline (12.5 µg/mL)	general expression host, which provides six rare codon tRNAs	Novagen®
Rosetta-gami™(DE3)	Δ*ara–leu7697* Δ*lacX74* Δ*phoA PvuII phoR araD139 ahpC galE galK rpsL* (DE3) F '[*lac*$^+$ *lacI*^{q} *pro*] *gor522* ::Tn10 *trxB* pRARE (Cam^R, Kan^R, Str^R, Tet^R)	Chloramphenicol (34 µg/mL) Kanamycin (15 µg/mL) Streptomycin (50 µg/mL) Tetracycline (12 5 µg/mL)	general expression host, with enhanced disulfide bond formation in *E. coli* cytoplasm; provides six rare codon tRNAs	Novagen®

DH10Bac™	F- *mcrA* Δ(*mrr-hsdRMS-mcrBC*) φ80*lacZ*Δ*M15* Δ*lacX74 recA1 endA1 araD139* Δ (*ara-leu*)*7697 galU galK* λ- *rpsL nupG* /pMON14272 / pMON7124	Kanamycin (50 µg/mL)	production of recombinant baculovirus DNA for protein expression in insect cells	Invitrogen™
ccdB Survival™	F- *mcrA* Δ(*mrr-hsdRMS-mcrBC*) Φ80*lacZ*Δ*M15* Δ*lacX74 recA1 ara*Δ*139* Δ(*ara-leu*)*7697 galU galK rpsL* (Str^R) *endA1 nupG tonA*::*Ptrc-ccdA*	Chloramphenicol (15-30 µg/mL)	CcdB-resistant strain used for propagation of vectors containing the ccdB gene	Invitrogen™
Deinococcus radiodurans R1 (DSM 20539)	Type strain, isolated from irradiated ground pork and beef	none	RNA isolation	DSMZ[3]; (Brooks & Murray, 1981)
Streptomyces avermitilis MA-4680 (DSM 46492)	Type strain, isolated from soil (Japan)	none	RNA isolation	DSMZ; (Burg *et al.*, 1979; Kim & Goodfellow, 2002)
Aspergillus niger	--	none	gDNA isolation, RNA isolation	IMBS culture collection

[3] DSMZ: Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (German Collection of Microorganisms and Cell Cultures), Braunschweig, Germany

2.8 Plasmids

Plasmids used during this work are summarised in **Table 2.3**.

Table 2.3: Plasmids used in this project.

Plasmid	Size [kbp]	Main Features	Source/ Reference
pMAL™-c2G	6.7	P_{tac}, *malE* (-signal sequ.), *lacI*^{q}, Genenase™ I cleavage site, Amp^R	NEB®
pMAL™-p2G	6.72	P_{tac}, *malE* (+signal sequ.), *lacI*^{q}, Genenase™ I cleavage site, Amp^R	NEB®
pET-32a(+)	5.9	T7*lac*, Amp^R	Novagen®
pET-40b(+)	6.2	T7*lac*, *dsbC* (N), His_6 (C), Kan^R	Novagen®
pETDuet™	5.4	T7, 2 MCS, Amp^R	Novagen®
pETDuet™-Trx-HST-*Nco*	5.75	Modified pETDuet™; *trx* in MCS1, His_6 (C)	T.S. Loo
pETDuet™-DsbC-HST-*Nco*	6.13	Modified pETDuet™; *dsbC* in MCS1, His_6 (C)	T.S. Loo
pETDuet™-MalE-HST-*Nco*	6.53	Modified pETDuet™; *malE* in MCS1, His_6 (C)	T.S. Loo
pKS_OmpA_*Dra*_nosigpep	6.77	T7; OmpA-leader sequ. (N), His_6 (C), Amp^R, *Dra*PNGase coding region (50-1962)	Dr. G.E. Norris/Jessie Green
pTUM4	5.9	*dsbA*, *dsbC*, *fkpA*, *surA*, Cm^R	(Schlapschy *et al.*, 2006)
pENTR™/SD/D-TOPO®	2.6	*att*L1, *att*L2, TOPO recognition sites 1+2	Invitrogen™
pENTR™/TEV/D-TOPO®	2.6	*att*L1, *att*L2, TOPO recognition sites 1+2	Invitrogen™
pDEST™8	6.5	AcMNPV polyhedrin promoter, *ccd*B, *att*R1+2, Tn7, pUC ori, Ap^R, Gm^R, Cm^R	Invitrogen™
pDEST™10	6.7	As pDEST™8, His_6 (N)	Invitrogen™
pDEST™15	7.0	T7 promotor, *ccd*B, , Cm^R between *att*R1+2 for counter selection, Amp^R, GST (N)	Invitrogen™
pDEST™17	6.4	T7 promotor, *ccd*B, , Cm^R between *att*R1+2 for counter selection, Amp^R, His_6 (N)	Invitrogen™
pOPH6	5.9	T7; OmpA-leader sequ. (N), His_6 (C), Amp^R, PNGase F coding region	(Loo *et al.*, 2002)

2.9 Measurement of the Optical Density of Bacterial Cultures (OD_{600})

To determine the optical density of liquid bacterial cultures a spectrophotometer (SmartSpec™ Plus, Bio-Rad) was set to the wavelength of 600 nm and a zero calibrated using the appropriate sterile medium used to grow the bacterial culture. An aliquot of the culture to be measured was transferred into a 1 mL cuvette and the OD_{600} determined. Generally, OD_{600} readings above 0.3 are not accurate. Therefore, the culture was diluted in case of readings $OD_{600} > 0.3$ using the appropriate sterile medium.

2.10 Polymerase Chain Reaction (PCR)

Polymerase chain reactions were performed to amplify specific DNA sequences for subsequent cloning into plasmids or for analytical purposes (2.11).

Oligonucleotides (2.12) were designed based on available sequence data and synthesised specifically for the DNA sequence to be amplified. Appropriate restriction sites were included in case of subsequent cloning requiring restriction endonuclease digest.

The purpose of the PCR determined the type of DNA polymerase used in the reaction. For analytical PCR *Taq* DNA polymerase (Roche Applied Science) was employed. If high fidelity transcription was essential either KOD or *Pwo* DNA polymerase (Novagen®; Roche Applied Science, respectively) was used. These enzymes exhibit 3´-5´ exonuclease (proofreading) activity, resulting in lower mutation frequencies, which is important to obtain mutation-free DNA fragments for subsequent cloning procedures and protein expression.

The standard reaction set ups for these three enzymes are summarised in **Table 2.4**.

Table 2.4: Standard PCR set ups for *Taq*, *Pwo* and KOD DNA polymerase.

Component	*Taq* DNA Polymerase	*Pwo* DNA Polymerase	KOD DNA Polymerase	Final concentration
	Volume [µL]			
10× Reaction Buffer[4]	5.0	5.0	5.0	1×
25 mM $MgCl_2$	-	-	3.0	1.5 mM
dNTPs[5]	5.0	5.0	5.0	0.2 mM (each)
Sense (5′) Primer (10 pmol/µL)	2.0	2.0	2.0	0.4 µM
Anti-Sense (3′) Primer (10 pmol/µL)	2.0	2.0	2.0	0.4 µM
Template DNA[6]	1.0	1.0	1.0	-
DNA polymerase[7]	1.0	0.2	0.4	1.0 U
H_2O_{pure}	34.0	34.8	31.6	-
Total Volume	50.0	50.0	50.0	-

The reaction mixture was prepared on ice in 0.2 µL thin-walled PCR tubes. The volumes given in **Table 2.4** were scaled to the amount of product required. Where appropriate, a master mixture was prepared.

The general thermal cycle used is given in **Table 2.5**.

Table 2.5: Thermal profile used for amplification of DNA fragments using a Biometra TGradient Thermocycler.

Cycles		Temperature	Time
Initial Denaturation			
1×		94°C	5 min
Amplification			
	Denaturation	94°C	1 min
25×-33×	Annealing[8]	50-65°C	45 sec
	Elongation	72°C	1 min/kbp
Final Elongation			
1×		72°C	10 min

[4] *Taq* PCR buffer with MgCl2; *Pwo* PCR buffer with MgSO4; KOD PCR buffer
[5] Nucloetide mix containing 10 mM of each nucleotide
[6] ≤ 200 ng genomic DNA; 1-10 ng bacterial DNA; 0.1-1 ng plasmid DNA
[7] *Taq* DNA polymerase 1 U/µL; *Pwo* DNA polymerase 5.0 U/µL; KOD DNA polymerase 2.5 U/µL
[8] The optimal annealing temperature depends on the melting temperature of the primers used

The results of all PCR reactions were analysed using agarose gel electrophoresis (2.19) and the PCR products were purified from agarose gels (2.13).

2.11 Whole Cell PCR Screening of E. coli Transformants (Colony PCR)

The transformation of a ligation reaction (2.15) into competent *E. coli* cells usually results in a mixture of transformants, some of which will contain the plasmid with the desired DNA fragment and some that will contain only the re-ligated empty plasmid. Colony PCR was used as a tool to quickly screen a large number of transformants without the necessity of prior plasmid isolation (2.18).

Transformants that appeared on selective solid media after overnight incubation at 37°C were 'picked' off the plates using the tips of sterile toothpicks. The cells were then resuspended in sterile microcentrifuge tubes containing 200 µL of culture medium containing the appropriate antibiotics and incubated for 5-15 h at 37°C with shaking. After this incubation, 5 µL of each transformant culture were spotted on plates made up with fresh selective media for later reference. The cells were then harvested in a bench-top centrifuge (2 min, 14000 *g*), resuspended in 50 µL sterile H_2O_{pure} and boiled for 5 minutes in a water bath to lyse the cells. 1 µL of this mixture was used as template in subsequent PCR reactions using primers specific for the inserted DNA fragment, primers that bind just outside of the inserted fragment in the plasmid, or a combination of both. The fragments resulting from these PCRs were then separated using agarose gel electrophoresis and visualised using ethidium bromide staining (2.19). No PCR fragment is expected for transformants containing only the empty plasmid. Transformants harbouring the plasmid containing the desired DNA insertion should produce a fragment whose size is determined by the primer combinations used.

The composition of a typical Colony PCR sample was as follows. Usually a master-mix was prepared containing all the necessary components except for the template, which was specific for each reaction.

10× PCR buffer	0.5 µL
dNTPs (10 mM)	0.1 µL
Primer-fwd (10 µM)	0.25 µL
Primer-rev (10 µM)	0.25 µL
Template	1.0 µL
Taq DNA polymerase (5 U/µL)	0.02 µL
H_2O_{pure}	2.88 µL

For the PCR program set up refer to section 2.10.

2.12 Oligonucleotides for PCR

Table 2.6: List of relevant oligonucleotides used for cloning in this project.

Nr.[9]	**Name of Oligonucleotide**	**5'→3'-nucleotide sequence**
O1	AnigerPNGase.fwd	CACCATGCTGGTCTCTTTCAGTGTCGC
O2	AnigerPNGase.rev	CTAGCTGTCAGTATCCGATACAAC
O3	Dra_full_pET40mod_NcoI_fwd	GTAACCATGGGCAGCAAGGACACTCGCTCA
O4	DraPNGdomain_pET32a_NcoI_fwd	GAAAACCATGGGCGGCGAGTACCTGAGTTGGGAA
O5	DraPNGdomain_pET32a_BamHI_rev	GGAAGGATCCCTACTGCTTGACGTTCGGTTTG
O6	Dra_trunc123_GWN.fwd	CACCGAAAACCTGTATTTTCAGGGCGCGCTCGGCAAATTGCTCG
O7	Dra_trunc591_GWN.rev	CTAGCTGTTCCACAGCTTGACGCC
O8	Dra_trunc643_GWN.rev	TCAGCCACGCTCGGCGTAATAGAC
O9	SavPNGblunt	AAGCACACCGCCGAGGCCACGCCGT
O10	SavPNG-R2	AAAGGATCCATTAACAGTCGCTGCGGTCACGCGTCAA
O11	Sav_GWN_fwd	CACCAAGCACACCGCCGAGGCCACGCCGT
O12	Sav_GWN_rev	TCAGCAGTCGCTGCGGTCACGCGTCAAC
O13	Sso_pENTR_NtermHis_fwd	CACCCAGACTAGCTCTAGTATCTCGCATC
O14	Sso_pENTR_NtermHis_rev	TTATACTATAATTCTAAGGAAATGTATG
O15	PNG-W59Q-Fwd	AAAACTTGTGATGAACAGGATCGTTATGCCAAT
O16	PNG-W59Q-Rev	ATTGGCATAACGATCCTGTTCATCACAAGTTTT
O17	PNG-I82Q-Fwd	ACGAAATAGGACGCTTTCAGACTCCATATTGGGTGG
O18	PNG-I82Q-Rev	CCACCCAATATGGAGTCTGAAAGCGTCCTATTTCGT

[9] The numbers given here will be used in the following to refer to specific oligonucleotides.

O19	PNG-I82R-Fwd	ACGAAATAGGACGCTTTCGTACTCCATATTGGGTGG
O20	PNG-I82R-Rev	CCACCCAATATGGAGTACGAAAGCGTCCTATTTCGT
O21	PNG-W207Q-Fwd	GAGGTTGTGCAGAACAGTGCTTCAGAACACA
O22	PNG-W207Q-Rev	TGTGTTCTGAAGCACTGTTCTGCACAACCTC
O23	PNG-V257K-Fwd	CCCGGGAATGGCAAAACCAACACGTATAGATGTACTGAATAAT
O24	PNG-V257K-Rev	ATTATTCAGTACATCTATACGTGTTGGTTTTGCCATTCCCGGG
O25	PNG-V257N-Fwd	CCCGGGAATGGCAAATCCAACACGTATAGATGTACTGAATAAT
O26	PNG-V257N-Rev	ATTATTCAGTACATCTATACGTGTTGGATTTGCCATTCCCGGG
O27	AnigerPNGase.internal.rev	GCCTGTAACTCTGATCTCGAA

2.13 Purification of PCR Products from Agarose Gels (Vogelstein & Gillespie, 1979)

For a variety of downstream applications such as restriction nuclease digests (2.14) and cloning, DNA fragments were purified directly from TAE agarose gels (2.19) using the Perfectprep® Gel Cleanup Kit following to the manufacturer's instructions (Eppendorf AG). The only alteration to the protocol provided with the kit was the staining and excision procedure. The gel was not completely stained in ethidium bromide in order to minimise contamination with ethidium bromide and to avoid exposure of the DNA to UV light which could potentially lead to undesirable mutations. Instead, only the two lanes containing DNA standards located adjacent to the sample lane were cut off including a thin slice of the sample lane. These slices were stained, exposed to UV light and the position of the desired DNA fragment marked with toothpicks. The stained, marked gel slices were then aligned with the unstained part of the gel and the DNA fragment excised by cutting along the markings using a scalpel.

2.14 DNA Hydrolysis with Restriction Endonuclease

Restriction endonucleases are enzymes that recognise and cut unique palindromic DNA sequences leaving either complementary overhangs ('sticky ends') or blunt ended DNA fragments.

Restriction endonuclease digests were performed to generate cohesive overhangs or blunt ends on both plasmid DNA and DNA fragments for subsequent ligations (2.15). Restriction endonucleases were also used for the analytical digestion of plasmids or DNA fragments to confirm the presence of inserts. This restriction site mapping procedure allowed the fast analysis of newly generated recombinant plasmids before DNA sequencing (2.21).

A typical restriction digest contained the following components:

DNA	1.0 µg
10× Reaction buffer	2.5 µL
Restriction enzyme	1.0 U
H_2O_{pure}	up to 25.0 µL

The reactions were generally incubated for 3-4 h at the appropriate temperature. Buffers used were specific for each restriction enzyme and were supplied with the enzyme. One unit of enzyme activity is defined as the concentration of enzyme that completely cleaves 1 µg λ-DNA in 1 h at the enzyme specific incubation temperature, usually 37°C.

Following incubation, sample buffer for agarose gel electrophoresis was added as restriction products for subsequent cloning were routinely purified from agarose gels. Analytical restriction digests were also analysed using AGE (2.19).

2.15 Ligation of DNA-fragments

DNA ligation is the formation of a phosphodiester bond between 3′-hydroxyl- and 5′-phosphate ends of double stranded DNA. This bond formation can occur between ends of the same DNA fragment (e.g. religation of linearised plasmid DNA) or the ends of two separate DNA fragments, which have complementary ends (e.g. linearised plasmid and insert in restriction cloning protocols). The ligation reaction is catalysed by the ATP-dependent enzyme DNA ligase, most commonly used in molecular biology is T4 DNA ligase.

The ratio used between linearised plasmid and insert was 1:3 for sticky end ligations and 1:5 for blunt end ligations. Standard 10 µL-ligation reactions were performed using up to 1 µg total DNA, 1 µL 10× ligation buffer and 1-3 U T4 DNA ligase (Roche Applied Science). The reactions were either incubated overnight at 4°C or for 1 h at room temperature. The ligation mixture was then transformed (2.17) into a standard cloning *E. coli* strain (2.7, **Table 2.2**).

2.16 Preparation of Chemically Competent Cells of E. coli (Hanahan, 1983)

The *E. coli* strains were grown at 37°C in 50 mL LB broth (2.5.1) in presence of antibiotics if required until the culture reached an OD_{600} ~0.5. After 15 minutes incubation on ice the cells were harvested (20 min, 2,600 *g*, 4°C) and the cell pellet resuspended in 18 mL RF1 buffer. After 30 minutes incubation on ice, the cells were pelleted again (as before) and resuspended in 4 mL RF2 buffer. This cell suspension was then dispensed into sterile microcentrifuge tubes in 50 µL aliquots and immediately snap-frozen in liquid nitrogen.

Competent cells prepared with this method could be stored for several months at -80°C.

RF1 Buffer:

RbCl	100.0 mM
$MnCl_2$	50.0 mM
Potassium acetate	30.0 mM
$CaCl_2$	10.0 mM

The pH was adjusted to 5.8 using acetic acid.

RF2 Buffer:

RbCl	10.0 mM
MOPS	10.0 mM
$CaCl_2$	75.0 mM
Glycerol	15.0% (v/v)

The pH was adjusted to 5.8 using NaOH.

Both solutions were sterilised by passing through a 0.22 μm filter into previously autoclaved bottles.

2.17 Transformation of Plasmid-DNA into E. coli (Inoue et al., 1990)

An aliquot (50 μL) of chemically competent *E. coli* cells (2.16) was mixed with 50 to 250 ng of plasmid DNA and incubated on ice for 30 minutes. This incubation step allows the plasmid DNA to adsorb to the cell surface. The passive take up of the DNA into the cells was induced by subjecting the cells to a heat shock at 42°C for 90 sec. After the heat shock, the cells were immediately placed on ice for 5 minutes, followed by the addition of 500 μL of LB broth or S.O.C. medium to each tube and incubation at 37°C for 30-60 minutes, depending on the antibiotic(s) encoded on the transformed plasmid and strain specific antibiotic resistances. After this incubation period transformants were selected by plating them on LB agar or growing them in LB broth containing the appropriate antibiotic selection for 12-15 h.

S.O.C. Medium:

Tryptone	2.0% (w/v)
Yeast extract	0.5% (w/v)
NaCl	10.0 mM
KCl	2.5 mM
$MgCl_2$	10.0 mM
$MgSO_4$	10.0 mM
Glucose	20.0 mM

2.18 Small Scale Isolation of Plasmid DNA

Plasmid DNA was isolated from *E. coli* cells for several purposes, including DNA sequencing (2.21), restriction endonuclease digest (2.14) and storage. Plasmids were isolated using the High Pure Plasmid Isolation Kit (Roche Applied Biosciences) according to the manufacturers' instructions. The method employed by this kit is based on the alkaline lysis method described by Birnboim and Doly, 1979 (Birnboim & Doly, 1979).

2.19 Agarose Gel Electrophoresis (AGE)

DNA fragments were separated according to their size by agarose gel electrophoresis (AGE) in submerged horizontal gels using the Sub-Cell® System (Bio-Rad).

Routinely, 1% Agarose (w/v) gels were used to analyse DNA samples. The agarose was dissolved in TAE buffer by heating until the solution was homogenous. The gel was prepared by pouring the liquid agarose into a horizontal tray and insertion of a comb to form the sample wells. After the agarose solidified, the gel was submerged in TAE running buffer. Samples were mixed with an appropriate volume of loading buffer and transferred into the sample wells. The electrophoresis was performed using a constant voltage of 100 V for approximately 45 minutes. For subsequent detection the gel was stained in a solution containing 2 µg/mL ethidium bromide in H_2O_{pure}. The

DNA-ethidium bromide complex was visualised by exposure to UV light (254 nm) using the GelDoc gel imaging system (Bio-Rad).

TAE Buffer:

Tris	40.0 mM
Acetic acid	20.0 mM
EDTA	2.0 mM
pH 8.0	

6x Sample Buffer:

Tris	60.0 mM
EDTA	60.0 mM
Glycerol	60.0% (v/v)
Orange G	0.2% (w/v)
Xylene Cyanol FF	0.05% (w/v)

2.20 Quantification of Nucleic Acids

The quantity and purity of DNA and RNA preparations was determined using the NanoDrop™ 1000 Spectrophotometer (Thermo Fisher Scientific). This spectrophotometer measures the absorption of a 1 µL nucleic acid sample at 260 nm and 280 nm against an appropriate blank measurement.

The concentration of a DNA sample is calculated as follows: A260 = 1 = 50 µg/mL. The purity of the DNA sample is shown by the A260/A280 ratio, where a ratio of between 1.7 and 2 is indicative of a good quality DNA preparation.

RNA concentrations are calculated with A260 = 1 = 40 µg/mL. The A260/A280 ratio for good quality RNA should be between 1.8 and 2.

Alternatively, DNA concentrations were estimated using agarose gel electrophoresis (2.19). Dilutions of a DNA preparation were separated on an agarose gel and the DNA stained with ethidium bromide. The dilution showing the minimal, just visible fluorescence was considered to contain approximately 2 ng DNA.

2.21 DNA Sequence Analysis

Fragments of DNA generated by PCR (2.10) or PCR generated DNA fragments ligated into plasmid DNA were analysed by DNA sequencing to ensure the absence of any unwanted PCR derived mutations.

DNA sequencing was carried out by the Genome Service provided by The Allan Wilson Centre for Molecular Ecology and Evolution. For details on the equipment used and the methods and materials employed by this service refer to http://awcmee.massey.ac.nz/genome-sequencing.htm.

The samples were prepared according to the instructions given by the sequencing service. Briefly, the concentration of the DNA stock was determined (2.20) and the samples premixed using the following concentrations:

Template:	300-450 ng/ 15 µL for plasmids 2 ng/100 bp/ 15 µL for PCR products
Primer:	3.2 pmol/ 15 µL

The results provided were then analysed using the program 'Sequencing Scanner Software V1.0', provided free of charge from Applied Biosystems Inc., Foster City, USA.

2.22 Determination of Protein Concentration

Protein concentrations were measured using either the Bradford protein assay or UV absorption at 280 nm.

2.22.1 Bradford Protein Assay (Bradford, 1976)

The Bradford method for the determination of protein concentrations is based on the binding of the dye Coomassie Blue G-250 to proteins. Coomassie Blue G-250, generally blue in colour, turns brown-red when dissolved in strong

acids. However, upon binding to a positively charged protein the blue colour is restored due to a shift in the pK_a of the bound dye. Therefore, the intensity of the blue colour is dependent on the protein concentration and can be measured at 595 nm. The concentration of unknown protein solutions was determined using a standard curve. This standard curve was drawn from the absorption of protein standards at 595 nm in the range of 0.1-2 mg/mL of BSA. 100 µL of protein solution was mixed with 1 mL of Bradford reagent and incubated at room temperature for 10 minutes before measuring the absorbance at 595 nm.

Bradford Reagent (5×):

Coomassie brilliant blue G-250	0.1 g
Ethanol (95%)	50.0 mL
Phosphoric acid (concentrated)	100.0 mL
H_2O_{pure}	up to 200.0 mL

2.22.2 Protein Concentration Determination using UV Absorption

This method is based on the absorbance of the aromatic amino acid side chains of tyrosine and tryptophan residues and the presence of disulphide bonds. Since the amount of these two amino acids varies enormously between proteins, the extinction coefficient ($E_{280}^{1\,mg/mL}$) varies accordingly. Therefore, absorption at 280 nm can only give an estimate of the exact protein concentration. However, if the extinction coefficient for a pure protein is known, then this method provides very accurate measurements.

The extinction coefficient for PNGase F has been determined previously as $E_{280}^{1\%} = 18$ (Loo, 2000; Tarentino *et al.*, 1990). All measurements made for samples containing PNGase F were corrected with this figure using the following equation:

$$\text{PNGase F (mg/mL)} = \frac{\text{A280}}{1.8} \times \text{Dilution Factor}$$

2.23 SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE; (Laemmli, 1970))

Polyacrylamide gel electrophoresis was performed to separate proteins on the basis of their molecular weight as described by Laemmli using Mini-PROTEAN® II cells (Bio-Rad). The Laemmli method includes SDS in the buffer system and proteins are denatured by boiling them in buffer containing SDS and a reducing agent. This treatment leads to proteins with a uniform charge-to-mass ratio proportional to their molecular weight. Hence, these proteins separate according to their molecular weight.

Generally aliquots of protein samples were mixed with the appropriate amount of loading buffer, boiled for 5 minutes in a water bath and aliquots of these mixtures loaded into the wells of the polyacrylamide gel. The electrophoresis was performed at a constant voltage of 200 V for approximately 40 minutes. By this time the loading dye front had usually reached the bottom of the resolving gel. The gel was then carefully removed from the set up, fixed and stained in staining solution for approximately 20 minutes. The gel background was destained in destain solution until clear.

The molecular weight of the proteins was estimated using a protein standard (Bio-Rad; Fermentas) containing proteins of defined molecular weight, which was loaded onto the same gel.

The following Tables **Table 2.7** and **Table 2.8** show volumes of given stock solutions used for the preparation of SDS-polyacrylamide separation gels of different acrylamide percentages and the 4% stacking gel. The size of the gels was 7.5 × 10 cm, the thickness 0.75 mm. Unless stated otherwise, 12% separation gels were used. After pouring the separation gel solution into the gel chamber the solution was topped with an overlay of butanol to obtain a straight edge between separation and stacking gel. After 30 minutes to 1 h the butanol was removed and replaced by the stacking gel solution. Wells were formed by inserting a 10- or 15-tooth comb into the stacking gel solution before significant polymerisation had occurred. After polymerisation the gels were wrapped in a damp paper towel and plastic wrap and stored at 4°C until use.

Table 2.7: Preparation of the separating gel solutions for SDS-PAGE.

Component	8%	10%	12%	15%
H_2O_{pure}	5.3 mL	5.25 mL	4.5 mL	3.75 mL
1.5 M Tris-HCl, pH 8.8	2.5 mL	2.5 mL	2.5 mL	2.5 mL
10% (w/v) SDS	0.1 mL	0.1 mL	0.1 mL	0.1 mL
40% Acrylamide:Bis[10]	2.0 mL	2.5 mL	3.0 mL	3.75 mL
Mix solution and degas for ≥ 15 min				
10% ammonium persulfate (w/v)	0.1 mL	0.1 mL	0.1 mL	0.1 mL
TEMED	5.0 µL	5.0 µL	5.0 µL	5.0 µL
Total volume	10.0 mL	10.0 mL	10.0 mL	10.0 mL

Table 2.8: Stacking gel preparation for SDS-PAGE.

Component	4%
H_2O_{pure}	5.3 mL
0.5 M Tris-HCl, pH 6.8	2.5 mL
10% (w/v) SDS	0.1 mL
40% Acrylamide:Bis[10]	1.0 mL
Mix solution and degas for ≥ 15 min	
10% ammonium persulfate (w/v)	0.1 mL
TEMED	10.0 µL
Total volume	10.0 mL

5× Electrode (Running) Buffer:

Tris	15.0 g
Glycine	72.0 g
SDS	5.0 g
H_2O_{pure}	up to 1.0 L

[10] Acrylamide:Bis ready-to-use solution 40% (19:1)

10× SDS-Loading Buffer:

0.5 M Tris-HCl, pH 6.8	2.0 mL
Glycerol	2.0 mL
10% (w/v) SDS	3.2 mL
DTT	0.77 g
0.1% (w/v) BPB	0.8 mL

Coomassie Brilliant Blue R 250 or G 250 Staining Solution:

Methanol	40.0%
Acetic Acid	10.0%
Coomassie Brilliant Blue R/G250	0.1% (w/v)

Destain solution was prepared in the same way leaving out the Coomassie Brilliant Blue R 250 or G 250 dye.

2.24 Western Blot

2.24.1 Electrophoretic Transfer of Proteins on Membranes (Matsudaira, 1987; Towbin et al., 1979)

For the electrophoretic transfer of proteins from polyacrylamide gels onto nitrocellulose or polyvinylidene difluoride (PVDF) membranes the semi-dry blotting method was employed using the Mini Trans-Blot® Electrophoretic Transfer Cell with appropriate power supply (Bio-Rad).

The polyacrylamide gel electrophoresis was performed as described above (2.23). After electrophoresis the gel was equilibrated for 20 minutes in transfer buffer. Fibre pads, membrane and Whatman filter paper were cut to fit the size of the gel and were soaked together with the fibre pads in transfer buffer for 10 minutes. The blotting sandwich was assembled in the gel holder cassette in transfer buffer in the following order: 1 fibre pad, 3 layers Whatman paper,

polyacrylamide gel, membrane, 3 layers of Whatman paper and another fibre pad. A glass tube was carefully rolled over the sandwich to ensure no air bubbles were trapped between the layers, which would interfere with the transfer. The gel holder cassette was closed and inserted into the electrode module. The electrode module holding the gel cassette was then placed in the buffer tank together with the Bio-Ice cooling unit, which had been filled with water and placed at -20°C until needed. The tank was completely filled with transfer buffer and a stirring bar was added to maintain even buffer temperature and ion distribution. The transfer was performed by applying a constant voltage of 80 V. The duration of the transfer was chosen depending on the size of the target protein and was usually 1.5-3.5 h. After electrophoretic transfer the blotting sandwich was carefully disassembled, the membrane subjected to immunoblot procedures (2.24.2) and the gel stained with Coomassie to evaluate the transfer efficiency.

Transfer Buffer:

Tris	25.0 mM
Glycine	192.0 mM
SDS	0.1% (w/v)
Methanol	10.0% (v/v)

2.24.2 Immunodetection of Immobilised Proteins on Membranes

The membrane carrying the transferred proteins was placed in 10 mL of blocking solution to block nonspecific binding sites and incubated for 1 hour at room temperature with agitation, followed by overnight incubation at 4°C and 1 h incubation at room temperature. The membrane was then washed three times with PBS-Tween and subsequently incubated with the primary antibody solution for 1 hour at room temperature with agitation. This was followed by three washes for 5 minutes each with PBS-Tween, before the blot was placed in the secondary antibody solution and incubated for 1 hour with agitation at room temperature. This incubation was followed again by 3 wash steps for 5 minutes each with PBS-Tween. Then, the chemiluminescent blotting substrate (Roche)

was applied onto the membrane and bands were visualised as described in 2.24.3.

PBS-Tween:

Sodium phosphate, pH 7.2	10.0 mM
NaCl	0.9% (w/v)
Tween-20	0.1%

Blocking solution:

PBS-Tween with 3% (w/v) BSA

Primary antibody solution:

Antibody dilution in PBS-Tween according to manufacturer's instructions

Secondary antibody solution:

Antibody dilution 1:2,500 in PBS-Tween (Anti-mouse IgG conjugated to horse radish peroxidise)

2.24.3 Chemiluminescent Visualisation of Immobilised Proteins

The visualisation of proteins that have been immobilised and immunolabelled on membranes was performed using the BM Chemiluminescence Blotting Substrate (POD; Roche). The basis of this detection system is the oxidative reaction catalysed by horseradish peroxidise (POD or HRP), which is bound to the secondary antibody (2.24.2). This enzyme catalyzes the oxidation of luminol in presence of hydrogen peroxide, resulting in an activated intermediate reaction product, which decays to the ground state by emitting light. The light emission is enhanced by 4-iodophenol, which acts as a radical transmitter between the oxygen radical formed in the reaction and luminol. The detection solution was prepared by mixing 10 mL of solution A

(luminescence substrate solution) with 100 µL of solution B (starting solution) (ratio of 1:100). After the solution reached room temperature (15 - 25 °C) the blot was covered completely with the substrate solution and incubated for 1 minute. Excess substrate was drained off and the blot placed on a transparent plastic sheet and covered with another sheet, taking care that no air bubbles were trapped in between. The bands of the protein standard were marked with a phosphorescent marker, the blot placed in a dark box (Intelligent Dark Box II, LAS-1000, Fujifilm) and multiple exposures of 10 seconds were taken. Images of each interval were recorded for up to 5 minutes and the image (or images) with the best exposure was (were) saved.

2.25 In-Gel Tryptic Digest for Protein ID by Mass Spectrometry

The method used here is based on the protocol described by (Shevchenko *et al.*, 1996).

Coomassie stained protein bands were excised from PAGE gels and washed for 1 h in 100 mM NH_4HCO_3 in a microcentrifuge tube. The solution was replaced by 25-35 µL acetonitrile and the mixture incubated for 10 minutes at room temperature to dehydrate and shrink the gel pieces. The acetonitrile was removed and the gel pieces dried using a speed-vacuum centrifuge for 10 minutes. Gel pieces were rehydrated in 150 µL 10 mM DTT in 100 mM NH_4HCO_3 and incubated for 1 h at 56°C. After cooling to room temperature the DTT solution was replaced with 150 µL 55 mM iodacetamide in 100 mM NH_4HCO_3 followed by 45 minutes incubation at room temperature in the dark. The solution was replaced with 150 µL 100 mM NH_4HCO_3 and the gel pieces incubated 10 minutes at room temperature. The gel pieces were dehydrated again by 10 minutes incubation in 150 µL acetonitrile, which was removed in a speed vacuum centrifuge until gel pieces were dry. Gel pieces were rehydrated in 25-35 µL digestion buffer (12.5 ng trypsin in 50 mM NH_4HCO_3) for 45 minutes on ice. The digestion buffer was replaced by 10 µL 50 mM NH_4HCO_3 and the digest incubated overnight at 37°C. The gel pieces were spun down, the supernatant removed and saved in a microcentrifuge tube, and 20 µL 20 mM

NH_4HCO_3 added to the gel pieces. After 10 minutes incubation the supernatant was transferred into the same tube as before. 25 µL of 5% formic acid, 50% acetonitrile were added to the gel pieces followed by 20 minutes incubation at room temperature. Supernatant was removed and saved again in the same tube as before. The formic acid extraction was repeated twice more and the collected supernatants completely dried in a speed vacuum centrifuge. Before MALDI-TOF MS analysis, the dried tube content was dissolved in 5 µL 0.5% TFA and the peptides purified using a ZipTip® pipette tip (Millipore).

The matrix used for MALDI-TOF MS was prepared following the Rapid Evaporation Method (Shevchenko *et al.*, 1996). The matrix solution was prepared by dissolving 2.5 mg of nitrocellulose and 10 mg α-cyano-4-hydroxy-trans-cinnamic acid (HCCA, Sigma) in 0.25 mL of acetone followed by the addition of 0.25 mL isopropanol. 1 µL matrix solution was applied to the sample target plate and left to dry. 1 µL of the peptide sample was pipetted onto the matrix and left to dry at room temperature. The sample target plate was inserted into a Micromass® M@LDI mass spectrometer with a time-of-flight analyser (Waters®, USA) and analysed in reflection mode.

The result was then analysed using Mascot (www.matrixscience.com).

2.26 Determination of Deglycosylating Activity

For the determination of Peptide:*N*-glycanase activity, two assays were used.

2.26.1 Gelshift Assay

In order to test different possible PNGase substrates the Gelshift assay was employed. In this assay different glycoproteins, usually ovalbumin, RNase B and α-1-acid glycoprotein, were incubated with protein preparations and analysed on SDS-PAGE (2.23). As PNGases cleave the glycan moiety of glycoproteins (given they are suitable substrates), PNGase activity can be detected on SDS-gels as a mobility shift of the glycoprotein bands. In case of PNGase activity, the mobility of the glycoprotein increases as compared to the native form due to the

removal of the glycan. RNase B (17. kDa) is a high-mannose glycoprotein, carrying a single *N*-glycan. α-1-acid glycoprotein is a 183 amino acid protein with five highly sialylated complex-type *N*-linked glycans that represent 45% of its 43 kDa mass. Hen egg ovalbumin, the major protein in egg white, is a 386 amino acid containing protein with a molecular weight of 45 kDa. A single, heterogeneous high-mannose carbohydrate side chain is linked to Asn293.

Typically, 40 µL of substrate solution (1 mg/mL) were incubated overnight with the enzyme sample (various concentrations and purification states) at the appropriate temperature. Each substrate was tested in its native and denatured form as PNGases usually prefer or exclusively act on denatured proteins. The glycoproteins were denatured by adding 0.05% (w/v) SDS followed by 5 minutes boiling in a water bath. The total volume of the assay mixture was 100 µL. The reaction result was analysed by SDS-PAGE (15% acrylamide; 2.23).

2.26.2 Reverse Phase (RP)-HPLC Based PNGase Activity Assay

PNGase activity was measured using a discontinuous assay based on the deglycosylation of a hen egg ovalbumin-derived 11-mer glycopeptide (**Figure 2.1**; (Norris *et al.*, 1994a)). This glycopeptide is a well established substrate for the measurement of deglycosylation activity by various PNGases. It carries an *N*-linked complex, biantennary oligosaccharide with nine uniformly distributed hybrid and high-mannose glycoforms.

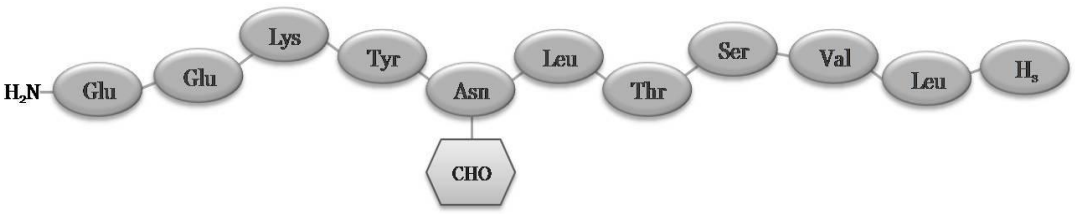


Figure 2.1: The ovalbumin glycopeptide.
'CHO' represents the *N*-linked glycan moiety. 'H_s' indicates the homoserine lactone at the C-terminus, which is results from the conversion of a methionine residue during CNBr-digest of ovalbumin.

The detection of deglycosylation activity is based on the difference in hydrophobicity exhibited by substrate and product. The product shows a higher hydrophobicity than the substrate due to the loss of the hydrophilic glycan moiety. The difference in hydrophobicity can be detected as the difference in retention time on a C18-HPLC column, where hydrophilic compounds elute earlier than hydrophobic ones.

In a typical assay, 5 µL of enzyme at an appropriate, defined concentration was incubated with 45 µL of substrate. The substrate concentration varied depending on the purpose of the assay. Both enzyme and substrate dilutions were present in the following buffer: 5 mM EPPS (pH 8.5), 1x Roche Mini Complete Protease Inhibitor, 1 mM EDTA. After an appropriate incubation time the reaction was stopped by boiling for 3 minutes in a water bath. Before loading onto a C18-HPLC column (C18, Jupiter® 5 µm 300 Å, 250×4.6 mm) 50 µL of reaction buffer were added and the reaction mixture was centrifuged for 30 minutes at 14,000 *g*. The reaction products were separated using a 15 minute gradient from 80% solvent A (0.1% TFA in pure water), 20% solvent B (0.08% TFA in acetonitrile) to 60% solvent A (0.1% TFA in pure water), 40% solvent B (0.08% TFA in acetonitrile) at a flow rate of 1 mL/min. Products in the eluent were detected at 214 nm. The data were analysed using Chromeleon™ Client software.

Kinetic studies of PNGase F and its mutants (Chapter 8) required a more sensitive detection method as the substrate concentrations used in these kinetic studies were too low to be accurately detected at 214 nm. Therefore, the 11-mer ovalbumin glycopeptide was fluorescently labelled with fluorescein isothiocyanate (FITC). The labelling and assay procedures are described in Chapter 8, section 8.2.2.

Section I

Putative PNGases

Chapter 3: Identification and Bioinformatical Analyses of Putative PNGases

Chapter 4: Gene Expression Analyses

Chapter 5: Cloning and Expression of Genes Encoding Putative PNGases

I believe there is no philosophical high-road in science, with epistemological signposts. No, we are in a jungle and find our way by trial and error, building our road behind us as we proceed.

Max Born (1882-1970), German Physicist.

Chapter 3

Identification and Bioinformatical Analyses of Putative PNGases

3 Identification and Bioinformatical Analyses of Putative PNGases

3.1 Introduction

This chapter describes the identification of amino acid sequences homologous to PNGase F or PNGase A and *At*. The aim was to identify new and reasonable target sequences for subsequent cloning, expression, purification and characterisation, including crystallisation. Furthermore, a proposal for a classification of PNGases is presented based on the differences between members of the EC class 3.5.1.52 (PNGases).

PNGase F is well characterised, mainly in terms of substrate specificity and three-dimensional structure (1.2.1.1), but for a long time it remained the only protein in this group of PNGases as no homologues were present in databases. Recently this changed with the addition of a homologous sequence predicted from the sequenced genome of the bacterium *D. radiodurans*. To identify additional homologues BLAST searches were performed on a regular basis and identified a small number of new sequences with some similarities to PNGase F. A selection of these sequences was analysed using bioinformatic programs. At the start of this project the *D. radiodurans* homologue was the only one and therefore the target for this PNGase group.

PNGase A and PNGase *At* have been characterised to some degree as mentioned in section 1.2.1.2, but no three-dimensional structure has yet been published and nothing is known about the catalytic mechanism employed by this group of PNGases. These two facts make these PNGases an interesting target as they appear to be entirely different from the other two groups in terms of amino acid sequence, but still catalyse the same reaction. There are considerable numbers of PNGase A and PNGase *At* homologues present in databases. To analyse the phylogenetic distribution of these homologues and identify interesting target sequences for this project, BLASTp searches were performed and the chosen targets were analysed using bioinformatic programs.

Type III PNGases (1.2.1.3; **Table 1.2**) are readily identifiable by sequence similarity and well characterised structurally, so they were not analysed further in this project.

3.2 Methods

3.2.1 Identification of PNGase F-type proteins

Protein sequences homologous to PNGase F were identified using the protein-protein-BLAST BLASTp algorithm (Altschul *et al.*, 1990). The search was performed against the non-redundant protein database, which contains all non-redundant GenBank CDS translations, RefSeq proteins (RefSeq protein sequences from NCBI's Reference Sequence Project), PDB sequences (Sequences derived from the 3-dimensional structure from Brookhaven Protein Data Bank), SwissProt sequences (SWISS-PROT protein sequence database) and sequences from PIR (Protein Information Resource) and PRF (Protein Research Foundation). In addition, BLASTp queries are automatically analysed for the presence of conserved domains by searching the **C**onserved **D**omain **D**atabase (CDD) (Marchler-Bauer *et al.*, 2009). CDD includes NCBI-curated domains as well as data mainly resourced from SMART (**S**imple **M**odular **A**rchitecture **R**esearch **T**ool), Pfam (**P**rotein **Fam**ilies Database), COGs (**C**lusters of **O**rthologous **G**roups of proteins), PRK (**Pr**otein **K**©lusters) and TIGRPFAM (**T**he **I**nstitute for **G**enomic **R**esearch's database of **p**rotein **fam**ilies).

Homologous sequences obtained from BLASTp were selected for the generation of a multiple protein sequence alignment using the following criteria: *(i)* E-value below zero; *(ii)* defined sourced organism (i.e. sequences derived from unidentified organisms were excluded); *(iii)* the presence of either conserved PNGase F domain, Peptide-*N*-glycosidase F, C-terminal (pfam09113) domain or Peptide-*N*-glycosidase F, N-terminal (pfam09112) domain. The multiple protein sequence alignment was generated using CLUSTAL W2

(default settings) at the EMBL-EBI server (Larkin *et al.*, 2007). Prior to multiple sequence alignment, protein sequences containing additional domains were truncated to the conserved PNGase F-like domain.

3.2.2 Identification of PNGase A and PNGase At-type proteins

Protein sequences homologous to PNGase A and PNGase *At* were identified with the protein-protein-blast BLASTp using the same search parameters as described for PNGase F (3.2.1). BLASTp searches were performed for both PNGase A and PNGase *At*.

For the generation of a comprehensive multiple protein sequence alignment, sequences with E-values lower than that of the other characterised homologue were chosen, i.e. if the BLASTp search was performed with the PNGase A protein sequence as query, sequences that showed a lower E-value than PNGase *At* were selected for the multiple alignment and vice versa. This selection procedure was used based on the assumption that sequences 'between' the only two proven examples of this PNGase type are most likely to be actual PNGases themselves and therefore should show a sequence pattern that might permit the identification of essential residues common to all homologues. Identical plant paralogues were identified and only one of the identical sequences was included in the multiple sequence alignment. In the case of sequences from different strains of the same organism only one sequence was included in the alignment. The consensus sequence of this extensive alignment (Appendix 1, **Figure 10.1**) was then compared with the consensus sequence obtained by aligning PNGase A and PNGase *At* with the target protein sequences of this project, *A. niger*, *S. avermitilis* and *S. solfataricus*.

3.3 Results & Discussion

3.3.1 Identification of PNGase F-type proteins

To identify amino acid sequences similar to PNGase F a BLASTp search was performed using the complete PNGase F sequence, including signal peptide, as query (gi:148719; (Tarentino *et al.*, 1990)). Sequences were selected from the initial result as described in 3.2.1. The selected sequences contained regions that were identified as being similar to the C-terminal domain of PNGase F which is present in the conserved domain database (Marchler-Bauer *et al.*, 2009). All of the selected sequences except one (*Flavobacterium bacterium* BBFL7) had additional N-terminal domains not seen for PNGase F. These extensions were removed to obtain sequences mainly containing the PNGase F-like C-terminal domain, and the truncated sequences used as BLASTp queries to determine their similarity to PNGase F. The results of these BLASTp searches are summarised in **Table 3.1**.

Table 3.1: BLASTp results.
For organisms marked with an asterisk amino acid sequences were truncated to the sequence region containing the conserved C-terminal PNGase F-like domain. These sequences were then used as query for a BLASTp search. Values given correspond to their similarity to PNGase F sequence gi:148719 (Tarentino *et al.*, 1990).

	Expect	**Identities**	**Positives**	**Gaps**
***F. bacterium* BBFL7**	5×10^{-102}	174/341 (51%)	234/341 (68%)	6/341 (1%)
***D. radiodurans* R1*[11]**	6×10^{-4}	72/291 (24%)	111/291 (38%)	39/291 (13%)
***D. rerio*[12]**	5×10^{-6}	61/252 (24%)	98/252 (38%)	31/252 (12%)
***S. salar*[13]**	7×10^{-7}	59/250 (23%)	99/250 (39%)	29/250 (11%)
***P. pacifica* SIR-1*[14]**	4×10^{-7}	38/111 (34%)	50/111 (45%)	14/111 (12%)
***C. intestinalis*[15]**	0.001	43/150 (28%)	63/150 (42%)	20/150 (13%)

As indicated by **Table 3.1**, PNGase F appears to be a protein with a very limited number of homologues. Initially it appeared that in terms of

[11] *Deinococcus radiodurans* R1 (Bacteria)
[12] *Danio rerio* (Zebrafish)
[13] *Salmo salar* (Atlantic salmon)
[14] *Plesiocystis pacifica* SIR-1 (Bacteria)
[15] *Ciona intestinalis* (Sea squirt)

phylogenetic distribution homologues were restricted to bacteria, but during the last four years homologues have been identified in other phylogenetic groups as well. Interestingly, all of the organisms containing sequences that match the selection criteria are adapted to marine habitats. The occurrence of a PNGase F-like protein in the fish species *D. rerio* and *S. salar* could indicate that they received the PNGase F-like sequences *via* horizontal gene transfer from marine bacteria. The *Cytophaga-Flavobacterium* group constitute the largest bacterial group in marine water and members are also present in freshwater systems (Glockner *et al.*, 1999). Common infections of salmonid fish (salmon, trout) and zebrafish are caused by *Flavobacterium psychrophilum* and *Flavobacterium columnare*, respectively (Moyer & Hunnicutt, 2007; Nematollahi *et al.*, 2003). This is indirect support for the hypothesis that the PNGase F homologues found in *D. rerio* and salmon may have been obtained *via* horizontal gene transfer from these microorganisms. As these *Flavobacteria* fish-pathogens constitute a significant threat to the fishing industry, their genomes have been sequenced, but contain no PNGase F homologues. *F. bacterium* BBFL7 is also a marine species and could also be a possible source of genetic material for horizontal gene transfer. In general, however, it is almost impossible to pinpoint putative source organism(s) as the vast majority of microorganisms, marine or not, have yet to be identified and genetically characterised.

As mentioned earlier, PNGase F has been characterised in terms of structure, possible active site residues and other residues critical for its activity (substrate binding, environment of catalytic residues). To determine if these residues are conserved between PNGase F and its closest homologues (listed in **Table 3.1**), a multiple sequence alignment was generated using CLUSTAL W2 (3.2.1). **Figure 3.1** shows the multiple amino acid sequence alignment where potentially critical residues are highlighted in terms of their importance, proposed or proven, for enzyme activity or structure. Functionally important residues can be expected to be under stronger selective pressure than those involved in more general functions.

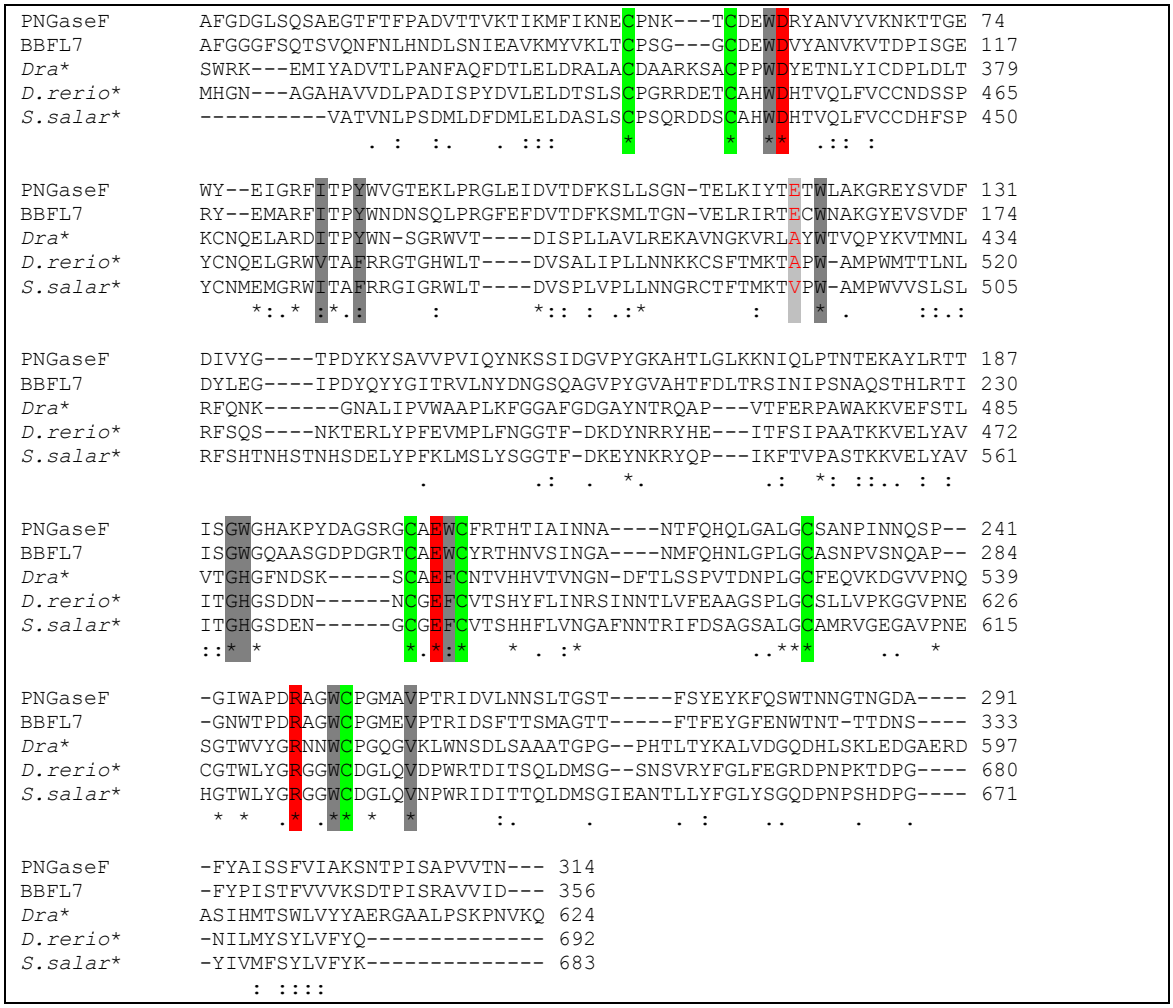


Figure 3.1: CLUSTAL W2 Multiple Sequence Alignment for PNGase F and related sequences.
PNGase F: *F. meningosepticum* (gi: 148719); BBFL7: *Flavobacterium bacterium* BBFL7 (gi: 89891048); *Dra*: *D. radiodurans* R1 (gi: 15807985); *D. rerio* (gi: 33417217); *S. salar* (gi: 223648018). Complete sequences were used for PNGase F (excluding signal sequence) and BBFL7. Sequences marked with an asterisk (*) were truncated to their PNGase F-related domains. Highlighted red are the conserved residues with proposed catalytic function. Highlighted light grey is residue Glu118 (PNGase F numbering) which has been proposed as a catalytic residue, but is not conserved among the homologues. Highlighted green are the cysteine residues that form disulfide bridges in PNGase F. Highlighted dark grey are residues with other proposed functions such as substrate binding or forming the environment necessary for catalytic activity.

The alignment in **Figure 3.1** shows that two of the three residues found to be essential by Kuhn *et al.*, Asp60 and Glu206, are conserved (Kuhn *et al.*, 1995). A third residue suggested to be important for activity, Glu118, is substituted with alanine in both *D. radiodurans* and *D. rerio* and with valine in the *S. salar* homologue. In a structure with *N,N'*-diacetylchitobiose bound in the active site, Glu118 was shown to form hydrogen bonds with *O6* of the second GlcNAc, water molecules 146 and 349 and *OG* of Ser155. Its mutation to

glutamine resulted in a dramatic decrease in relative activity to 0.1% of the wildtype (Kuhn *et al.*, 1995). As Glu118 appears to be mainly involved in substrate binding, through the formation of hydrogen bonds, it is not clear why its substitution with glutamine should result in such a dramatic decrease in enzyme activity. The glutamine residue should still be able to form the hydrogen bonds that are thought to be formed by glutamate. The main change here is a change in charge, which should not affect the active site residues and consequently the enzymatic activity. Interestingly, Kuhn *et al.* have provided no evidence that this mutant protein is correctly folded. Although they reported that they had solved the three dimensional structure of this and other mutants, no mutant structures have been published or deposited in the PDB. Furthermore, *N,N'*-diacetylchitobiose is not a good model substrate. The natural substrate for PNGase F is a glycoprotein or peptide where the glycan is covalently linked to the asparagine side chain by what is essentially an amide bond. There is no guarantee that a free disaccharide will bind into the active site in the same orientation as it would when part of a much larger glycoprotein molecule. Therefore the substitution of Glu118 observed in *D. radiodurans* and the two fish species does not necessarily imply that these proteins cannot be active PNGases. It might however reflect a difference in substrate specificity.

A residue that is likely to play an important role in catalysis is Arg248, which is conserved among the homologues. The hypothesis is that Arg248 forms a hydrogen bond to the carbonyl oxygen of the *N*-glycosidic linkage, making the asparagine-carbonyl carbon more susceptible to a nucleophilic attack by a weak nucleophile such as water, the most likely nucleophile in PNGase F.

Besides these possible catalytic residues, there are several residues that are thought to play an important role in forming a hydrophobic environment around the active site residues and in substrate binding. All except one of these residues are conserved or conservatively substituted.

All cysteine residues that form disulfide bonds in PNGase F (Cys51-Cys56, Cys204-Cys208, Cys231-Cys252) are conserved in all homologues. This indicates that the overall fold may be preserved, and that these disulfide bonds are essential for the stability of the structure.

Overall, the conservation of the residues that have been predicted to play a role in the catalytic mechanism of PNGase F supports the hypothesis that these homologues proteins may also function as PNGases.

3.3.2 Bioinformatical Characterisation of Deinococcus radiodurans Putative PNGase

To obtain further information about the PNGase F-like protein from *D. radiodurans*, which was chosen as the target for heterologous expression and functional and structural characterisation, bioinformatics analyses were carried out. When this work began, the *D. radiodurans* putative PNGase was the only recognised homologous sequence and was therefore the main target for investigation. It already had been annotated as a putative *N*-glycosidase.

The gene (DRA0325) encoding the putative PNGase F-like protein is located on chromosome II, which contains mainly genes involved in amino acid utilisation, cell envelope formation, and transport functions (White *et al.*, 1999). The ORF is 1965 bp long and encodes a 654-amino acid protein.

A summary of the bioinformatics analyses of *D. radiodurans* putative PNGase (*Dra*PNGase) is shown in **Table 3.2**.

A signal peptide and cleavage site location was predicted using the program SignalP 3.0 (Bendtsen *et al.*, 2004). When running SignalP 3.0 the organism group can be specified as Gram-positive or Gram-negative bacteria, or eukaryotes. As *D. radiodurans* is unusual in terms of its cell wall and shows traits of both Gram-negative and Gram-positive bacteria, SignalP 3.0 was run using each of these options. Both analyses predicted the presence of a signal peptide with the most likely cleavage site being located between residues 30 and 31. This indicates that the protein is secreted, which is consistent with the extracellular location of native PNGase F. The numbering of the *Dra*PNGase protein sequence used throughout this thesis will refer to the mature protein unless stated otherwise.

Table 3.2: Summary of the bioinformatics characterisation of putative *D. radiodurans* PNGase.
Number of amino acids, molecular weight and predicted pI are based on the mature protein (i.e. lacking the predicted signal peptide) and were predicted using the program ProtParam (Gasteiger *et al.*, 2003).

Number of amino acids	Molecular weight [kDa]	Predicted pI	Signal peptide	Conserved domains[16]
624	66.8	6.39	Yes (30 aas)	Yes (PA, and C-terminal PNGase)

In addition to the signal peptide, a region of hydrophobic amino acid residues (31 to 50 of the unprocessed *Dra*PNGase) was predicted to be a potential transmembrane segment using the 'DAS-Transmembrane prediction server (Cserzo *et al.*, 1997). This suggests that the extracellular protein could be anchored in the cell membrane.

Two conserved domains were identified within the sequence, a Protease-Associated (PA)-domain (or PA superfamily) and the C-terminal PNGase F-like domain as shown in **Figure 3.2**.

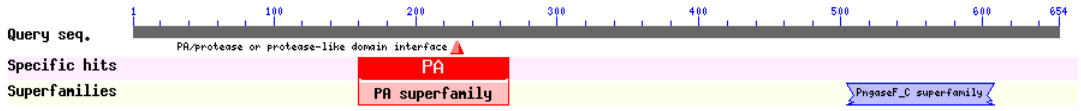


Figure 3.2: Putative conserved domains.
The numbering given in this diagram relates to the complete ORF, including signal peptide. PA domain: residues 169-256. PNGaseF_C domain: residues 469-644.

The PA-domain is characterised as an insert domain that has been found in various proteases. Its function and significance, however, remains unclear. Accordingly, most sequences showing some similarity to the N-terminal half of the protein in the BLASTp search are proteases or protease-domain containing proteins. However, no highly conserved amino acid sequences could be identified.

[16] Conserved domains as identified by CDD (NCBI)

3.3.2.1 Secondary Structure Prediction and Fold-Recognition

A structure prediction for *Dra*PNGase was performed using the remote homology modelling server Phyre (**P**rotein **h**omology/analog**y** **r**ecognition **e**ngine; http://www.sbg.bio.ic.ac.uk/phyre/; (Bennett-Lovsey *et al.*, 2008; Kelley & Sternberg, 2009)).

In Phyre, the secondary structure of a protein is predicted by the three independent secondary structure prediction programs Psi-Pred (McGuffin *et al.*, 2000), SSPro (Pollastri *et al.*, 2002) and JNet (Cole *et al.*, 2008). From these predictions, a consensus prediction is generated based on the confidence values given by each program for each position of the query sequence. The query profile (generated by Phyre) and secondary structure is then scanned against the Phyre fold library, which consists of known protein structures deposited in the Structural Classification of Proteins (SCOP) database (Murzin *et al.*, 1995) and newer Protein Data Bank (PDB) (Berman *et al.*, 2000) depositions. **Table 3.3** summarises the results of the secondary structure prediction for *Dra*PNGase and comparison with PNGase F, the highest scoring fold-recognition match. A graphical view of the secondary structure prediction is shown in Appendix 2.

Table 3.3: Consensus secondary structure prediction result (Phyre) for *Dra*PNGase and comparison with PNGase F.
The C-terminal domain comprises the last 304 residues of *Dra*PNGase, also used for the alignment shown in **Figure 3.1**. The Phyre-confidence level ranges from 1 (low) to 9 (high). The PNGase F secondary structure was experimentally determined by Kuhn *et al.* (1994) and Norris *et al.* (1994b).

	***Dra*PNGase**		**PNGase F**	
	Complete (-signal sequence)	C-terminal domain	predicted	experimental
α-Helix	12.8%	6.2%	2.9%	7.6%
β-Strand	27.7%	33.2%	47.1%	48.1%
Coil	59.5%	60.6%	50%	44.3%
Average Confidence value	6.8	6.8	--	--

The secondary structure prediction for the complete *Dra*PNGase suggests that almost 60% of the protein exists in a coil conformation, connecting the sequence regions that are predicted to adopt a defined α-helical structure or β-strand conformation with the majority being the latter. It was to be expected that the PNGase F-homologous C-terminal part of *Dra*PNGase shows more β-strand than helical conformation, as PNGase F was shown to fold into two domains each with an eight-stranded antiparallel β-jelly roll fold (1.2.1.1; (Kuhn *et al.*, 1994; Norris *et al.*, 1994b)). Based on amino acid sequence, *Dra*PNGase was predicted to contain almost 15% less β-strand therefore more coil content. This discrepancy can probably be explained by the relatively low sequence identity of only 24%, which in the Phyre program is considered very low. However, the fold-recognition scan identified PNGase F as the best match with an E-value of 5.1×10^{-19} if the complete *Dra*PNGase sequence is used as the query and with an E-value of 5.7×10^{-25} if only the C-terminal domain is used as the query. In both cases the 'Estimated Precision' was given at 100%. This score shows the relation between a reported E-value and the empirical frequency of error as determined by the Phyre developers (Kelley & Sternberg, 2009). According to this, an estimated precision score of 100% indicates that 100% of sequences in a test-set that received this score were true homologues. A 3D-model of *Dra*PNGase was then constructed based on the PNGase F structure. This model showed a high degree of similarity of the C-terminal domain of *Dra*PNGase to PNGase F in terms of secondary structure. **Figure 3.3** shows the superposition of the *Dra*PNGase model and PNGase F.

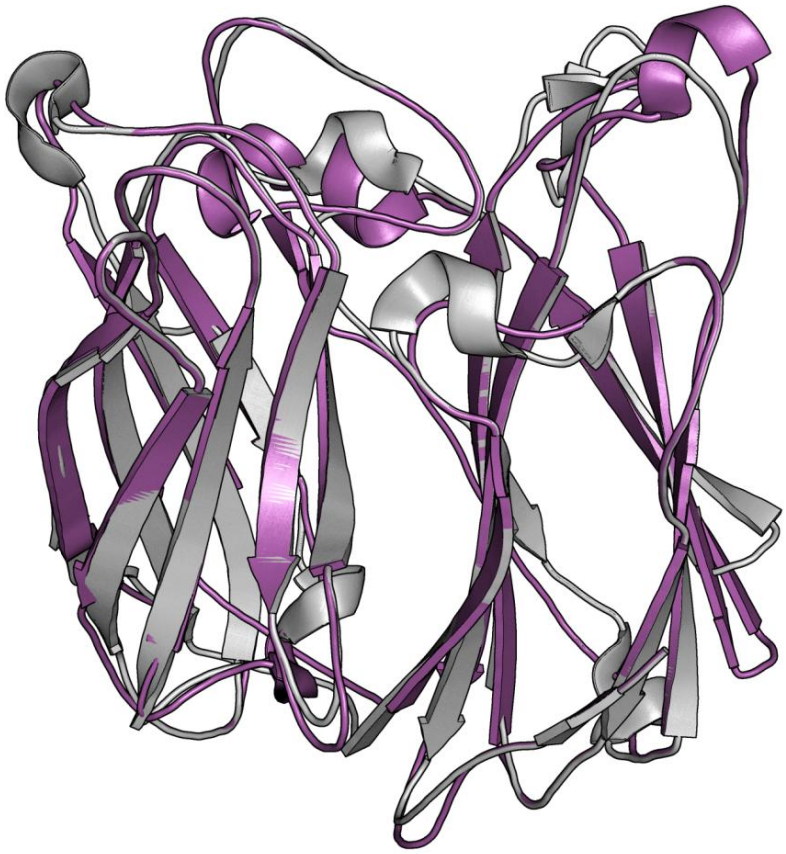


Figure 3.3: Superposition of PNGase F and the *Dra*PNGase-model.
The model was generated by Phyre following the fold-recognition scan. Shown in grey is PNGase F (PDB ID: 1PGS; (Norris *et al.*, 1994b)), and displayed in magenta is the *Dra*PNGase-model. Superposition was performed in PyMOL (DeLano, 2002).

An overlay of the main active site residues that have been identified for PNGase F is shown in **Figure 3.4**.

Overall, the results of the secondary structure prediction and particularly the fold-recognition scan indicate a high likelihood that the C-terminal domain of the putative *Dra*PNGase actually functions as a PNGase.

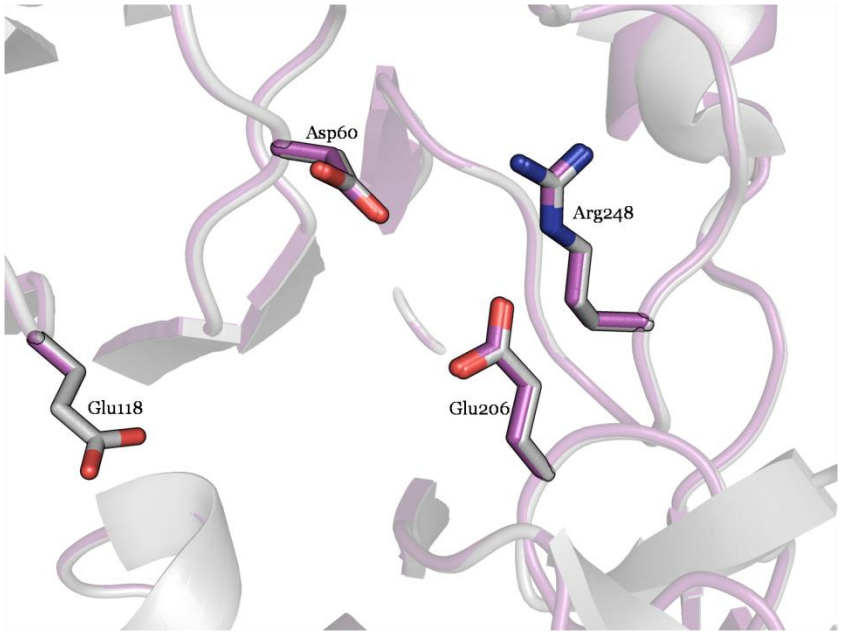


Figure 3.4: Active site superposition.
The active site is located 'on top' of the molecules as displayed in **Figure 3.3**, and is shown here looking down on the 'top' of the protein. Shown in grey is PNGase F and displayed in magenta is the *Dra*PNGase model. The residue numbering refers to PNGase F. Residue Glu118 is substituted with an alanine residue in *Dra*PNGase.

3.3.3 Identification of PNGase A/At-type proteins

Originally the vast majority of homologues in this group of PNGases were identified in plants, which usually have several paralogues, and in fungi. However, at the beginning of this project, two sequences which originated in two different phylogenetic groups were identified, one in bacteria and one in archaea. The bacterial homologue was identified in the actinomycete *Streptomyces avermitilis* MA-4680 and the archaeal one in *Sulfolobus solfataricus* P2. The number of identified homologues has increased in the last few years, although the largest number of homologues is still found in plants and fungi. Initially, the main focus of this project was the characterisation of these two putative PNGases to establish that they possess PNGase activity and to investigate the phylogenetic distribution of type II PNGases. Later, the fungal homologue from *Aspergillus niger* was also included as a type II PNGase target.

To determine the phylogenetic distribution and conserved residues within these PNGases, a BLASTp search was performed and a multiple sequence

alignment generated as described above (3.2.2). The initial BLASTp search using PNGase A as query sequence resulted in 100 hits from 54 different organisms with the following taxonomic distribution: 10 plants, 34 fungi, 7 bacteria and 3 archaea. Most plant species contain between three and seven paralogues, whereas fungi and bacteria usually have only one copy. Interestingly, all but one of the bacterial homologues were found in actinomycetes. Actinomycetes belong to the class *Actinobacteria* and are characterised by their filamentous growth, which resembles the mycelia/hyphae formed by certain fungi. Considering the number of bacterial genomes that have been so far sequenced, it may be that PNGase A-like proteins have a specific role in actinomycetes. Taking into account that most type II PNGases have been identified in fungi, these proteins may have a role in hyphae- or spore-development. The only bacterial homologue that does not belong to this genus is from *Acidobacterium capsulatum*, which is proposed to belong to the new and diverse phylogenetic group *Acidobacteria* (Kishimoto *et al.*, 1991; Ludwig *et al.*, 1997; Quaiser *et al.*, 2003).

Figure 3.5 shows a multiple amino acid sequence alignment of the two characterised type II PNGases and the target proteins chosen for this project from *S. avermitilis*, *S. solfataricus* and *A. niger*. For direct comparison two consensus sequences are given (for the most conserved part of the sequence only) in **Figure 3.5**; on top is the consensus sequence obtained from the alignment of the 50 selected homologues (3.2.2, Appendix 1), while on the bottom the consensus sequence for the alignment presented here is displayed.

```
PNGase At      ------MLVSFGVAFYLVSLLFSPARALLEVFEVYQPVPTGHGS------ 38
A.niger        ------MLVSFSVAFYLVSLLFSPVRAVLEVFEVYQPVPTGHGS------ 38
PNGase A       ---------------EPTPLHDTPPTVFFEVTKPIEVPKTKP-------- 27
Sav            ------MSMLVGVILVASTLLGASPAPAAGKHTAEATPSAEPAPPAEFGT 44
Sso            MRKNIALILLFSILAGIIVVPISSSQTSSSISHPLILGNISVLNSGKIPY 50
                                   :  :.

PNGase At      -------------VGCNEEVLLMDHVFGYSYGEPYVGIYEPPNC---TFD 72
A.niger        -------------VGCNEAILLMDHVFGYSYGEPYVGIYEPPNC---TFD 72
PNGase A       ---------------CSQLILQHDFAYTYGQAPVFANYTPPSDCPSQTFS 62
Sav            DWHDPLTAAPPIGKPATRSCQVTVAEAQFRDFTPYRGTYAPPRGCGDRWS 94
Sso            DPTYYSFEAYQIHPPNVTPVVIRIATNAVFNNSGLTPYIVHVNIPKGNYS 100
                                                          :.

                                                          :
PNGase At      TVRINFTVT-SKGRQYDRLALMYLGDTEVFRTSTAEPTTDGIIWTYIKDM 121
A.niger        TVRLNLTVT-SNGTQYDRLALMYLGDTEVFRTSTAEPTTNGIIWTYIKDM 121
PNGase A       TIVLEWKAT-CRRRQFDRIFGVWLGGVEILRSCTAEPRPNGIVWTVEKDI 111
Sav            RVVLRLDGK-VRGRQFDRLGYLHIGGVEILRTSTPEPSPDGIEWSVEKDV 143
Sso            MEILNVSIAESNGAQYDRPVYIFANGVPIFWGSTQEFFNS----TAETDV 146
                  :.       .  *:**   :  ...  ::  .* *   .    :  .*:
```

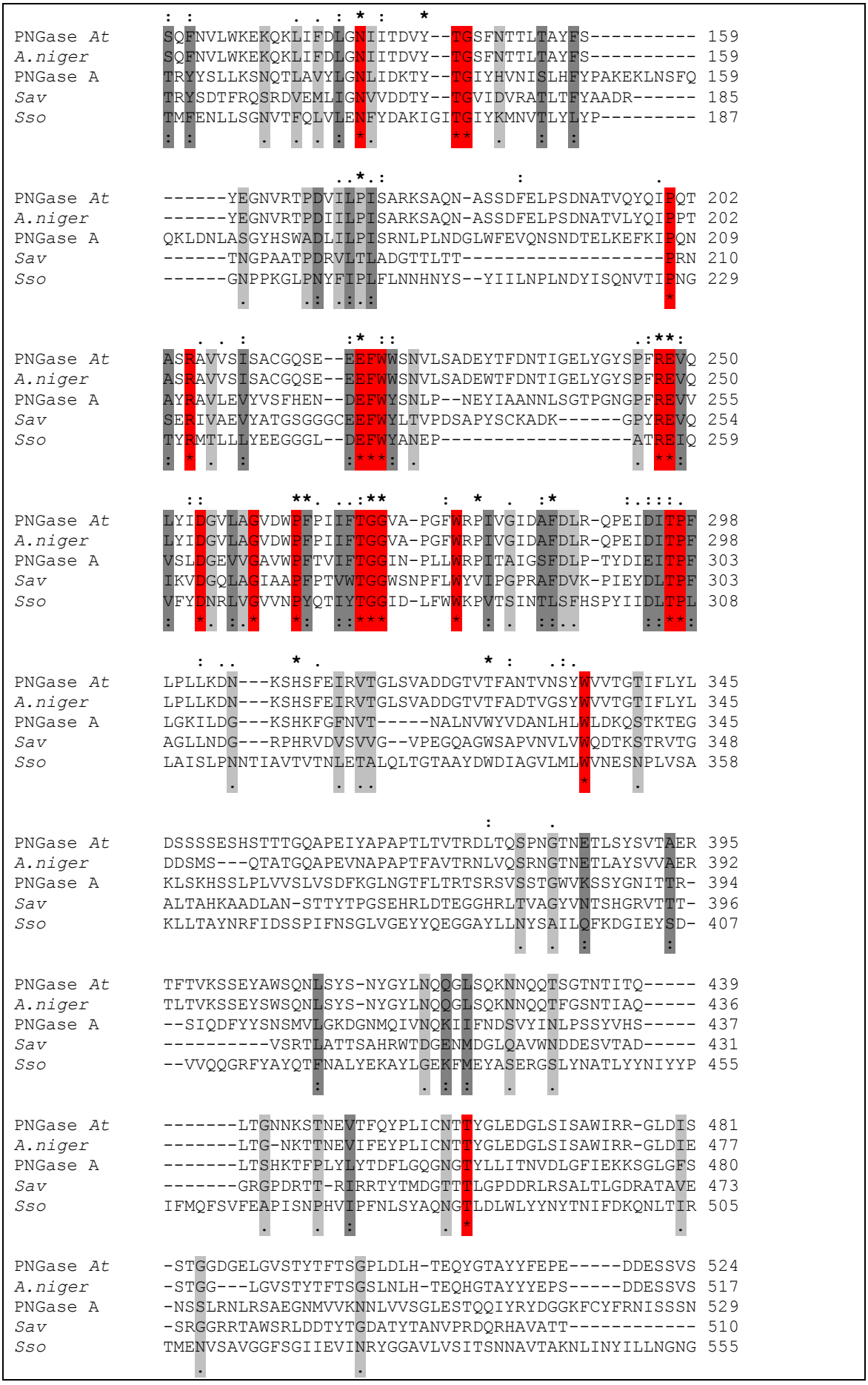

PNGase At SQFNVLWKEKQKLIFDLGNIITDVY--TGSFNTTLTAYFS---------- 159
A.niger SQFNVLWKEKQKLIFDLGNIITDVY--TGSFNTTLTAYFS---------- 159
PNGase A TRYYSLLKSNQTLAVYLGNLIDKTY--TGIYHVNISLHFYPAKEKLNSFQ 159
Sav TRYSDTFRQSRDVEMLIGNVVDDTY--TGVIDVRATLTFYAADR------ 185
Sso TMFENLLSGNVTFQLVLENFYDAKIGITGIYKMNVTLYLYP--------- 187
PNGase At ------YEGNVRTPDVILPISARKSAQN-ASSDFELPSDNATVQYQIPQT 202
A.niger ------YEGNVRTPDIILPISARKSAQN-ASSDFELPSDNATVLYQIPPT 202
PNGase A QKLDNLASGYHSWADLILPISRNLPLNDGLWFEVQNSNDTELKEFKIPQN 209
Sav ------TNGPAATPDRVLTLADGTTLTT----------PRN 210
Sso ------GNPPKGLPNYFIPLFLNNHNYS--YIILNPLNDYISQNVTIPNG 229
PNGase At ASRAVVSISACGQSE--EEFWWSNVLSADEYTFDNTIGELYGYSPFREVQ 250
A.niger ASRAVVSISACGQSE--EEFWWSNVLSADEWTFDNTIGELYGYSPFREVQ 250
PNGase A AYRAVLEVYVSFHEN--DEFWYSNLP--NEYIAANNLSGTPGNGPFREVV 255
Sav SERIVAEVYATGSGGGCEEFWYLTVPDSAPYSCKADK------GPYREVQ 254
Sso TYRMTLLLYEEGGGL--DEFWYANEP----------ATREIQ 259
PNGase At LYIDGVLAGVDWPFPIIFTGGVA-PGFWRPIVGIDAFDLR-QPEIDITPF 298
A.niger LYIDGVLAGVDWPFPIIFTGGVA-PGFWRPIVGIDAFDLR-QPEIDITPF 298
PNGase A VSLDGEVVGAVWPFTVIFTGGIN-PLLWRPITAIGSFDLP-TYDIEITPF 303
Sav IKVDGQLAGIAAPFPTVWTGGWSNPFLWYVIPGPRAFDVK-PIEYDLTPF 303
Sso VFYDNRLVGVVNPYQTIYTGGID-LFWWKPVTSINTLSFHSPYIIDLTPL 308
PNGase At LPLLKDN---KSHSFEIRVTGLSVADDGTVTFANTVNSYWVVTGTIFLYL 345
A.niger LPLLKDN---KSHSFEIRVTGLSVADDGTVTFADTVGSYWVVTGTIFLYL 345
PNGase A LGKILDG---KSHKFGFNVT-----NALNVWYVDANLHLWLDKQSTKTEG 345
Sav AGLLNDG---RPHRVDVSVVG--VPEGQAGWSAPVNVLVWQDTKSTRVTG 348
Sso LAISLPNNTIAVTVTNLETALQLTGTAAYDWDIAGVLMLWVNESNPLVSA 358
PNGase At DSSSSESHSTTTGQAPEIYAPAPTLTVTRDLTQSPNGTNETLSYSVTAER 395
A.niger DDSMS---QTATGQAPEVNAPAPTFAVTRNLVQSRNGTNETLAYSVVAER 392
PNGase A KLSKHSSLPLVVSLVSDFKGLNGTFLTRTSRSVSSTGWVKSSYGNITTR- 394
Sav ALTAHKAADLAN-STTYTPGSEHRLDTEGGHRLTVAGYVNTSHGRVTTT- 396
Sso KLLTAYNRFIDSSPIFNSGLVGEYYQEGGAYLLNYSAILQFKDGIEYSD- 407
PNGase At TFTVKSSEYAWSQNLSYS-NYGYLNQQGLSQKNNQQTSGTNTITQ----- 439
A.niger TLTVKSSEYSWSQNLSYS-NYGYLNQQGLSQKNNQQTFGSNTIAQ----- 436
PNGase A --SIQDFYYSNSMVLGKDGNMQIVNQKIIFNDSVYINLPSSYVHS----- 437
Sav ----------VSRTLATTSAHRWTDGENMDGLQAVWNDDESVTAD----- 431
Sso --VVQQGRFYAYQTFNALYEKAYLGEKFMEYASERGSLYNATLYYNIYYP 455
PNGase At -------LTGNNKSTNEVTFQYPLICNTTYGLEDGLSISAWIRR-GLDIS 481
A.niger -------LTG-NKTTNEVIFEYPLICNTTYGLEDGLSISAWIRR-GLDIE 477
PNGase A -------LTSHKTFPLYLYTDFLGQGNGTYLLITNVDLGFIEKKSGLGFS 480
Sav -------GRGPDRTT-RIRRTYTMDGTTTLGPDDRLRSALTLGDRATAVE 473
Sso IFMQFSVFEAPISNPHVIPFNLSYAQNGTLDLWLYYNYTNIFDKQNLTIR 505
PNGase At -STGGDGELGVSTYTFTSGPLDLH-TEQYGTAYYFEPE-----DDESSVS 524
A.niger -STGG---LGVSTYTFTSGSLNLH-TEQHGTAYYYEPS-----DDESSVS 517
PNGase A -NSSLRNLRSAEGNMVVKNNLVVSGLESTQQIYRYDGGKFCYFRNISSSN 529
Sav -SRGGRRTAWSRLDDTYTGDATYTANVPRDQRHAVATT---------- 510
Sso TMENVSAVGGFSGIIEVINRYGGAVLVSITSNNAVTAKNLINYILLNGNG 555

Figure 3.5: CLUSTAL W2 Multiple Sequence Alignment for PNGase A and PNGase *At* and three putative type II PNGases targeted in this project.
Conserved residues are highlighted according to their degree of conservation. Red: identical residue (*). Dark grey: conserved substitution (:). Light grey: semi-conserved substitution (.).

In a multiple amino acid sequence alignment, identical amino acids, or the presence of very conservative substitutions, that is, the substitution of amino acids whose side chains have similar biochemical properties, suggest that this region or residue has structural or functional importance. For this group of PNGases, neither structural studies nor functional characterisations of specific residues have been carried out. Therefore, sequence conservation is the only way to predict regions of the structure and perhaps specific residues that may be important for function. To assess the possibility of the putative PNGases being catalytically active, the conservation of residues that are known to be most frequently involved in enzyme catalysis such as histidine, aspartate, glutamate, lysine, cysteine, arginine, serine, threonine, tyrosine and tryptophan was investigated to try to predict both a catalytic mechanism and a substrate binding site among these homologues. Holliday *et al.* (2009) analysed the functional role of amino acid residues in enzymes using information available from the MACiE (**M**echanism, **A**nnotation and **C**lassification **i**n **E**nzymes) and CSA (**C**atalytic **S**ite **A**tlas) databases (Holliday *et al.*, 2005; Holliday *et al.*, 2009; Porter *et al.*, 2004). They found that in hydrolases (EC class 3 enzymes) the residues with the highest catalytic propensity are histidine followed by aspartate, glutamate and cysteine. Stabilisation of reaction intermediates or other residues (generally through electrostatic interactions) is usually provided by residues such as tryptophan, tyrosine and arginine. Several of these potentially important residues are conserved in the homologues shown, with the highest degree of conservation, especially when the initial alignment of 50

homologues is included, being between residues 210 to 303 (PNGase A numbering). However, without any further information from structural or mutational studies it is difficult to speculate which of these residues might be responsible for catalytic activity and substrate binding.

3.3.4 Bioinformatical Characterisation of Selected PNGase A Homologues

3.3.4.1 Streptomyces avermitilis Putative PNGase

S. avermitilis MA-4680 was originally isolated and characterised in 1978 in Japan (Burg *et al.*, 1979). It is a Gram-positive, filamentous, conidia-forming organism belonging to the genus *Streptomyces* within the eubacterial class *Actinobacteria*. Members of this genus are the most important industrial producers of antibiotics and other secondary metabolites, which are used in human and veterinary medicine and agriculture, as well as of anti-parasitic agents, herbicides, pharmacologically active metabolites and several enzymes used in the food industry (Demain, 1999). *S. avermitilis* is best known for the production of avermectin, an anti-parasitic agent that is widely used to rid livestock of worm and insect infestations and to protect large numbers of people from river blindness in sub-Saharan Africa. The complete genome sequence was published by Ikeda *et al.* in 2003, describing a 9.02561 Mbp genome with a high GC-content of 70% and at least 7,500 ORFs (Ikeda *et al.*, 2003).

This putative peptide:*N*-glycanase (*Sav*PNGase) is encoded by a 1,626 bp ORF that includes the coding sequence for an export signal sequence. The gene product is a 541 amino acid protein which includes a predicted 24 amino acid N-terminal signal sequence (SignalP 3.0; (Bendtsen *et al.*, 2004)). The calculated molecular weight of the mature protein is 56.4 kDa with a pI of 5.88. The mature protein contains 6 cysteine residues that potentially form disulfide bonds, common for secreted proteins.

Table 3.4: Summary of bioinformatic characterisation of putative *S. avermitilis* PNGase.
Number of amino acids, molecular weight and predicted pI are based on the mature protein and were predicted using the program ProtParam (Gasteiger *et al.*, 2003).

Number of amino acids	Molecular weight [kDa]	Predicted pI	Signal peptide	*N*-glycosylation[17]
517	56.4	5.88	Yes (24 aas)	Yes (1)

Two *N*-glycosylation sequons are present in *Sav*PNGase at positions 360 and 387, the latter one being predicted to be probably glycosylated (NetNGlyc 1.0 Server; (Gupta *et al.*, 2004)).

3.3.4.2 Sulfolobus solfataricus Putative PNGase

S. solfataricus P2 is the model organism for the archaeal domain of crenarchaeotes. It is an obligate aerobic archaeon, which grows in hot (~80°C) and acidic (pH 2-4) environments, and was first isolated from a solfataric field near Naples, Italy (Zillig *et al.*, 1980). Its genome sequence was published in 2001 (She *et al.*, 2001).

The ORF SSO2552 encoding the putative peptide:*N*-glycanase (*Sso*PNGase) is 1833 bp long and codes for a 610 amino acid protein that is predicted to be secreted as it contains an N-terminal secretory signal sequence with the most likely cleavage site being between residues 25 and 26 (Signal P 3.0, for eukaryotic signal peptides). The exact cleavage site is difficult to predict as the three programs available are designed to predict eukaryotic, Gram-negative or Gram-positive signal sequences. However, it has been observed for experimentally determined cleavage sites that the eukaryotic type is preferred by *S. solfataricus* (Albers & Driessen, 2002). The predicted molecular weight of the mature, probably secreted, protein is 65.99 kDa with a calculated pI of 4.87. It does not contain any cysteine residues.

[17] In brackets the number of probable *N*-glycosylation sequons is given.

Table 3.5: Summary of bioinformatic characterisation of putative *S. solfataricus* PNGase.
Number of amino acids, molecular weight and predicted pI are based on the protein lacking the predicted signal peptide and were predicted using the program ProtParam (Gasteiger *et al.*, 2003).

Number of amino acids	Molecular weight [kDa]	Predicted pI	Signal peptide	*N*-glycosylation[18]
585	66	4.87	Yes (25 aas)	Yes (17)

S. solfataricus putative PNGase is predicted to be highly *N*-glycosylated (NetNGlyc 1.0 Server; (Gupta *et al.*, 2004)). Of the 21 typical *N*-glycosylation sequons Asn-X-Ser/Thr present in the protein, 17 are predicted to be glycosylated.

Interestingly, the ORF SSO2551, directly upstream of the putative *Sso*PNGase, codes for a putative serine protease. The intergenic space between these two genes is rather small (4 nucleotides) suggesting that they may be transcribed simultaneously and functionally coupled. Parallels can be drawn with *Dra*PNGase and some other PNGase F-like proteins, where a protease-associated domain is actually part of the PNGase ORF.

3.3.4.3 Aspergillus niger Putative PNGase

A. niger is a filamentous fungus growing aerobically in soil, litter, compost, decaying plant material and generally on organic matter. It is one of the most important organisms used in biotechnology, producing citric acid and several commercial enzymes such as glucoamylase (Schuster *et al.*, 2002). The genome sequence of *A. niger* CBS 513.88 was published by Pel *et al.* in 2007 (Pel *et al.*, 2007).

The putative PNGase from *A. niger* (*Ani*PNGase) is encoded by a 1713 bp ORF that contains one intron of 57 bp (4.3.3). The primary translation product (551 residues) is predicted to contain a secretion signal sequence that is cleaved between residues 21 and 22. The mature protein contains four cysteines, which

[18] In brackets the number of probable *N*-glycosylation sequons is given.

may be involved in disulfide formation. A summary of some basic features of *Ani*PNGase is presented in **Table 3.6**.

Table 3.6: Summary of bioinformatic characterisation of putative *A. niger* PNGase.
Number of amino acids, molecular weight and predicted pI are based on the mature protein and were predicted using the program ProtParam (Gasteiger *et al.*, 2003).

Number of amino acids	Molecular weight [kDa]	Predicted pI	Signal peptide	*N*-glycosylation[19]
530	58.5	4.31	Yes (21 aas)	Yes (6)

*Ani*PNGase contains a total of 13 *N*-glycosylation sequons with six of these being predicted to be actually modified (NetNGlyc 1.0 Server; (Gupta *et al.*, 2004)).

3.3.5 Classification

Based on the bioinformatical results obtained here and previous functional and structural studies that have been discussed in the Introduction (Chapter 1) it appears reasonable to organise PNGases into three types. The reasoning for the proposed classification has already been described in the Introduction (1.2.1) together with the rationalisation for a classification scheme, i.e. to avoid confusion between proteins belonging to the different types. Therefore, just a brief recapitulation of the main arguments shall be presented here.

The PNGase family (EC 3.5.1.52) can be divided into three types mainly based on their primary amino acid sequence, but also on their subcellular localisation, phylogenetic distribution (to date) and physiological function (if known). For two types, type I and III, crystal structures have been solved, which revealed very different structures and obvious differences in their catalytic mechanism. However, despite these evident differences the same overall reaction is being catalysed by all members of this family, which raises the question of how these proteins developed. The theory of convergent evolution

[19] In brackets the number of probable N-glycosylation sequons is given.

provides the most feasible explanation (1.2.1). The proposed classification is shown again in **Table 3.7**.

Table 3.7: Proposed classification of peptide:*N*-glycanases (EC 3.5.1.52).

Type	Main characteristics	Examples of enzyme	Source
I	secreted; bacterial or possibly bacterial origin	PNGase F	*F. meningosepticum*
II	secreted/exoplasmic; archaea, bacteria, fungi, plants	PNGase A PNGase *At*	*P. amygdalus* (sweet almond) *A. tubingensis*
III	cytoplasmic, proteasome-associated; ubiquitous in eukaryotes; not found in bacteria or archaea	yPng1p mPNGase hPng1p	*S. cerevisiae* *Mus musculus* *Homo sapiens*

Chapter 4

Gene Expression Analyses

4 Gene Expression Analyses

4.1 Introduction

Bioinformatic analyses can provide valuable theoretical information about gene sequences and proteins as demonstrated in Chapter 3. However, these methods are only the start of the process for obtaining evidence of a gene's function and its product. Results obtained using these computational methods have to be proven experimentally as the presence of a gene sequence in an organism does not necessarily mean that this gene is actually transcribed into messenger-RNA, the basic prerequisite for protein production. It has to be established that the target genes are actively transcribed and are not silent DNA sequences (cryptic genes, pseudogenes; (Hall *et al.*, 1983; Harrison & Gerstein, 2002)). Generally, if an organism expends energy on transcribing a gene into mRNA, translation follows.

One of the most powerful and sensitive methods for gene expression analysis is Reverse Transcriptase (RT)-PCR. This method permits the detection of minuscule amounts of mRNA present in a sample or an organism.

To demonstrate the transcription of the putative PNGase genes into mRNA in the native strains *A. niger*, *S. avermitilis* and *D. radiodurans*, total RNA was isolated from these organisms and subjected to qualitative RT-PCR analysis using gene specific primers.

4.2 Methods

4.2.1 Cultivation of Aspergillus niger

A. niger was obtained as actively growing culture from the culture collection of the Institute of Molecular Biosciences.

This organism was grown on malt extract agar plates or in liquid malt-extract medium (2.5.5). The cultures were incubated at 30°C for approximately 2-3 days.

4.2.2 Initiation and Cultivation of Streptomyces avermitilis MA-4680

S. avermitilis MA-4680 was obtained as a vacuum dried culture from the 'Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH' (German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany).

The dried culture was rehydrated for 30 minutes in 0.5 mL GYM *Streptomyces* medium (2.5.2). The cells were resuspended in the medium and used as inoculum for a 20 mL liquid culture and streak plate. A single colony from the streak plate was grown in 5 mL liquid medium and prepared as glycerol stocks for storage at -80°C (2.4).

In subsequent experiments *S. avermitilis* was grown for 2-4 days at 30°C on GYM or Oatmeal agar plates or in each liquid medium (2.5.2, 2.5.3).

4.2.3 Initiation and Cultivation of Deinococcus radiodurans R1

D. radiodurans R1 was obtained as a vacuum dried culture from the 'Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH' (German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany).

The dried culture was rehydrated for 30 minutes in 0.5 mL *Corynebacterium* medium (2.5.4). The cells were resuspended in the medium and used as inoculum for a 20 mL liquid culture and streak plate. A single colony of the streak plate was grown in 5 mL liquid medium and prepared as glycerol stock for storage at -80°C (2.4).

4.2.4 Extraction of genomic DNA from Aspergillus niger

Genomic DNA from *A. niger* was isolated using the genomic DNA isolation reagent DNAzol® (Invitrogen™). This reagent, a guanidine-detergent solution, allows the selective ethanol-precipitation of DNA from a cell lysate. After an ethanol washing step and re-solubilisation, the precipitated DNA was ready to use for downstream applications.

A. niger was grown on malt extract agar (2.5.5) as described above (4.2.1). Mycelia and spores (approximately 50 mg) were scraped off the agar plate and suspended in 1 mL DNAzol® reagent. The isolation was performed according to the manufacturer's instructions.

4.2.5 Isolation of total RNA

4.2.5.1 General Considerations and Precautions for RNA Work

When working with RNA great care has to be taken to keep the working environment RNase free. RNases are present everywhere and very stable, therefore all materials, solutions and equipment used for RNA isolation and any downstream applications had to be specially treated to remove RNases, as autoclaving alone does not completely remove RNase activity. Aseptic techniques were always used to avoid contaminations. A 'RNA work only'-area was set up with dedicated labware, pipettes and other equipment. Also, all chemicals and kits used for RNA work were stored in a separate area.

The main sources for RNase contaminations are hands/skin and airborne moulds or other particles. To avoid contamination from these sources gloves were worn at all times and changed frequently. If not in use, the working space designated for RNA work was covered with plastic foil at all times. Before starting any RNA work the foil was removed and the bench surface sprayed with a 0.05% diethyl pyrocarbonate (DEPC) solution.

Non-disposable glassware was baked at 250°C overnight. Plastic ware was first rinsed with 0.1 M NaOH/1 mM EDTA followed by washing with DEPC-treated H_2O_{pure}.

Solutions were treated with the addition of 0.05% DEPC, incubated overnight at room temperature and then autoclaved twice.

DEPC reacts with histidine residues of proteins and inactivates RNases. As it also reacts with RNA, DEPC has to be removed by heat treatment before use of treated solutions or materials.

4.2.5.2 Isolation of total RNA

The Illustra™ RNAspin Mini Isolation Kit (GE Healthcare) was used for isolation of total RNA. The procedure was performed mainly as described in the manufacturer's instructions. The homogenisation and lysis of cells was achieved by grinding approximately 100 mg cells to a fine powder in liquid nitrogen using mortar and pestle ensuring that the sample stayed frozen at all times.

The DNase treatment procedure provided in the protocol was altered as it did not remove DNA efficiently. To ensure complete removal of DNA contamination the DNase treatment was repeated at least twice and the incubation time was extended from the suggested 15 minutes to 30 minutes.

The concentration of isolated RNA was determined using the NanoDrop™ 1000 spectrophotometer.

4.2.6 Reverse Transcriptase (RT)-PCR

The identification of specific mRNAs within the previously isolated total RNA (4.2.5.2) was performed using the SuperScript™ II One-Step RT-PCR System with Platinum® *Taq* DNA polymerase (Invitrogen™). With this system, cDNA synthesis and PCR amplification can be performed in one step as it combines a Reverse Transcriptase and a *Taq* DNA polymerase. The *Taq* DNA

polymerase is complexed with an inhibitory antibody, which blocks the polymerase activity during cDNA synthesis and is removed in the initial denaturation step in the PCR cycle. The reaction mix was prepared on ice in a nuclease-free, thin-walled 0.2 mL PCR tube containing the components given in **Table 4.1**.

Table 4.1: Composition of a RT-PCR reaction mixture using SuperScript™ II One-Step RT-PCR System with Platinum® *Taq* DNA polymerase (Invitrogen™).

Component	Volume
2× Reaction Mix[20]	25.0 µL
Template RNA	x µL
Sense (5′) Primer (10 pmol/µL)	2.0 µL
Anti-Sense (3′) Primer (10 pmol/µL)	2.0 µL
SuperScript II RT/Platinum® *Taq* Mix	2.0 µL
H_2O_{pure}	up to 50.0 µL

To ensure the RNA preparation was free of any DNA contamination a control reaction was performed by substituting the SuperScript™ II RT/Platinum® *Taq* Mix with 2 units of Platinum® *Taq* DNA polymerase (Invitrogen™).

The reaction mixes were placed in a pre-warmed thermocycler (Biometra) and a program was started as given in **Table 4.2**.

[20] Mix includes dNTPs, Mg^{2+} and stabilisers at optimised concentrations

Table 4.2: Thermal profile used for a one-step Reverse Transcriptase-PCR. Enzyme: SuperScript™ II One-Step RT-PCR System with Platinum® *Taq* DNA polymerase (Invitrogen™).

Cycles		**Temperature**	**Time**
cDNA Synthesis			
1×		50°C	30 min
Initial Denaturation			
1×		94°C	5 min
Amplification			
	Denaturation	94°C	1 min
25×-33×	Annealing	55-62°C	45 sec
	Elongation	68°C	1 min/kbp
Final Elongation			
1×		68°C	10 min

The result of the RT-PCR was analysed using Agarose Gel Electrophoresis.

4.3 Results & Discussion

4.3.1 Transcriptional Analysis of the Putative D. radiodurans PNGase

Total RNA isolated from a liquid culture of *D. radiodurans* was used for RT-PCR. The oligonucleotides O6 and O8 (**Table 2.6**) were used to amplify the ORF from the cDNA that was produced in the initial reverse transcriptase step of the reaction. With this primer combination a PCR product of 1585 bp should be obtained. **Figure 4.1** shows the result of this RT-PCR reaction.

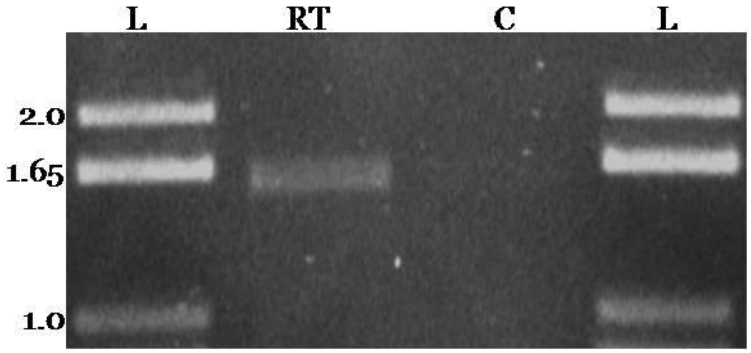


Figure 4.1: RT-PCR result for the putative *D. radiodurans* PNGase.
Primer used: O6 and O8. Annealing temperature used: 58°C. RT: RT-PCR; C: DNA control reaction; L: DNA ladder with fragment lengths given in kbp.

The size of the specific DNA fragment visible in lane 'RT' is in agreement with the expected size of the amplification product. The absence of any PCR product in the control reaction (lane 'C') demonstrates that the RT-PCR product must be derived from mRNA/cDNA and is not a result of DNA contamination.

4.3.2 Transcriptional Analysis of the Putative S. avermitilis PNGase

Total RNA isolated from a liquid culture of *S. avermitilis* was used for RT-PCR. The oligonucleotides O11 and O12 were employed to amplify the ORF from the cDNA that was derived by reverse transcription of mRNA in the first step of the reaction. With this primer combination the expected size of the PCR product was 1558 bp. **Figure 4.2** shows the result of the RT-PCR.

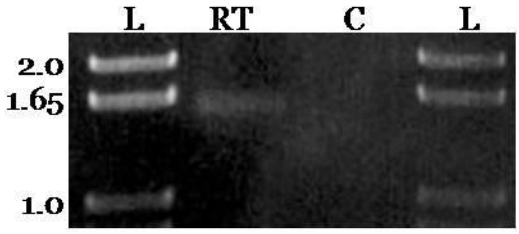


Figure 4.2: RT-PCR result for *S. avermitilis*.
Primer used: O11 and O12. Annealing temperature used: 62°C. RT: RT-PCR; C: DNA control reaction; L: DNA ladder with fragment lengths given in kbp.

In lane 'RT' a specific reaction product with the expected size was detected. No amplification product was obtained in the DNA control reaction (lane 'C').

This proves that no DNA was present in the RNA preparation and that the PCR product obtained in the RT-PCR reaction resulted from *Sav*PNGase-mRNA.

4.3.3 Genomic and Transcriptional Analysis of the Putative A. niger PNGase

In case of *A. niger* it was required to initially analyse the ORF encoding the putative PNGase as it was unknown which *A. niger* strain was present in the IMBS culture collection. After confirmation of the presence and sequence of the ORF at the genomic level, the transcriptional analysis was performed to ensure that the gene was in fact expressed in *A. niger*.

4.3.3.1 Amplification of the Putative A. niger PNGase ORF from Genomic DNA

Total genomic DNA was isolated from *A. niger* (mycelia and spores) in order to determine the presence of the putative PNGase gene in the strain obtained from the IMBS culture collection and to confirm the homology of its nucleotide sequence to the sequence deposited in the public databases.

The isolated genomic DNA was used to amplify the putative PNGase ORF by PCR using gene specific oligonucleotides O1 and O2, which were designed from publicly available sequence data and include the features required for subsequent TOPO® cloning.

The result of the amplification of the putative PNGase ORF is shown in **Figure 4.3**. A specific PCR product with the expected size of 1713 bp was obtained.

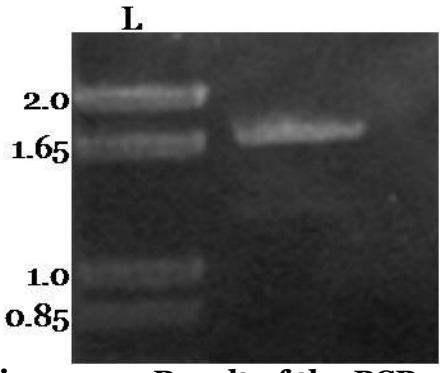


Figure 4.3: Result of the PCR amplification of the putative *A. niger* PNGase ORF.
L: DNA Ladder with fragment lengths given in kbp.

The PCR product was gel purified and cloned into pENTR/SD/D-TOPO for subsequent DNA sequencing and sequence analysis (4.3.3.3).

4.3.3.2 Transcriptional Analysis using RT-PCR

Total RNA was isolated from *A. niger* mycelia and spores and used for a one-step RT-PCR. Two primer sets were used to amplify the putative *A. niger* PNGase in the PCR step of the reaction. To ensure the absence of any DNA contamination in the isolated RNA, which could lead to false positive results, control reactions were performed as described above (4.2.6).

Figure 4.4 shows the result of the RT-PCRs and the respective control reactions. In the reactions 'RT-1' and 'C-1' gene specific primers that bind to the ORF's 3′- and 5′-termini (O1, O2; **Table 2.6**) were used. In contrast to the PCR product resulting from the genomic DNA template, the transcript obtained from cDNA was expected to be 57 bp smaller due to the loss of an intron. In the RT-PCR reaction a specific reaction product was obtained with the expected size of approximately 1,656 bp. No PCR product was obtained in the control reaction omitting the reverse transcriptase step, indicating the absence of DNA contaminants. In the reactions 'RT-2' and 'C-2' the gene specific 5′-primer (O1) was combined with a reverse, internal oligonucleotide (O27). The RT-PCR product was as expected approximately 1 kbp, and no product was obtained in the control reaction.

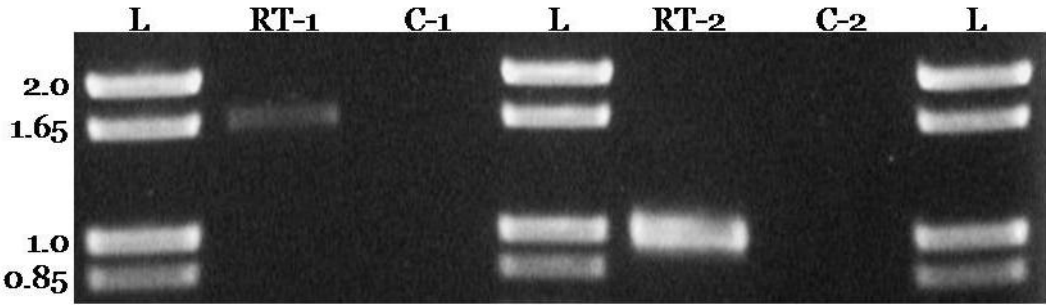


Figure 4.4: RT-PCR result for *A. niger*.
Annealing temperature used: 55°C. RT-1: RT-PCR using primer set O1+O2; C-1: DNA control PCR using primer set O1+O2; RT-2: RT-PCR using primer set O1+O27; C-2: DNA control PCR using primer set O1+O27; L: DNA ladder with fragment lengths given in kbp.

To analyse the nucleotide sequence and to confirm the absence of the intron, the RT-PCR product obtained in 'RT-1' was cloned into the vector pENTR/TEV/D-TOPO and then subjected to DNA sequencing.

4.3.3.3 Sequence Analysis of the Putative A. niger PNGase

DNA sequences were obtained for the PCR products from genomic DNA and cDNA (2.21) and compared to known nucleotide sequences for the putative *A. niger* PNGase.

The nucleotide sequence of both PCR fragments was subjected to a BLASTn search optimised for highly similar sequences (megablast). **Table 4.3** shows the result of the BLASTn that was acquired using the cDNA sequence of the putative PNGase gene from the IMBS *A. niger* strain.

Table 4.3: Result of a BLASTn (megablast) searching for highly similar sequences.
The cDNA sequence of the IMBS A. *niger* strain was used as query sequence.

Accession	Description	Total score	Query coverage	E-value	Max identity
XM_001390176.1	*Aspergillus niger* CBS 513.88 hypothetical protein (An03g03300) partial mRNA	2,897	100%	0.0	98%
AM270052.1	*Aspergillus niger* contig An03c0110, complete genome	2,900	100%	0.0	98%
U96923.1	*Aspergillus niger* peptide-N4-(N-acetyl-beta-D-glucosaminyl) asparaginase amidase N (pngN) mRNA, complete cds	1,997	100%	0.0	86%

The first two sequences both originate from the *A. niger* strain CBS 513.88 and represent the same ORF (Pel *et al.*, 2007). The third sequence, now wrongly assigned as *Aspergillus niger* peptide-N4-(N-acetyl-beta-D-glucosaminyl) asparaginase amidase N (pngN), is in fact the sequence that was originally published as PNGase *At* in 1997 by Ftouhi-Paquin (Ftouhi-Paquin *et al.*, 1997). Indeed, *A. tubingensis* and *A. niger* are closely related, but the BLASTn result and the sequence alignment clearly show sequence variations. The taxonomic relationship between *A. niger* and *A. tubingensis* has been discussed by Schuster *et al.* (Schuster *et al.*, 2002). The differences at the molecular level between these two Aspergilli have been demonstrated in an array of studies using different methods (Bussink *et al.*, 1991; de Graaff *et al.*, 1994; Gielkens *et al.*, 1997; Megnegneau *et al.*, 1993; Parenicova *et al.*, 1997; Parenicova *et al.*, 2001; Varga *et al.*, 1993). Despite the absence of phenotypic differences, it has been established that *A. niger* and *A. tubingensis* are indeed different species.

To demonstrate the similarity of the *A. niger* sequences amongst members of the same species, and the differences to the PNGase *At* sequence, a multiple alignment of the relevant ORFs was generated using ClustalW2 (**Figure 4.5**; (Larkin *et al.*, 2007).

```
Ani*_genomic      ATGCTGGTCTCTTTCAGTGTCGCATTCTACCTAGTGTCTCTACTATTTTCCCCAGTGCGG 60
Ani_ATCC1015      ATGCTGGTCTCTTTCAGTGTCGCATTCTACCTAGTGTCTCTACTATTTTCCCCAGTGCGG 60
Ani*_cDNA         ATGCTGGTCTCTTTCAGTGTCGCATTCTACCTAGTGTCTCTACTATTTTCCCCAGTGCGG 60
An03g03300        ATGCTGGTCTCTTTCAGTGTCCCATTCTACCTGGTGTCTCTACTATTTTCCCCAGTGCGG 60
"PNGase_N"        ATGCTGGTCTCTTTCGGTGTCGCATTCTACCTAGTATCTCTACTGTTTTCTCCAGCGCGG 60
PNGase_At         ATGCTGGTCTCTTTCGGTGTCGCATTCTACCTAGTATCTCTACTGTTTTCTCCAGCGCGG 60
                  *************** ***** ********** ** ******** ***** **** ****

Ani*_genomic      GCTGTTCTAGAGGTCTTCGAGGTATACCAGCCCGTCCCGACAGGCCATGGGAGCGTTGGT 120
Ani_ATCC1015      GCTGTTCTAGAGGTCTTCGAGGTATACCAGCCCGTCCCGACAGGCCATGGGAGCGTTGGT 120
Ani*_cDNA         GCTGTTCTAGAGGTCTTCGAGGTATACCAGCCCGTCCCGACAGGCCATGGGAGCGTTGGT 120
An03g03300        GCTGTTCTAGAGGTCTTCGAGGTATACCAGCCCGTCCCGACAGGCCATGGGAGCGTTGGT 120
"PNGase_N"        GCTCTTCTAGAGGTCTTCGAGGTATACCAGCCCGTCCCCACAGGTCATGGCAGCGTTGGG 120
PNGase_At         GCTCTTCTAGAGGTCTTCGAGGTATACCAGCCCGTCCCCACAGGTCATGGCAGCGTTGGG 120
                  *** ********************************** ***** ***** ********

Ani*_genomic      TGCAATGAGGCAATCCTCCTGATGGACCATGTCTTCGGCTATAGCTATGGTGAACCATAC 180
Ani_ATCC1015      TGCAATGAGGCAATCCTCCTGATGGACCATGTCTTCGGCTATAGCTATGGTGAACCATAC 180
Ani*_cDNA         TGCAATGAGGCAATCCTCCTGATGGACCATGTCTTCGGCTATAGCTATGGTGAACCATAC 180
An03g03300        TGCAATGAGGCAATCCTCCTGATGGACCATGTCTTCGGCTATAGCTATGGGGAACCATAC 180
"PNGase_N"        TGCAATGAGGAAGTCCTCCTGATGGACCATGTCTTCGGCTATAGCTATGGGGAACCATAC 180
PNGase_At         TGCAATGAGGAAGTCCTCCTGATGGACCATGTCTTCGGCTATAGCTATGGGGAACCATAC 180
                  ********** * ************************************ *********

Ani*_genomic      GTCGGTAGGAAACAGTGCTGGACAGATGATCCGATGAGTTCATAAGCTAAATATGAACCA 240
Ani_ATCC1015      GTCGGTAGGAAACAGTGCTGGACAGATGAGCCGATGAGTTCATAAGCTAAATATGAACCA 240
Ani*_cDNA         GTCGG------------------------------------------------------- 185
An03g03300        GTCGG------------------------------------------------------- 185
"PNGase_N"        GTCGG------------------------------------------------------- 185
PNGase_At         GTCGG------------------------------------------------------- 185
                  *****

Ani*_genomic      GGGATCTACGAGCCACCGAATTGTACCTTTGACACCGTTCGCCTCAATCTCACTGTCACT 300
Ani_ATCC1015      GGGATCTACGAGCCACCGAATTGTACCTTTGACACCGTTCGCCTCAATCTCACTGTCACT 300
Ani*_cDNA         --GATCTACGAGCCACCGAATTGTACCTTTGACACCGTTCGCCTCAATCTCACTGTCACT 243
An03g03300        --GATCTACGAGCCACCGAATTGTACCTTTGACACCGTTCGCCTCAATCTCACTGTCACT 243
"PNGase_N"        --GATCTACGAACCACCAAACTGTACCTTTGACACCGTTCGCATCAATTTCACTGTCACT 243
PNGase_At         --GATCTACGAACCACCAAACTGTACCTTTGACACCGTTCGCATCAATTTCACTGTCACT 243
                    ********* ***** ** ********************** ***** ***********

Ani*_genomic      TCCAATGGCACACAGTATGATCGCCTGGCGCTTATGTACTTAGGGGACACAGAGGTGTTC 360
Ani_ATCC1015      TCCAATGGCACACAGTATGATCGCCTGGCGCTTATGTACTTAGGGGACACAGAGGTGTTC 360
Ani*_cDNA         TCCAATGGCACACAGTATGATCGCCTGGCGCTTATGTACTTAGGGGACACAGAGGTGTTC 303
An03g03300        TCCAAGGGCACACAGTATGATCGCCTGGCGCTTATGTACTTAGGGGACACAGAGGTGTTC 303
"PNGase_N"        TCCAAGGGCAGACAGTATGATCGTCTGGCGCTCATGTACTTAGGGGACACAGAGGTGTTC 303
PNGase_At         TCCAAGGGCAGACAGTATGATCGTCTGGCGCTCATGTACTTAGGGGACACAGAGGTGTTC 303
                  ***** **** ************ ******** ***************************

Ani*_genomic      CGAACATCAACTGCTGAACCAACGACCAACGGAATCATCTGGACCTATATCAAAGACATG 420
Ani_ATCC1015      CGAACATCAACTGCTGAACCAACGACCAACGGAATCATCTGGACCTATATCAAAGACATG 420
Ani*_cDNA         CGAACATCAACTGCTGAACCAACGACCAACGGAATCATCTGGACCTATATCAAAGACATG 363
An03g03300        CGAACATCAACTGCTGAACCAACGACCAACGGAATCATCTGGACCTATATCAAAGACATG 363
"PNGase_N"        CGAACATCAACTGCTGAACCAACGACCGACGGCATCATCTGGACCTATATCAAAGACATG 363
PNGase_At         CGAACATCAACTGCTGAACCAACGACCGACGGCATCATCTGGACCTATATCAAAGACATG 363
                  *************************** **** ***************************

Ani*_genomic      TCTCAGTTCAACGTACTGTGGAAAGAAAAACAGAAATTGATCTTTGATCTTGGAAACATC 480
Ani_ATCC1015      TCTCAGTTCAACGTACTGTGGAAAGAAAAACAGAAATTGATCTTTGATCTTGGAAACATC 480
Ani*_cDNA         TCTCAGTTCAACGTACTGTGGAAAGAAAAACAGAAATTGATCTTTGATCTTGGAAACATC 423
An03g03300        TCTCAGTTCAACGTACTGTGGAAAGAAAAACAGAAATTGATCTTTGATCTTGGAAACATC 423
"PNGase_N"        TCTCAGTTCAACGTACTATGGAAAGAAAAACAGAAATTGATCTTCGATCTTGGCAACATC 423
PNGase_At         TCTCAGTTCAACGTACTATGGAAAGAAAAACAGAAATTGATCTTCGATCTTGGCAACATC 423
                  ***************** ************************* ******** ******

Ani*_genomic      ATTACTGATGTCTACACCGGATCTTTCAATACCACTCTAACGGCGTATTTCTCGTATGAG 540
Ani_ATCC1015      ATTACTGATGTCTACACCGGATCTTTCAATACCACTCTAACGGCGTATTTCTCATATGAG 540
Ani*_cDNA         ATTACTGATGTCTACACCGGATCTTTCAATACCACTCTAACGGCGTATTTCTCGTATGAG 483
An03g03300        ATTACTGATGTCTACACCGGATCTTTCAATACCACTCTAACGGCGTATTTCTCGTATGAG 483
"PNGase_N"        ATTACTGATGTCTACACCGGCTCTTTCAACACCACCCTAACAGCCTACTTCTCATACGAG 483
PNGase_At         ATTACTGATGTCTACACCGGCTCTTTCAACACCACCCTAACAGCCTACTTCTCATACGAG 483
                  ******************** ******** ***** ***** ** ** ***** ** ***
```

```
Ani*_genomic      GGCAATGTCAGAACCCCAGATATTATTCTTCCAATATCTGCGCGCAAATCCGCACAAAAT 600
Ani_ATCC1015      GGCAATGTCAGAACCCCAGATATTATTCTTCCAATATCTGCGCGCAAATCCGCACAAAAT 600
Ani*_cDNA         GGCAATGTCAGAACCCCAGATATTATTCTTCCAATATCTGCGCGCAAATCCGCACAAAAT 543
An03g03300        GGCAATGTCAGAACCCCTGATATTATTCTTCCAATATCTGCGCGCAAATCCGCACAAAAT 543
"PNGase_N"        GGCAACGTCAGGACCCCAGACGTTATTCTCCCAATATCTGCTCGCAAATCCGCCCAGAAC 543
PNGase_At         GGCAACGTCAGGACCCCAGACGTTATTCTCCCAATATCTGCTCGCAAATCCGCCCAGAAC 543
                  ***** ***** ***** **  ******* ********** ********** ** **

Ani*_genomic      GCATCAAGTGACTTTGAGCTTCCATCTGACAACGCCACGGTGCTATATCAGATCCCTCCG 660
Ani_ATCC1015      GCATCAAGTGACTTTGAGCTTCCATCTGACAATGCCACGGTGCTATATCAGATCCCTCCG 660
Ani*_cDNA         GCATCAAGTGACTTTGAGCTTCCATCTGACAACGCCACGGTGCTATATCAGATCCCTCCG 603
An03g03300        GCATCAAGTGACTTTGAGCTTCCATCTGACAATGCCACGGTGCTCTATCAAATCCCTCCG 603
"PNGase_N"        GCATCAAGCGACTTCGAACTTCCATCTGACAACGCCACAGTGCAATATCAGATCCCCCAG 603
PNGase_At         GCATCAAGCGACTTCGAACTTCCATCTGACAACGCCACAGTGCAATATCAGATCCCCCAG 603
                  ******** ***** ** ************** ***** ****  ***** ***** * *

Ani*_genomic      ACAGCATCCCGTGCAGTTGTGTCTATTTCTGCATGTGGCCAGTCAGAAGAAGAATTCTGG 720
Ani_ATCC1015      ACAGCATCCCGTGCAGTTGTGTCTATTTCTGCATGTGGCCAGTCAGAAGAAGAATTCTGG 720
Ani*_cDNA         ACAGCATCCCGTGCAGTTGTGTCTATTTCTGCATGTGGCCAGTCAGAAGAAGAATTCTGG 663
An03g03300        ACAGCATCCCGTGCAGTTGTGTCTATTTCTGCATGTGGCCAGTCAGAGGAAGAATTCTGG 663
"PNGase_N"        ACAGCATCCCGTGCAGTCGTGTCCATTTCTGCCTGTGGCCAATCCGAGGAAGAATTCTGG 663
PNGase_At         ACAGCATCCCGTGCAGTCGTGTCCATTTCTGCCTGTGGCCAATCCGAGGAAGAATTCTGG 663
                  ***************** ***** ******** ******** ** ** ************

Ani*_genomic      TGGTCCAACGTTCTCTCTGCCGATGAGTGGACCTTCGACAATACCATTGGTGAGCTGTAC 780
Ani_ATCC1015      TGGTCCAACGTTCTCTCTGCCGATGAGTGGACCTTCGACAATACCATTGGTGAGCTGTAC 780
Ani*_cDNA         TGGTCCAACGTTCTCTCTGCCGATGAGTGGACCTTCGACAATACCATTGGTGAGCTGTAC 723
An03g03300        TGGTCCAACGTTCTCTCTGCCGATGAGTGGACCTTCGACAATACCATTGGTGAGCTGTAC 723
"PNGase_N"        TGGTCCAACGTCCTCTCTGCCGATGAGTATACCTTCGACAATACCATTGGCGAGCTATAC 723
PNGase_At         TGGTCCAACGTCCTCTCTGCCGATGAGTATACCTTCGACAATACCATTGGCGAGCTATAC 723
                  *********** ****************  ******************** ***** ***

Ani*_genomic      GGGTACTCTCCATTCCGTGAAGTCCAGCTTTATATCGACGGCGTCCTTGCTGGCGTGGAC 840
Ani_ATCC1015      GGGTACTCTCCATTCCGTGAAGTCCAGCTTTATATCGACGGCGTCCTTGCTGGCGTGGAC 840
Ani*_cDNA         GGGTACTCTCCATTCCGTGAAGTCCAGCTTTATATCGACGGCGTCCTTGCTGGCGTGGAC 783
An03g03300        GGGTACTCCCCATTCCGTGAAGTCCAGCTTTATATCGACGGCGTCCTTGCTGGCGTGGAC 783
"PNGase_N"        GGGTACTCCCCATTCCGTGAAGTCCAGCTCTATATCGACGGCGTCCTTGCCGGCGTGGAC 783
PNGase_At         GGGTACTCCCCATTCCGTGAAGTCCAGCTCTATATCGACGGCGTCCTTGCCGGCGTGGAC 783
                  ******** ******************** ******************** *********

Ani*_genomic      TGGCCGTTCCCCATAATCTTCACCGGCGGTGTCGCTCCAGGGTTCTGGCGTCCTATTGTG 900
Ani_ATCC1015      TGGCCGTTTCCCATAATCTTCACCGGCGGTGTCGCTCCAGGGTTCTGGCGTCCTATTGTG 900
Ani*_cDNA         TGGCCGTTTCCCATAATCTTCACCGGCGGTGTCGCTCCAGGGTTCTGGCGTCCTATTGTG 843
An03g03300        TGGCCGTTTCCCATAATCTTCACCGGCGGTGTCGCTCCAGGGTTCTGGCGTCCTATTGTG 843
"PNGase_N"        TGGCCATTCCCCATCATCTTCACCGGCGGTGTCGCGCCAGGATTCTGGCGTCCTATCGTA 843
PNGase_At         TGGCCATTCCCCATCATCTTCACCGGCGGTGTCGCGCCAGGATTCTGGCGTCCTATCGTA 843
                  ***** ** ***** ******************** ***** ************** **

Ani*_genomic      GGAATTGATGCTTTTGACCTACGCCAACCAGAGATCGATATCACACCCTTCCTTCCCTTG 960
Ani_ATCC1015      GGAATTGATGCTTTTGACCTACGCCAACCAGAGATCGATATCACACCCTTCCTTCCCTTG 960
Ani*_cDNA         GGAATTGATGCTTTTGACCTACGCCAACCAGAGATCGATATCACACCCTTCCTTCCCTTG 903
An03g03300        GGAATTGATGCTTTTGACCTACGCCAACCTGAGATCGATATCACACCCTTCCTTCCCTTG 903
"PNGase_N"        GGAATCGACGCTTTCGACCTACGCCAACCCGAGATCGACATCACACCCTTCCTCCCCTTG 903
PNGase_At         GGAATCGACGCTTTCGACCTACGCCAACCCGAGATCGACATCACACCCTTCCTCCCCTTG 903
                  ***** ** ***** ************** ******** ************** ******

Ani*_genomic      CTCAAGGATAACAAGTCGCATTCTTTCGAGATCAGAGTTACAGGCTTGAGTGTCGCAGAT 1020
Ani_ATCC1015      CTCAAGGATAACAAGTCGCATTCTTTCGAGATCAGAGTTACAGGCTTGAGTGTCGCAGAT 1020
Ani*_cDNA         CTCAAGGATAACAAGTCGCATTCTTTCGAGATCAGAGTTACAGGCTTGAGTGTCGCAGAT 963
An03g03300        CTCAAGGATAACAAGTCGCATTCTTTCGAGATCAGAGTTACAGGCTTGAGTGTCGCAGAT 963
"PNGase_N"        CTCAAGGACAACAAGTCCCATTCGTTCGAGATCAGAGTTACAGGCCTGAGCGTTGCCGAC 963
PNGase_At         CTCAAGGACAACAAGTCCCATTCGTTCGAGATCAGAGTTACAGGCCTGAGCGTTGCCGAC 963
                  ******** ******** ***** ********************** **** ** ** **

Ani*_genomic      GACGGAACAGTGACTTTCGCCGACACAGTTGGCTCCTACTGGGTGGTCACCGGCACTATA 1080
Ani_ATCC1015      GACGGAACAGTGACTTTCGCCGACACAGTTGGCTCCTACTGGGTGGTCACCGGCACTATA 1080
Ani*_cDNA         GACGGAACAGTGACTTTCGCCGACACAGTTGGCTCCTACTGGGTGGTCACCGGCACTATA 1023
An03g03300        GACGGAACAGTGACTTTCGCCGACACAGTTGGCTCCTACTGGGTGGTCACCGGCACTATA 1023
"PNGase_N"        GACGGAACAGTAACATTCGCCAATACAGTCAACTCCTACTGGGTAGTCACCGGCACTATC 1023
PNGase_At         GACGGAACAGTAACATTCGCCAATACAGTCAACTCCTACTGGGTAGTCACCGGCACTATC 1023
                  *********** ** ****** * *****   ************ **************
```

```
Ani*_genomic      TTCCTTTACTTGGACGACTCCATGTCTCAA---------ACCGCAACCGGCCAGGCGCCC 1131
Ani_ATCC1015      TTCCTTTACTTGGACGACTCCATGTCTCAA---------ACCGCAACCGGCCAGGCGCCC 1131
Ani*_cDNA         TTCCTTTACTTGGACGACTCCATGTCTCAA---------ACCGCAACCGGCCAGGCGCCC 1074
An03g03300        TTCCTTTACTTGGACGACTCCATGTCTCAA---------ATCGCAACCGGCCAGGCGCCC 1074
"PNGase_N"        TTCCTTTACTTGGACTCCTCCTCCTCTGAATCACACAGCACAACCACCGGCCAAGCGCCC 1083
PNGase_At         TTCCTTTACTTGGACTCCTCCTCCTCTGAATCACACAGCACAACCACCGGCCAAGCGCCC 1083
                  ***************  ****   *** **         *   * ******** ******

Ani*_genomic      GAGGTGAACGCCCCGGCGCCAACATTCGCCGTTACGCGGAATCTTGTCCAGAGTCGGAAC 1191
Ani_ATCC1015      GAGGTGAACGCCCCGGCGCCAACATTCGCCGTTACGCGGAATCTTGTCCAGAGTCGGAAC 1191
Ani*_cDNA         GAGGTGAACGCCCCGGCGCCAACATTCGCCGTTACGCGGAATCTTGTCCAGAGTCGGAAC 1134
An03g03300        GAGGTGAACGCCCCGACGCCAACATTCGCCGTTACGCGGAATCTTGTCCAGAGTCGGAAC 1134
"PNGase_N"        GAAATCTACGCCCCGGCGCCCACCCTCACCGTCACACGGGATCTCACCCAGAGTCCAAAC 1143
PNGase_At         GAAATCTACGCCCCGGCGCCCACCCTCACCGTCACACGGGATCTCACCCAGAGTCCAAAC 1143
                  **  *  ******** **** **  ** **** ** *** ****   ********  ***

Ani*_genomic      GGGACAAACGAAACATTGGCATACTCTGTCGTGGCAGAAAGAACATTGACAGTGAAGTCT 1251
Ani_ATCC1015      GGGACAAACGAAACATTGGCATACTCTGTCGTGGCAGAAAGAACATTGACAGTGAAGTCT 1251
Ani*_cDNA         GGGACAAACGAAACATTGGCATACTCTGTCGTGGCAGAAAGAACATTGACAGTGAAGTCT 1194
An03g03300        GGGACAAACGAAACATTGGCATACTCTGTCGTGGCGGAAAGAACATTGACAGTGAAGTCT 1194
"PNGase_N"        GGGACCAACGAAACACTATCATACTCCGTCACAGCCGAACGAACATTCACCGTAAAGTCC 1203
PNGase_At         GGGACCAACGAAACACTATCATACTCCGTCACAGCCGAACGAACATTCACCGTAAAGTCC 1203
                  ***** ********* *  ******* ***   ** *** ******* ** ** *****

Ani*_genomic      TCTGAATATTCATGGAGTCAAAACCTCTCATACTCGAACTACGGATATCTGAACCAGCAA 1311
Ani_ATCC1015      TCTGAATATTCATGGAGTCAAAACCTCTCATACTCGAACTACGGATATCTGAACCAGCAA 1311
Ani*_cDNA         TCTGAATATTCATGGAGTCAAAACCTCTCATACTCGAACTACGGATATCTGAACCAGCAA 1254
An03g03300        TCTGAATATTCATGGAGTCAAAACCTCTCATACTCGAACTACGGATATCTGAACCAGCAA 1254
"PNGase_N"        TCCGAATACGCATGGAGCCAAAACCTCTCCTACTCAAACTACGGATATCTAAACCAGCAA 1263
PNGase_At         TCCGAATACGCATGGAGCCAAAACCTCTCCTACTCAAACTACGGATATCTAAACCAGCAA 1263
                  ** *****  ******* *********** ***** ************** *********

Ani*_genomic      GGATTGAGCCAGAAAAACAATCAACAGACCTTCGGCTCTAACACGATCGCTCAGCTTACT 1371
Ani_ATCC1015      GGATTGAGCCAGAAAAACAATCAACAGACCTTCGGCTCTAACACGATCGCTCAGCTTACT 1371
Ani*_cDNA         GGATTGAGCCAGAAAAACAATCAACAGACCTTCGGCTCTAACACGATCGCTCAGCTTACT 1314
An03g03300        GGATTGAGCCAGAAAAACAATCAACAGACCTTCGGCTCTAACACGATCGCTCAGCTTACT 1314
"PNGase_N"        GGACTCAGCCAGAAAAACAACCAACAGACCTCCGGCACTAACACCATCACCCAGCTTACC 1323
PNGase_At         GGACTCAGCCAGAAAAACAACCAACAGACCTCCGGCACTAACACCATCACCCAGCTTACC 1323
                  *** * ************** ********** **** ******* *** * ********

Ani*_genomic      GGAAACAAAA---CTACGAACGAGGTCATTTTCGAGTATCCGCTGATTTGTAACACGACG 1428
Ani_ATCC1015      GGAAACAAAA---CTACGAACGAGGTCATTTTCGAGTATCCGCTGATTTGTAACACGACG 1428
Ani*_cDNA         GGAAACAAAA---CTACGAACGAGGTCATTTTCGAGTATCCGCTGATTTGTAACACGACG 1371
An03g03300        GGAAACAGAA---CTACGAACGAGGTCACGTTCGAGTATCCGCTGATTTGTAACACGACG 1371
"PNGase_N"        GGGAACAACAAATCTACAAATGAGGTCACCTTCCAGTACCCTCTAATCTGTAACACAACG 1383
PNGase_At         GGGAACAACAAATCTACAAATGAGGTCACCTTCCAGTACCCTCTAATCTGTAACACAACG 1383
                  ** ****  *   **** ** *******  *** **** ** ** ** ******** ***

Ani*_genomic      TATGGACTCGAAGATGGACTTTCCATTAGTGCCTGGATCCGCAGAGGCCTAGATATCGAG 1488
Ani_ATCC1015      TATGGACTCGAAGATGGACTTTCCATTAGTGCCTGGATCCGCAGAGGCCTAGATATCGAG 1488
Ani*_cDNA         TATGGACTCGAAGATGGACTTTCCATTAGTGCCTGGATCCGCAGAGGCCTAGATATCGAG 1431
An03g03300        TATGGACTCGAAGATGGACTTTCCATTAGTGCCTGGATCCGCAGAGGCCTAGATATCGAG 1431
"PNGase_N"        TACGGCCTCGAAGATGGCCTTTCCATTAGTGCCTGGATCCGCAGAGGCCTGGATATTAGT 1443
PNGase_At         TACGGCCTCGAAGATGGCCTTTCCATTAGTGCCTGGATCCGCAGAGGCCTGGATATTAGT 1443
                  ** ** *********** ******************************** *****

Ani*_genomic      TCGACCGGTGG---------GCTCGGGGTCTCGACTTATACTTTTACCTCCGGGTCTTTG 1539
Ani_ATCC1015      TCGACCGGTGG---------GCTCGGGGTCTCGACTTATACTTTTACCTCCGGGTCTTTG 1539
Ani*_cDNA         TCGACCGGTGG---------GCTCGGGGTCTCGACTTATACTTTTACCTCCGGGTCTTTG 1482
An03g03300        TCGACCGGTGG---------GCTCGGGGTCTCGACTTATACTTTTACCTCCGGGTCTTTG 1482
"PNGase_N"        TCGACCGGTGGTGATGGGGAGCTCGGCGTCTCGACGTATACGTTTACCTCGGGGCCATTG 1503
PNGase_At         TCGACCGGTGGTGATGGGGAGCTCGGCGTCTCGACGTATACGTTTACCTCGGGGCCATTG 1503
                  ***********         ****** ******** ***** ******** *** * ***

Ani*_genomic      AATCTGCATACAGAACAACATGGAACAGCGTATTATTATGAACCCTCTGACGATGAGAGT 1599
Ani_ATCC1015      AATCTGCATACAGAACAACATGGAACAGCGTATTATTATGAACCCTCTGACGATGAGAGT 1599
Ani*_cDNA         AATCTGCATACAGAACAACATGGAACAGCGTATTATTATGAACCCTCTGACGATGAGAGT 1542
An03g03300        GATCTGCATACAGAACAACATGGAACAGCGTATTATTATGAACCCTCTGACGATGAGAGT 1542
"PNGase_N"        GATCTGCATACGGAACAATATGGAACGGCGTATTATTTCGAACCGGAGGATGATGAGAGT 1563
PNGase_At         GATCTGCATACGGAACAATATGGAACGGCGTATTATTTCGAACCGGAGGATGATGAGAGT 1563
                   ********** ****** ******* **********  *****    ** *********
```

```
Ani*_genomic      TCGGTCTCGTATGGTGAAACGTTCGATGTCTTTGGAAGCAATGCCGGTGGAGTTCAGTAC 1659
Ani_ATCC1015      TCGGTCTCGTATGGTGAAACGTTCGATGTCTTTGGAAGCAATGCCGGTGGAGTTCAGTAC 1659
Ani*_cDNA         TCGGTCTCGTATGGTGAAACGTTCGATGTCTTTGGAAGCAATGCCGGTGGAGTTCAGTAC 1602
An03g03300        TCGGTCTCGTATGGTGAAACGTTCGATGTCTTTGGAAGCAATGCCGGTGGAGTTCAATAC 1602
"PNGase_N"        TCCGTCTCGTATGGTGAAACTGTTGATGTCTGGGGAAGTAATGCTGGGGGAGTTGAGTAC 1623
PNGase_At         TCCGTCTCGTATGGTGAAACTGTTGATGTCTGGGGAAGTAATGCTGGGGGAGTTGAGTAC 1623
                  ** ****************  * *******  ***** ***** ** ****** * ***

Ani*_genomic      TTTCGGAACGTCCATGCGGTAAATGGTACTGTTGTATCGGATACTGACAGCTAG 1713
Ani_ATCC1015      TTTCGGAACGTCCATGCGGTAAATGGTACTGTTGTATCGGATACTGACAGCTAG 1713
Ani*_cDNA         TTTCGGAACGTCCATGCGGTAAATGGTACTGTTGTATCGGATACTGACAGCTAG 1656
An03g03300        TTTCGGAACGTCCATGCGGTAAATGGTACTGTTGTATCGTATACTGACAGCTAG 1656
"PNGase_N"        GCGAGGAATGTTCGTGCGGTAAATGGGACGGTTGTTTCGGATACTGAGAGCTAG 1677
PNGase_At         GCGAGGAATGTTCGTGCGGTAAATGGGACGGTTGTTTCGGATACTGAGAGCTAG 1677
                      **** ** * ************ ** ***** *** ******* ******
```

Figure 4.5: Multiple sequence alignment of nucleotide sequences of putative *A. niger* PNGases and PNGase *At*.
The alignment was obtained using ClustalW2 at EBI. Sequences labelled *Ani**_genomic and *Ani**_cDNA are the sequences acquired for the *A. niger* strain from the IMBS culture collection. Sequence Ani_ATCC1015 was obtained from (http://genome.jgi-psf.org/Aspni5/Aspni5.home.html). Sequence *An*03g03300 originated in *A. niger* strain CBS 513.88. "PNGase N" is the sequence apparently wrongly assigned as PNGase from *A. niger* (pngN). Sequence PNGase *At* is the sequence taken from the original paper. Red, identical for all *A. niger* sequences, but different to PNGase *At*; Green, different in one *A. niger* sequence; Blue, specific variation in *A. niger**; Shaded grey, intron region.

Several conclusions can be drawn from the sequence analyses presented so far:

(i) The putative *A. niger* PNGase gene was successfully amplified from genomic DNA and cDNA.

(ii) DNA sequencing and comparison of both PCR products showed the presence of a 57 bp intronic region (bp 186-242) in the 1713 bp genomic sequence, that was expected from previous studies on PNGase *At* (Ftouhi-Paquin *et al.*, 1997). The *Ani**_cDNA is 1656 bp long, 21 nucleotides less compared to PNGase *At*, but consistent with the other *A. niger* sequences.

(iii) The absence of the intronic region in the *Ani**_cDNA sequence proves that the DNA fragment obtained in the RT-PCR reaction was indeed derived from mRNA/cDNA, showing the active transcription of the ORF in native *A. niger*.

(iv) The cDNA nucleotide sequence is highly conserved amongst different *A. niger* strains, with maximum sequence identities of 99% (*Ani**_cDNA ↔ ATCC 1015) and 98% (*Ani**_cDNA ↔ CBS 513.88 *An*03g003300). There appear to be two specific nucleotide

substitutions in the *A. niger** strain, G210T (in the intron) and T633C. These substitutions are not due to PCR errors as they were found in both PCR and RT-PCR products that have been obtained in independent reactions.

(v) Although the sequences are still very similar, the degree of conservation between the putative *A. niger* PNGase and PNGase *At* is, at 86% identity, considerably lower than that between *A. niger* strains. This demonstrates, that the reassignment of PNGase *At* as *A. niger* PNGase (pngN) does indeed appear to be wrong as already stated in accession XM_001390176.1/GI:145235128 ("Remark: the ORF encoded protein is almost identical to a peptide-N4-(N-acetyl-beta-D-glucosaminyl) asparaginase amidase N (pngN) of *A. tubingensis* which is wrongly assigned to *A. niger* (compare TREMBL:U96923_1 with PUBMED-ID: 9312552.")).

To analyse the effects of the nucleotide substitutions, the sequences of *Ani**_cDNA and *A. niger* ATCC1015 were translated into their amino acid sequences ('Translate', ExPASy) and aligned with the protein sequences for *An*03g03300 and PNGase *At*. **Figure 4.6** shows the ClustalW2 alignment of these amino acid sequences.

```
Ani*            MLVSFSVAFYLVSLLFSPVRAVLEVFEVYQPVPTGHGSVGCNEAILLMDHVFGYSYGEPY 60
ATCC1015        MLVSFSVAFYLVSLLFSPVRAVLEVFEVYQPVPTGHGSVGCNEAILLMDHVFGYSYGEPY 60
An03g003300     MLVSFSVPFYLVSLLFSPVRAVLEVFEVYQPVPTGHGSVGCNEAILLMDHVFGYSYGEPY 60
PNGase_At       MLVSFGVAFYLVSLLFSPARALLEVFEVYQPVPTGHGSVGCNEEVLLMDHVFGYSYGEPY 60
                *****.*.**********.**:********************** :***************

Ani*            VGIYEPPNCTFDTVRLNLTVTSNGTQYDRLALMYLGDTEVFRTSTAEPTTNGIIWTYIKD 120
ATCC1015        VGIYEPPNCTFDTVRLNLTVTSNGTQYDRLALMYLGDTEVFRTSTAEPTTNGIIWTYIKD 120
An03g003300     VGIYEPPNCTFDTVRLNLTVTSKGTQYDRLALMYLGDTEVFRTSTAEPTTNGIIWTYIKD 120
PNGase_At       VGIYEPPNCTFDTVRINFTVTSKGRQYDRLALMYLGDTEVFRTSTAEPTTDGIIWTYIKD 120
                ***************:*:****:* ************************:*********

Ani*            MSQFNVLWKEKQKLIFDLGNIITDVYTGSFNTTLTAYFSYEGNVRTPDIILPISARKSAQ 180
ATCC1015        MSQFNVLWKEKQKLIFDLGNIITDVYTGSFNTTLTAYFSYEGNVRTPDIILPISARKSAQ 180
An03g003300     MSQFNVLWKEKQKLIFDLGNIITDVYTGSFNTTLTAYFSYEGNVRTPDIILPISARKSAQ 180
PNGase_At       MSQFNVLWKEKQKLIFDLGNIITDVYTGSFNTTLTAYFSYEGNVRTPDVILPISARKSAQ 180
                **************************************************:***********

Ani*            NASSDFELPSDNATVLYQIPPTASRAVVSISACGQSEEEFWWSNVLSADEWTFDNTIGEL 240
ATCC1015        NASSDFELPSDNATVLYQIPPTASRAVVSISACGQSEEEFWWSNVLSADEWTFDNTIGEL 240
An03g003300     NASSDFELPSDNATVLYQIPPTASRAVVSISACGQSEEEFWWSNVLSADEWTFDNTIGEL 240
PNGase_At       NASSDFELPSDNATVQYQIPQTASRAVVSISACGQSEEEFWWSNVLSADEYTFDNTIGEL 240
                *************** **** ****************************:*********

Ani*            YGYSPFREVQLYIDGVLAGVDWPFPIIFTGGVAPGFWRPIVGIDAFDLRQPEIDITPFLP 300
ATCC1015        YGYSPFREVQLYIDGVLAGVDWPFPIIFTGGVAPGFWRPIVGIDAFDLRQPEIDITPFLP 300
An03g003300     YGYSPFREVQLYIDGVLAGVDWPFPIIFTGGVAPGFWRPIVGIDAFDLRQPEIDITPFLP 300
PNGase_At       YGYSPFREVQLYIDGVLAGVDWPFPIIFTGGVAPGFWRPIVGIDAFDLRQPEIDITPFLP 300
                ************************************************************

Ani*            LLKDNKSHSFEIRVTGLSVADDGTVTFADTVGSYWVVTGTIFLYLDDSMS---QTATGQA 357
ATCC1015        LLKDNKSHSFEIRVTGLSVADDGTVTFADTVGSYWVVTGTIFLYLDDSMS---QTATGQA 357
An03g003300     LLKDNKSHSFEIRVTGLSVADDGTVTFADTVGSYWVVTGTIFLYLDDSMS---QIATGQA 357
PNGase_At       LLKDNKSHSFEIRVTGLSVADDGTVTFANTVNSYWVVTGTIFLYLDSSSSESHSTTTGQA 360
                ****************************:**.**************.* *   . :*****

Ani*            PEVNAPAPTFAVTRNLVQSRNGTNETLAYSVVAERTLTVKSSEYSWSQNLSYSNYGYLNQ 417
ATCC1015        PEVNAPAPTFAVTRNLVQSRNGTNETLAYSVVAERTLTVKSSEYSWSQNLSYSNYGYLNQ 417
An03g003300     PEVNAPTPTFAVTRNLVQSRNGTNETLAYSVVAERTLTVKSSEYSWSQNLSYSNYGYLNQ 417
PNGase_At       PEIYAPAPTLTVTRDLTQSPNGTNETLSYSVTAERTFTVKSSEYAWSQNLSYSNYGYLNQ 420
                **: **:**::***:*.** ********:***.****:********:***************

Ani*            QGLSQKNNQQTFGSNTIAQLTG-NKTTNEVIFEYPLICNTTYGLEDGLSISAWIRRGLDI 476
ATCC1015        QGLSQKNNQQTFGSNTIAQLTG-NKTTNEVIFEYPLICNTTYGLEDGLSISAWIRRGLDI 476
An03g003300     QGLSQKNNQQTFGSNTIAQLTG-NRTTNEVTFEYPLICNTTYGLEDGLSISAWIRRGLDI 476
PNGase_At       QGLSQKNNQQTSGTNTITQLTGNNKSTNEVTFQYPLICNTTYGLEDGLSISAWIRRGLDI 480
                *********** *:***:**** *::**** *:*************************

Ani*            ESTGG---LGVSTYTFTSGSLNLHTEQHGTAYYYEPSDDESSVSYGETFDVFGSNAGGVQ 533
ATCC1015        ESTGG---LGVSTYTFTSGSLNLHTEQHGTAYYYEPSDDESSVSYGETFDVFGSNAGGVQ 533
An03g003300     ESTGG---LGVSTYTFTSGSLDLHTEQHGTAYYYEPSDDESSVSYGETFDVFGSNAGGVQ 533
PNGase_At       SSTGGDGELGVSTYTFTSGPLDLHTEQYGTAYYFEPEDDESSVSYGETVDVWGSNAGGVE 540
                .****   ***********.*:*****:*****:**.***********.**:*******:

Ani*            YFRNVHAVNGTVVSDTDS 551
ATCC1015        YFRNVHAVNGTVVSDTDS 551
An03g003300     YFRNVHAVNGTVVSYTDS 551
PNGase_At       YARNVRAVNGTVVSDTES 558
                * ***:******** :
```

Figure 4.6: Multiple sequence alignment of amino acid sequences of three putative *A. niger* PNGase and PNGase *At*.
The alignment was produced using ClustalW2 at EBI. Red, amino acid substitution in PNGase *At*; Green: substitution in one *A. niger* sequence; Shaded grey: cleavage site in PNGase *At* for α- and β-subunit formation. * = conserved residue; : = conserved substitution; . = semi-conserved substitution.

At the amino acid level, the putative *A. niger** PNGase is completely identical to the ATCC 1015 protein and 98% identical to the *An*03g003300 (CBS 513.88) sequence with only eight amino acid substitutions in the 551 amino acid

sequence. The alignment of the *A. niger** and PNGase *At* sequences shows 91% sequence conservation. There are several notable differences between those sequences. Firstly, the PNGase *At* sequence contains seven additional amino acid residues, two separate three-residue insertions and one single amino acid insertion. Secondly, a stretch of ten amino acids (D343-A353, *Ani** numbering) shows the highest degree of variations clustered together. These variations include the insertion of three additional amino acids and four amino acid substitutions in PNGase *At*. Interestingly, this stretch of sequence is directly C-terminal to the cleavage site for the formation of the α- and β-subunit of mature PNGase *At*. Cleavage occurs between T356 and T357 (PNGase *At* numbering). T356 in PNGase *At* aligns with A353 in *A. niger* PNGase. The enzyme responsible for subunit formation has not been identified, but it appears to be a protease specific to the *A. tubingensis* extract from which PNGase *At* was originally isolated (Ftouhi-Paquin *et al.*, 1997). This was confirmed after recombinant PNGase *At* was shown to be expressed as a single chain protein in both a baculovirus expression system and *Aspergillus awamori*. Both forms, cleaved and uncleaved, show identical specific activity (Ftouhi Paquin *et al.*, 1998).

Chapter 5

Cloning and Expression of Genes Encoding Putative PNGases

5 Cloning and Expression of Genes Encoding Putative PNGases

5.1 Introduction

This chapter describes some of the attempts that have been made to express, purify and characterise the selected putative PNGases from *D. radiodurans*, *S. avermitilis*, *S. solfataricus* and *A. niger*.

For *Dra*PNGase, *Sav*PNGase and *Sso*PNGase recombinant protein expression in *E. coli* was initially tried and the results obtained for *Dra*PNGase and *Sav*PNGase will be described first. Following the *E. coli* expression system, the baculovirus expression system (BVES) was used for protein expression in insect cells. This system was tried for all of the four targets mentioned above, because it had been used successfully for the recombinant expression of the PNGase from *A. tubingensis*, PNGase *At* (Ftouhi Paquin *et al.*, 1998).

5.2 Methods

5.2.1 Detection of Sugars in Glycoconjugates

The detection of sugars in glycoconjugates, such as *N*-glycosylated peptides and proteins, was carried out to show the glycans had been removed by the activity of the various PNGase preparations. The DIG Glycan Detection Kit (Roche Applied Science) was used for this purpose. This detects glycoproteins that have been immobilised on nitrocellulose or PVDF membranes (2.24.1). First, hydroxyl groups in sugars are oxidised to aldehydes using sodium metaperiodate. Digoxigenin (DIG) is then covalently linked to these aldehyde groups *via* a chemical spacer group, and is subsequently detected by a digoxigenin specific antibody conjugated with alkaline phosphatase.

The samples to be analysed were treated and incubated as described in 2.26.1 (Gel shift assay), separated by SDS-PAGE (2.23) and then transferred to nitrocellulose membranes.

The detection procedure was performed according to the manufacturer's instructions (Method B).

In case of PNGase activity no or at least a less intense substrate band is expected in the assay sample containing both substrate and enzyme. In contrast, for the control sample containing the substrate glycoprotein only, detection of the substrate glycoconjugate would be expected.

5.2.2 pMAL™ Protein Fusion and Purification system

Genes cloned into a vector of the pMAL™-series (NEB®) are inserted downstream of the *E. coli* gene *malE*, which encodes the maltose binding protein (MBP). This translational fusion between the MBP and the target gene may result in increased solubility of the recombinant protein, as well as providing a tag for affinity purification and detection. The pMAL™-p2G vector encodes the complete MBP, including the signal peptide, leading to the export of the fusion protein into the periplasm. This vector is preferred for fusion proteins requiring disulfide bond formation, such as secreted proteins.

The fusion to MBP allows affinity chromatography to be used to purify the target protein on the basis of the affinity of MBP for amylose. A Genenase™ I cleavage site just upstream of the insertion facilitates the separation of the MBP target protein.

5.2.3 Affinity Purification of MalE-Fusion-proteins

For the purification of MalE-fusion-proteins the natural affinity of MBP for amylose is exploited. Amylose resin (NEB®) was poured into a glass Econo Column® (1.0×7.0 cm, Bio-Rad) to a column volume of 2.0 mL. The column was

washed with 8 column volumes of equilibration buffer, then the cell lysate was loaded at gravity flow followed by washing with 12 column volumes of equilibration buffer. The fusion-protein was eluted by applying equilibration buffer containing 10 mM maltose, and 0.5 mL fractions were collected and analysed by SDS-PAGE.

Equilibration Buffer:

Tris	20.0 mM
NaCl	0.2 M
EDTA	1.0 mM
pH 7.4	

The amylose resin was regenerated using the following sequence of washes:

H_2O_{pure}	3.0 CV
0.1% SDS	3.0 CV
H_2O_{pure}	1.0 CV
Equilibration buffer	3.0 CV

5.2.4 Detection of MalE-Fusion-protein on Nitrocellulose Membranes

For the detection of MalE fusion-proteins cell extracts or amylose affinity chromatography fractions were separated using SDS-PAGE and subsequently transferred onto nitrocellulose membranes. Anti-MBP antiserum from rabbit (NEB®) was the primary antibody. A horseradish peroxidase (POD)-labelled secondary anti-rabbit antibody (anti-rabbit IgG (whole molecule)-peroxidase antibody produced in goat; Sigma-Aldrich) was used to detect the bound primary antibody by chemiluminescence (2.24.3).

Following the protein transfer, the membrane was incubated overnight in PBS blocking buffer with slow shaking. The membrane was washed 3 × 10 minutes with shaking in PBS-Tween buffer followed by a 2 h incubation

at room temperature with the anti MBP-antiserum at a 1/10,000 dilution. The membrane was again washed 3 times with PBS-Tween buffer and subsequently incubated for 1 h at room temperature with the anti-rabbit-POD conjugate. Following another 3 washes with PBS-Tween, detection was carried out using BM Chemiluminescence Blotting Substrate (POD; Roche Applied Science) and a Fujifilm Intelligent Dark Box II (Fujifilm Corp.).

Phosphate Buffered Saline (PBS) Buffer:

K-Phosphate buffer	10.0 mM
NaCl	0.5% (w/v)
pH 7.2	

PBS-Tween Buffer:

Tween-20 0.1% (v/v) in PBS buffer

Blocking Buffer:

BSA 3.0% (w/v) in PBS-Tween buffer

5.2.5 TOPO®- and Gateway®-Cloning

The TOPO®- and Gateway®-Cloning Kits were purchased from Invitrogen™. The protocols and procedures used and described here were derived from the following three user manuals:

1. pENTR™ Directional TOPO® Cloning Kits (Version F)
2. *E. coli* Expression System with Gateway® Technology (Version E)
3. Baculovirus Expression System with Gateway® Technology (Version E)

5.2.5.1 Directional TOPO® Cloning

Directional TOPO® cloning exploits the duplex DNA binding and cleaving characteristics of topoisomerase I from Vaccinia virus, that have been described by (Cheng & Shuman, 2000; Shuman, 1991; Shuman, 1994). Briefly, a blunt-end PCR product is generated using a forward primer with the 5′ nucleotide sequence CACC. The TOPO®-charged cloning vectors contain a complementary overhang (GTGG), which anneals the 5′ end of the PCR product, ensuring that it is in the correct orientation. The PCR primers were designed according to the manufacturers' guidelines. Once a PCR product was obtained using KOD DNA polymerase (2.10), a TOPO® cloning reaction was set up containing the following components:

Fresh PCR product	0.5 to 4.0 µL (0.5:1–2:1 molar ratio PCR product:TOPO® vector)
Salt solution (supplied)	1.0 µL
H_2O_{pure}	to a final of 5.0 µL (before addition of vector)
TOPO® vector	1.0 µL

The reaction was incubated for 30 minutes at room temperature after which 2 µL of the reaction mixture were used to transform (2.17) chemically competent *E. coli* One Shot® TOP10 (Invitrogen™). Transformants were analysed by colony PCR (2.11) and subsequent DNA sequencing (2.21).

5.2.5.2 Cloning using Gateway® Technology

The Gateway® technology is a recombinational cloning system that allows the fast transfer of a target gene between different vectors and expression systems. The reaction processes and proteins involved in the site-specific recombination of the bacteriophage lambda both into and out of the *E. coli* chromosome are used in this system in an *in vitro* reaction (Hartley *et al.*, 2000; Landy, 1989). The key components are the specific attachment (*att*) sites (*attB*, *attP*, *attL*, *attR*) to which the recombination proteins bind and between

which the recombination occurs. **Figure 5.1** shows the two possible recombination reactions.

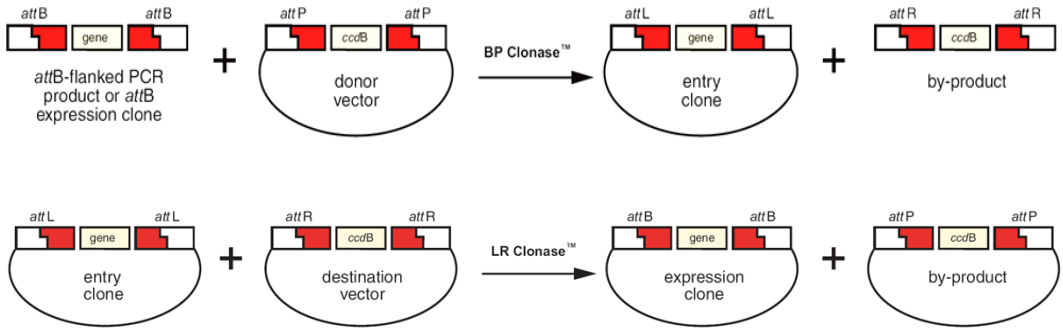


Figure 5.1: The BP- and LR reactions employed in the Gateway® Technology.
(Graphics taken from Gateway® Technology User Manual, Version E, Invitrogen™).

For the generation of expression vectors the LR reaction was used to transfer genes of interest from *attL*-containing TOPO® entry vectors into different *attR*-containing Gateway® destination vectors. A standard LR reaction was performed in a 1.5 mL microcentrifuge tube at room temperature and contained the following components:

Entry clone (50-150 ng/reaction)	1.0-7.0 µL
Destination vector (150 ng/µL)	1.0 µL
TE Buffer, pH 8.0	up to a final of 8.0 µL (before addition of enzyme)
LR Clonase™ II enzyme mix	2.0 µL

The reaction mixture was incubated at 25°C for up to 18 h. After the incubation, 1 µL Proteinase K was added followed by 10 minutes incubation at 37°C. 1 µL of this reaction was used to transform chemically competent Library Efficiency® DH5α *E. coli* cells (Invitrogen™). Transformants were analysed by colony PCR (2.11) and subsequent DNA sequencing (2.21).

5.2.6 Insect Cell Culture and Baculovirus Expression System (BVES)

For the production of recombinant proteins in eukaryotic cells the baculovirus expression system was used. This system is based on the generation of baculovirus particles carrying the target gene in their chromosome and the subsequent infection of insect cells with these virus particles. The protocols and procedures used and described here for the generation of recombinant virus particles, growth and maintenance of insect cell cultures and recombinant protein production were derived mainly from the following user manuals provided by the manufacturer (Invitrogen™):

Growth and Maintenance of Insect Cell Lines (Version K)
Baculovirus Expression System with Gateway® Technology (Version E)
Bac-to-Bac® Baculovirus Expression System (Version E)

The following figure indicates the main steps involved in the production of a target protein using the BVES with Gateway® Technology.

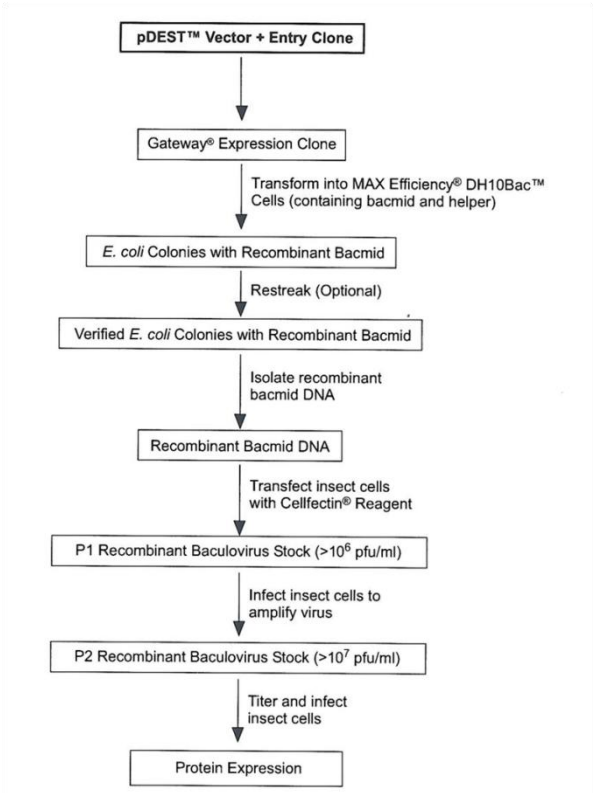


Figure 5.2: Experimental outline for the production of a recombinant target protein using the BVES with Gateway® Technology.
(adapted from 'Bac-to-Bac® Baculovirus Expression System' User Manual, Version E, Invitrogen™).

Sf9 cells that are pre-adapted to serum-free medium and suspension culture methods were used throughout these experiments. To avoid bacterial and fungal contamination all experiments were carried out under sterile conditions in a biohazard cabinet that had been exposed to at least 30 minutes of UV radiation and sprayed with a 70% (v/v) ethanol solution. All equipment that was used in the biohazard cabinet during the experiments was sprayed with a 70% (v/v) ethanol solution just before transfer to the cabinet. Disposables (tips, microcentrifuge tubes etc.), bottles and Erlenmeyer flasks were autoclaved twice before use.

5.2.6.1 Initiation and Maintenance of Spodoptera frugiperda (Sf9) cells

For initiation of Sf9 cells from frozen stocks, a vial containing 1 mL of 1.5×10^7 Sf9 cells was removed from liquid nitrogen storage and quickly transferred to a 37°C water bath until almost thawed. The cell suspension was then transferred to an Erlenmeyer flask containing 27 mL of pre-warmed serum-free medium Sf-900 II SFM. The flask was transferred to a 27°C incubator and incubated under gentle orbital shaking at 150 rpm until the cell density reached $> 2 \times 10^6$ viable cells/mL. Cells were then subcultured by seeding flasks with 3-5 $\times 10^5$ viable cells/mL.

Stock cultures were maintained as 50 mL cultures in 250 mL Erlenmeyer flasks. These cultures were grown until cell density reached 2-3 $\times 10^6$ viable cells/mL and then diluted into fresh medium to a density of 3×10^5 cells/mL. This subculturing procedure was generally performed twice weekly to maintain cells in optimal condition.

The cell density was determined using a haemocytometer. Cell viability was determined by adding 0.1 mL of a 0.4% trypan blue solution to 1 mL of culture. Cells that take up the stain, and therefore appear blue, are considered as not viable. Cell viability was calculated as number of viable cells divided by the total number of cells within the haemocytometer grid.

5.2.6.2 Transfection of Sf9 cells and Preparation of Viral Stocks

Bacmid DNA to be used for transfection of insect cells was purified using a PureLink® HQ Mini Plasmid Purification Kit, which produces DNA free from contaminants that might interfere with the transfection reagent and so decrease transfection efficiency. Cells used for transfections were in a range of 1.5-2.5 × 10^6 cells/mL at a viability of ≥95%.

Transfections were performed in 6-well plates (1 well/bacmid) under the following general conditions:

Number of cells	9.0 × 10^5 cells/well
Bacmid DNA	1.0-2.0 µg
Cellfectin® II reagent	6.0 µL

The cells were added to 2 mL Grace's Medium (unsupplemented) and allowed to attach for ~30 minutes at room temperature. Cellfectin® reagent and bacmid DNA were each mixed separately with 100 µL Grace's Medium (unsupplemented). These mixtures were then combined, mixed gently and incubated for 15-30 minutes at room temperature. After the incubation, the DNA-Cellfectin® mixture was added slowly to the cells and the plates incubated at 27°C for 3-4 h. The transfection mixture was then removed from the cells and replaced with 2 mL Sf-900 II SFM. Cells were incubated at 27°C for 3-5 days until signs of infection became visible. The medium was removed from the cells, centrifuged to remove cells and cell debris and the clarified supernatant kept at 4°C in the dark as P1 viral stock.

Amplification of the P1 viral stock was carried out in a 10 mL suspension culture at a cell density of 2 × 10^6 cells/mL. Cells were infected with 0.4 mL P1 viral stock (multiplicity of infection (MOI; ratio virus:Sf9 cells) of ~ 0.1) and incubated for 48-72 h at 27°C. Cells and cell debris were removed by centrifugation and the supernatant (= P2 viral stock) was stored at 4°C in the dark.

5.2.6.3 Determination of Virus Titres - Plaque Assay

The determination of the P2 viral stock titre was performed in 6-well plates, with 2 plates being required for every stock to be titered. 2 mL of a 5×10^5 cells/mL suspension were transferred into each well and the plates incubated at room temperature for 1 h. The medium was subsequently replaced with 1 mL of a serial dilution (in SF-900 II SFM) of the P2 viral stock to be titered. Dilutions used were 10^{-4} to 10^{-8} and a negative control containing no virus was included in each plate (each with duplicate). Following 1 h incubation at room temperature to allow infection of the cells, the stock dilutions were replaced with 2 mL of plaquing medium. The plates were left for 1 h at room temperature to allow the agarose overlay to harden and then moved to a 27°C humidified incubator where they were incubated for 7-10 days until plaques were visible. Plaques were visualised by staining with neutral red and counted. The following formula was used to calculate the viral titre:

$$Titre\ (pfu/mL) = number\ of\ plaques \times dilution\ factor \times \frac{1}{mL\ of\ inocculum\ per\ well}$$

Plaquing medium (volumes per assay):

Sf-900 SFM (1.3×)	30.0 mL
4% Agarose, melted	10.0 mL

The medium was equilibrated at 40°C prior to use.

The P2 viral stock was used for infection of Sf9 cells in protein expression trials and expression optimisation experiments.

5.3 Results & Discussion

5.3.1 D. radiodurans putative PNGase (DRA0325)

5.3.1.1 Expression and Purification of Full-Length DraPNGase

The vector used for expression of *Dra*PNGase in *E. coli*, pKS_OmpA_Dra_nosigpep, was constructed previously by Jessie Green (Dr. G.E. Norris, personal communication). Briefly, *D. radiodurans* genomic DNA and appropriate oligonucleotides were used to obtain a PCR product of 1965 bp comprising the complete ORF for *Dra*PNGase with exception of the first 150 bp coding for the signal peptide and the stretch of hydrophobic amino acids (3.3.2). Restriction sites for *Eco*RI (5′) and *Bam*HI (3′) were incorporated into the primer sequences and used to ligate the PCR fragment into the compatibly restricted vector pKS_OmpA_His (Loo *et al.*, 2002). This vector contains the OmpA leader sequence from *E. coli* that directs the protein to the periplasmic space of *E. coli*. The presence of a signal sequence and 12 cysteines in the native protein were reasons for expressing this protein in the *E. coli* periplasm, which, because of its oxidising environment, is more suited for disulfide bond formation. The expressed protein contains an N-terminal OmpA leader sequence and a C-terminal hexa-histidine tag. Three additional amino acid residues (GIL-) will remain at the N-terminus after successful protein export into the periplasm, and 9 additional residues will remain at the C-terminus, including the His_6-tag (-RDPHHHHHH). The final, mature protein has a predicted molecular weight of 66.16 kDa.

Small scale protein expression trials testing different *E. coli* strains, incubation temperatures and IPTG concentrations for induction led to identification of optimal expression conditions: *E. coli* RosettaBlue (DE3) grown in LB medium at 22°C and using 1 mM IPTG to induce protein production.

A 50 mL overnight culture of *E. coli* RosettaBlue (DE3) freshly transformed with pKS_OmpA_Dra_nosigpep was used to inoculate 3 L of LB medium containing the appropriate antibiotics (12.5 µg/mL tetracycline, 34 µg/mL chloramphenicol, 100 µg/mL ampicillin). The culture was grown in a “Mini

fors" fermenter (stirrer 400 rpm, aeration 1-2 lpm) at 37°C until an OD_{600} of ~0.5 was reached. The culture was then induced with 1 mM IPTG, immediately cooled down to 22°C, grown overnight for about 19 h and harvested at OD_{600} ~3.0. The cells were washed once with PBS then frozen in two equal parts at -80°C.

Cells from 1.5 L culture were resuspended in IMAC binding buffer (20 mM Na-phosphate-buffer, pH 7.4; 10 mM imidazole; 0.5 M NaCl) and lysed by 2 passages through a French press (~7 kpsi). The lysate was centrifuged (25 min, 30,000 *g*, 4°C) to remove unlysed cells and the insoluble cell debris, and the resulting supernatant was filtered through a 0.8 µm filter to remove any residual insoluble particles. A column (BioRad) with a column volume of 5 ml was packed with 'Chelating Sepharose Fast Flow' resin (GE Healthcare), which was then charged with Ni^{2+} ions and equilibrated in IMAC binding buffer (20 mM Na-phosphate pH 7.4, 10 mM imidazole, 0.5 M NaCl) according to the manufacturer's instructions. After loading the sample (~50 mL cell lysate supernatant) the following program was used to remove the unbound proteins and elute the bound His_6-tagged-protein: *(i)* 5 CV wash step with IMAC binding buffer; *(ii)* 5 CV wash step with 20 mM Na-phosphate pH 7.4, 0.5 M NaCl, 50 mM imidazole; *(iii)* linear gradient from 50-300 mM imidazole in 15 CV; *(iv)* 5 CV wash step with 20 mM Na-phosphate pH 7.4, 0.5 M NaCl, 500 mM imidazole. Chromatography was performed at 4°C using an Äkta FPLC system (Amersham Pharmacia Bioscience). *Dra*PNGase eluted at an imidazole concentration between ~140-230 mM (**Figure 5.3**).

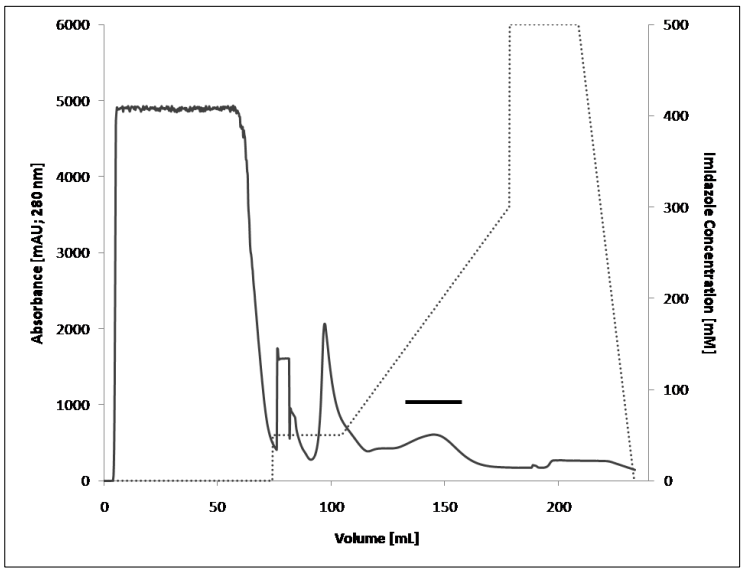


Figure 5.3: IMAC chromatogram of *Dra*PNGase.
Solid line: absorbance at 280 nm [mAU]; dotted line: imidazole concentration gradient [mM]. The bar indicates the peak corresponding to *Dra*PNGase.

Figure 5.4 shows the SDS-PAGE analysis of the IMAC purification of *Dra*PNGase. Six elution fractions (lanes 6 to 10 and one not shown in **Figure 5.4**) contained almost pure *Dra*PNGase. These fractions containing the putative PNGase were pooled and used in subsequent experiments.

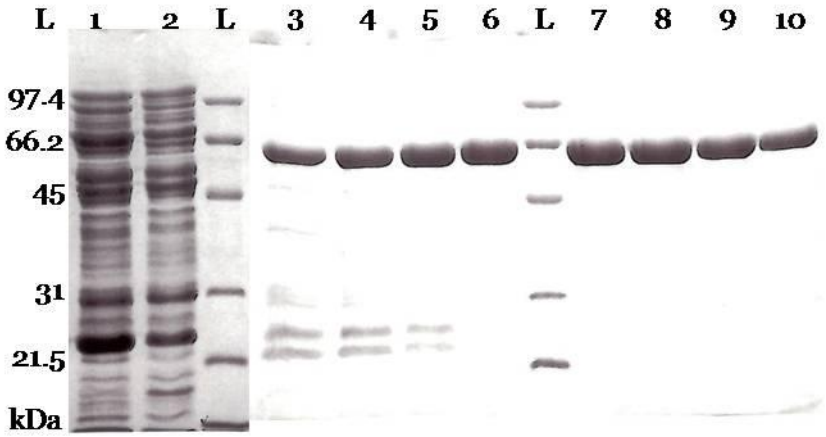


Figure 5.4: SDS-PAGE analysis of IMAC purification of *Dra*PNGase
(Coomassie G250 stained). L: Mw protein marker; Lanes 1 and 2: 2 fractions corresponding to peak eluting at ~100 mL (**Figure 5.3**); Lanes 3-10: 8 fractions corresponding to peak marked with a black bar in **Figure 5.3**.

The purified protein was then analysed by MALDI-TOF/MS to ensure that it is indeed the putative PNGase from *D. radiodurans*. The Coomassie-stained

bands in lanes 9 and 10 (**Figure 5.4**) were excised from the polyacrylamide gel, subjected to in-gel tryptic digest and the tryptic peptides analysed as described in 2.25. In the Mascot (Matrix Science) analysis of the tryptic peptides the top score was the 'probable *N*-glycosidase' from *D. radiodurans* strain R1 with a sequence coverage of 24%. This result confirmed that the protein purified was indeed *Dra*PNGase.

5.3.1.2 Determination of PNGase Activity of Full-Length DraPNGase

In order to investigate possible PNGase activity of the enzyme, SDS-PAGE-gelshift assays (2.26.1) were carried out on intact glycoproteins with a range of glycan structures. These were ovalbumin (complex bianntenary), RNase B (high-mannose) and α-1 acid glycoprotein (complex tetraantennary glycan chains) in their native and denatured forms. The assays contained 20 µL (protein concentration not determined) of the purified putative *Dra*PNGase and 0.4 mg/ml substrate (final concentration). The effect of reducing agent and protease inhibitors was tested by including 2 mM DTT and/or 1× EDTA-free complete mini protease inhibitor (Roche). The mixtures were incubated overnight at 25°C, then aliquots were analysed by SDS-PAGE (15% polyacrylamide). Activity was detected only for denatured RNase B, and was independent of presence or absence of DTT or protease inhibitor.

In order to determine the pH optimum for the enzyme, the reaction was repeated for native and denatured RNase B over a pH range from 4 to 9 (pH 4 and 5: 10 mM Na-acetate; pH 6-9: 10 mM BTP). Purified *Dra*PNGase was dialysed against buffers at pH 4, 5, 6, 7, 8 and 9 and substrate was prepared using the appropriate buffer. Assays were performed as described above. As already seen in the first assays, no deglycosylation of the native RNase B was observed at any pH. For denatured RNase B PNGase activity was virtually absent at pH 4 to 6 with a slight activity at pH 7. The best results were achieved at pH 9 in presence of protease inhibitor (lane 12, **Figure 5.5**). However, only partial deglycosylation was observed. This could be a result of a too short incubation time or denaturation of the *Dra*PNGase.

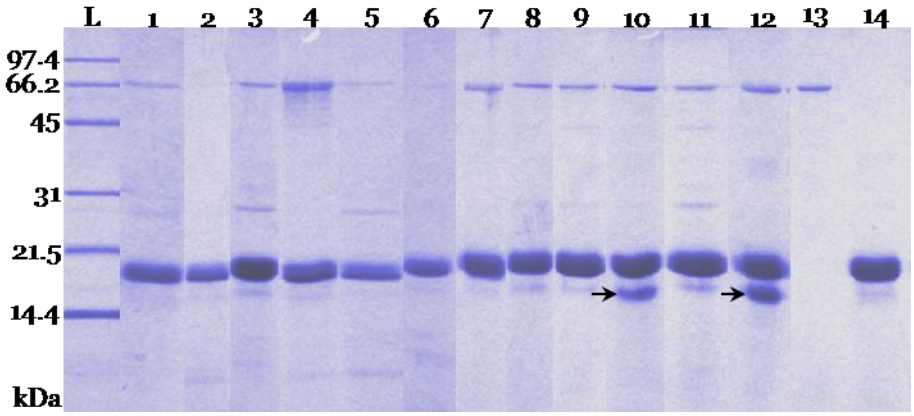


Figure 5.5: Determination of PNGase activity of putative *Dra*PNGase at different pH using native (n) and denatured (dn) RNase B as substrates.
Lane L: Mw protein marker; lane 1: pH 4 n; lane 2: pH 4 dn; lane 3: pH 5 n; lane 4: pH 5 dn; lane 5: pH 6 n; lane 6: pH 6 dn; lane 7: pH 7 n; lane 8: pH 7 dn; lane 9: pH 8 n; lane 10: pH 8 dn; lane 11: pH 9 n; lane 12: pH 9 dn; lane 13: *Dra*PNGase only at pH 8; lane 14: RNase B (dn) only at pH 8.

To test if extended incubation time and/or increased *Dra*PNGase concentration would result in complete processing of the substrate, six increasing *Dra*PNGase concentrations were used at 25°C for 19.5 and 43.5 h (data not shown). With increasing *Dra*PNGase concentration, deglycosylation of denatured RNase B increased, although the reaction never went to completion, regardless of the incubation time. However, after these incubation times, control samples that only contained *Dra*PNGase showed some degradation of the enzyme. This could indicate that the *Dra*PNGase preparation might be contaminated with proteases or that only a small fraction of the enzyme is actually correctly folded and catalytically active.

In order to confirm that the increased mobility of the observed lower band is actually due to deglycosylation of RNase B rather than proteolysis, a Western blot was performed followed by staining of glycan-containing proteins (5.2.1). Assays were performed as described above and proteins separated by SDS-PAGE were transferred onto a nitrocellulose membrane. The result is shown in **Figure 5.6**.

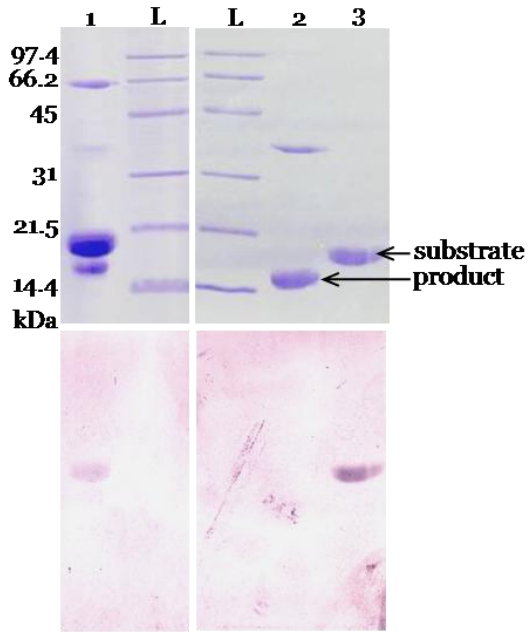


Figure 5.6: Digoxygenin (DIG) labelling of glycosylated RNase B as confirmation of the deglycosylating activity of putative *Dra*PNGase.
Lanes L: Mw protein marker; lane 1: dn RNase B incubated with putative *Dra*PNGase; lane 2: dn RNase B incubated with PNGase F (control); lane 3, dn RNase B. Top: Coomassie stained polyacrylamide gel; Bottom: Western blot.

The result showed that the lower band does not contain any glycan chains as it was not stained by DIG (**Figure 5.6**, lane 1). However, it should be noted that the transfer of the proteins in this part of the blot appears less effective than for the other proteins shown in lanes 2 and 3. The strongly stained band of unprocessed RNase B visible in the polyacrylamide gel in lane 1 appears to be very weak on the corresponding western blot. This may indicate that the protein amount transferred from the lower, weaker stained band (probably deglycosylated RNase B) is below the detection limit. Unfortunately, this experiment could not be repeated as there was insufficient *Dra*PNGase remaining from this preparation (5.3.1.1), and all following *Dra*PNGase isolations using the vector and methods described failed to show the activity observed here. A possible reason for this is that in the later experiments the protein might not have folded properly, although it was still soluble. It is well known that one of the major bottlenecks in recombinant protein production is the inability of the expressed, foreign protein to reach their native conformation when expressed in bacteria (Baneyx & Mujacic, 2004; Gasser *et al.*, 2008). Failing to fold correctly is generally a result of a combination of different events occurring in the host cell, including bottlenecks in transcription and translation,

undertitration of chaperones, improper codon usage, inefficient export (if protein is targeted to periplasm) and the inability to form correct disulfide bonds (Gasser *et al.*, 2008). Un- or partly folded proteins are prone to aggregation through exposed, normally buried hydrophobic patches and these aggregates are then usually deposited as inclusion bodies within the cell (Speed *et al.*, 1996; Villaverde & Carrio, 2003). Even if it is soluble there is still the possibility that the folding is not entirely correct. *Dra*PNGase contains 12 cysteine residues and therefore might require the formation of up to 6 disulfide bonds for correct folding, which might be problematic in *E. coli*, even if the protein is targeted to the periplasm as often the concentration of chaperones catalysing disulfide bonding in the *E. coli* periplasm (DsbA/DsbB: disulfide formation; DsbC/DsbD: rearrangement of non-native to native disulfides) cannot cope with the amount of recombinant protein produced (Gasser *et al.*, 2008; Nakamoto & Bardwell, 2004). It is also possible (and possibly more likely) that the export of the protein into the periplasmic space was inefficient, leading to incorrect folding of the protein within the cell. It has been observed for other secretory recombinant proteins that the over-expression of such a protein can lead to a blockage of the Sec-translocation machinery, inhibiting the translocation of the recombinant protein and endogenous secretory proteins, eventually resulting in cell death (Fu *et al.*, 2005). This possibility, however, is unlikely to be the case here as the cells grew as expected. It could be possible though that the translocation machinery reached saturation and therefore some recombinant protein remained in the cytoplasm, unable to reach its native conformation.

Another possibility is that the observed activity was indeed the result of proteolysis. However, the fact that only one specific product was observed, instead of a ladder of degradation products, one would expect from non-specific proteolysis makes this possibility less likely. Furthermore, it did not make any difference to the reaction if protease inhibitor was present or not. It is also unlikely that the product seen was produced by a cross-contamination with PNGase F, which has been used as a positive control. As obvious from **Figure 5.6**, and previous observations (T.S. Loo, personal communication), PNGase F is very efficient in deglycosylating denatured RNase B even at low concentrations. If PNGase F was present in the *Dra*PNGase preparation,

complete deglycosylation of denatured RNase B would have been expected. However, both proteolysis and PNGase F contamination cannot be ruled out at this stage.

Despite the latter preparations showing no apparent PNGase activity, crystallisation trials were set up with purified *Dra*PNGase (at 10 mg/mL) using the sitting-drop vapour diffusion method. The crystal screens used were Molecular Dimensions Structure screens 1 & 2, Hampton Research Crystal screen 1 & 2 and Molecular Dimensions crystallisation screen PACT*premier*™. All trials were set up in 96-well plates and incubated at room temperature. Molecular Dimensions Structure screens 1 & 2 and Hampton Research Crystal screen 1 & 2 were set up also for incubation at 10°C. However, no crystals or lead conditions were identified, which may be another indication that the protein was improperly folded.

Assuming that the main problem might have been incorrect folding of recombinant *Dra*PNGase, there are many possible ways to encourage correct folding of *Dra*PNGase in *E. coli*. Several such methods were tried for the full-length *Dra*PNGase. However, none of these trials led to a positive result:

(i) Co-expression of four periplasmic chaperones from the helper plasmid pTUM4 (DsbA, DsbC, FkpA, SurA), (Schlapschy *et al.*, 2006) to support correct protein folding in the periplasm

(ii) Co-expression of *Dra*PNGase with different chaperones/solubility promoting proteins (TrxA, DsbC, MalE) using the vector pETDuet (Novagen®; contains two multiple cloning sites)

(iii) Vector containing *dsbC* ORF for N-terminal fusion; DsbC is a periplasmic chaperone that promotes protein folding and contains a leader sequence for periplasmic localisation

From these methods (*(i)-(iii)*), soluble full-length *Dra*PNGase was obtained when *co*-expressed with thioredoxin using the vector pETDuet in *E. coli* Rosetta-gami B cells and 0.1 mM IPTG to induce protein expression (data not shown). The protein was purified using IMAC and SEC and activity assays were been performed, but no deglycosylation activity was detected using the gelshift assay (data not shown). Several attempts were made to crystallise the purified protein without success.

It might appear strange to express a protein that most likely requires the formation of disulfide bonds together with the thioredoxin TrxA. Reduced TrxA and TrxC react with disulfides in substrate proteins, leaving them reduced while becoming oxidised themselves in the process. The thioredoxin reductase TrxB recycles oxidised TrxA/C by reducing their active site disulfides using NADPH. However, it has been shown that in *trxB* mutants (such as *E. coli* Origami or Rosetta-gami strains) the function of TrxA and TrxC is reversed from reductases to oxidases due to their accumulation in a disulfide-bonded form in the absence of TrxB (Baneyx & Mujacic, 2004; Stewart *et al.*, 1998).

5.3.1.3 Cloning, Expression, Purification and Characterisation of a Truncated DraPNGase

As mentioned earlier, compared to PNGase F the putative *Dra*PNGase contains an additional N-terminal domain, which includes a protease-associated domain (3.3.2). In order to test whether the PNGase F-like domain exhibits PNGase activity when expressed separately without the N-terminal protease-associated domain, a truncated version of *Dra*PNGase (*Dra*PNGase-trunc) comprising only the PNGase F-like domain was cloned, expressed, purified and analysed for PNGase activity.

An expression vector was prepared containing only the coding sequence for amino acids 286 to 654 of the complete *Dra*PNGase (numbering corresponds to protein including predicted signal peptide). The truncation position is located in the N-terminal half of a predicted long helical region that forms the connection between the two putative domains, i.e. the PA-domain and PNGase F-like domain (Appendix 2, **Figure 10.3**). Even though this truncation might lead to the interruption of this predicted helix, it should not affect the folding of the PNGase F-like domain assuming it folds independently in a similar way to PNGase F (3.3.2.1; Appendix 2, **Figure 10.3**).

The expression vector used was a modified version of the vector pET32a(+), which contains an rTEV protease recognition site immediately upstream of *Dra*PNGase allowing removal of the thioredoxin-His_6-tag after IMAC purification, leaving only two additional amino acids at the N-terminus of

*Dra*PNGase-trunc. Oligonucleotides were designed accordingly, incorporating restriction sites for *Nco*I (5′) and *Bam*HI (3′), and used for PCR amplification of the target sequence employing pKS_OmpA_*Dra*_nosigpep as template for the reaction (primer combination: O4 + O5, **Table 2.6**). The resulting PCR product was purified by agarose gel electrophoresis and extraction, restricted with *Nco*I and *Bam*HI and ligated into pET32a(+)_*trxA*_His_6_rTEV, which had been cut using the same restriction enzymes, using T4 DNA ligase. The ligation reaction was then transformed into RbCl-competent *E. coli* XL1 Blue cells and colonies were analysed by colony PCR. The plasmid of a positive colony was isolated and transformed into the *E. coli* protein expression strain Origami B (DE3) for a small scale expression trial. This strain was chosen as it carries the *trxB* and *gor* mutations that provide a disulfide bond formation-promoting environment in the cytoplasm, which is expected to be important for *Dra*PNGase. Cells were cultured at 37°C until OD_{600} ~0.5, then protein expression was induced by the addition of 1 mM IPTG followed by overnight incubation at 25°C. The SDS-PAGE analysis of this expression trial (**Figure 5.7**) showed that the protein was expressed at a high level and that approximately 50% of the recombinant protein was soluble. The fusion protein produced with this vector has a predicted molecular weight of 52.9 kDa.

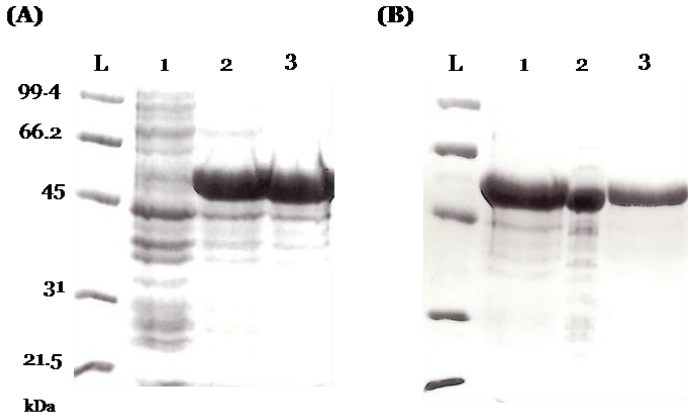


Figure 5.7: SDS-PAGE analysis of a small scale expression trial for *Dra*PNGase-trunc.
Panel A: normalised (for OD_{600}) whole cell extract samples at different time points. Lane L: Mw protein marker; lane 1: before induction; lane 2: 3 h post induction; lane 3: ~20 h post induction. Panel B: fractionation of the 20 h sample into soluble and insoluble fractions. Lane L: Mw protein marker; lane 1: whole cell extract; lane 2: soluble fraction; lane 3: insoluble fraction.

Following the expression trial, a larger scale protein expression experiment was performed to obtain enough protein for PNGase activity assays and crystallisation trials. For this, freshly transformed *E. coli* Origami B (DE3) cells (10 mL) were used to inoculate 1 L of LB medium in a "Mini fors" fermenter (stirrer 400 rpm, aeration 1-2 lpm). Cells were grown at 37°C until an OD_{600} of ~0.6 was reached. The culture was then induced with 1 mM IPTG, immediately cooled down to 25°C and grown overnight for about 19 h. Cells were then harvested, resuspended in IMAC binding buffer (20 mM Na-phosphate, 0.5 M NaCl, 20 mM imidazole; pH 7.4) and lysed with three passages through a French press (~7 kpsi). The lysate was centrifuged and the supernatant filtered (0.8 µM filter) to remove larger insoluble particles and then loaded onto an IMAC column (CV: 8 mL). Chromatography was performed using the following protocol (the elution buffers used consisted of IMAC binding buffer containing imidazole at the concentration indicated): 5 CV wash (IMAC binding buffer); 5 CV 50 mM imidazole; 2 CV 75 mM imidazole; 1 CV 100 mM imidazole; 1 CV 150 mM imidazole; 2 CV 200 mM imidazole; 1 CV 300 mM imidazole; 5 CV 500 mM imidazole. The SDS-PAGE analysis of this chromatography is presented in **Figure 5.8**.

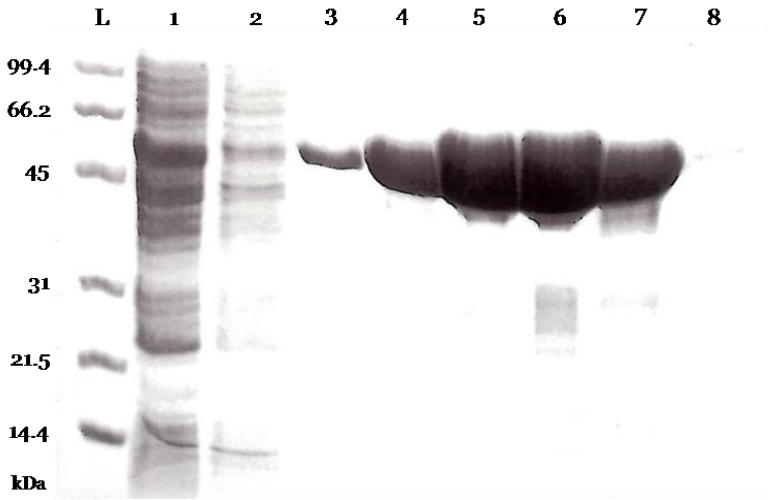


Figure 5.8: SDS-PAGE analysis of IMAC for *Dra*PNGase-trunc.
Lane L: Mw protein marker; lane 1: flow-through; lane 2: wash; lane 3: 50 mM imidazole; lane 4: 75 mM; lane 5: 100 mM; lane 6: 150 mM; lane 7: 200 mM; lane 8: 300 mM.

The protein concentrations of the main elution fractions (50-200 mM imidazole) were determined using the Bradford assay (2.22.1) and gave a total

amount of ~100 mg soluble *Dra*PNGase-trunc. Fractions were combined and the N-terminal TrxA-His_6-tag was removed using rTEV protease resulting in a 37.7 kDa protein. Residual uncleaved *Dra*PNGase-trunc, the tag and rTEV protease (which contains a His_6-tag) were separated from the cleaved protein by IMAC. The fraction containing unbound proteins, i.e. mainly the cleaved *Dra*PNGase-trunc, was then concentrated and subjected to SEC (Superdex 200 10/300GL) to remove minor contaminants and desalt the protein sample. The results of this chromatography are shown in **Figure 5.9**. Proteins were eluted using 20 mM HEPES buffer (pH 7.4) over 1.5 column volumes (35.6 mL).

The elution fractions marked with a black bar (**Figure 5.9**) were combined, concentrated and used to set up crystallisation screens in 96-well plates using the vapour-diffusion sitting drop method (21°C; 1:1 ratio of sample (at 10 mg/mL) and motherliquor). The following screens were tested: Molecular Dimensions Structure screens 1 & 2, Hampton Research Crystal screen 1 & 2 and Molecular Dimensions crystallisation screen PACT*premier*™. The crystallisation screens were inspected regularly, but no crystals or lead conditions were obtained.

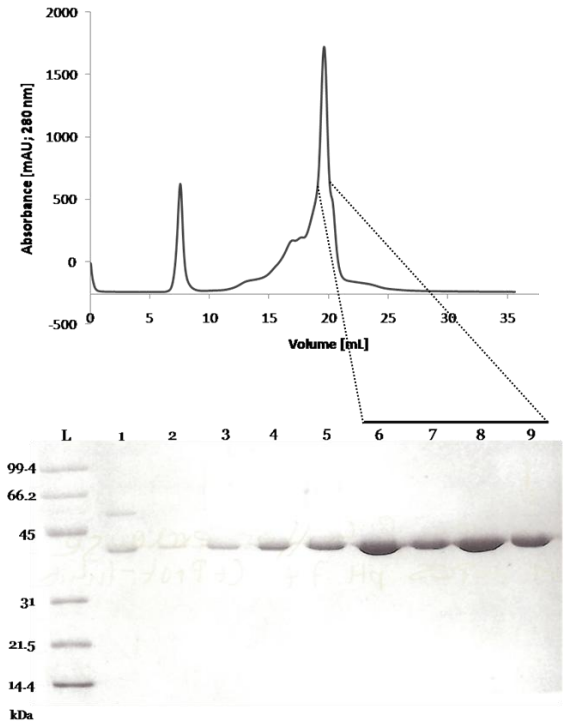


Figure 5.9: SEC of *Dra*PNGase-trunc (after rTEV cleavage).
The top panel shows an example chromatogram of this purification, which comprised multiple runs due to sample size and concentration. The bottom panel shows the SDS-PAGE analysis. Indicated with dashed lines and the black bar are the fractions that were combined for further experiments.

Purified *Dra*PNGase-trunc was also tested for PNGase activity using the SDS-PAGE-gelshift assays (2.26.1; 5.3.1.2) with native and denatured RNase B as substrates at pH 8. No PNGase activity was observed.

These results, combined with the results obtained for the full-length *Dra*PNGase, could indicate that the *E. coli* expression system might not be suitable to produce correctly folded *Dra*PNGase, full-length or truncated, due to the reasons discussed above (5.3.1.2). However, the good expression levels of soluble protein for both forms of *Dra*PNGase may also indicate that the problem is not incorrect folding. As mentioned earlier, misfolded proteins usually tend to aggregate leading to the formation and deposition of inclusion bodies in the cell.

It is also possible that *Dra*PNGase recognises substrates different to those used in the assays. This, however, is rather unlikely due to its similarity to PNGase F and the possible activity seen for the full-length protein (5.3.1.2).

Given these results and the enzyme kinetic results obtained for PNGase F towards the end of this project, the possibility that structural differences between *Dra*PNGase and PNGase F might be responsible for the apparent inactivity of *Dra*PNGase was also considered. As described in Chapter 3, the residues that have been directly associated with catalytic activity or substrate binding in PNGase F are conserved in *Dra*PNGase, with Glu118 (PNGase F numbering) being the exception. Glu118 in PNGase F is substituted with an alanine (Ala451) in *Dra*PNGase[21]. However, based on the PNGase F enzyme kinetics results presented in Chapter 8, two other amino acid substitutions in *Dra*PNGase could interfere with PNGase F-like activity. In *Dra*PNGase the residue aligning with Trp207 in PNGase F is a phenylalanine (Phe500) and, possibly more important, residue Trp191 (PNGase F) aligns with a histidine (His489).

[21] Site-directed mutagenesis of Ala451 to glutamate was performed, but led to the production of only insoluble *Dra*PNGase-A451E (data not shown).

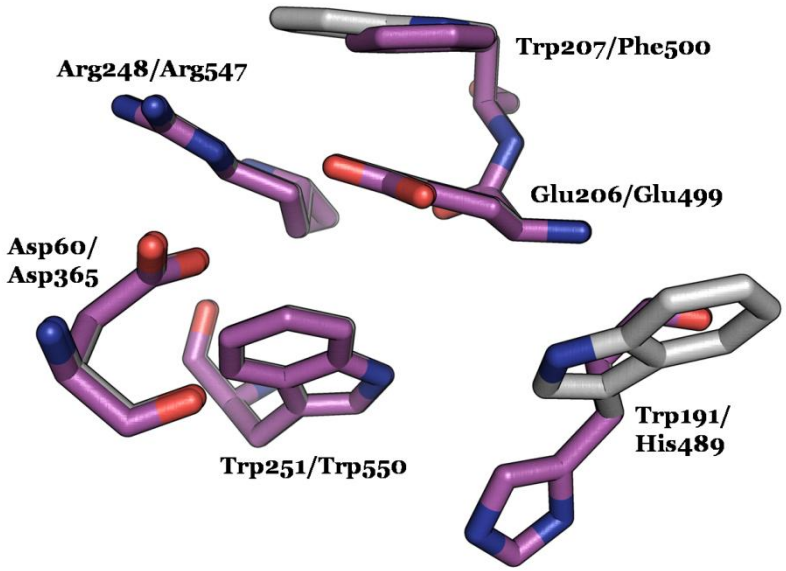


Figure 5.10: Superposition of the active site residues of PNGase F and *Dra*PNGase.
Shown in grey: PNGase F structure 1PGS (Norris *et al.*, 1994b)). Shown in magenta: Model of *Dra*PNGase generated in a fold recognition scan using the Phyre server (3.3.2.1). Residue numbers show PNGase F numbering first followed by *Dra*PNGase numbering (numbers are for the mature enzymes, i.e. without signal sequences).

As described in Chapter 8, the mutations of Trp207 and Trp251 to glutamine in PNGase F led to a strong decrease in catalytic activity (k_{cat}), possibly as a result of a decreased hydrophobicity around the proposed catalytic residue Glu206. Another tryptophan residue, Trp191 (PNGase F), is in close proximity to Glu206. The substitution of Trp207 with a phenylalanine might affect catalytic efficiency due to phenylalanine's slightly smaller, but still hydrophobic, side chain. However, the effect of this conservative substitution would be expected to be minimal. The other substitution, W191H (PNGase F numbering), could however, have a more profound negative effect on PNGase F activity. In the model shown in **Figure 5.10** the histidine residue in *Dra*PNGase points away from the main active site residues Glu206, Asp60 and Arg248. However, at this stage this is only a model and it could be possible that this histidine assumes a different conformation in the native *D. radiodurans* protein, which brings it closer to the active site, especially Glu206. This could possibly interfere with PNGase F activity in a similar way as shown for the W207Q and W251Q mutants. At this stage there is no proof for this theory, but the strong negative effect of mutating the tryptophans in the vicinity of Glu206 on PNGase F

catalytic activity suggests that substitution of Trp191 with a histidine could also impair activity.

Finally, it is also possible that this protein is simply not a PNGase, but has adopted a different function such as glycan binding and transport.

5.3.2 S. avermitilis MA-4680 putative PNGase (Sav1567)

5.3.2.1 Cloning, Expression and Purification of SavPNGase

As mentioned earlier (3.3.4.1), *Sav*PNGase is predicted to be secreted. Therefore, to promote correct folding of the protein, *Sav*PNGase was inserted into the vector pMAL-p2g in order to express a maltose binding protein-(MBP)-*Sav*PNGase-translational fusion protein (MBP N-terminal) where MBP contains a signal sequence resulting in export of the fusion protein to the periplasm. Furthermore, the fusion protein can be purified by affinity chromatography using amylose resin and the MBP-tag can be cleaved from the *Sav*PNGase using the protease Genenase™ I.

For cloning of the *Sav*PNGase into the vector pMAL-p2g, a 1.6 kbp DNA fragment was amplified starting with codon 25 and containing a 3′-terminal *Bam*HI restriction site (primer combination: O9 + O10, **Table 2.6**). The vector was prepared by sequential restriction using the enzymes *Sna*BI (blunt) and *Bam*HI. Ligation and transformation of the ligation reaction into *E. coli* XL1 Blue led to transformants carrying a plasmid showing the expected bands in an analytical restriction digest (data not shown).

Small scale expression trials were performed in order to determine the best expression conditions for soluble *Sav*PNGase. Different incubation temperatures (25°C, 37°C), IPTG concentrations for induction of protein production (0.3 mM, 0.5 mM and 1 mM) and induction at different cell densities (OD_{600}) were tested (data not shown). The best conditions found were expression in *E. coli* TB1 at 37°C adding 1 mM IPTG at OD_{600} ~0.4 followed by incubation for 5 to 6 hours (data not shown). A protein with the expected

molecular weight for the MBP-*Sav*PNGase-fusion protein of ~101 kDa (MBP 42.5 kDa; *Sav*PNGase 58 kDa) was expressed in soluble form (**Figure 5.11**).

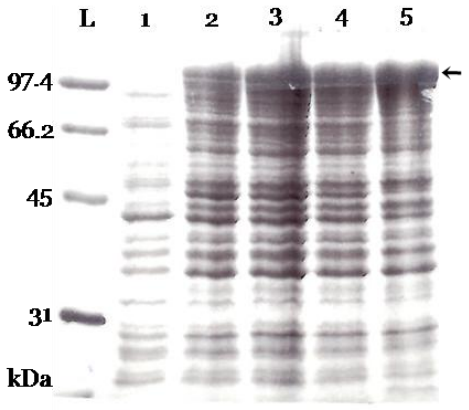


Figure 5.11: SDS-PAGE analysis of *Sav*PNGase small scale expression trial. Whole cell extracts (normalised for OD_{600}) of cells at different time points. Lane L: Mw protein marker; lane 1: uninduced; lane 2: 3 h post induction (p.i.); lane 3: 5 h p.i.; lane 4: 6 h p.i.; lane 5: 22 h p.i.. Separation of cell lysate into soluble and insoluble fractions was performed later and showed that most of the recombinant *Sav*PNGase was soluble.

For purification of MBP-*Sav*PNGase, 100 mL of freshly transformed *E. coli* TB1 were grown under the conditions determined in the small scale expression trial. Cells were harvested 5.5 hours after induction, resuspended in equilibration buffer (5.2.3) and then lysed using 3 French press passages (~7 kpsi). After removal of the cell debris by centrifugation, the supernatant (crude extract) was loaded by gravity flow onto a column containing ~2 mL of pre-equilibrated amylose resin (NEB) and chromatography was performed as described in 5.2.3. Elution fractions were analysed by SDS-PAGE, but, in contrast to the small scale expression trial, no protein band with the expected molecular weight of 101.5 kDa was observed. Instead, two dominant proteins with estimated molecular weights of ~42 kDa and ~60 kDa were present (**Figure 5.12**).

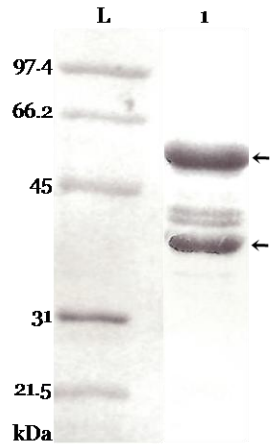


Figure 5.12: Amylose affinity chromatography purification of MBP-*Sav*PNGase.
Lane L: Mw protein marker, lane 1: one of several elution fractions (other elution fractions showed the same band pattern). Arrows indicate the two dominant protein bands at ~42 and ~60 kDa.

There are three possibilities for the appearance of these two main bands (besides some weaker contaminating protein bands): *(i)* MBP is proteolytically cleaved during the course of purification, *(ii) Sav*PNGase is auto-proteolytically cleaved to form a heterodimer, as was initially suggested, but later disproved for PNGase *At* (Ftouhi-Paquin *et al.*, 1997; Ftouhi Paquin *et al.*, 1998) or *(iii)* proteolysis by *E. coli* proteases occurred because no protease inhibitor was added. The latter possibility was initially ruled out when the same SDS-PAGE band pattern was seen after cell lysis and purification were repeated in the presence of protease inhibitor (EDTA-free Mini Complete Protease Inhibitor cocktail, Roche).

To further investigate the nature of the two main proteins, the elution fractions (except for the two main fractions that would be used for activity measurements) were combined and re-chromatographed using amylose affinity chromatography. In case *(i)* only the band of about 42 kDa (MBP) should appear in SDS-PAGE and the *Sav*PNGase should be present in the unbound fraction (flow-through). In the second case *(ii)*, self-cleavage and heterodimer-formation, the same pattern of two major bands at approximately 42 (*Sav*PNGase-C-term.) and 60 kDa (MBP + *Sav*PNGase-N-term.) should be seen. SDS-PAGE analysis showed that both protein bands were still present indicating that these proteins might form a non-covalent complex under native conditions. However, even though the presence of protease inhibitor during cell

lysis and purification did not change the observed SDS-PAGE band pattern, it could still not be completely ruled out that the recombinant protein was proteolytically degraded after cell lysis. The result of the re-chromatography could also indicate that both proteins contained the MBP-tag. This would suggest that *Sav*PNGase was degraded by some specific protease.

In order to determine if one or both fragments (60 and 42 kDa) contained the MBP, an aliquot of one elution fraction obtained from the purification was treated with Genenase™ I. The recognition site for this protease is located between the MBP and *Sav*PNGase. If the 60 kDa protein contained the MBP with the N-terminal domain of *Sav*PNGase, Genenase™ I should cleave the protein into two fragments, the MBP (42.5 kDa) and the N-terminal domain of the PNGase. If the two bands represent the *Sav*PNGase and MBP, Genenase™ I treatment should not alter the band pattern.

Genenase™ I treatment resulted in cleavage of the 60 kDa protein into a ~42 kDa (possibly MBP) fragment and a ~22 kDa (possibly N-terminus of PNGase) fragment (**Figure 5.13** (A), lane 1). This result indicated that the 60 kDa band comprises the MBP and an N-terminal part of *Sav*PNGase and led to the conclusion that the smaller protein at ~42 kDa represents the C-terminal domain of the *Sav*PNGase.

To further confirm this conclusion, a Western blot was performed using anti-MBP antiserum as the primary antibody, and anti-rabbit-PDO-antibody as secondary antibody. The Western blot and the corresponding SDS-PAGE gel are shown in **Figure 5.13**.

The Western blot showed that both protein bands contained the maltose binding protein as well as some weaker bands in between these two main bands and above the ~60 kDa band. The positive control (**Figure 5.13**, lane 6) demonstrated that the result is not due to non-specific binding of one of the antibodies, as only one band with the expected size appeared for this sample. No antibody binding was shown for the negative control (Figure 5.13, lane 7). This result indicates that the 101 kDa protein expressed at 37°C is most likely to be unstable and therefore susceptible to protease degradation rather than being specifically cleaved.

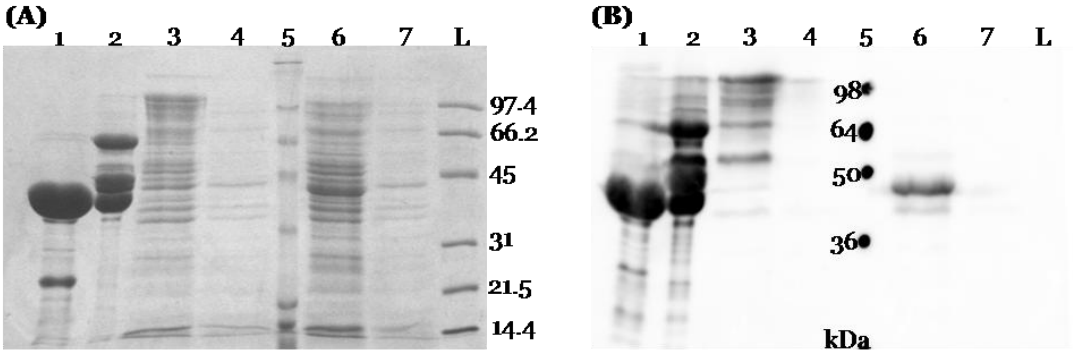


Figure 5.13: SDS-PAGE (A) and corresponding Western blot analysis (B) for *Sav*PNGase.
Lane 1: Amylose affinity chromatography elution fraction + Genenase™ I; lane 2: Amylose affinity chromatography elution fraction before Genenase™ I treatment; lane 3: *E. coli* TB1 (pMAL-p2g::*Sav*PNGase) crude extract induced; lane 4: *E. coli* TB1 (pMAL-p2g::*Sav*PNGase) crude extract uninduced; lane 5: pre-stained Mw protein standard: lane 6: positive control *E. coli* TB1 (pMAL-c2g::*Sso*PNGase[22]) crude extract induced; lane 7: control *E. coli* TB1 (pMAL-c2g::*Sso*PNGase[1]) crude extract uninduced; lane L: Mw protein standard.

Further expression trials were carried out at reduced temperatures (16°C and 25°C) in *E. coli* TB1 and *E. coli* Origami (DE3). Because the latter strain provides a less reducing environment in the cytoplasm due to deletion of thioredoxin reductase (*trxB*) and glutathione oxido-reductase (*gor*), it was trialled to see if this environment would have an expected beneficial effect on correct protein folding of an exported protein. MBP-*Sav*PNGase was detected in the insoluble cell fraction in both *E. coli* strains under all tested conditions.

It is possible that another problem with the pMAL-p2g expression system was the rather large size of the fusion protein. Therefore, *Sav*PNGase was cloned into bacterial expression vectors that contain smaller fusion partners such as the OmpA leader sequence (pET32a(+)_*ompA*_His_6_rTEV) and thioredoxin (pET32a(+)_*trxA*_His_6_rTEV) for periplasmic and cytoplasmic protein expression, respectively. Different expression conditions were investigated and although expression of *Sav*PNGase was detected when fused to TrxA and expressed in *E. coli* Origami (DE3) at 25°C, no soluble protein was obtained (data not shown).

In conclusion, the most convincing explanation for the results obtained here is that *Sav*PNGase was unable to assume its native conformation. Being fused to

[22] This vector expresses a MBP (lacking the signal sequence) plus additional five amino acids from the putative PNGase from *S. solfataricus*. A frame shift mutation was introduced into this vector as a result of incorrectly working restriction endonuclease *Sna*BI.

MBP, *Sav*PNGase was possibly held in solution by the highly soluble maltose binding protein, a fusion partner often used as solubility tag. *Sav*PNGase itself might not have been folded correctly and therefore been susceptible to proteolytic degradation as indicated by the Western blot experiment shown in **Figure 5.13**. This result showed that, while still partly present as an intact protein in the whole cell lysate (lane 3), proteolysis had already started to occur. With further processing of the whole cell extract further proteolysis led to almost complete degradation of the *Sav*PNGase part of the fusion protein. Only the N-terminal ~20 kDa region of the *Sav*PNGase appeared to be fairly stable as part of the ~60 kDa MBP-*Sav*PNGase fragment. While the C-terminus is susceptible to proteolysis, the MBP possibly protects the *Sav*PNGase N-terminus from rapid proteolytic degradation. In contrast to the *Dra*PNGase, there is no doubt that *Sav*PNGase failed to fold correctly under all tested conditions, making further analyses of this putative PNGase impossible. The factors possibly involved in incorrect folding of *Sav*PNGase are those previously described for *Dra*PNGase (5.3.1).

5.3.3 Summary of Results for Recombinant Protein Expression in E. coli and Insect Cells using Gateway® Technology

Several problems were encountered during expression of the putative PNGases from *D. radiodurans*, *S. avermitilis* and *S. solfataricus* in *E. coli*. Expression of *Sso*PNGase in *E. coli* was tried, but was unsuccessful (data not shown). This, however, was not unexpected due to the high probability of *Sso*PNGase being N-glycosylated (3.3.4.2). *E. coli* is generally not able to glycosylate proteins, and this post-translational modification can, however, be important for correct folding of glycoproteins. Following these problems another expression host (insect cells) was trialled for recombinant expression of these three proteins and the putative PNGase from *A. niger*. The latter protein is highly similar to PNGase *At* from *A. tubingensis*, which has been successfully expressed in insect cells using the baculovirus expression system (BVES; (Ftouhi Paquin *et al.*, 1998)). As the putative PNGases from *S. avermitilis* and *S. solfataricus* are type II, PNGase A/*At*-type PNGases (3.3.5), the successful

expression of PNGase *At* using the BVES led to the decision to perform expression experiments using this system. *Dra*PNGase was also included, but here two different truncated versions of this protein were cloned. The positions of these truncations were chosen based on a disorder prediction by the PONDR®-server (www.pondr.com; Appendix 3). At the N-terminus of *Dra*PNGase, 93 amino acid residues were removed (numbering based on the full-length protein, excluding the predicted signal sequence). Two different truncations points were chosen for the C-terminus, the first after residue 561 and the second after residue 613.

For recombinant protein expression in insect cells, the TOPO® and Gateway® Systems and parts of the Bac-to-Bac® Baculovirus expression system (Invitrogen™) were employed using the methods described in 5.2.5 and 5.2.6. The success of each cloning step was verified by colony PCR. The nucleotide sequences of the fragments inserted into the entry vectors were verified by DNA sequencing.

Table 5.1 summarises the intermediate steps and expression results obtained for recombinant protein production in *E. coli* and Sf9 cells using Gateway® technology.

Table 5.1: Summary of results obtained for recombinant protein production in *E. coli* and Sf9 cells using Gateway® technology.

			E. coli			**BVES**				
Protein	**PCR**	**pENTR**	**pDEST15**	**pDEST17**	**Expression (soluble)**	**pDEST8**	**pDEST10**	**Bacmid**	**Phage stock (P2 titre)**	**Expression**
***Dra*PNGase-561**	✓	✓	✓	✓	×	-	✓	✓	✓	×
***Dra*PNGase-613**	✓	✓	✓	✓	×	-	✓	✓	✓	×
***Sav*PNGase**	✓	✓	×	×	-	-	×	-	-	-
***Sso*PNGase**	✓	✓	✓	✓	×	-	✓	✓	✓	×
***Ani*PNGase**	✓	✓	-	-	-	✓	-	✓	✓	×

First, entry clones were generated using directional TOPO®-cloning (5.2.5.1). Primers were designed to generate PCR products that would be suitable for the fusion of N-terminal purification tags after transfer into appropriate Gateway® destination vectors, i.e. no start codon was included in the primer sequence. Only *Ani*PNGase was cloned for subsequent expression as an untagged protein as previously described for PNGase *At* (Ftouhi Paquin *et al.*, 1998). The inserts for the different targets were obtained using the following primer combinations (**Table 2.6**): O1 + O2 for *Ani*PNGase, O6 + O7 for *Dra*PNGase-561, O6 + O8 for *Dra*PNGase-613, O11 + O12 for *Sav*PNGase, and O13 + O14 for *Sso*PNGase. All targets were successfully inserted into pENTR vectors. For subsequent protein production in *E. coli*, inserts were transferred from the entry vectors into the destination vectors pDEST15 (N-terminal GST tag) and pDEST17 (N-terminal His_6 tag) as described in 5.2.5.2. This step was performed for all targets except *Ani*PNGase and was successful in all cases except for *Sav*PNGase (data not shown). It is not clear why this sequence was resilient to transfer from the entry vectors into any of the destination vectors as the sequence was correct. Small scale expression trials were performed using *E. coli* BL21-AI, but no soluble protein was obtained for either of the two *Dra*PNGase proteins and no recombinant protein could be detected for *Sso*PNGase (data not shown).

For protein expression using the BVES (**Figure 5.2**) the inserts were transferred from the entry vectors into either pDEST10 (N-terminal His_6 tag) or pDEST8 (no tag). These destination vectors were then transformed into *E. coli* DH10Bac cells, which contain the bacmid (baculovirus shuttle vector) and a helper plasmid. Recombinant bacmids are generated by transposing a mini-Tn*7* element from a donor plasmid (pDEST™ vectors) to the mini-*att*Tn*7* attachment site on the bacmid. The Tn*7* transposition functions are provided by the helper plasmid. Following verification of the bacmid, Sf9 cells were transfected and a P1 stock was isolated and amplified (5.2.6.2). The titres of the P2 stocks were determined (5.2.6.3) and the following values were obtained:

(i) *Dra*PNGase-561: 5.7×10^7 pfu/mL

(ii) *Dra*PNGase-613: 7.0×10^7 pfu/mL

(iii) *Sso*PNGase: 9.3×10^8 pfu/mL

(iv) *Ani*PNGase: 2.0×10^8 pfu/mL

The P2-stock titres were well above the value given as a guideline by the manufacturer (> 10^7 pfu/mL) and were used for infection of Sf9 cells in protein expression trials and expression optimisation experiments. After initial expression trials failed to show any recombinant protein the following variables were optimised (based on manufactures recommendations and (Farrell & Iatrou, 2004)):

(i) Cell density at time of infection (6×10^5, 1×10^6, 2×10^6 cells/mL)

(ii) Multiplicity of infection (MOI; 2.5, 5, 10 pfu/cell)

(iii) Time (samples taken after 2, 3, 4, 5 days post infection)

However, no recombinant protein was obtained for any target (data not shown). It is difficult to explain why no protein was produced as all steps leading up to the infection of the cells were successful, including generation of the recombinant bacmids that were used to infect the Sf9 cells. One possibility could be the insect cell line used. It is possible that the expression level of the recombinant proteins was too low to be detected and that the use of other cell lines such as High Five™ (ovarian cells of the cabbage looper, *Trichoplusia ni*), which can, according to the manufacturer, in some cases yield higher amounts of recombinant protein, may be required.

However, due to the numerous problems encountered during this project and time constraints, this line of research was abandoned and the focus of this work shifted to the generation, expression and characterisation of recombinant (r)PNGase F and its site-specific mutants (Section II).

Section II

rPNGase F and Its Site-Specific Mutants – Structural and Functional Characterisation

Chapter 6: rPNGase F Site-Specific Mutants: Generation, Expression and Purification

Chapter 7: Structural Characterisation of rPNGase F

Chapter 8: Kinetic Characterisation of rPNGase F and its Site-Specific Mutant Proteins

Anybody who has been seriously engaged in scientific work of any kind realizes that over the entrance to the gates of the temple of science are written the words: 'Ye must have faith.'

Max Planck (1858-1947), German Physicist.

Chapter 6

rPNGase F Site-Specific Mutants: Generation, Expression and Purification

6 PNGase F Site-Specific Mutants: Generation, Expression and Purification

6.1 Introduction

PNGase F has been isolated and characterised previously by different groups as described in Chapter 1 (1.2.1.1). Site-specific mutagenesis studies facilitated the location of the active site within the whole molecule and identified the residues involved in substrate binding and catalysis (Kuhn *et al.*, 1995). Kuhn *et al.* proposed that Asp60, Glu206 and H_2O were involved. However, the exact catalytic mechanism has not been described so far. Based on these results and preliminary results by Loo *et al.* (personal communication Dr. G.E. Norris), a catalytic mechanism for PNGase F was proposed as shown in **Figure 6.1**.

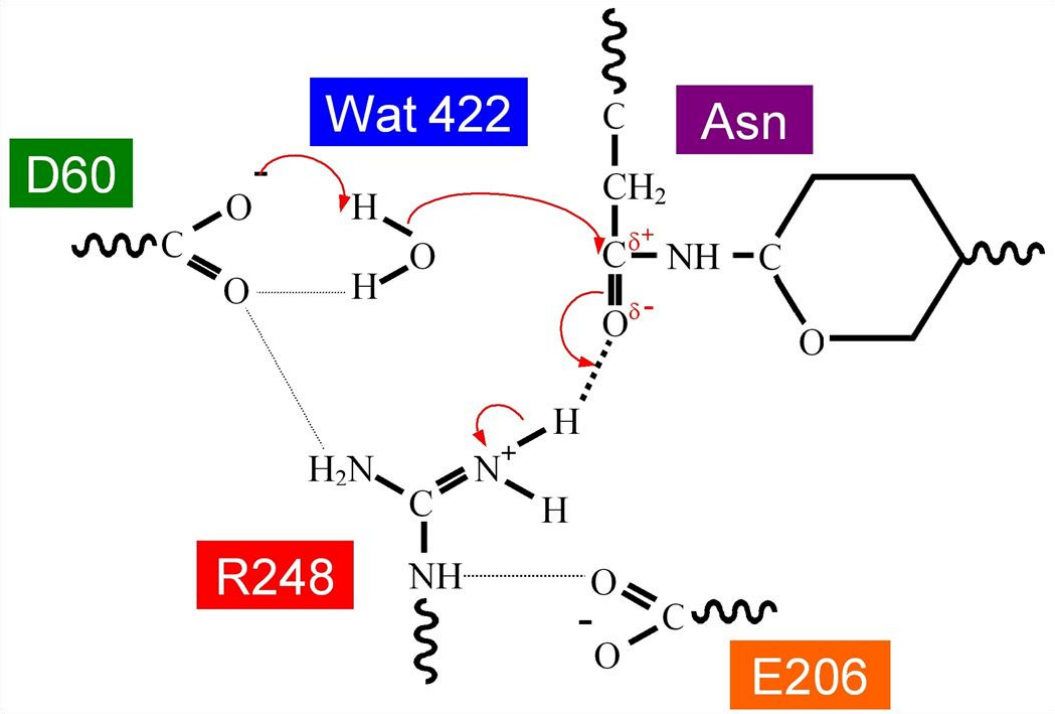


Figure 6.1: Proposed mechanism for PNGase F.
The numbering of the water is based on the PNGase F structure published by Norris *et al.* (1994).

In this mechanism, Arg248 forms a hydrogen bond to the carbonyl oxygen of the *N*-glycosidic linkage, making the Asn-carbonyl carbon more susceptible for a nucleophilic attack from an OH^- -ion. This hydroxide ion is probably formed by a proton transfer from a bound water molecule to the catalytically essential

residue Asp60, which can then transfer the hydrogen to the nitrogen of the glycosidic linkage, leading to its cleavage and formation of the glycosylamine as an intermediate. This mechanism relies on the pK_a of Asp60 being raised from the usual 4.5-5.4 to ~8.5, the pH optimum of the reaction. The proposed stabilising role of Glu206 (Kuhn *et al.*, 1995) probably involves the formation of a hydrogen bond between its carboxylate and the *NH* of the guanidinium group of Arg248. This bonding in turn could stabilise the hydrogen bond between Arg248 and the carbonyl oxygen of the glycosidic linkage by increasing the electron density on the R248 guanido group and therefore lowering its pK_a. The function of Glu206 cannot only be its ability to form these hydrogen bonds, as mutating this residue to glutamine, which theoretically can form the same H-bonds, abolishes more than 99% of the activity. There must therefore be a charge related function as well.

To test the validity of this proposed catalytic mechanism, a variety of site-specific mutant proteins was produced for further characterisation of the catalytic mechanism employed by PNGase F. The mutations selected for this study and the rationale for their selection are summarised in **Table 6.1**.

Table 6.1: Mutations introduced into the PNGase F ORF.

Mutation	Proposed function	Rationale for mutation
W59Q	Generation of a hydrophobic environment in/around the active site, Asp60 in particular	Exchange of a hydrophobic residue with a polar residue, testing the importance of the hydrophobic environment on the catalytic mechanism.
D60C	It is thought that Asp60 accepts an H^+ from a water molecule to activate a bound water, which acts as the nucleophile; this is possible only if its pK_a is raised from the usual 4.5-5.4 as the optimal pH of the reaction is 8.5.	Investigating the role of Asp60 in being able to accept a proton from the bound water. Cysteine should not be able to accept an H^+ at pH 8.5, especially if the environment artificially raises the pK_a as proposed.
I82Q **I82R**	Generation of a hydrophobic environment in/around the active site.	Investigating the importance of the hydrophobic environment of Asp60 on the catalytic mechanism.
W86F	Substrate binding	Investigate the importance of the ability to H-bond

W207Q	Generation of a hydrophobic environment in/around the active site.	Investigating the importance of the hydrophobic environment of Asp60 on the catalytic mechanism.
R248K	Arg248 forms a hydrogen bond to the carbonyl oxygen of the *N*-glycosidic linkage, making the Asn-carbonyl carbon more susceptible to a nucleophilic attack.	Probing the architecture of the active site.
R248Q		Probing the effect of charge on the mechanism by exchanging the charged residue with a polar residue.
W251Q	Generation of a hydrophobic environment in/around the active site	Investigating the importance of the hydrophobic environment of Asp60 on the catalytic mechanism.
V257N		
V257K		

6.2 Methods

6.2.1 Generation of Site-Specific Mutants of rPNGase F

In vitro site-specific mutagenesis was performed based on the method used in the QuikChange® Site-Directed Mutagenesis Kit (Stratagene).

2 primers containing the desired mutation were designed and used to amplify the complete vector containing the gene of interest. The primers were 25-45 bases long, with a melting temperature T_m of ≥78°C and had the mutation in the middle of the sequence with approximately 10-15 bases on either side. The reaction mix was prepared on ice in a nuclease-free, thin-walled 0.2 mL PCR tube containing the components given in **Table 6.2**.

Table 6.2: Composition of a Mutagenesis-PCR reaction using KOD DNA-polymerase

Component	Volume
10× KOD reaction buffer	1.0 µL
25 mM $MgSO_4$	0.4 µL
Template DNA (10 ng/µL)	1.0 µL
Sense (5′) Primer (10 µM)[23]	1.2 µL
Anti-Sense (3′) Primer (10 µM)[1]	1.2 µL
dNTPs (2 mM)	1.0 µL
KOD DNA polymerase	1.0 µL
H_2O_{pure}	to 10.0 µL

The program used for the amplification is shown in **Table 6.3.**

Table 6.3: Thermal profile used for site-specific mutagenesis of PNGase F.

Cycles		Temperature	Time
Initial Denaturation			
1×		94°C	5 min
Amplification			
	Denaturation	94°C	30 sec
16×	Annealing	55°C	45 sec
	Elongation	72°C	1 min/kbp
Final Elongation			
1×		72°C	10 min

Following the amplification, 10 U of the restriction endonuclease *Dpn*I were added directly to the reaction mixture, which was then incubated for 1 h at 37°C. *Dpn*I specifically digests methylated and hemimethylated DNA and therefore only degrades the dam methylated parental DNA used as template, leaving behind the newly synthesised DNA containing the mutation.

3 µL of the reaction mixture were then used to transform chemically competent *E. coli* XL1-Blue cells. Plasmids isolated from resulting transformants were analysed by DNA sequencing to confirm the successful introduction of the desired mutation.

[23] For primer sequences refer to **Table 2.6**.

6.2.2 Production of Recombinant PNGase F and PNGase F Site-Specific Mutant Proteins

Chemically competent *E. coli* BL21 (DE3) were transformed with the vector OPH6 containing the ORF for the PNGase F wildtype protein or its site specific mutants. The transformation mixture was transferred into 50 mL LB broth containing 100 μg/mL ampicillin and incubated with shaking (200 rpm) at 37°C for 12-15 h. This culture was then used as a whole to inoculate 2.5 L LB broth (pre-warmed to 37°C), which has been prepared and autoclaved in a Minifors Benchtop Fermentation system (Infors AG, Bottmingen, Switzerland). Ampicillin was added to a final concentration of 100 μg/mL. The culture was stirred at 350 rpm and aerated with filtered compressed air. The temperature setting was decreased from 37°C to the protein expression temperature of 22°C ~10 minutes after inoculation as the fermenter was tap water cooled and took between 3-4 h depending on the ambient temperature of the water. When necessary, the cooling process was accelerated by attaching ice bags to the glass cylinder containing the culture medium. Protein expression was induced by the addition of 0.5 mM IPTG when the culture had reached an OD_{600} of at least 0.5, followed by incubation for ~15-20 h.

The cells were harvested at 4,500 *g* for 30 minutes at 4°C (Sorvall Evolution RC) and the cell pellets obtained transferred to a 50 mL screw lid tube, then resuspended in ~18 mL IMAC binding buffer (6.2.3.1).

Cell lysis was performed by 3 passes through a French press (SLM Aminco) at 6 kpsi.

Insoluble cell debris was removed by centrifugation at 30,000 *g* (4°C) for 30 minutes. The cell free supernatant was carefully decanted and the remaining pellet discarded. For subsequent purification using IMAC, the supernatant was filtered through a 0.8 μm filter (Sartorius AG) to remove as many insoluble particles as possible.

6.2.3 Purification of Recombinant PNGase F and PNGase F Site-Specific Mutant Proteins

For the purification of PNGase F and its site-specific mutants from whole cell lysates, column chromatographic methods were used.

6.2.3.1 Immobilised Metal Affinity Chromatography (IMAC)

The cell lysate obtained in 6.2.2 was subjected to IMAC using the Äkta™ Explorer Chromatography System (GE Healthcare). Recombinant PNGase F and its mutants contain a C-terminal hexahistidine-tag, which allows efficient purification using this affinity purification method. IMAC can be performed using a variety of divalent metal ions, such as Fe, Co, Ni, Cu and Zn, which bind tightly to metal chelating groups like iminidoacetic acid (IDA). These chelators are coupled to the column matrix *via* a spacer arm. The purification relies on the interaction of the metal ion with exposed basic amino acid side chains on proteins, especially histidine. In these experiments only nickel was used to 'charge' a 5 mL HiTrap™ Chelating HP.

Cell lysate (~35 mL) was loaded onto the column, followed by a wash step to remove unbound proteins. **Figure 6.2** shows the gradient profile applied for elution of PNGase F and its mutants.

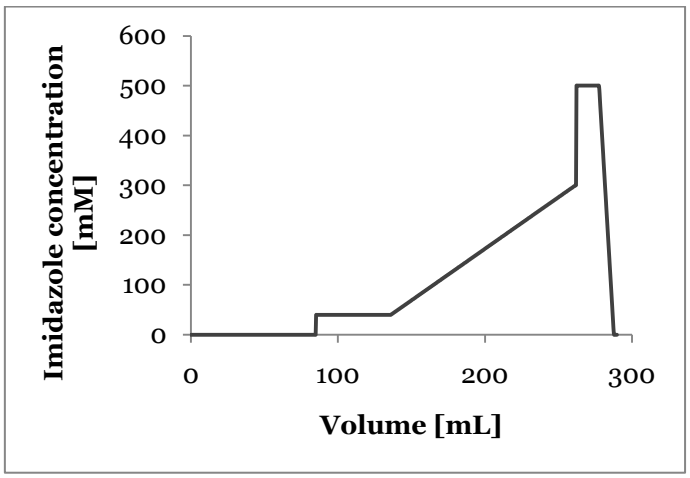


Figure 6.2: Gradient profile used for the IMAC purification of PNGase F and its site specific mutants.

Fractions were collected and analysed by SDS-PAGE. Fractions containing the protein of interest were pooled, concentrated using Vivaspin™ centrifugal concentrators (Sartorius) and subjected to size exclusion chromatography (SEC; 6.2.3.2) for further purification.

IMAC-Binding Buffer (Eluent A):

EPPS	0.1 M
NaCl	0.5 M
Imidazole	0.01 M
pH 8.5	

IMAC-Elution Buffer (Eluent B):

EPPS	0.1 M
NaCl	0.5 M
Imidazole	0.5 M
pH 8.5	

6.2.3.2 Size Exclusion Chromatography (SEC)

SEC separates globular macromolecules according to their molecular weight. Different column materials can be used that have beads with different pore sizes and hence different exclusion limits. The separation depends on the pore size, the bead size and also the shape of the beads. Molecules that are small relative to the pore size can enter deeply into the pores compared to larger molecules and therefore elute at different rates.

SEC was performed to remove residual minor contaminations from protein preparations that could not be removed by IMAC. A Superdex™ 75 10/300 GL column with an optimal separation range of globular proteins between 3-70 kDa was used for that purpose. The predicted molecular weight for recombinant PNGase F is ~36 kDa and it therefore falls into the optimal separation range of Superdex™ 75 10/300 GL. Besides further purification, this step was also used

to desalt the sample derived from IMAC. The buffer used was 50 mM EPPS, pH 8 and the elution of the proteins was performed isocratically over 1.5 column volumes. Fractions were analysed by SDS-PAGE, pooled where appropriate and concentrated. The final PNGase F preparations were pure to degrees between ~95-99% of total protein content. Aliquots of these preparations were stored at -80°C.

6.2.3.3 Reverse Phase (RP)-HPLC Purification

Reverse-Phase HPLC was used to obtain very pure protein for downstream applications like mass spectrometry (6.2.4). A Jupiter® 5 µm C18 300 Å (Phenomenex®) was loaded with 100-200 µL of protein solution obtained from SEC (6.2.3.2), which has been diluted to a concentration between 1-4 mg/mL. The mobile phases used were H_2O_{pure} with 0.1% TFA (Eluent A) and acetonitrile with 0.08% TFA (Eluent B). To elute bound proteins a gradient was applied over 50 minutes from the starting conditions (10% A + 90% B) to 60% B eluent. The elution profile was monitored by an UVD-340S Photo Array Detector at 214 and 280 nm. The peaks corresponding to PNGase F and its mutants were collected, frozen at -80°C and freeze dried.

6.2.4 Mass Spectrometry

Masses of the RP-purified wildtype and the mutant proteins (6.2.3.3) were determined by Mass spectrometry. These experiments were performed on a Bruker APEX-Q FTMS (9.4 T, Dual source II) in the Laboratory of Molecular Structure Characterization, Institute of Microbiology, Academy of Sciences of the Czech Republic in Prague by Dr. Petr Novák.

6.2.5 Circular Dichroism Spectrometry of Purified Recombinant PNGase F and PNGase F Site-Specific Mutants

Circular Dichroism (CD) spectrometry experiments were performed on a Chirascan™ CD spectrometer equipped with a peltier temperature controller for precise temperature control of the sample cell (Applied Photophysics, UK). Precision cells made of Quartz SUPRASIL® with 0.1 mm pathlength (Hellma®, Germany) were used to collect CD spectra. Processing of the raw data such as averaging and smoothing was performed using Pro-Data-Viewer, which was provided by the instruments' manufacturer. To avoid cross contamination between different samples, the cell was cleaned after each sample with nitric acid. This treatment and buffer degassing aid to prevent the formation of small bubbles in the cell during data collection. Concentrated purified protein was diluted in 5 mM EPPS, pH 8.5 (degassed) to a concentration of 1.0 mg/mL which was verified using the NanoDrop® spectrophotometer. The experimental conditions and instrument settings used for all samples are summarised in **Table 6.4**.

Table 6.4: Experimental conditions for CD

Parameter	Condition
Protein concentration	1.0 mg/mL
Pathlength	0.1 mm
Wavelength range	260 nm – 180 nm
Temperature	22°C
Time per point	0.25 s
Bandwidth	1 nm
Step size	1 nm
Repeats	5

The five scans collected for each sample were averaged and the baseline was subtracted. The resulting spectra were smoothed using an appropriate smoothing window size. This should be as high as possible without distorting the spectrum, i.e. the residuals should be distributed randomly around zero.

For temperature destabilisation studies, the temperature of the cell holder was increased by 10°C increments. One scan was taken per temperature point.

6.3 Results & Discussion

6.3.1 Generation of Site-Specific Mutations in the PNGase F ORF

The vector pOPH6 was used as template for the generation of site-specific mutations using the method described in 6.2.1. In this plasmid the PNGase F coding region (except the signal sequence) is placed downstream from the T7 promoter and ribosome binding site. Furthermore, it contains the coding regions for the OmpA secretion signal sequence (5′) and a hexahistidine tag (3′). The resulting expression product therefore carries an additional 12 amino acids: G-I-P at the N-terminus and L-D-P-His_6 at the C-terminus (Loo *et al.*, 2002). This study and subsequent unpublished work also established that PNGase F and some mutants can be expressed heterologously in *E. coli* using this plasmid.

The plasmids containing the mutations D60C, W86F, R248Q, R248R and W251Q were already generated prior to the start of this work (T.S. Loo).

The generation of PNGase F site specific mutants was performed for the mutations W59Q, I82Q, I82R, W207Q, V257N and V257K. Plasmids generated were submitted for DNA sequencing (2.21), which in each case confirmed the successful introduction of the desired mutation as shown in **Figure 6.3**.

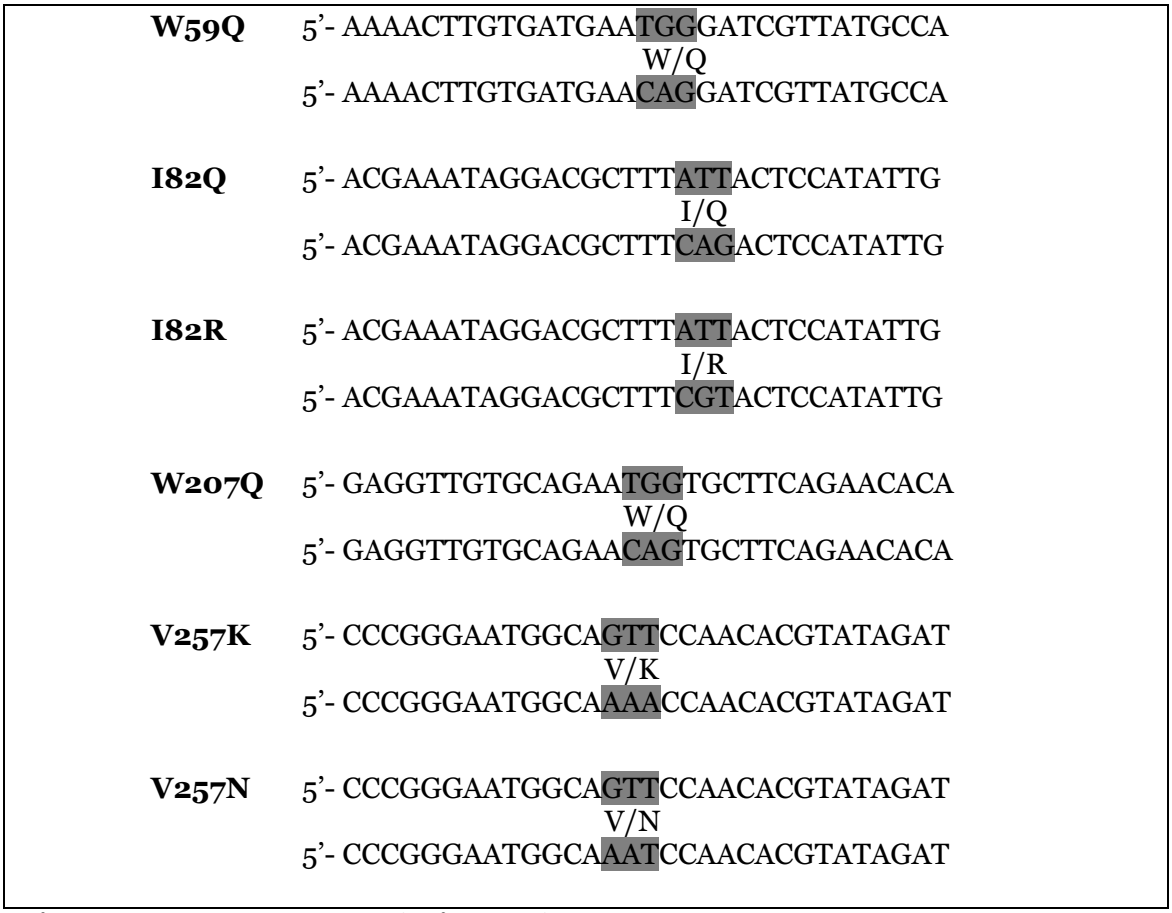


Figure 6.3: Sequence analysis results.
The mutations introduced into the PNGase F coding region are shaded grey.

6.3.2 Recombinant Expression and Purification

It had been previously shown that PNGase F could be expressed in *E. coli* BL21 (DE3) (Loo *et al.*, 2002). However, previous attempts to express several mutant proteins had proved unsuccessful (personal communication G.E. Norris, T.S. Loo). Therefore, small scale expression trials were performed for the wildtype protein and a selection of mutants to confirm the reproducibility of the system especially for the PNGase F mutants. For these trials, 10 mL cultures were grown under the growth conditions described in section 6.2.2. For each protein, an equal amount of cells was removed from the culture, pelleted and the periplasmic fraction extracted using the Polymyxin B method (Sahalan & Dixon, 2008; Tarrago-Trani & Storrie, 2004). Extracted His_6-tagged PNGase F molecules were bound to and eluted from ~50 µL Ni-charged IMAC resin to concentrate the protein. The proteins tested were produced in a soluble form

and exported into the periplasm. From this result it was concluded that soluble PNGase F was produced in reasonable quantities and at least partly exported into the periplasm. Therefore, no periplasmic extraction was performed for the large-scale purifications.

For the large-scale purification of PNGase F wildtype and mutants a two-step purification protocol was established based on the presence of the C-terminal His_6-tag and the proteins' molecular weight (~36 kDa) (6.2.3). The results for the purification of the various PNGase F proteins are summarised in **Table 6.5** and an example of each a typical IMAC and SEC chromatogram is shown in **Figure 6.4**. It should be noted that a considerable amount of protein was lost during the purification process as fractions were selected for purity rather than quantity.

Table 6.5: Two-step purification of PNGase F wildtype and mutants.

Protein	Protein concentration [mg/L culture]	
	IMAC	**SEC S75**
Wildtype	14.7	7.81
W59Q	17.7	8.58
D60C	5.68	3.69
I82Q	10.2	5.62
I82R	8.02	4.75
W86F	7.94	3.31
W207Q	13.8	7.14
R248K[24]	1.05	0.4
R248Q	3.08	1.06
W251Q	17.4	4.54
V257N	4.32	0.28

Figure 6.4 shows the IMAC chromatogram (Panel A) and a SEC chromatogram (Panel B) for the mutant protein W86F as an example of the purification process. The purifications of the other PNGase F mutants and the

[24] An unknown amount was lost during the purification process.

wildtype appeared very similar. In the first purification step, a concentrated cell lysate was applied onto the IMAC column and unbound or loosely bound proteins and contaminants were washed off by an intensive washing step. When an imidazole gradient was applied, PNGase F eluted between ~50 mM and 125 mM imidazole with an absorbance maximum at ~90 mM imidazole. When fractionated using SEC, PNGase F eluted in one peak after approximately 0.5 column volumes. PNGase F W86F eluted after ~12 mL buffer had passed through the column in a peak that was reasonably resolved from contaminants. As can be seen in **Figure 6.4 (B)**, the load was too high resulting in lower resolution of the desired peak than would probably have been achieved with a lower load.

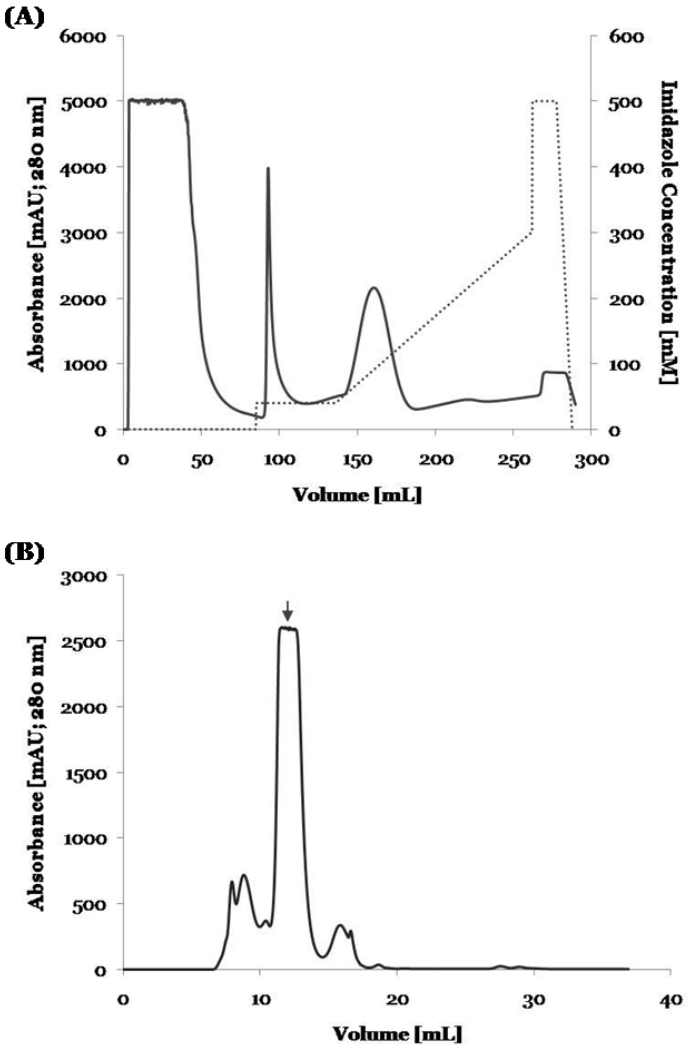


Figure 6.4: Two-step purification of PNGase F wildtype and mutants.
Shown here are the chromatograms for the mutant W86F. Panel A shows the IMAC chromatogram; solid line: Absorbance at 280 nm [mAU]; dotted line: imidazole gradient [mM]. Panel B shows the SEC chromatograms. The arrows indicate the peaks containing PNGase F.

Figure 6.5 shows a SDS-PAGE result for the purified PNGase F wildtype and mutants. This demonstrates the high purity of the proteins that was achieved with the two-step purification procedure described above.

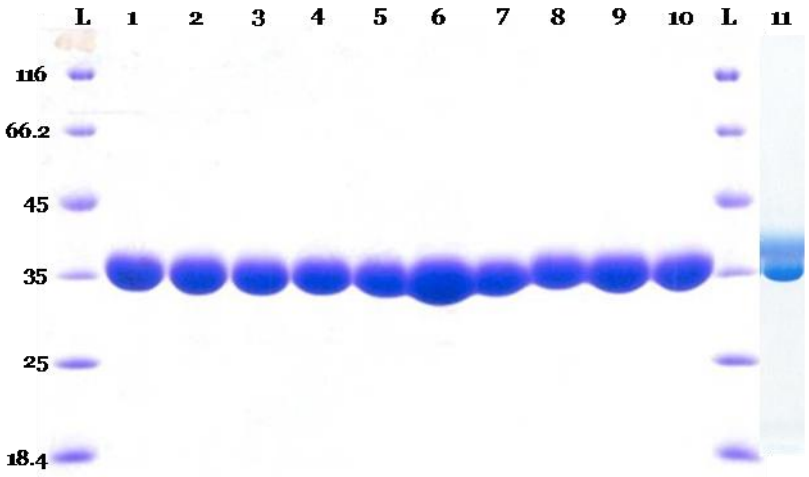


Figure 6.5: SDS-PAGE of PNGase F wildtype and site specific mutants.
L, Protein Ladder [kDa]; 1, wildtype; 2, W59Q; 3, D60C; 4, I82Q; 5, I82R; 6, W86F; 7, W207Q; 8, R248Q; 9, W251Q; 10, V257N; 11, R248K. Approximately 8 µg protein were loaded in lanes 1-10. R248K was run on a separate gel (the shadow above the protein band is an often seen residual shadow from the gel scanning process, not an actual band in the gel).

6.3.3 Mass Spectrometry Analysis

In order to prove the successful introduction of the desired mutations into PNGase F, the purified wildtype and the mutant proteins were analysed by mass spectrometry. With this method the exact masses of proteins can be determined and the mass shift resulting from the introduction of a different amino acid in the mutant proteins can be detected.

The expected mass shifts, the experimentally determined masses and the resulting experimental mass shifts between the PNGase F wildtype and the various mutants are summarised in **Table 6.6**. In all cases the experimental mass shift was in agreement with the expected mass shift, proving the successful introduction of the desired mutations into PNGase F.

Table 6.6: Mass spectrometry results for PNGase F and the mutant proteins (monoisotopic, MH+1).

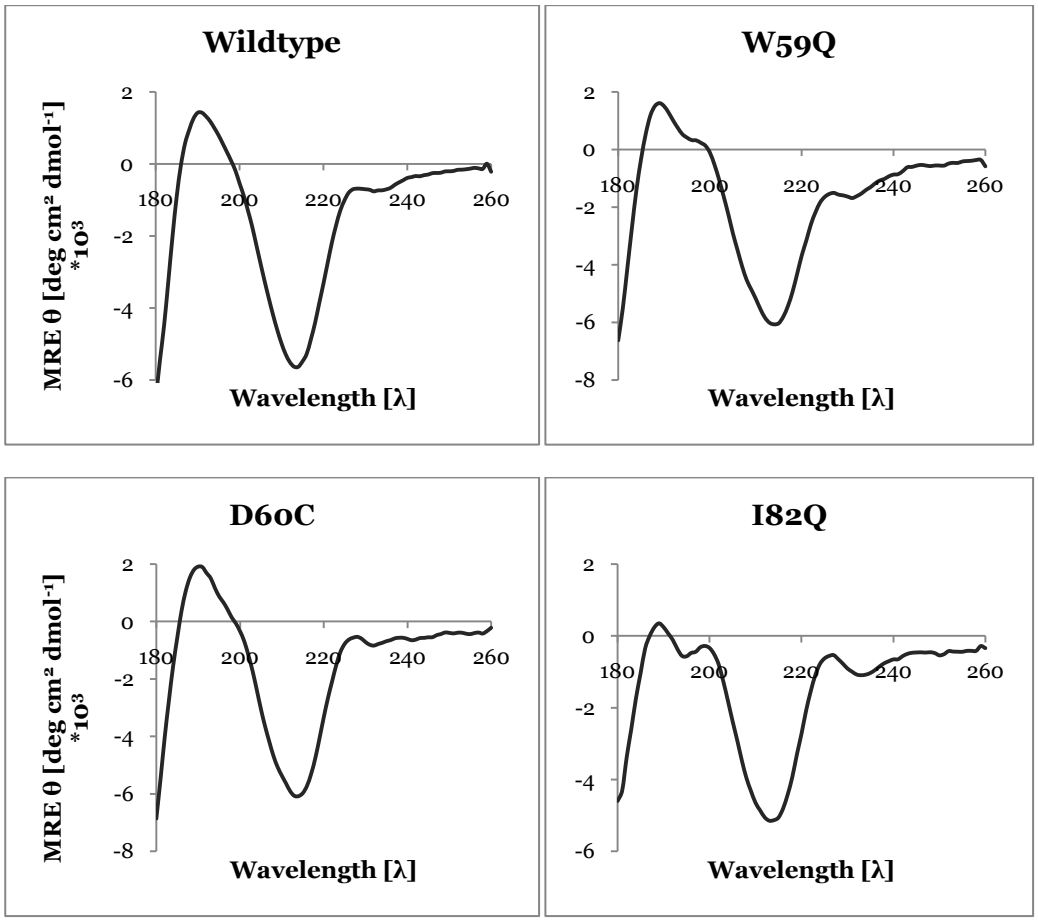
Protein	Expected mass shift [Da]	Experimental Mass [Da]	Experimental Mass shift [Da]
Wildtype	--	36223.85	--
W59Q	-58	36165.90	-57.95
D60C	-12	36211.87	-11.98
I82R	+43	36265.96	+42.11
I82Q	+15	36238.87	+15.02
W86F	-39	36184.90	-38.95
W207Q	-58	36165.90	-57.95
R248K	-28	36195.89	-27.96
R248Q	+30	36195.82	+28.03
W251Q	-58	36165.84	-58.01
V257N	+15	36238.87	+15.02

6.3.4 Circular Dichroism Analysis

Circular Dichroism (CD) is a spectroscopic method that can be used for the determination of secondary structure elements present in proteins, i.e. α-helix, β-stand and β-turns and random coil. The electronic transitions of protein backbone peptide bonds in different confirmations produce differential absorption spectra for left- and right-handed circularly polarised light in the far-UV (below 240) (Kelly *et al.*, 2005). Generally, α-helical proteins show negative bands at 222 nm and 208 nm and a positive band at 193 nm. Well defined antiparallel β-sheets (β-helices) have a negative band at 218 nm and a positive band at 195 nm and disordered proteins show a very low ellipticity above 210 nm and negative bands near 195 nm (Greenfield, 2006).

Figure 6.6 shows the CD spectra for PNGase wildtype and all site-specific mutants except mutant R248Q for which not enough pure protein was available. All proteins show the landmark bands characteristic for proteins with a high secondary structure content of β-strands, which has been shown by X-ray diffraction techniques for native PNGase F (Kuhn *et al.*, 1994; Norris *et al.*, 1994b) and for the recombinant protein in this study (Chapter 7). Another

noteworthy feature was the band appearing at approximately 225 nm, which was most likely due to contributions of aromatic amino acid side chains to the CD spectra. Aromatic side chains have characteristic peaks between 260 nm and 320 nm, but have also been shown to significantly contribute to CD spectra in the far-UV below 250 nm. This is especially so for tryptophan and tyrosine (Kelly *et al.*, 2005; Krittanai & Johnson, 1997; Woody, 1994; Yanagida *et al.*, 2008). Woody (1994) stated that these contributions can be significant, especially in proteins of low helical content (Woody, 1994). These findings are particularly relevant for PNGase F as it contains a total of nine tryptophans, fourteen phenylalanines and sixteen tyrosines and the X-ray structure contains very little helical structural features (Kuhn *et al.*, 1994; Norris *et al.*, 1994b). The different intensities seen for this peak at ~225 nm are possibly due to slight changes in the environment of tryptophans upon mutation. Tryptophan has also been found to have several transitions in the 190 to 210 nm region, which usually do not appear as distinct bands in CD spectra (Krittanai & Johnson, 1997). However, the minor transition at ~200 nm appearing in some of the spectra shown below may well be due to tryptophan.



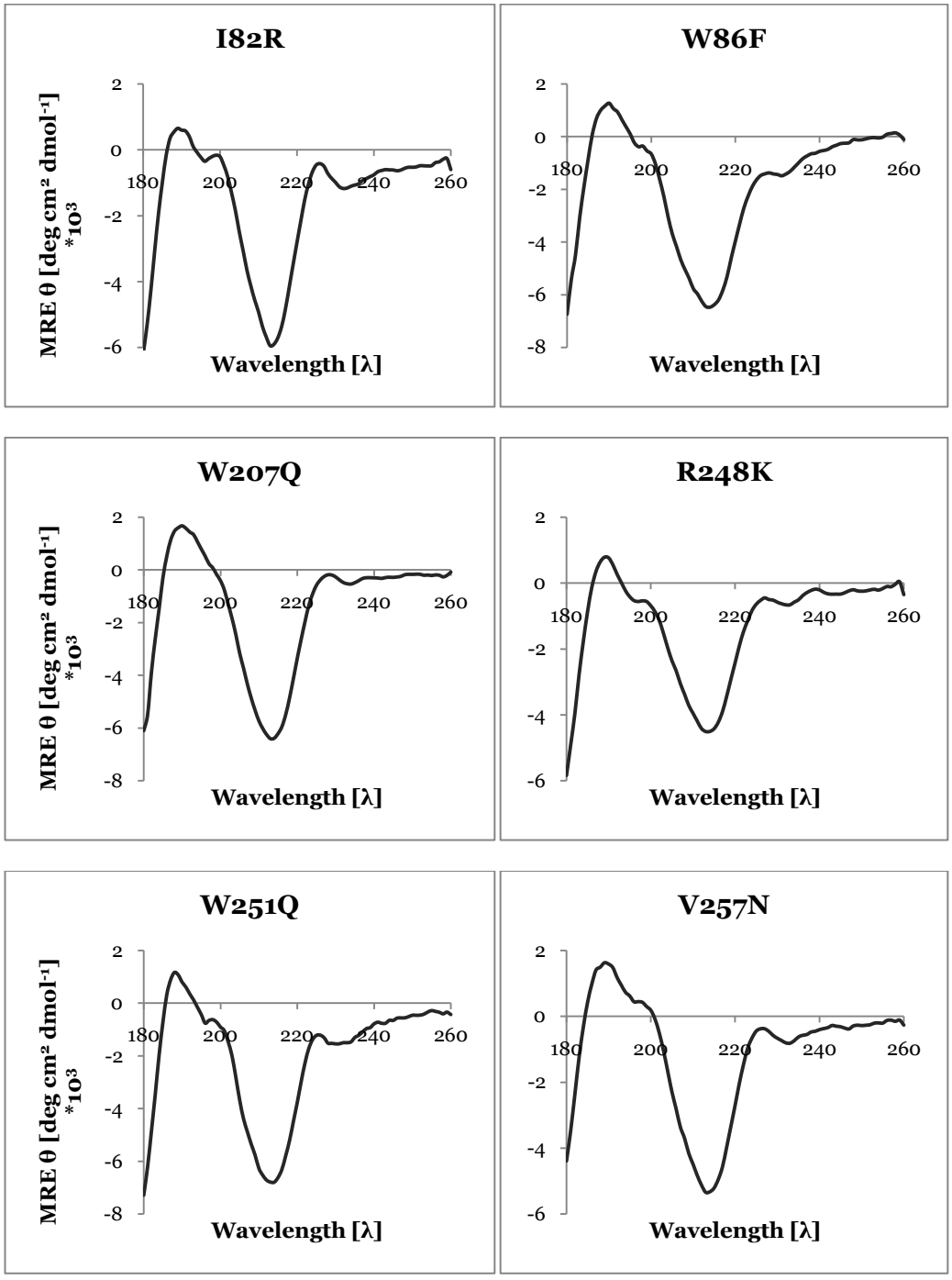


Figure 6.6: Circular Dichroism spectra.
Data were collected as milli-degrees and then converted into the Mean Residue Ellipticity (MRE, θ). The mean residue molecular weight for PNGase F was determined to be 111 Da. The spectrum for mutant W86F showed a baseline drift between 250 and 260 nm, where it should be close to zero. This was an instrumental factor that can occur when the CD spectrophotometer is first turned on (Greenfield, 2006). The W86F spectrum was the first sample to be scanned. The drift was corrected for by adding the average drift between 250 and 260 nm.

Several algorithms are available that estimate the secondary structure composition of a protein using reference datasets that consist of CD spectra of proteins whose structures have been solved by X-ray crystallography. Hence,

these techniques are of an empirical, comparative nature and therefore restricted by the structural features present in the proteins in the reference datasets used for analysis (Whitmore & Wallace, 2004). This poses problems for the analysis of the conformation of certain proteins, such as proteins with a majority of pure β-helices (synthetic polypeptides, amyloids), fibrous proteins (e.g. collagen) or coiled-coil proteins (Greenfield, 2006). In addition, the possible contribution of aromatic side chains to the spectra can lead to problems with secondary structure interpretations (Krittanai & Johnson, 1997). Therefore, in this case, the deconvolution of the CD results is more a tool to gauge the integrity of the folding for each mutant to the wildtype rather than being a tool for obtaining definite secondary structure contents.

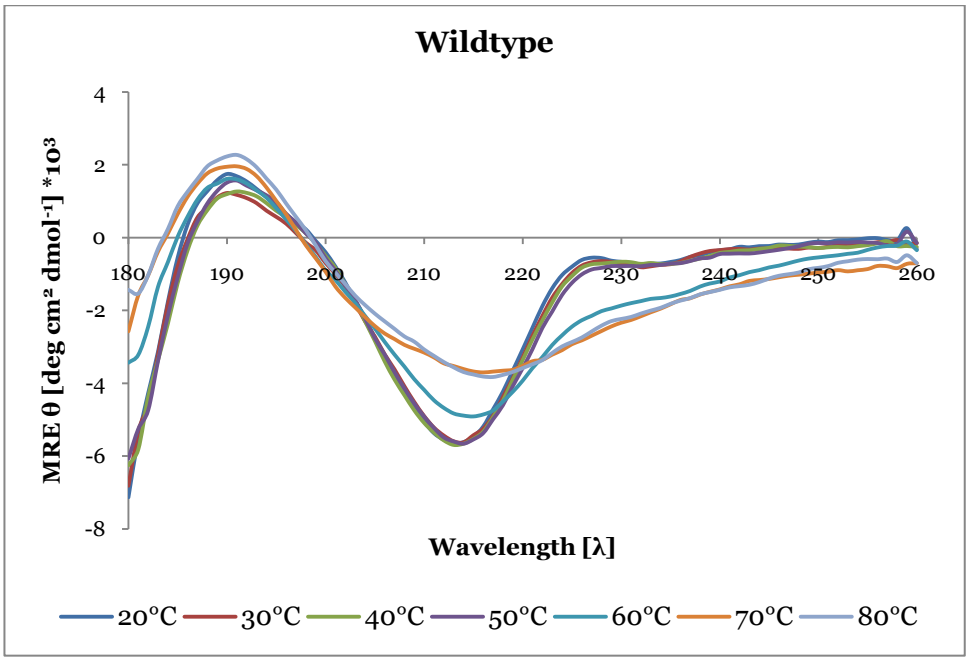
Two algorithms, CDSSTR (Compton & Johnson, 1986) and CONTIN (Provencher & Glockner, 1981; van Stokkum *et al.*, 1990), were chosen for the deconvolution of the spectra shown above. These algorithms were accessed *via* the web server DICHROWEB (Whitmore & Wallace, 2004). **Table 6.7** shows a summary of the CD data deconvolution.

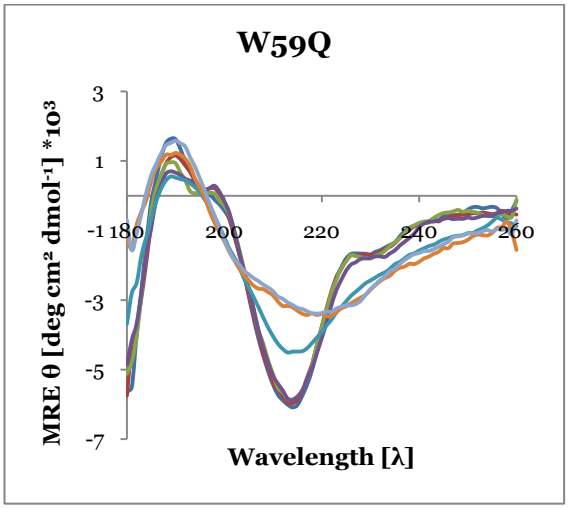
Table 6.7: CD data deconvolution.
The Results were generated using two algorithms available on the DICHROWEB web server.

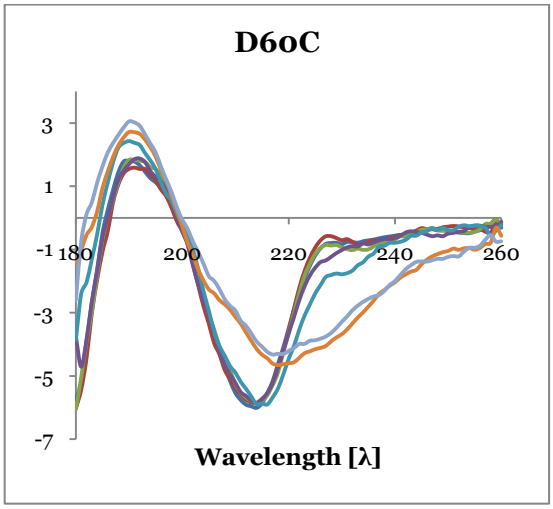
	CDSSTR				CONTIN			
	α-helix	β-sheet	Turns	Unordered	α-helix	β-sheet	Turns	Unordered
	[%]				[%]			
WT	2	41	25	30	4.7	41.5	24.8	29.1
W59Q	3	39	24	33	7.1	38.9	23.5	30.6
D60C	2	42	24	30	5.3	41.4	25.3	28
I82Q	2	41	24	32	4.9	42.2	24.1	28.8
I82R	2	42	24	32	5.2	41.6	25.1	28.1
W86F	5	35	21	39	9.4	33.6	23	33.9
W207Q	2	41	25	30	3.9	44.6	26.3	25.1
R248K	2	41	23	32	4.6	39.9	23.2	32.3
W251Q	3	37	25	35	7.4	39.3	24.6	28.7
V257N	2	45	24	28	4.8	42.3	24.7	28.2

Results from both algorithms showed that both wildtype and mutant proteins had very similar secondary structure (± 3%), indicating correct folding. The mutant W86F shows the highest deviation compared to the wildtype, but still exhibits the general main structural features.

To test the stability of the mutant proteins melting profiles of the different mutants and the wildtype sample were obtained. Decreased stability of a mutant protein would result in the loss of structural features at a lower temperature compared to the wildtype protein. This structural change can be detected using CD spectroscopy performed at increasing temperatures as shown in **Figure 6.7**.







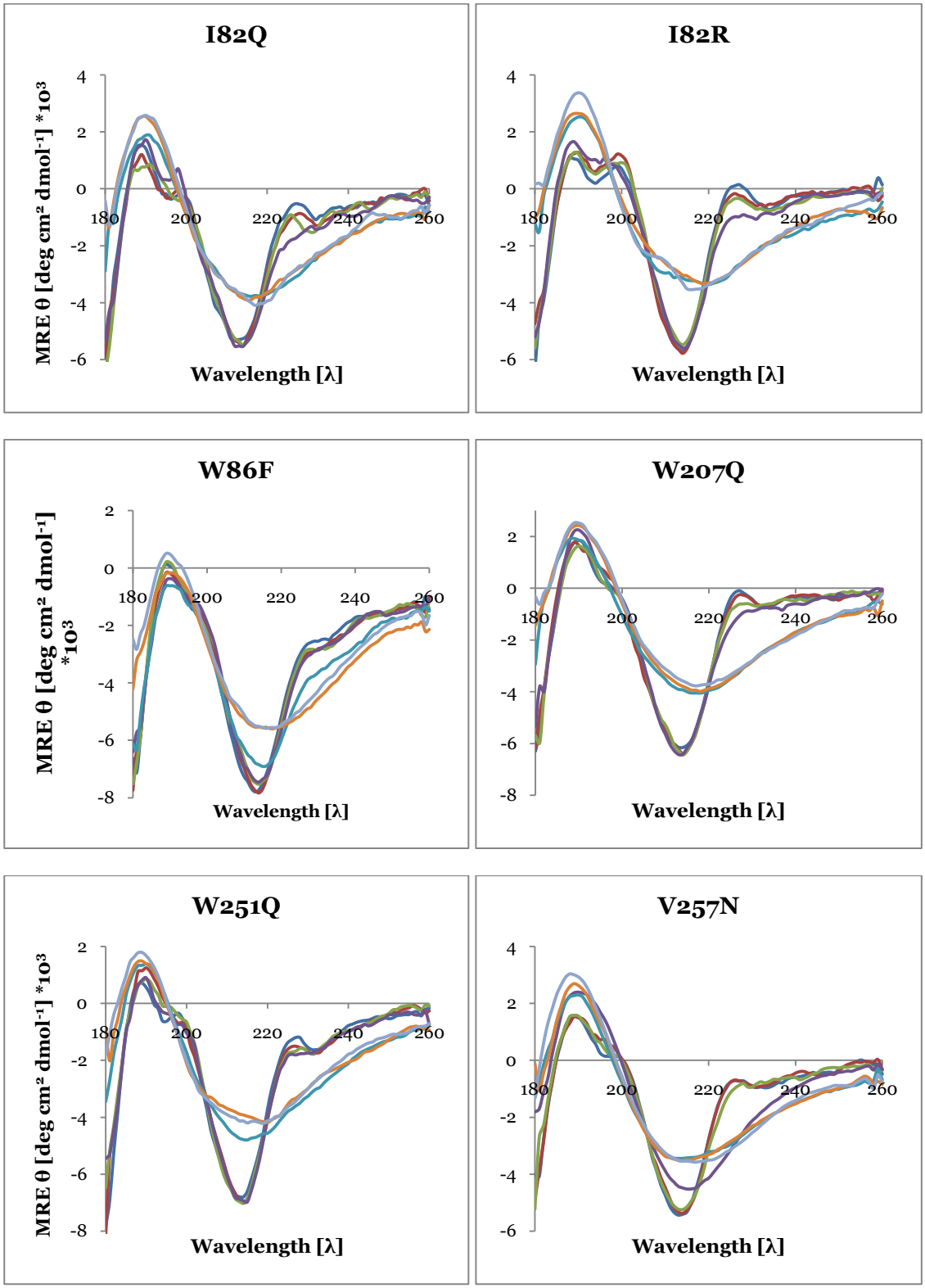


Figure 6.7: Protein stability studies at different temperatures.
For the colour code refer to the wildtype chart.

The PNGase F wildtype started to denature at 60°C. Denaturation increased with a 10°C increase to 70°C, but did not show extreme changes with another 10°C increase. The mutants W59Q, W86F and W251Q showed the same

denaturation pattern. The mutants I82Q, I82R a W207Q differed from the wildtype pattern in that they already reached the maximum level of denaturation at 60°C, indicating a possibly slight decrease in stability. The melting experiment on mutant V257N demonstrated that this protein had a decreased stability as it started to denature at 50°C instead of the more usual 60°C. Interestingly, the 60°C scan of mutant D60C showed only a minor change compared to the lower temperature scans, indicating a slightly increased stability for this mutant compared to the wildtype protein.

The deconvolution of the data collected at 80°C demonstrated the loss of defined structural features compared to the 22°C data as shown in **Table 6.8**. A 17% increase in unordered structure was determined for the wildtype. The mutant proteins had similar proportions of unordered structure at this temperature (± 3%), indicating a similar overall stability to temperature. One exception is the mutant W86F, which shows already an elevated amount of unordered features at 22°C. However, this doesn't seem to affect the overall stability, as complete denaturation is not evident until the temperature reaches 60°C, similar to the wildtype protein.

Table 6.8: CD data deconvolution of data collected at 80°C.
The results were generated using CDSSTR on the DICHROWEB web server. The unordered structure increase was calculated using results of CDSSTR deconvolution results in **Table 6.7**.

	α-helix	β-sheet	Turns	Unordered	Unordered Structure Increase
	[%]				
WT	5	30	18	47	17
W59Q	4	28	17	49	16
D60C	5	31	17	46	16
I82Q	7	21	19	52	20
I82R	5	30	18	46	14
W86F	3	30	21	44	5
W207Q	5	28	17	49	19
W251Q	5	26	17	50	15
V257N	4	30	18	46	18

Chapter 7

Crystallisation of rPNGase F

7 Crystallisation of rPNGase F

7.1 Introduction

The comparison of mutant structures with the wildtype structure in conjunction with kinetic data can provide valuable information regarding the function(s) of the mutated residue(s). While kinetic data indicate the importance of a residue by showing, for example, decreased enzymatic activity, structural data can explain the reason for changes in activity at a molecular level.

In order to obtain high-resolution structural information about the PNGase F mutants, crystallisation trials for the purified proteins were performed. Although the three-dimensional structure of native PNGase F has been solved previously, the recombinant wildtype PNGase F (rPNGase F) was crystallised to confirm its correct folding, and to form a basis for comparison with any mutant structures obtained.

7.2 Methods

7.2.1 Crystallisation trials

Initial crystallisation trials were carried out using one or more of the following crystallisation screens: Hampton Crystal Screens 1 & 2, Molecular Dimensions Structure Screens 1 & 2 and Molecular Dimensions PACT premier™. Trials were set up using sitting drop vapour diffusion in 96-well Intelli- or Crystalquick-plates (Art Robbins Instruments; Greiner). This was done either by hand or using the Mosquito® Crystallisation Robot (TPP Labtech). Generally, 1 µL of protein solution at 10 mg/mL was mixed with 1 µL of well solution if screens were set up by hand or 200 nL of each solution were

mixed if the robot was used. Plates were covered with ClearSeal Film (Hampton Research), inspected and incubated at room temperature.

Once potential crystallisation conditions had been identified, refinement screens were carried out using hanging drop vapour diffusion in 24-well VDX plates (Hampton Research). Well solutions were made up using Optimize reagents, StockOptions Kits (both Hampton Research) and, if necessary, self-made stock solutions that were filtered through a 0.2 μm filter to remove any particles. Protein and well solutions were mixed (1 μL of each) on a siliconised glass circle cover slide (22 mm; Hampton Research), which was then inverted over 500 μL of mother liquor. The well was sealed with petroleum jelly (Shell). All non-sealed consumables, such as pipette tips, were dusted with compressed air prior to use.

7.2.2 Data Collection & Processing

Suitable crystals were extracted from the drop with a nylon loop attached to a magnetic crystal mount (Hampton Research) and transferred into cryo-protectant solution consisting of 20% glycerol in mother liquor. After ~1 minute the mounted crystal was transferred onto the goniometer and frozen in a stream of gaseous nitrogen at 120K. To assess the level of diffraction by a crystal, two images were taken at 0 and 90° using an exposure time of 5-10 minutes. Diffraction data were collected using a MicroMax-007 microfocus rotating anode (copper) generator and an R-AXIS IV++ imaging plate area detector (both Rigaku). Data were collected and initially processed using the software CrystalClear 1.3.6 (Rigaku).

The Xia2 program suite incorporating XDS for spot integration (Kabsch, 1993; Winter, 2010) was used to process the diffraction data (images).

SCALA (Evans, 2006; Potterton *et al.*, 2003) was used to scale together multiple observations of reflections, and to merge multiple observations into average intensities. SCALA generates an output file that provides several important measures of data quality, which were carefully inspected: *(i)* R_{merge} reports the quality of the experimental diffraction data as the average

discrepancy of multiple measurements of the same reflection (the lower the discrepancy the lower R_{merge}); *(ii)* an $\sum I/\sum SigI$ value greater than 2 (in the highest resolution shell) indicates the resolution limit. It measures the mean intensity in a resolution shell relative to its standard deviation and is a measure of signal-to-noise ratio; *(iii)* the completeness of the dataset was inspected for each resolution shell. Dataset completeness should be as close to 100% as possible.

The averaged intensities were then read into the program TRUNCATE (French & Wilson, 1978; Potterton *et al.*, 2003), which reads reflection data files of averaged intensities (SCALA outputs) and produces an mtz reflection data file containing a mean amplitude (F) and standard deviations. A number of important measures of data quality generated by this program were thoroughly examined: *(i)* the Wilson plot was inspected to confirm a normal distribution of intensities as a function of resolution; *(ii)* Twinning statistics were analysed for the possibility of crystal twinning.

Finally, FREERFLAG (Brunger, 1997) was used to flag a random 10% of reflections, which would not be used in refinement and thus provide an indication of structure improvement (Free R factors, R_{free}).

Molecular replacement was used for structure solution and was performed using the program PHASER (McCoy *et al.*, 2007). In this method initial phase information is obtained by using the phases of a molecule with similar structure, usually a protein that has a sequence similarity with the target protein greater than 30%.

Model building was carried out using COOT (Emsley & Cowtan, 2004) followed by structure refinement. Refinements were done in REFMAC5 (Murshudov *et al.*, 1997), with each round comprising 10 cycles of maximum likelihood restrained refinement.

Model validation was performed using MolProbity (Davis *et al.*, 2007).

PyMol was used to generate graphical presentations of structural features (DeLano, 2002).

7.3 Results & Discussion

7.3.1 Crystallisation of Recombinant Wildtype PNGase F & Mutant W251Q

The recombinant PNGase F wildtype was crystallised using the hanging-drop method (7.2.1) in the previously determined conditions 25% PEG 4000, 0.2 M ammonium sulfate ($(NH_4)_2SO_4$) and 0.1 M sodium acetate at pH 4.5 (T.S. Loo, personal communication). The protein was made up to a concentration of 10 mg/mL in 5 mM EPPS at pH 8.5. One crystal appeared approximately two months after set up (**Figure 7.1**). This rather long crystallisation time appears to be a general characteristic of PNGase F since the native PNGase F protein in different crystallisation conditions took a similarly long time to form crystals in a previous study (Norris *et al.*, 1994b).

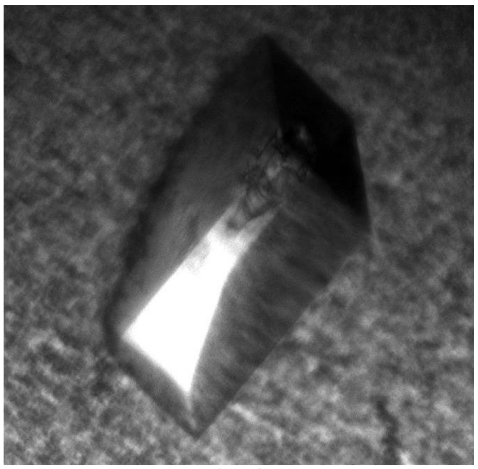


Figure 7.1: Crystal of rPNGase F.
Conditions: 25% PEG 4000, 0.2 M $(NH_4)_2SO_4$, 0.1 M sodium acetate, pH 4.5.

For the PNGase F mutants, initial sitting-drop vapour diffusion crystal screens were performed in 96-well plates (7.2.1) as well as hanging-drop experiments using the wildtype crystallisation conditions. The protein concentration generally used was 10 mg/mL. Lead conditions from the initial screens were refined using the hanging-drop method in VDX plates. However, only the mutant W251Q formed two crystals after approximately two months (**Figure 7.2**). Both crystals were tested for diffraction (7.2.2) and were found to not diffract.

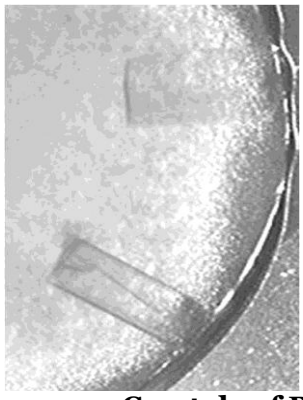


Figure 7.2: Crystals of PNGase F mutant W251Q.
Conditions: 25% PEG 4000, 0.2 M $(NH_4)_2SO_4$, 0.1 M sodium acetate, pH 4.5.

7.3.2 Data Collection & Processing for Recombinant Wildtype PNGase F

The collection of raw diffraction data was performed using CrystalClear (Rigaku). Two initial images were collected at 0 and 90° and indicated a monoclinic unit cell with a = 41.01 Å, b = 90.76 Å, c = 49.0 Å and the angles α, γ = 90° and β = 112.22°. Further data were collected using the strategy recommended by CrystalClear (Rigaku) to obtain a redundancy of 4. A total of 408 images were collected over a total oscillation angle of 122.4° at a detector to crystal distance of 100 mm using an exposure time of 8 minutes per image.

The MATTHEWS_COEF (Kantardjieff & Rupp, 2003; Matthews, 1968) was used to determine the number of molecules and the solvent content in the asymmetric unit. This analysis showed that there was one molecule in the asymmetric unit and determined the solvent content to be 48.66%, which is close to the average of ~47%, determined in 2002 by Kantardjieff & Rupp. With one molecule in the asymmetric unit the space group $P12_11$ ($P2_1$) dictates the presence of two molecules per unit cell. **Table 7.1** summarises the results of the data collection.

Table 7.1: Data collection statistics.
Values in parentheses are for the outermost shell.

Parameter	
Wavelength [Å]	1.542
Space group	$P12_11$ (monoclinic; Nr. 4)
Unit cell dimensions [Å]	
a	41.0
b	90. 76
c	48.8
α, β, γ [°]	90, 112.2, 90
Number of observations[25]	182516
Number of unique reflections	41972
Resolution [Å]	40.41-1.57
Mosaicity [°]	1.49°
Multiplicity	4.0 (2.9)
Completeness [%]	96.6 (82.2)
R_{merge}[26]	0.057 (0.154)
I/SigI	16.6 (5.1)

7.3.3 Molecular Replacement

The calculation of the distribution of electron density in a protein crystal requires a Fourier transform using structure factors derived from the diffraction data. Each structure factor is a vector with an amplitude and a phase. The measured x-ray intensities are proportional to the square of the structure factor amplitude. As diffraction data can only provide the intensities of reflections, the phase information is lost. Molecular replacement can be used to determine the phases of an unknown structure using phase information from a known, homologous structure (Adams *et al.*, 2009). As the structure for native PNGase F has been solved (PDB ID: 1PGS; (Norris *et al.*, 1994b)), this structure (without water molecules) was used to phase the data set collected from the recombinant protein crystal. The amino acid sequence of the recombinant PNGase F has been published and was shown to have eight strain specific (CDC3552) amino acid substitutions compared to the native PNGase F protein encoded by *F. meningosepticum* strain ATCC 33958 (Loo *et al.*, 2002;

[25] Observations to 1.45 Å resolution. Resolution was cut off at 1.57 Å due to incomplete data at higher resolution. All other values given include data up to 1.54 Å resolution.
[26] $R_{merge} = \sum_{hkl} \sum_i \left| I_{i,hkl} - \langle I_{hkl} \rangle \right| / \sum_{hkl} \sum_i I_{i,hkl}$, where $I_{i,hkl}$ is the intensity of the i^{th} observation of reflection *hkl* and $\langle I_{hkl} \rangle$ is the mean intensity of all observations of *hkl*.

Tarentino *et al.*, 1990). Six of these residues are also different in the native protein used for crystal structure determination of 1PGS. As the two proteins are almost identical, initial phase determination by molecular replacement using structure 1PGS (excluding the water molecules) as a model was performed using PHASER (McCoy *et al.*, 2007) within the CCP4i suite of programs. This rotation and translation search gave one solution corresponding to the one molecule expected in the asymmetric unit (McCoy *et al.*, 2007). From the diffraction pattern the space group was determined to be P121 (Nr. 3) as there were no systematic absences observed. However, as no solution could be found when the space group was fixed to P121 a number of different molecular replacement programs including AMORE (Navaza, 1994), MOLREP (Vagin & Teplyakov, 1997), MrBUMP (Keegan & Winn, 2007) were used without producing a solution. When subgroups of P121 were allowed for using PHASER, however, a clear solution was obtained. For the space group $P12_11$ every uneven numbered reflection along the *k*-axis should be absent. This pattern was, however, not observed, when the data was examined using the program 'hklview'. While this is puzzling, it could be due to the rather high mosaicity of 1.49°.

The resulting model showed good Z-scores for the rotational (RFZ) and translational (TFZ) functions (RFZ = 33.8 and TFZ = 19.4) and a log-likelihood gain (LLG) of 3765, indicating the solution was correct. The initial R factor was 35.97%.

7.3.4 Structure Refinement

During the refinement process the calculated structure factors, which are based on the model, are compared with the experimental data. This process involved several rounds of refinements using REFMAC5 (Murshudov *et al.*, 1997) followed by manual structure building in COOT (Emsley & Cowtan, 2004) using $2|F_o| - |F_c|$ electron density maps.

Overall, the density was of high quality as shown in **Figure 7.3**, so that all residues could be easily fitted into the electron density with the exception of parts of one loop comprising residues 163 to 169. These residues appear to form

a mobile loop on the surface of the molecule somewhat distant from the active site. Alternative conformations (two each) were found for eight amino acid side chains and all of them were assigned with an occupancy of 0.5. These residues were K41, E90, L169, K170, S201, S240, M255 and S297.

Chemically sensible water molecules were added to the structure in positions with σ-levels greater than 1σ. Furthermore, electron density was observed to which could be fitted glycerol and acetate molecules, and sulfate ions. Overall, 10 glycerol molecules were added with varying degrees of occupancy (0.5-1) as judged by inspection of the difference electron density map, which highlights features present or absent in the observed structure but not in the model used for phasing. Glycerol was not used for determination of structure 1PGS, but was used here as cryo-protectant (20% (v/v) in mother liquor). Two acetate molecules and two sulfate ions were added all with an occupancy of 1. Both components were present in the mother liquor of the original crystallisation conditions. The fact that they were observed here but not in the original model is evidence of the high quality of the data, and the correctness of the molecular replacement solution.

The values for R and R_{free} were used to follow the improvement of the model. R is a measure of the agreement between the crystallographic model and the experimental x-ray diffraction data, i.e. it is a measure of how well the refined structure matches the observed data (Morris *et al.*, 1992). R_{free} measures the agreement between observed and modelled structure factor amplitudes for a 'test' set of randomly selected reflections that is omitted in the modelling and refinement process. Therefore it is an indicator of the accuracy of models, showing if diffraction data are over- or mis-interpreted (Brunger, 1992).

Final refinement statistics are given in **Table 7.2**.

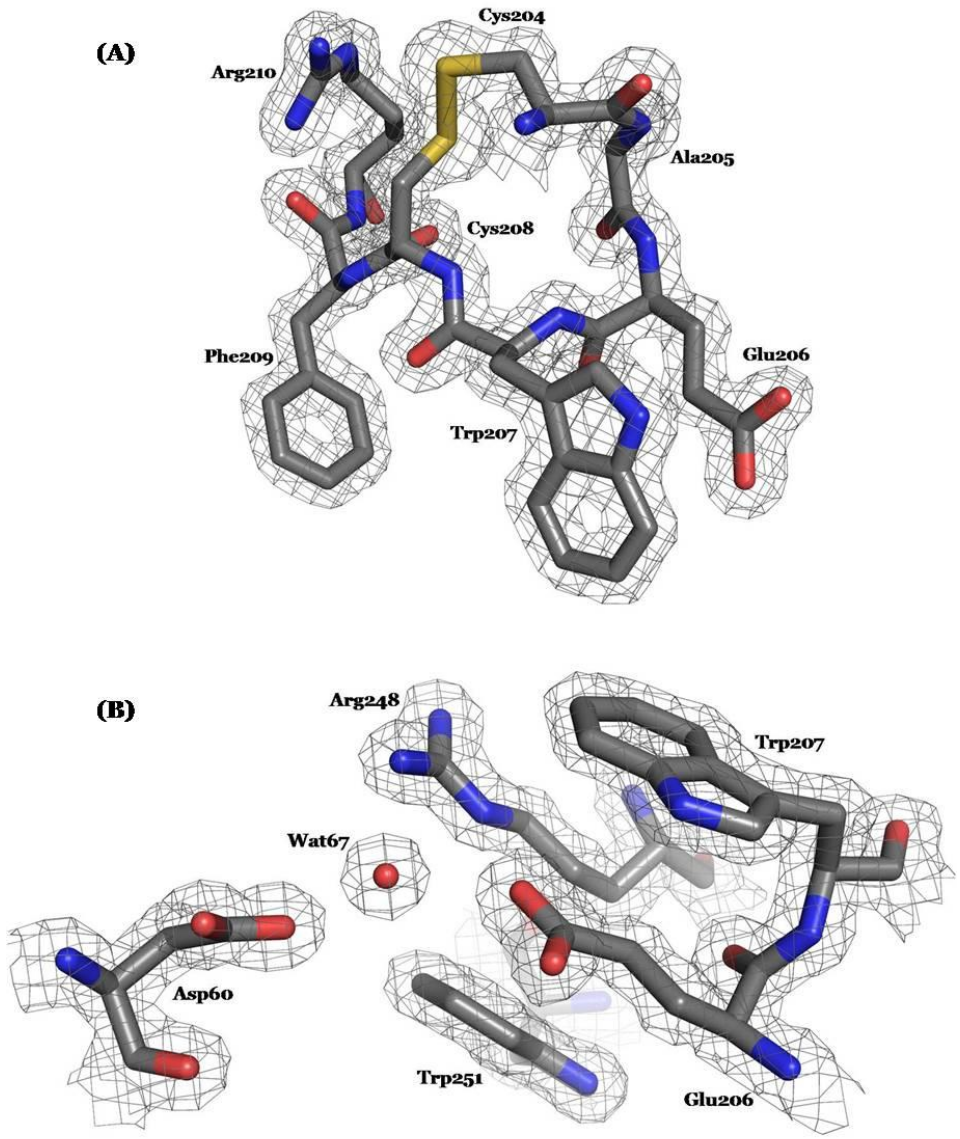


Figure 7.3: Two regions of the final electron density map calculated at a resolution of 1.54 Å.
The $2F_o$-F_c map is contoured at 1.5σ. Panel (A) shows the loop joined by disulphide bond 204-208 that includes one of the proposed catalytic residues, Glu206. Panel (B) shows five active site residues and the proposed water nucleophile Wat67 (This will be discussed in Chapter 8).

The final R and R_{free} values for the structure model were 15.1% and 17.8% respectively.

7.3.5 Ramachandran Plots

In a polypeptide the main chain bonds N-C_α and C_α-C are relatively free to rotate. These rotations are described by the torsion angles phi and psi, respectively. However, in a protein some rotations around these bonds can lead to conformations that cause steric clashes between atoms. The Ramachandran plot shows areas of allowed phi and psi angles, i.e. angles that do not lead to conformations causing sterical hindrance between atoms. The two main regions of allowed angles correspond to the α-helical (lower left hand side) and β-sheet (higher left hand side) conformations. In crystallography the Ramachandran plots are used to check that the torsion angles of the polypeptide backbone are within the allowed regions as these angles are not restrained during the refinement process.

The Ramachandran plots generated for the final model of PNGase F are shown in **Figure 7.4**.

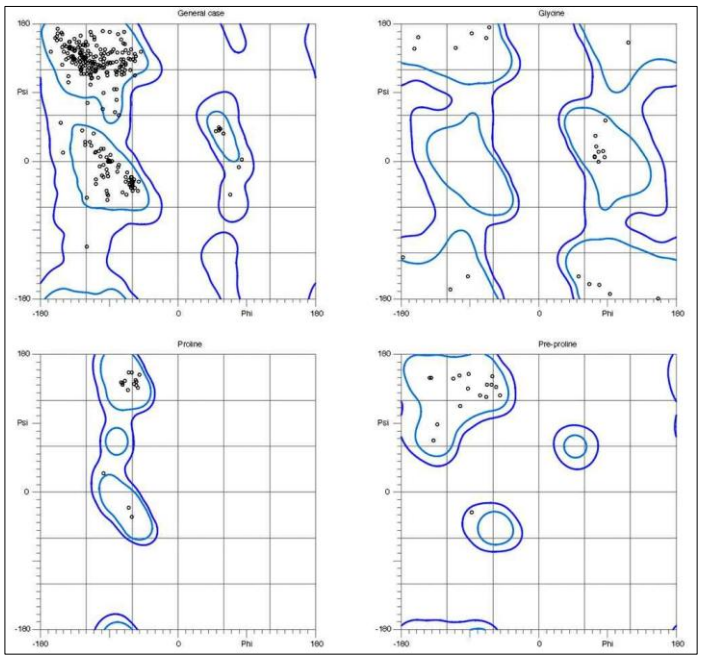


Figure 7.4: Ramachandran plot for the refined model of recombinant PNGase F.
Top left plot: General case. Top right plot: Glycines. Bottom left: Prolines. Bottom right: Pre-prolines. The plots were generated by the MolProbiety server (Davis *et al.*, 2007).

In the 'General case' plot, which includes all residues, the majority of residues fall into the higher left hand quadrant. This was to be expected as the

previously reported structures showed that PNGase F consists of two all-β domains (Kuhn *et al.*, 1994; Norris *et al.*, 1994b).

7.3.6 Statistical Validation

Table 7.2 summarises results of the structure refinement and parameters that are used to show the accuracy of a crystallographic model.

Table 7.2: Refinement statistics.

Parameter	
R factor[27] [%]	15.1
R_{free} [%]	17.8
Rmsd from ideal geometry	
Bond length [Å]	0.02
Bond angles [°]	1.82
Number of protein atoms	2489
Number of H_2O molecules	424
Number of other molecules	
Glycerol	10
Acetate	2
Sulfate	2
Ramachandran plot	
Outliers [%]	0
Allowed [%]	100
Most favoured [%]	96.8
Poor rotamer [%]	2.26
Overall B-value [Å^2]	13.1

Six poor rotamers were identified using the MolProbiety server (Davis *et al.*, 2007). These residues (Asn4, Asp21, Lys41, Lys153, Leu169, Asn285) are located on the surface of the molecule. For all except the Asn285 side chain, electron density was poor and/or indicated a possible alternative conformation suggesting that these residues were disordered. The density for Asn285 is well defined and the rotamer used fits it extremely well. This residue is held in position *via* a hydrogen bond between its *ND2* and Thr287 *OG1* and two hydrogen bonds from its *OD1* atom to water molecules Wat137 and Wat361.

[27] R and $R_{free} = \sum_{hkl} | F_{obs} - F_{calc} | / \sum_{hkl} F_{obs}$; R and R_{free} differ in the set of reflections they are calculated from.

7.3.7 The Overall Structure of Recombinant PNGase F

The overall structure of recombinant PNGase F proved very similar to the structures determined previously for the native protein (Kuhn *et al.*, 1994; Norris *et al.*, 1994b). Superposition of the final structure model with native PNGase F (PDB ID: 1PGS; **Figure 7.5**) using SSM on the EBI server (Krissinel & Henrick, 2004) resulted in a rmsd value of 0.36 Å between the C_α-atoms over the 314 residues. For the graphical presentation SSM (based on (Krissinel & Henrick, 2004)) was performed in COOT (Emsley & Cowtan, 2004) and the final picture prepared in PyMOL (DeLano, 2002).

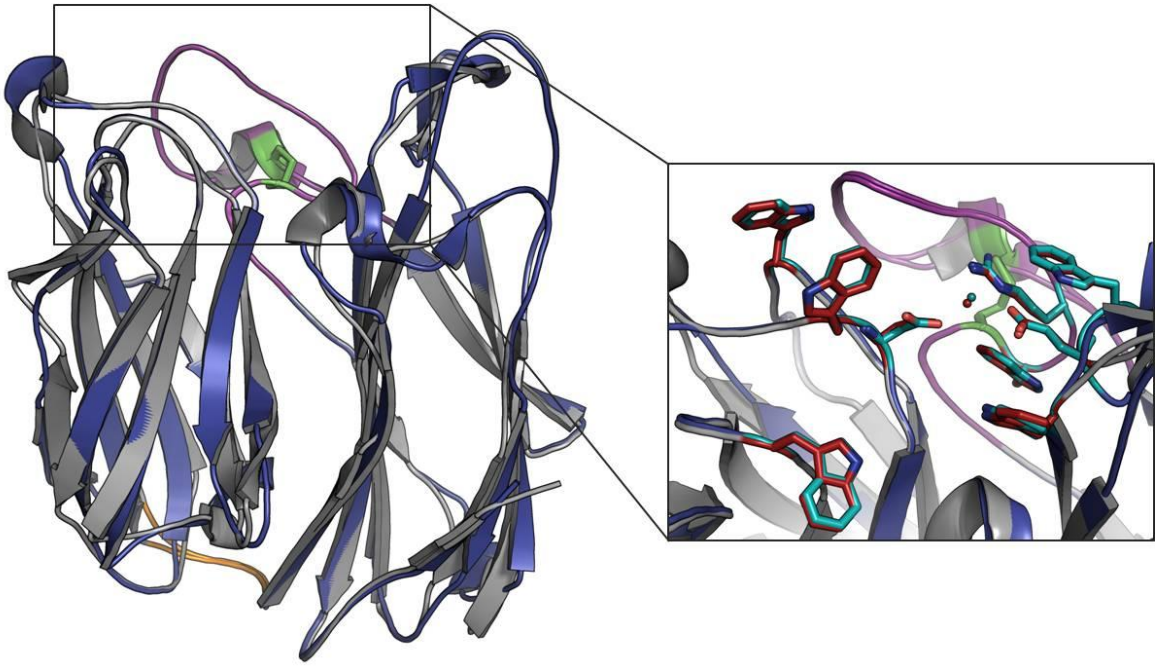


Figure 7.5: Superposition of rPNGase F with 1PGS.
Domain 1 is shown on the left and domain 2 on the right. Blue: 1PGS; Grey: rPNGase F; Orange: loop connecting domains 1 and 2 on the bottom of the molecule; Magenta: loop connecting domains on the top with disulphide bridge (green) joining two parts of this extended loop. The close up shows the active site area flanked with six tryptophan residues and the three proposed catalytic residues Asp60, Glu206 and Arg248 with the connecting water molecule. Cyan: 1PGS; Red: rPNGase F.

Details of the crystal structure for native PNGase F have been described previously by Norris *et al.* (1994) and in Chapter 1 (1.2.1.1). Briefly, the molecule consists of two eight-stranded antiparallel β-sandwich domains with jelly roll topology that include residues 1-135 (Domain 1) and 142-315 (Domain 2), respectively. The two domains are connected on the bottom of the molecule *via* a short peptide comprising residues 136-141 in 1PGS and 135-141 in rPNGase F. The second important loop that connects the two domains on the top of the

molecule, and also forms parts of the active site groove, is also well conserved as would be expected. This loop (residues 227-249) is an interstrand connection of domain 2 that reaches across to domain 1 tying the two domains together. It forms a double loop in which the first part (residues 227-249) extends to domain 1 and returns to domain 2. The second part formed by residues 250-257 is connected to the first part by a disulfide bridge (231-252). All three disulphide bridges present in 1PGS (51-56, 204-208, 231-252) were also found in rPNGase F. As shown in the close-up view in **Figure 7.5** the positions of the tryptophan residues that are thought to be involved in either substrate binding and/or generation of a hydrophobic environment in the active site are almost identical in both molecules. This is also the case for the three proposed catalytic residues Asp60, Glu206 and Arg248. These residues are connected by a conserved water molecule, Wat423 in 1PGS and Wat67 in rPNGase F. As mentioned above, a difference in six amino acid residues was shown between 1PGS and the amino acid sequence deduced from the nucleotide sequence of rPNGase F (Loo *et al.*, 2002). All of these differences were also identified in the crystal structure model of rPNGase F by analysis of the electron density map and difference electron density map. These substitutions (T39A, I149V, G168A, A219S, T269I, S281N 1PGS vs. rPNGase F) are located on the protein surface distant from the active site and do not appear to affect enzyme activity.

The presence of these important structural features and the preservation of the overall fold showed that rPNGase F was correctly folded. It also confirmed that the results obtained from circular dichroism experiments (6.3.4) were reliable, which is important in the case of the mutant proteins given that no crystal structures could be obtained to unequivocally demonstrate their correct folding.

7.3.8 Implications from Glycerol Molecules in the Active Site

It has been observed previously that PNGase F is reversibly inhibited by glycerol (personal communication Dr. G.E. Norris). As mentioned above glycerol was used as cryo-protectant prior to collection of crystal diffraction data (7.2.2, 7.3.4). During the refinement process electron density for ten glycerol

molecules was observed, three of them (GOL1, GOL7, GOL10) located in or close to (<8 Å) the active site pocket on top of the molecule as shown in **Figure 7.6**.

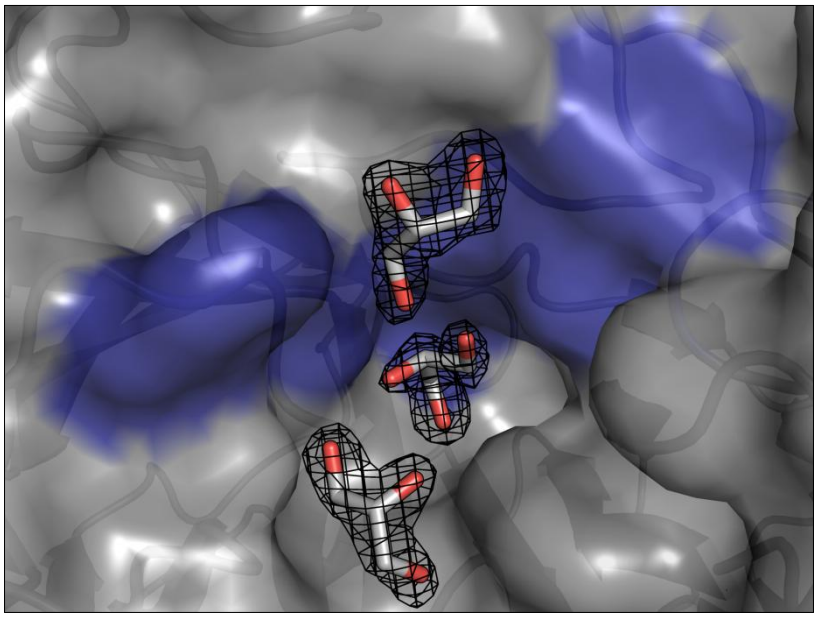


Figure 7.6: Electron density for three glycerol molecules bound to the active site.
From top to bottom: GOL1, GOL10, GOL7. The surface of rPNGase F is shown in grey with the active site area highlighted in blue.

A glycerol somewhat resembles the structure of trioses, which are classed as carbohydrates. The natural substrates of PNGase F are glyco-peptides or -proteins, and as has been shown previously, both the glycan and the peptide moiety affect activity (Altmann *et al.*, 1995; Fan & Lee, 1997). Thus, interactions of the glycerol molecules with PNGase F may indicate how the sugar moiety of glycopeptides binds to the active site. The structure of native PNGase F complexed with *N,N'*-diacelylchitobiose has been reported (PDB ID: 1PNF, (Kuhn *et al.*, 1995). This structure was obtained by soaking the crystal in a large excess of *N,N'*-diacelylchitobiose. **Figure 7.7** shows a two-dimensional representation of the contacts between the three glycerols and rPNGase F residues. This plot was generated using LIGPLOT (Wallace *et al.*, 1995). LIGPLOT uses the PDB-file of the crystal structure model to predict the hydrogen bonds and/or hydrophobic contacts of one or several specified components based on distance and geometry. In this plot, contacts made by the glycerol molecules GOL1, GOL7 and GOL10 (renumbered in the plot, refer to figure legend) are shown.

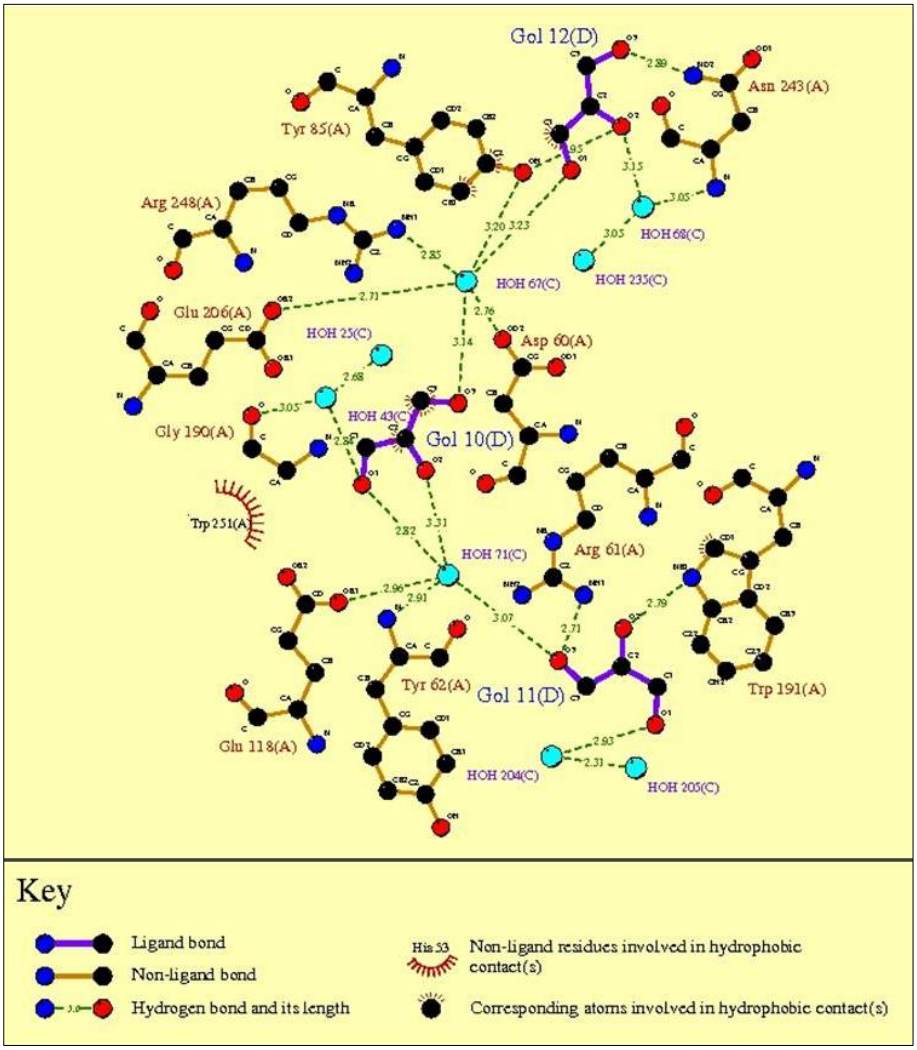


Figure 7.7: Interactions of glycerol molecules with rPNGase F and water molecules.
This plot was drawn using the program LIGPLOT (Wallace *et al.*, 1995). Two glycerol molecules were renumbered for this presentation: Gol11 is GOL7 in the final PDB-file and Gol12 is GOL1. Hydrogen bonds drawn by LIGPLOT are automatically generated using HBPLUS (McDonald & Thornton, 1994), which calculates all potential hydrogen bonds using distances and angles defined by Baker & Hubbard (Baker & Hubbard, 1984). Therefore, some water molecules show more than the possible four hydrogen bonds.

The LIGPLOT shows that GOL12 (GOL1 in PDB-file numbering) forms hydrogen bonds with Asn243 *via* its O_3 and with Tyr85-OH *via* O_2 (2.95 Å). It also might interact with two water molecules (Wat67, Wat68), although these bonds are possibly quite weak as the bond distances of 3.23 and 3.15 Å are relatively long for hydrogen bonds. The maximum distances as assigned by Baker & Hubbard (1984) are 3.5 Å for donor···acceptor distance or 2.5 Å where a hydrogen atom can be assigned (i.e. donor-H···acceptor). Of the three glycerol molecules GOL10 is located the deepest in the active site pocket (see **Figure**

7.6) and is positioned there through hydrophobic contacts with Trp251 and hydrogen bonds to three water molecules (Wat43, Wat67, Wat71). GOL11 (GOL7 PDB numbering) forms hydrogen bonds with two rPNGase F side chain atoms (NE_1 of Trp191, NH_1 of Arg61) and two water molecules.

Several residues involved in binding of the glycerol molecules were also found to make contacts to *N,N'*-diacetylchitobiose either directly or *via* connecting water molecules (**Figure 1.5**; (Kuhn *et al.*, 1995)). The most interesting similarity, however, is the position of Wat67, which is basically identical in both complexed structures and makes the same contacts to the protein. In the glycerol complexed structure it also makes a 3.23 Å hydrogen bond with the O_1 hydroxyl oxygen of GOL1 (PDB numbering). In the *N,N'*-diacetylchitobiose complex structure the equivalent water (Wat346) forms a hydrogen bond (3.29 Å) to the O_1 of the reducing end GlcNAc, which replaces (according to Kuhn *et al.* (1995)) the *ND* atom of the asparagine in the natural substrate. **Figure 7.8** shows a stereo diagram of GOL1 bound in the active site of rPNGase F.

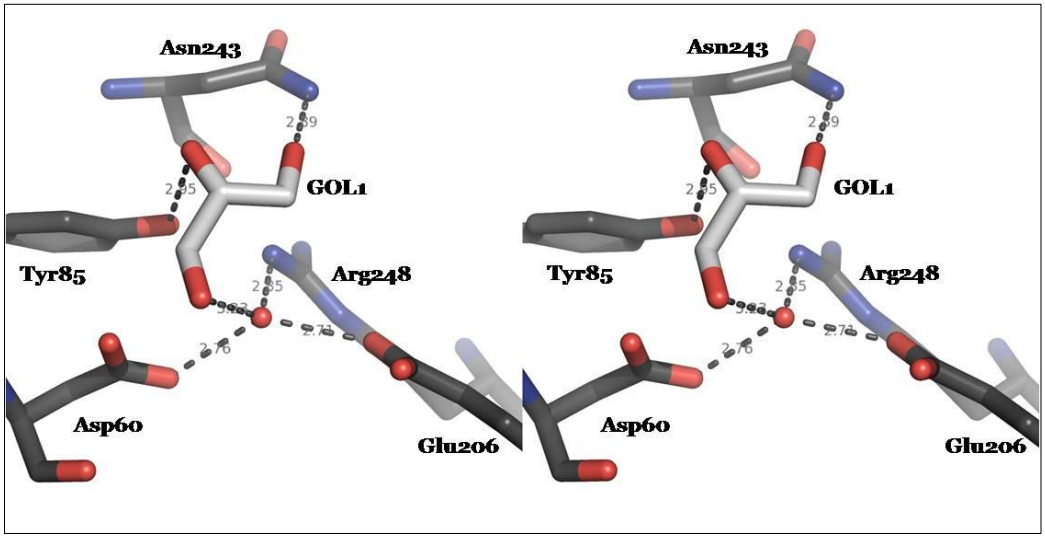


Figure 7.8: Stereo diagram of GOL1 bound in the active site of rPNGase F. The water molecule shown is Wat67.

Kuhn *et al.* further concluded that residue Asp60 is of primary importance for the catalytic mechanism as its OD_1 atom hydrogen bonds to the O_1 atom of the reducing end GlcNAc of *N,N'*-diacetylchitobiose (2.64 Å) and appeared to be essential for catalytic activity. Here, the distance of Asp60 OD_1 to the O_1 hydroxyl oxygen of GOL1 is within the allowed range (3.21 Å), but in a

geometrically unfavourable position when O_1 is considered as hydrogen donor and Asp60 OD_1 as the acceptor (angle C_1-O_1···OD_1 = 70.7°). Therefore, this was not considered as a hydrogen bond in LIGPLOT. The role of Asp60 in catalysis will be further discussed in Chapter 8.

Another result of this structural study revealed that the water (Wat422 in 1PGS; (Norris *et al.*, 1994b)) predicted to be the nucleophile for the cleavage of the amide bond between the asparagine side chain and the proximal GlcNAc (as described in Section 6.1) was not present in the glycerol complex structure. Instead it was replaced by GOL1. However, another water molecule (Wat423) equivalent to Wat67 was also found in structure 1PGS in almost the same position. **Figure 7.9** shows the water molecules Wat422 and Wat423 from structure 1PGS (green) superposed onto the glycerol complexed structure of rPNGase F.

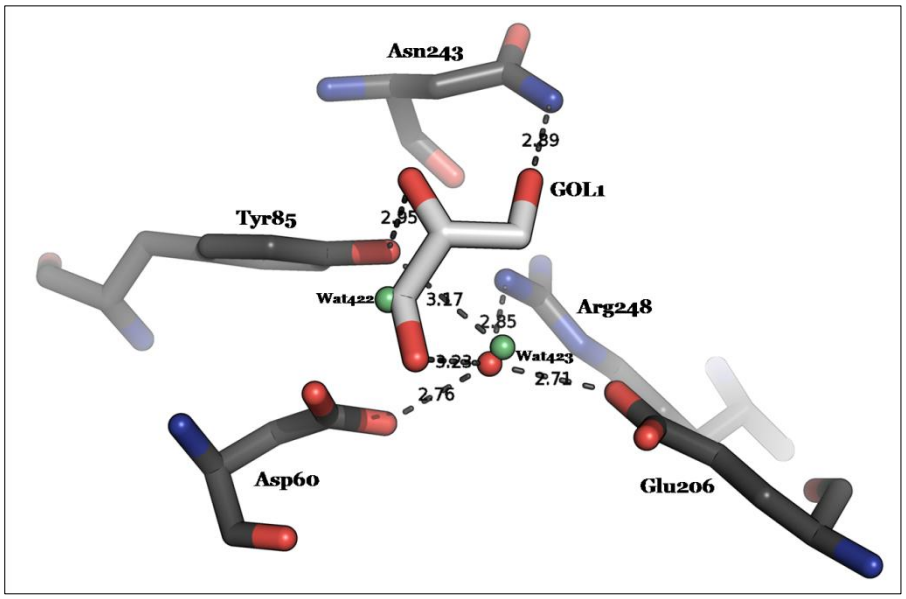


Figure 7.9: Replacement of Wat422 with GOL1.
Water molecules from structure 1PGS are shown in green. Wat67 is shown in red.

Two main conclusions can be drawn from the results described above:

(i) The presence of three glycerol molecules in the active site of rPNGase F explains the previous observation that glycerol acts as an inhibitor of PNGase F. This inhibition could either be due to the fact that the proposed nucleophile Wat422 (1PGS) is displaced by glycerol and therefore not available for catalysis or it could simply be due to a blockage of the active site by glycerol molecules. Of

those two possibilities the latter appears more likely as glycerol is far smaller than a glyco-peptide or –protein and can more easily access the active site. It is unlikely that the substrate would be able to bind at all to the active site as all or most of the possible binding sites will be already occupied by glycerol molecules as indicated by the similarity of side chains involved in hydrogen bonding of both the *N,N'*-diacetylchitobiose and glycerol within the active site.

(ii) The absence of the proposed water nucleophile (Wat422) in the glycerol complexed structure suggests that this might not be the nucleophile. Another water molecule, which is present in all PNGase F structures, complexed or not, seems to be the more likely catalytic water. This water (Wat67 in rPNGase; Wat423 in 1 PGS; Wat346 in the *N,N'*-diacetylchitobiose complex structure) is bound tightly between the main proposed catalytic residues Asp60, Glu206 and Arg248. It is possible that it could come into close contact with the scissile bond of the natural substrate as indicated by the two complexed structures. However, these structures can only provide an indication as to where the carbohydrate moiety of a glycopeptide is likely to bind as the natural substrates of PNGase F are far more complex than the observed ligands. It also might be possible that some rearrangements are required for such a complex molecule to access the active site. *Co*-crystallisation of PNGase F with a non-cleavable substrate analogue could provide valuable information on the binding mode of a complete glycopeptide to PNGase F. Three such substrate analogues have been reported in the literature: a high-mannose *C*-glycopeptide (Wang *et al.*, 1997), a high-mannose glycopeptide analogue containing a glucose-asparagine linkage (Deras *et al.*, 1998) and complex oligosaccharide *N*-linked to the side chain of a glutamine (Haneda *et al.*, 2001). To further elucidate the catalytic mechanism, crystallisation of the enzyme complexed with a transition state analogue is required. So far no such analogue has been produced although attempts were made with the synthesis of *N*-glycosyl phosphonamidates (Ferro *et al.*, 1998) and *C*-glycopeptides (Lenz, 2003).

Chapter 8

Kinetic Characterisation of rPNGase F Site-Specific Mutants

8 Kinetic Characterisation of rPNGase F Site-Specific Mutants

8.1 Introduction

Despite two independent structure analyses (Kuhn *et al.*, 1994; Norris *et al.*, 1994b) no definite catalytic mechanism has been established for PNGase F. To test whether the hypothesised catalytic mechanism described in the introduction to Chapter 6 (6.1) may be true, several mutations were introduced into PNGase F and their effect analysed by enzyme kinetic studies.

8.2 Methods

8.2.1 Preparation of PNGase Substrate Ovalbumin Glycopeptide (Norris et al., 1994a)

a. Cyanogen Bromide Digest of Hen Egg White Ovalbumin

Hen egg white ovalbumin (12 g; Sigma grade 2) was dissolved in 120 mL 50% (v/v) aqueous formic acid. Cyanogen bromide (CNBr; 2.7 g) was dissolved in acetonitrile and added to the ovalbumin solution under argon. This reaction mixture was incubated overnight at room temperature with stirring.

CNBr was removed from the mixture by rotary evaporation. Water (400 mL) was added to the mixture followed by the reduction of the total volume back to the original 120 mL. This step was repeated once to remove the last traces of CNBr.

b. Precipitation

Remaining, undigested protein and insoluble peptides were removed from the reaction mixture by the addition of an aqueous solution of trichloroacetic acid (TCA) to a final concentration of 5%. The precipitate was removed by centrifugation at 10,000 *g* for 15 minutes.

The supernatant was then extracted three times with diethyl ether to remove TCA. After removal of the ether, the sample was concentrated to approximately 10 mL by rotary evaporation.

Acetic acid was added to a final concentration of 0.5% (v/v) and insoluble material was removed from the solution by centrifugation at 14,000 *g* for 10 minutes.

c. Size Exclusion Chromatography & Reverse Phase-HPLC

Size exclusion chromatography (100×2.5 cm; Bio-Rad®, P-4, 50-100 mesh equilibrated in 0.1 M acetic acid) was used to separate glycopeptides. Aliquots of the supernatant (2.5 mL) obtained in the previous step were applied to the column, which was then eluted with 0.1 M acetic acid at a flow rate of 2 mL/min. Peptides eluting from the column were detected by monitoring the absorbance at 280 nm and 8 mL fractions were collected.

Each fraction was then analysed for the presence of glycopeptide using the phenol/sulphuric acid test for reducing sugars (Dubois *et al.*, 1956). 12.5 µL 80% (w/v) phenol and 1.25 mL concentrated H_2SO_4 were added to 0.5 mL sample and monitored for a colour change to orange/brown, which indicates presence of reducing sugars (Dubois *et al.*, 1956). Those fractions producing brown colour were pooled, lyophilised and the resulting product dissolved in water. Further purification was performed by RP-HPLC (250×10 mm, pore size 5 µm; C18; Jupiter series; Phenomenex). A 15 minute gradient was applied from 20% acetonitrile/ 0.08% TFA to 40% acetonitrile/ 0.08% TFA at a flow rate of 4 mL/min to elute glycopeptides. An elution profile was obtained by detection of peptides at 214 nm. Peak fractions were collected, lyophilised and analysed using the PNGase activity assay (2.26.2).

The glycopeptide-containing fraction was lyophilised and stored at -20°C until further use.

8.2.2 Preparation of Fluoresceine Isothiocyanate -labelled Substrate for PNGase F Activity Assay (adapted from (Hentz et al., 1997))

The standard assay for PNGase F involves the detection of the substrate and its deglycosylated product at 214 nm. It had already been shown that because of the low concentration of substrate required for the assay, detection at 214 nm was not sufficiently sensitive (Lenz, 2003). Therefore, the purified 11-mer ovalbumin glycopeptide (1296.4 Da) was labelled with fluoresceine isothiocyanate (FITC), which allowed fluorescent detection giving increased sensitivity. **Figure 8.1** shows a schematic illustration of the glycopeptide, indicating the possible labelling positions and the FITC molecule.

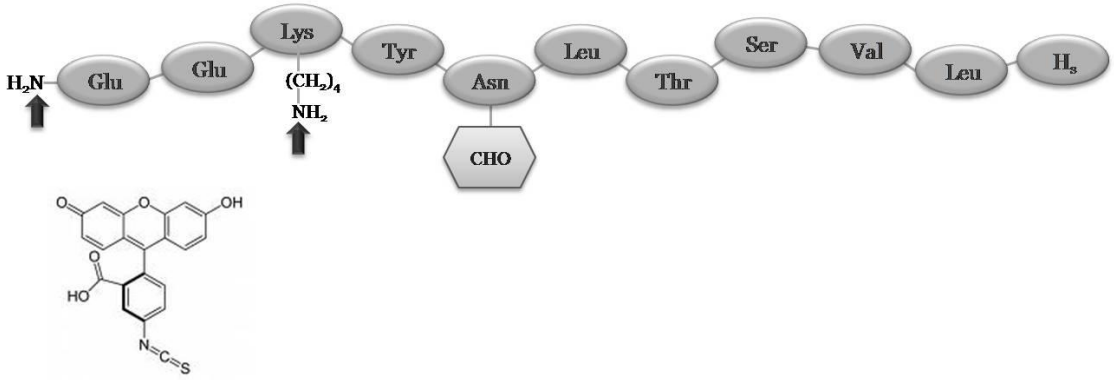


Figure 8.1: Schematic illustration of ovalbumin glycopeptide with FITC.
The arrows indicate the potential FITC labelling positions. CHO represents the *N*-glycan. H_s: Homoserine.

FITC reacts with amines and can therefore modify the α-amino group at the N-terminus and the ε-amino group of the lysine residue. A pH of 8.5 to 9.5 is required for the modification of lysine residues, whereas at neutral pH (~pH 7) the N-terminal amino group can be selectively modified. Here, the labelling reaction was performed at pH 7, resulting in a peptide selectively labelled at the N-terminus.

Approximately 100 mg of the purified ovalbumin glycopeptide was dissolved in 20 mL 0.1 M Na_2HPO_4/NaH_2PO_4 buffer (pH 7.0) in a 250 mL round bottom

flask, which was wrapped in tinfoil to protect the light sensitive FITC molecules. 2 mL of 0.5% FITC in acetone (w/v) were added drop wise with slow stirring to the ovalbumin peptide solution. The reaction mixture was incubated overnight at room temperature to allow the labelling reaction to complete. It had been previously observed that the reaction results in the appearance of two peaks in RP-HPLC representing two ovalbumin-FITC (Ova-FITC) isomers, the homoserine-lactone form and the 'open' form. To completely convert the labelled product to the 'open' form, the reaction mixture was lyophilised, resuspended in 0.1 M NH_4HCO_3 buffer (pH 8.5) and boiled in a water bath for 30 minutes. The solution was again lyophilised and the products analysed by subjecting a small quantity to RP-HPLC using a preparative C18 column (250×10 mm, pore size 5 µm; Jupiter; Phenomenex) and the following gradient at a flow rate of 4 mL/minute:

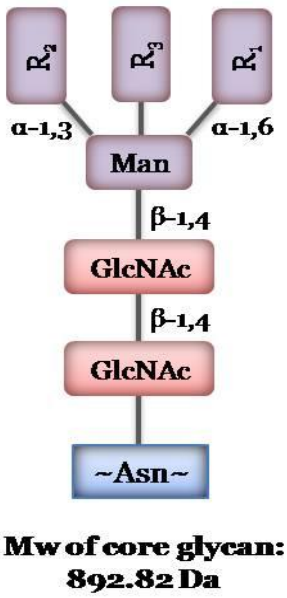
Isocratic flow, 80% solvent A (0.1% TFA in pure water), 20% solvent B (0.08% TFA in acetonitrile) for 2 minutes
Gradient to 60% A, 40% B over 5 minutes
Gradient to 30% A, 70% B over 10 minutes
Gradient to 100% B over 5 minutes
Isocratic flow at 100% B for 5 minutes
Gradient to 80% A, 20% B over 10 minutes

The elution was monitored by fluorescence at 520 nm and peaks were collected by following the emission profile (excitation: 495 nm, emission: 520 nm), lyophilised and analysed using the PNGase activity assay (2.26.2). After identification of the peak of interest the remaining FITC-labelled substrate was purified using the same RP-HPLC protocol. This purification step also ensured the removal of unlabelled ovalbumin peptide due to the difference in retention time (Ova: 7.6 min; Ova-FITC: 12.7 min under assay conditions). This was important as any unlabelled peptide would still be processed by the enzyme in activity assays without being detected by fluorescence. This would alter the kinetic constants of the enzyme as only FITC-labelled product was detected and used for the determination of catalytic rates.

Ova-FITC is unstable if stored at 4°C in biological buffers for extended periods of time, but is stable at -20°C. As multiple cycles of freeze-thawing

appeared to result in the re-conversion of the 'open' form to the lactone-form, aliquots were stored to minimise this effect.

As shown in **Figure 8.2**, ovalbumin exists in nine uniformly distributed glycoforms (hybrid and high-mannose) of varying molecular weights (Sharon, 1982). Therefore an average molecular weight was used for the calculation of molar substrate concentrations.



Group	Mw [Da]
I	1096.02
IIA	933.88
IIB	933.88
IIIA	730.68
IIIB	648.56
IIIC	771.74
IV	486.42
V	324.28
VI	162.14
Average	**676.4**

Figure 8.2: Hen egg white ovalbumin glycoforms.
Glycans of different molecular weights substitute positions R_1, R_2 and R_3.

The average molecular weight of the FITC-labelled ovalbumin substrate is therefore:

Peptide + core glycan + average glycan + FITC

= 1296.4 + 892.82 + 676.4 + 389.4 = **3255.02 Da**

8.2.3 Determination of PNGase F Activity

PNGase activity was determined as described in section 2.26.2 with slight modifications. The reaction volume was reduced to 30 µL (27 µL substrate + 3 µL enzyme) and 10 µL were loaded onto the RP-HPLC column. The reactions

were incubated at 37°C for varying amounts of time, depending on the purpose of the assay, before the enzyme was heat inactivated by transfer of the tubes into a boiling water bath for three minutes. For the determination of kinetic parameters, an incubation time of 4 minutes was chosen after a time course was run using two substrate concentrations (0.9 mg/mL, 0.225 mg/mL) for the wildtype enzyme. For the recording of reaction progress curves, incubation times ranged from 1 minute to 20 minutes.

The K_m for recombinant PNGase F using Ova-FITC as substrate had been determined previously to be 0.13 mg/mL (Lenz, 2003). This value was used as a guideline for the substrate concentrations employed for the determination of the kinetic parameters for rPNGase F and the rPNGase F mutant preparations used in this study. Final substrate concentrations used were: 0.0225 mg/mL, 0.045 mg/mL, 0.09 mg/mL, 0.225 mg/mL, 0.45 mg/mL, 0.675 mg/mL and 0.9 mg/mL. Three independent reactions were performed for each substrate concentration.

Different enzyme concentrations had to be used for the different mutants and the wildtype in order to be able to measure initial velocities using the constant 4 minute incubation time. Therefore, a reaction progress curve was recorded for each enzyme at the highest and lowest substrate concentration (0.9 mg/mL, 0.0225 mg/mL) to ensure that initial reaction velocities were measured, i.e. product appearance was within the linear range and represented less than ~15% conversion of the substrate. Time points taken were 1, 3, 5, 10 and 20 minutes. Only two substrate concentrations were analysed performing only single measurements for each time point of the time course in order to minimise substrate usage as this is very difficult and time consuming to produce in the quantities required. Enzyme concentrations are presented as mg/mL and/or μM. For the conversion from mg/mL to molar concentrations, the molecular weights determined for each protein by mass spectrometry were used (6.3.3).

8.2.3.1 Standard Curves

Enzymatic activity was determined by measuring the amount of product peptide produced in a reaction. This product was quantified by measuring the integrated area under the product peak. Standard curves were generated relating specified amounts of product to their integrated peak area. For this, an aliquot of Ova-FITC substrate (1 mg/mL) was completely deglycosylated with PNGase F. This solution was then diluted to different concentrations and 10 µL of each dilution were loaded onto the RP-HPLC column. The integrated area under the product peak was determined and plotted against the amount of product that was loaded onto the column. Different substrate preparations were used during this work and a standard curve was generated for each of them (**Figure 8.3**).

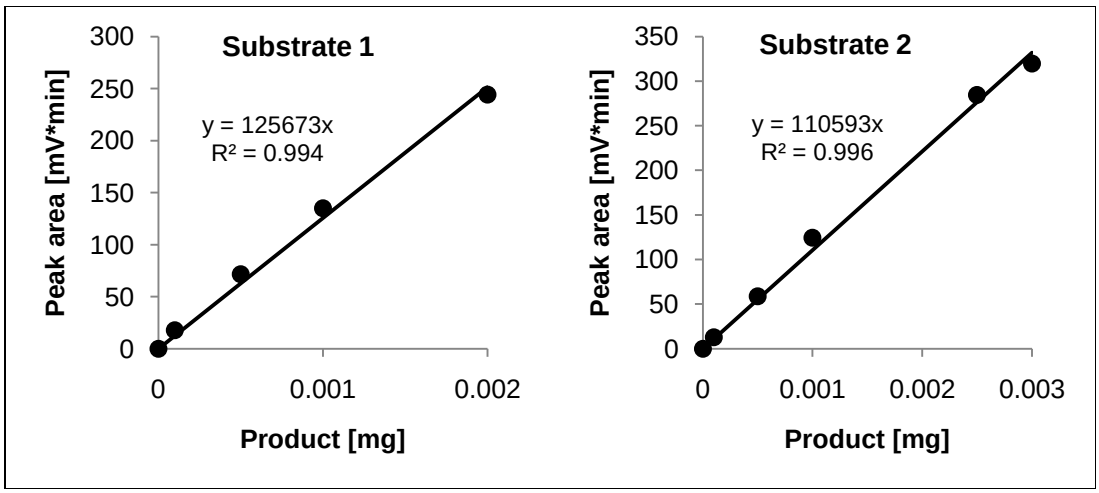


Figure 8.3: Standard curves.

As mentioned above (8.2.2), because hen egg white ovalbumin has nine different glycoforms of varying molecular weights, an average molecular weight was used for the calculation of molar concentrations. Due to the inherent error of using an average molecular weight, substrate concentrations and rates are given in mg/mL and mg/mL*min^{-1}, respectively. The kinetic parameters K_m and V_{max} are given in mg/mL and mg/mL*min^{-1}, respectively and in µM and µM*min^{-1}. For the calculation of k_{cat} and k_{cat}/K_m only the micromolar values were used.

8.2.3.2 Presentation of Kinetic Data and Determination of Kinetic Parameters

Experimental data were plotted and analysed using the software GraphPad Prism® 5. This program allows nonlinear curve-fitting to experimental data for different kinetic models (equations) as well as different methods for data transformation to fit a linear regression. The preferred and most accurate method for determining kinetic parameters, such as V_{max} and K_m, is nonlinear curve-fitting (Copeland, 2000), which was used here. Nonlinear regression gives more accurate results as linear transformations change the influences of the data values and the error structure of the model and the interpretation of any inferential results. Although somewhat outdated due to the availability of nonlinear regression, data obtained in activity assays were also transformed using the Hanes-Woolf transformation for comparative purposes. Data were plotted using a Hanes-Woolf plot, where $[S]/(v)$ is plotted against [S]. In this linear plot the slope is $1/V_{max}$, the y-intercept is K_m/V_{max} and the x-intercept is $-K_m$.

The values for k_{cat} and k_{cat}/K_m were calculated using K_m and V_{max} (micromolar values) determined using nonlinear regression.

The value of k_{cat}, also referred to as the turnover number of an enzyme, defines the number of catalytic turnover events that occur per time unit. Turnover numbers are typically reported in units of molecules of product produced per time unit per molecules of enzyme present. Hence, $k_{cat} = V_{max}/[E]$. The units of k_{cat} are reciprocal time. It defines the maximal velocity at which an enzymatic reaction can proceed at a fixed enzyme concentration and infinite availability of substrate. Changes in k_{cat} reflect perturbations of the chemical steps subsequent to substrate binding (Copeland, 2000).

The ratio k_{cat}/K_m defines the catalytic efficiency of an enzyme and is therefore generally used to compare the efficiencies of different enzymes or mutants of an enzyme to one another (Copeland, 2000). It combines the effectiveness of transformation of bound product and the effectiveness of substrate binding. However, it should be noted that the use of k_{cat}/K_m for comparing the catalytic efficiency of enzymes has been strongly questioned in recent years. Eisenthal *et al.* (2007) demonstrated that in a general case an

enzyme with a higher k_{cat}/K_m value can, at certain substrate concentrations, catalyse an identical reaction at lower rates than an enzyme with a lower k_{cat}/K_m value (Eisenthal *et al.*, 2007). Despite this, this ratio is still frequently used and will therefore be presented here.

The error of the experimental data was found to range between 1% and 5%.

8.3 Results & Discussion

Tables showing the raw data, i.e. the integrated product peak areas obtained during kinetic characterisation of PNGase F and its mutants, are presented in Appendix 4.

8.3.1 PNGase F Wildtype

As mentioned above (**8.2.3**) the K_m for recombinant PNGase F using Ova-FITC as substrate had been determined previously to be 0.13 mg/mL (Lenz, 2003). To test whether this value was correct or if the PNGase F prepared here differed from the previously isolated enzyme, the kinetic characteristics of the PNGase F wildtype protein were determined.

Figure 8.4 shows the reaction progress curve generated for PNGase F wildtype at a final enzyme concentration of 5.0×10^{-6} mg/mL (1.38×10^{-4} µM). The plot of product appearance versus reaction time shows that the deglycosylation of Ova-FITC was approximately linear with time for the first ~5 minutes. Hence, an incubation time of 4 minutes at the given enzyme concentration was considered to be suitable for the determination of kinetic parameters of wildtype PNGase F.

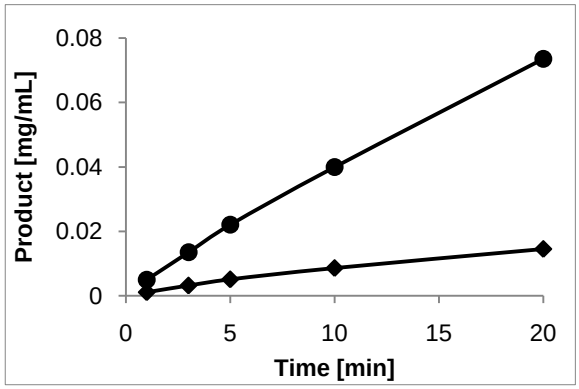


Figure 8.4: Reaction progress curve of wildtype PNGase F.
The final enzyme concentration was 5.0×10^{-6} mg/mL (1.38×10^{-4} µM). •, 0.9 mg/mL; ♦ 0.0225 mg/mL Ova-FITC.

Figure 8.5 shows the results of the reaction rate determination at different substrate concentrations. PNGase F appeared to exhibit substrate or product inhibition as indicated by the decline in reaction rate at high substrate concentrations of >~0.45 mg/mL. Therefore, the nonlinear regression model for substrate inhibition was used in the program GraphPad Prism® 5 to fit the curve to the experimental data and to determine K_m and V_{max}. This was also the model chosen by the software for the presented data when the Michaelis-Menten and the substrate inhibition models were compared. The K_m was calculated to be 0.2 mg/mL (55.3 µM) and V_{max} 5.9×10^{-3} mg/mL*min^{-1} (1.8 µM*min^{-1}). The K_m determined here for the same substrate is higher than the value obtained earlier by Lenz (2003). However, previously the Michaelis-Menten model was used to fit the experimental data although the data clearly showed the same decline in rate at higher substrate concentrations indicative of substrate inhibition. This would probably have significantly altered the kinetic values. When the Michaelis-Menten model was used to fit the data obtained here the following values were obtained: K_m = 0.07 mg/mL, V_{max} = 3.3×10^{-3} mg/mL*min^{-1}. This demonstrates the importance of choosing the best fitting model for nonlinear regression. Furthermore, Lenz (2003) did not specify which of the methods shown was used for K_m and V_{max} calculations, nonlinear regression (Michaelis-Menten) or linear regression of Lineweaver-Burk transformed data. When K_m was determined using the Hanes-Woolf transformation and linear regression, K_m was slightly lower compared to the K_m obtained by Lenz and considerably lower than the value calculated using nonlinear regression and the substrate inhibition model. As mentioned above, the most accurate way to determine K_m

and V_{max} is by nonlinear regression and therefore the values obtained using this method should be regarded as more accurate.

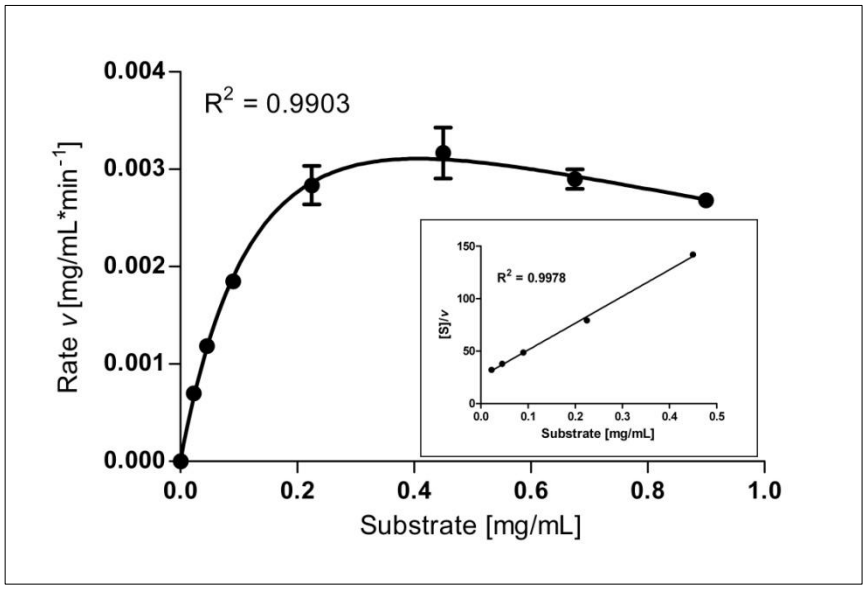


Figure 8.5: Kinetics of wildtype rPNGase F.
The plot shows the rate of generation of deglycosylated Ova-FITC as a function of substrate (Ova-FITC) concentration. The enzyme appears to be susceptible to substrate inhibition; therefore this model was used to fit the curve (nonlinear regression). Each data point represents the average of three independent reactions, and error bars show the standard deviation. The inset shows the Hanes-Woolf transformation of the five lowest substrate concentrations. For both plots, the R^2 value is given as a measure of the 'goodness of fit'. Both plots were generated using GraphPad Prism® 5.

The kinetic parameters determined for the recombinant PNGase F wildtype protein are summarised in **Table 8.1**.

Table 8.1: Kinetic parameters for wildtype rPNGase F.
Values for K_m and V_{max} were determined for both the substrate inhibition model and the Hanes-Woolf transformation. Values in µM and µM*min^{-1} are given in parentheses. k_{cat} and k_{cat}/K_m were calculated using the micromolar concentrations. Final enzyme concentration: 5.0×10^{-6} mg/mL (1.38×10^{-4} µM).

	Regression model/ Transformation		**k_{cat} [sec^{-1}]**	**k_{cat}/K_m [µM^{-1} sec^{-1}]**
	Substrate inhibition	**Hanes-Woolf**		
K_m [mg/mL]	0.2 (55.3 µM)	0.1 (30.8 µM)		
			217.4	3.9
V_{max} [mg/mL*min^{-1}]	5.9×10^{-3} (1.8 µM*min^{-1})	3.9×10^{-3} (1.2 µM*min^{-1})		

8.3.2 PNGase F W59Q

The progress curve generated for PNGase F mutant W59Q with two substrate concentrations is shown in **Figure 8.6**. At 4 minutes incubation time and using a final enzyme concentration of 7.5×10^{-4} mg/mL the reaction was still in the initial, linear phase. Hence, these conditions were used for the subsequent kinetic characterisations of this protein.

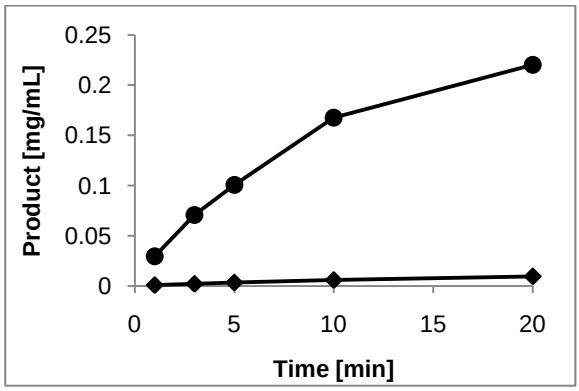


Figure 8.6: Reaction progress curve of PNGase F W59Q.
The final enzyme concentration was 7.5×10^{-4} mg/mL (0.0207 μM). •, 0.9 mg/mL; ♦ 0.0225 mg/mL Ova-FITC.

The graphical result of the kinetic characterisation of PNGase F W59Q is shown in **Figure 8.7** and indicates that higher substrate concentrations are needed for an accurate determination of its kinetic values. However, higher substrate concentrations could not be used because of the very limited availability of substrate due to the difficulties in preparing it and time constraints. A lower enzyme concentration was not used here because the resulting response would have been too low to be measured accurately. Despite these shortcomings approximate values for K_m and V_{max} were obtained. As for this mutant substrate inhibition was not apparent, the Michaelis-Menten model was used for nonlinear regression.

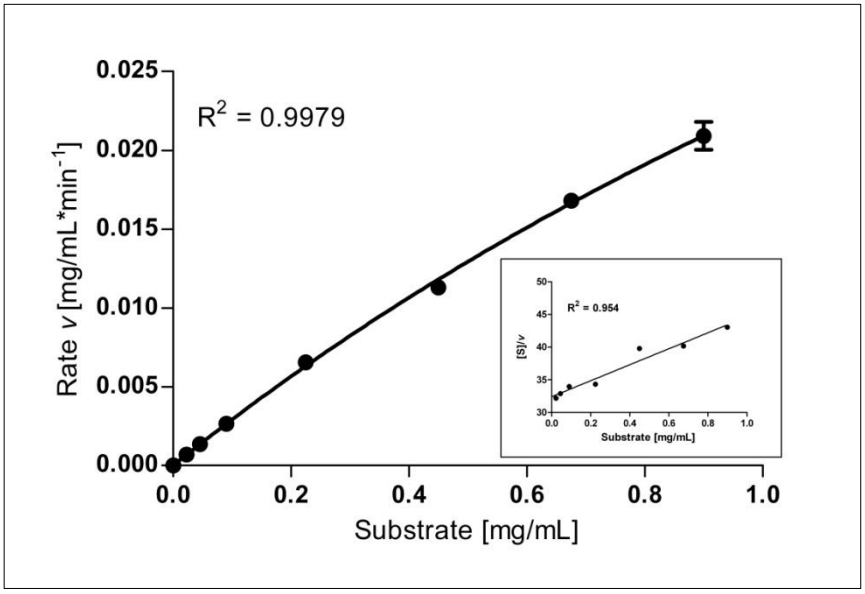


Figure 8.7: Kinetics of PNGase F W59Q.
The plot shows the rate of generation of deglycosylated Ova-FITC as a function of substrate (Ova-FITC) concentration. Each data point represents the average of two to three independent reactions, and error bars show the standard deviation. The inset shows the Hanes-Woolf transformation. R^2 values are given as a measure of the 'goodness of fit'. Both plots were generated using GraphPad Prism® 5.

The kinetic parameters for this mutant are summarised in **Table 8.2**. K_m showed a ~17-fold increase compared to the wildtype, indicating a decrease in substrate binding affinity. The k_{cat} also decreased ~10-fold indicating that the mutation also impaired the chemistry of the reaction subsequent to substrate binding. Both of these values lead to an overall decrease in catalytic efficiency by ~160-fold as shown by the ratio k_{cat}/K_m.

Table 8.2: Kinetic parameters for PNGase F W59Q.
Values for K_m and V_{max} were determined for both the Michaelis-Menten model and the Hanes-Woolf transformation. Values in µM and µM*min^{-1} are given in parentheses. k_{cat} and k_{cat}/K_m were calculated using the micromolar concentrations. Final enzyme concentration: 7.5×10^{-4} mg/mL (0.0207 µM).

	Regression model/ Transformation		**k_{cat} [sec^{-1}]**	**k_{cat}/K_m [µM^{-1} sec^{-1}]**
	Michaelis-Menten	**Hanes-Woolf**		
K_m [mg/mL]	3.0 (936.1 µM)	2.7 (813.8 µM)		
			22.7	0.02
V_{max} [mg/mL*min^{-1}]	0.1 (28.2 µM*min^{-1})	0.1 (25.1 µM*min^{-1})		

Trp59 was mutated to test its role in maintaining a hydrophobic environment around Asp60, which was predicted to be essential if Asp60 is to accept a proton from water at pH 8.5. It was postulated that mutating Trp to Gln would decrease the hydrophobicity in the active site. The far more dramatic increase of K_m was, however, unexpected and suggests that Trp59 is even more important for substrate binding than it is for enabling Asp60 to accept a proton. The importance of aromatic residues in carbohydrate binding is well known. The interactions between carbohydrates and aromatic amino acid side chains are also referred to as 'stacking-' or 'CH-π-interactions' and examples can be found in a variety of carbohydrate binding/-active proteins (Fernandez *et al.*, 2005; Patanjali *et al.*, 1984; Spiwok *et al.*, 2004).

8.3.3 PNGase F D60C

The progress curves that were recorded for the mutant D60C using a final enzyme concentration of 2.0×10^{-4} mg/mL are shown in **Figure 8.8** and demonstrate that product production is approximately linear with time for ~10 minutes under the specified conditions.

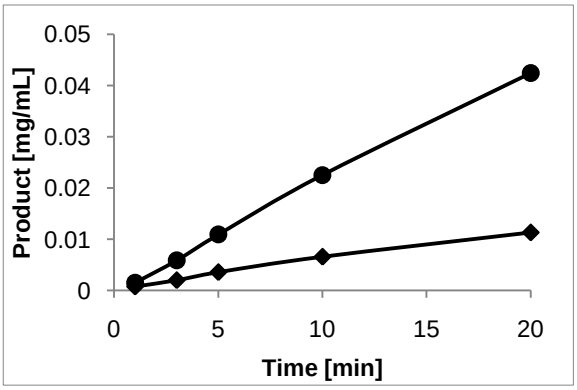


Figure 8.8: Reaction progress curve of PNGase F D60C.
The final enzyme concentration was 2.0×10^{-4} mg/mL (5.52×10^{-3} μM). •, 0.9 mg/mL; ♦ 0.0225 mg/mL Ova-FITC.

The graphical presentation of the kinetics of this mutant is shown in **Figure 8.9**. As seen for the PNGase F wildtype this protein also exhibits substrate inhibition. Here it becomes apparent from substrate concentrations >~0.225 mg/mL.

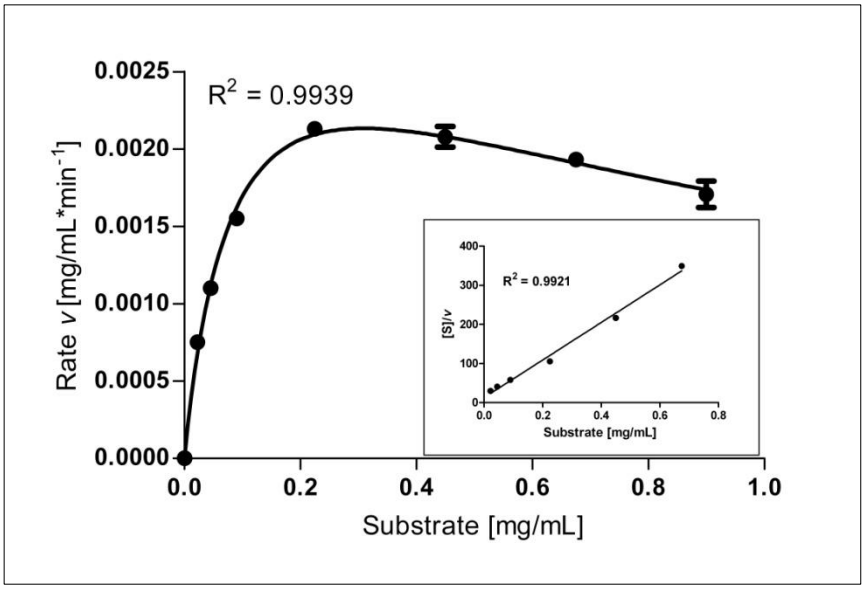


Figure 8.9: Kinetics of PNGase F D60C.
The plot shows the rate of generation of deglycosylated Ova-FITC as a function of substrate (Ova-FITC) concentration. Each data point represents the average of three independent reactions, and error bars show the standard deviation. The inset shows the Hanes-Woolf transformation. R^2 values are given as a measure of the 'goodness of fit'. Both plots were generated using GraphPad Prism® 5.

Table 8.3 summarises the values obtained for the different kinetic parameters. The overall catalytic efficiency represented by k_{cat}/K_m, decreased by a factor of ~35, which appears to be a result of the effect of this mutation on the reaction mechanism after formation of the enzyme-substrate complex as substrate binding seems to occur more effectively than in the wildtype (~2-fold lower K_m). The transformation of substrate to product, however, is less effective as indicated by the ~70-fold decrease in k_{cat}.

Table 8.3: Kinetic parameters for PNGase F D60C.
Values for K_m and V_{max} were determined for both the substrate inhibition model and the Hanes-Woolf transformation. Values in µM and µM*min^{-1} are given in parentheses. k_{cat} and k_{cat}/K_m were calculated using the micromolar concentrations. Final enzyme concentration: 2.0×10^{-4} mg/mL (5.52×10^{-3} µM).

	Regression model/ Transformation		**k_{cat} [sec^{-1}]**	**k_{cat}/K_m [µM^{-1} sec^{-1}]**
	Substrate inhibition	**Hanes-Woolf**		
K_m [mg/mL]	0.1 (27.5 µM)	0.03 (8.1 µM)		
			3.1	0.1
V_{max} [mg/mL*min^{-1}]	3.4×10^{-3} (1.0 µM*min^{-1})	2.1×10^{-3} (0.7 µM*min^{-1})		

These results were unexpected as, if the proposed mechanism is correct, the cysteine residue should, in the environment occupied by Asp60, remain protonated and be incapable of activating the water postulated to be the nucleophile in the reaction. It was expected that the mutation D60C would render the enzyme inactive. Indeed, k_{cat} has considerably decreased compared to the value measured for the wildtype enzyme, but among all the mutants tested it showed the fourth highest k_{cat} value. The fact that the Asp → Cys mutation led to a lower K_m might indicate that the cysteine residue is able to form an additional bond with the substrate. Sulfur-containing hydrogen bonds (SCHBs) have been found to play important roles in intramolecular interactions (Gregoret *et al.*, 1991; Zhou *et al.*, 2009). SCHBs are longer than hydrogen bonds involving oxygen and/or nitrogen because of sulfur's larger size and more diffuse electron cloud. The strength of SCHBs, however, is thought to be less than that of hydrogen bonds involving nitrogen and oxygen atoms. The average distance between a sulfhydryl group (donor) and amide nitrogen (acceptor) was found to be 3.65 Å and 3.51 Å for amide oxygen (Zhou *et al.*, 2009) compared to the maximum distance of 3.5 Å for O···O(N) hydrogen bonds (Baker & Hubbard, 1984). Furthermore, the sulfur atom can also participate in non-hydrogen interactions with amides, carbonyl groups and aromatic rings. However, without a crystal structure of this mutant and a better knowledge of how the native substrate binds to PNGase F, it is difficult to explain the higher binding affinity.

Considering these results, it is possible that an active site residue other than Asp60 activates the water molecule. The most likely candidate is Glu206, which has been shown to be very important for PNGase F activity (Kuhn *et al.*, 1995). The results obtained for mutants W207Q (8.3.6) and W251Q (8.3.9) and the close proximity of these residues to Glu206 support this possibility as it appears more important for Glu206 to be positioned in a hydrophobic environment than it does for Asp60. Furthermore, the replacement of the proposed catalytic Wat422 in 1PGS (Norris *et al.*, 1994b) by a glycerol molecule in the structure presented in Chapter 7, makes Wat423 (equivalent to Wat67 in rPNGase F) the likely candidate for the catalytic water. Wat67 is positioned between Asp60 and Glu206 at a distance of approximately 2.7 Å to either residue.

8.3.4 PNGase F I82Q

Figure 8.10 shows the reaction progress curves for the mutant I82Q at two substrate concentrations. The linearity of product appearance at 4 minutes indicates that the chosen conditions are suitable for the determination of kinetic parameters.

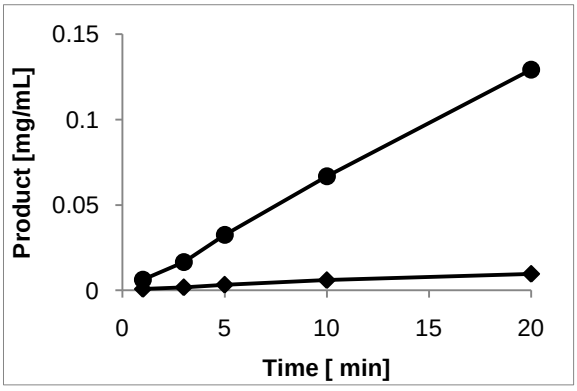


Figure 8.10: Reaction progress curve of PNGase F I82Q.
The final enzyme concentration was 5.0×10^{-5} mg/mL (1.38×10^{-3} μM). •, 0.9 mg/mL; ♦ 0.0225 mg/mL Ova-FITC.

Figure 8.11 shows the plots generated for the experimental data obtained for mutant I82Q. As for this mutant substrate inhibition was not observed, the Michaelis-Menten model was used for nonlinear regression.

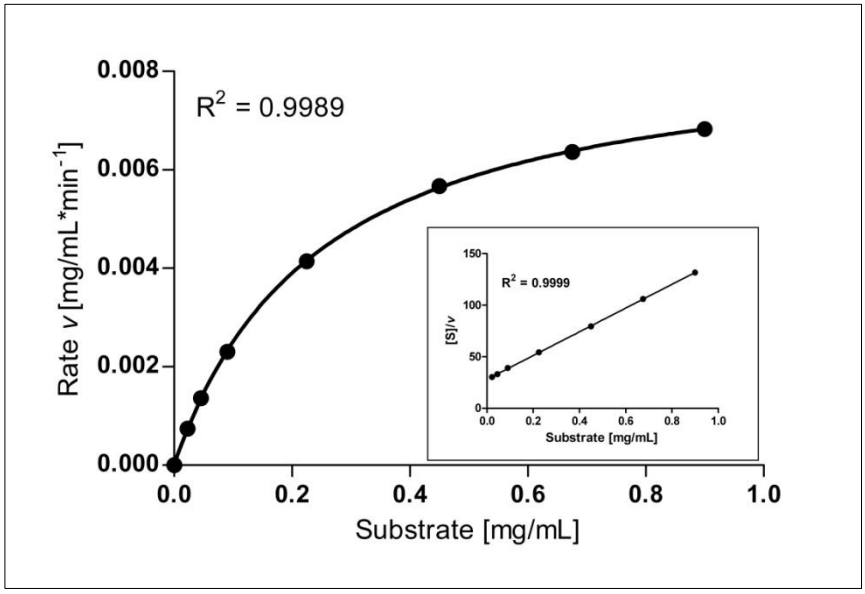


Figure 8.11: Kinetics of PNGase F I82Q.
The plot shows the rate of generation of deglycosylated Ova-FITC as a function of substrate (Ova-FITC) concentration. Each data point represents the average of three independent reactions, and error bars show the standard deviation. The inset shows the Hanes-Woolf transformation. R^2 values are given as a measure of the 'goodness of fit'. Both plots were generated using GraphPad Prism® 5.

Table 8.4 shows the kinetic values obtained for the mutant I82Q. The *turnover number* k_{cat}, decreased ~6.6-fold compared to the value obtained for the wildtype enzyme and K_m increased by ~50%, indicating a lower affinity of the enzyme for Ova-FITC. The combination of both values in the ratio k_{cat}/K_m showed that this mutant is ~9x less efficient than the wildtype.

Table 8.4: Kinetic parameters for PNGase F I82Q.
Values for K_m and V_{max} were determined for both the Michaelis-Menten model and the Hanes-Woolf transformation. Values in µM and µM*min^{-1} are given in parentheses. k_{cat} and k_{cat}/K_m were calculated using the micromolar concentrations. Final enzyme concentration: 5.0×10^{-5} mg/mL (1.38×10^{-3} µM).

	Regression model/ Transformation		**k_{cat} [sec^{-1}]**	**k_{cat}/K_m [µM^{-1} sec^{-1}]**
	Michaelis-Menten	**Hanes-Woolf**		
K_m [mg/mL]	0.3 (77.3 µM)	0.2 (74.87 µM)		
			32.7	0.4
V_{max} [mg/mL*min^{-1}]	8.8×10^{-3} (2.7 µM*min^{-1})	8.7×10^{-3} (2.7 µM*min^{-1})		

These results show that the introduction of a polar, but uncharged residue into the active site area does affect the catalytic efficiency of PNGase F, reducing it ~9-fold compared to the wildtype enzyme suggesting that the presence of the hydrophobic isoleucine in close proximity to Asp60 might be important. However, this mutant has the second highest overall catalytic efficiency among all mutants examined here. If the hydrophobic environment of Asp60 was essential as proposed, this mutation should have reduced the activity even more. This again points to the possibility that it is not Asp60 that accepts an H^+ from the catalytic water, as maintenance of its hydrophobic environment does not appear essential for activity.

Effects of this mutation will be further discussed in section 8.3.5.

8.3.5 PNGase F I82R

Figure 8.12 shows the reaction progress curve generated for this mutant at a final enzyme concentration of 0.1 mg/mL (2.76 μM), which is 2×10^4-times higher than the enzyme concentration required for the wildtype PNGase F. The plot of product appearance versus reaction time shows that the deglycosylation of Ova-FITC was approximately linear with time for the complete incubation time of 20 minutes. Hence, an incubation time of 4 minutes at the given enzyme concentration was within the initial phase of the enzyme reaction and therefore suitable for the determination of kinetic parameters of wildtype PNGase F.

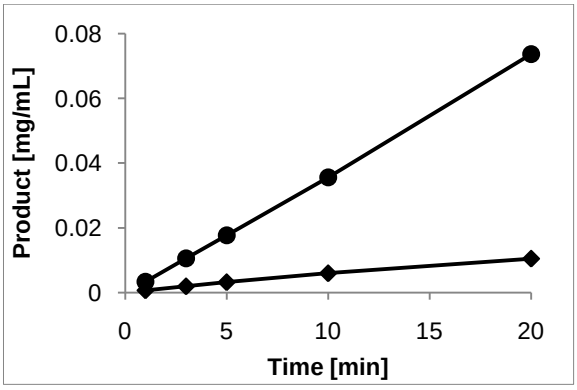


Figure 8.12: Reaction progress curve of PNGase F I82R.
The final enzyme concentration was 0.1 mg/mL (2.76 μM). •, 0.9 mg/mL; ♦ 0.0225 mg/mL Ova-FITC.

Figure 8.13 shows the graphical result of the kinetic characterisation of PNGase F I82R. As for this mutant substrate inhibition was not apparent, the Michaelis-Menten model was used for nonlinear regression.

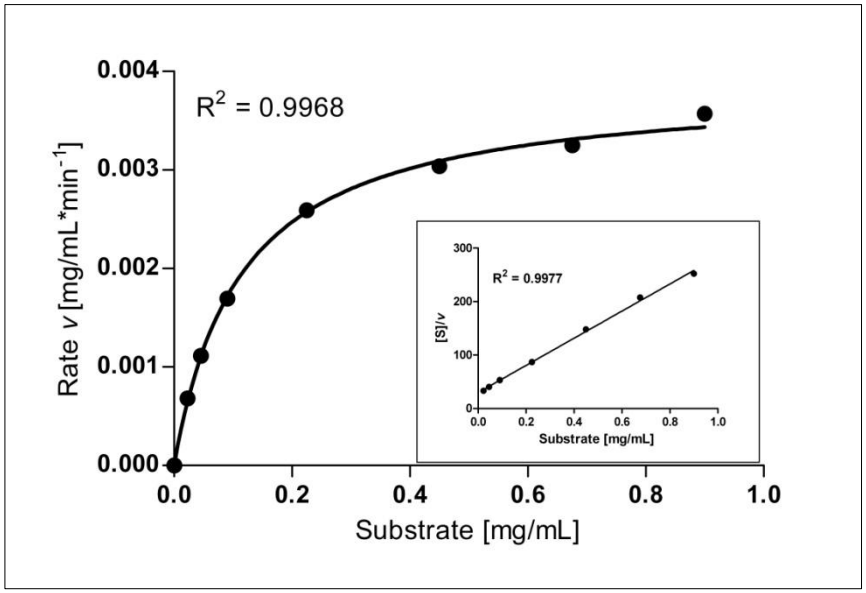


Figure 8.13: Kinetics of PNGase F I82R.
The plot shows the rate of generation of deglycosylated Ova-FITC as a function of substrate (Ova-FITC) concentration. Each data point represents the average of two to three independent reactions, and error bars show the standard deviation. The inset shows the Hanes-Woolf transformation. R^2 values are given as a measure of the 'goodness of fit'. Both plots were generated using GraphPad Prism® 5.

The kinetic parameters obtained for this mutant are summarised in **Table 8.5**. Although K_m decreased for this mutant compared to the wildtype enzyme, the very high enzyme concentration required to achieve any substrate conversion is reflected in the $\sim 31 \times 10^4$-fold decrease of k_{cat}. This mutation results in an enzyme whose overall catalytic efficiency (k_{cat}/K_m) is approximately 2×10^4-times lower than that of the wildtype'.

Table 8.5: Kinetic parameters for PNGase F I82R.
Values for K_m and V_{max} were determined for both the Michaelis-Menten model and the Hanes-Woolf transformation. Values in µM and µM*min^{-1} are given in parentheses. k_{cat} and k_{cat}/K_m were calculated using the micromolar concentrations. Final enzyme concentration: 0.1 mg/mL (2.76 µM).

	Regression model/ Transformation		**k_{cat} [sec^{-1}]**	**k_{cat}/K_m [µM^{-1} sec^{-1}]**
	Michaelis-Menten	**Hanes-Woolf**		
K_m [mg/mL]	0.1 (34.6 µM)	0.1 (36.25 µM)		
			7.0×10^{-3}	2.1×10^{-4}
V_{max} [mg/mL*min^{-1}]	3.9×10^{-3} (1.2 µM*min^{-1})	3.9×10^{-3} (1.2 µM*min^{-1})		

The effect of the introduction of a charged residue in place of the hydrophobic isoleucine rendered the enzyme almost inactive. A high concentration of enzyme was required to be able to measure any enzymatic activity. Hence, the overall catalytic efficiency of this mutant was the lowest of all mutants, with a residual activity of ~0.005% relative to the wildtype. *In silico* mutation of Ile82 to arginine showed that this may not be a result of changes in the hydrophobic environment of Asp60. **Figure 8.14** shows *in silico* models for both mutations of Ile82, I82R (panel (A)) and I82Q (panel (B)) with two rotamers for each introduced amino acid.

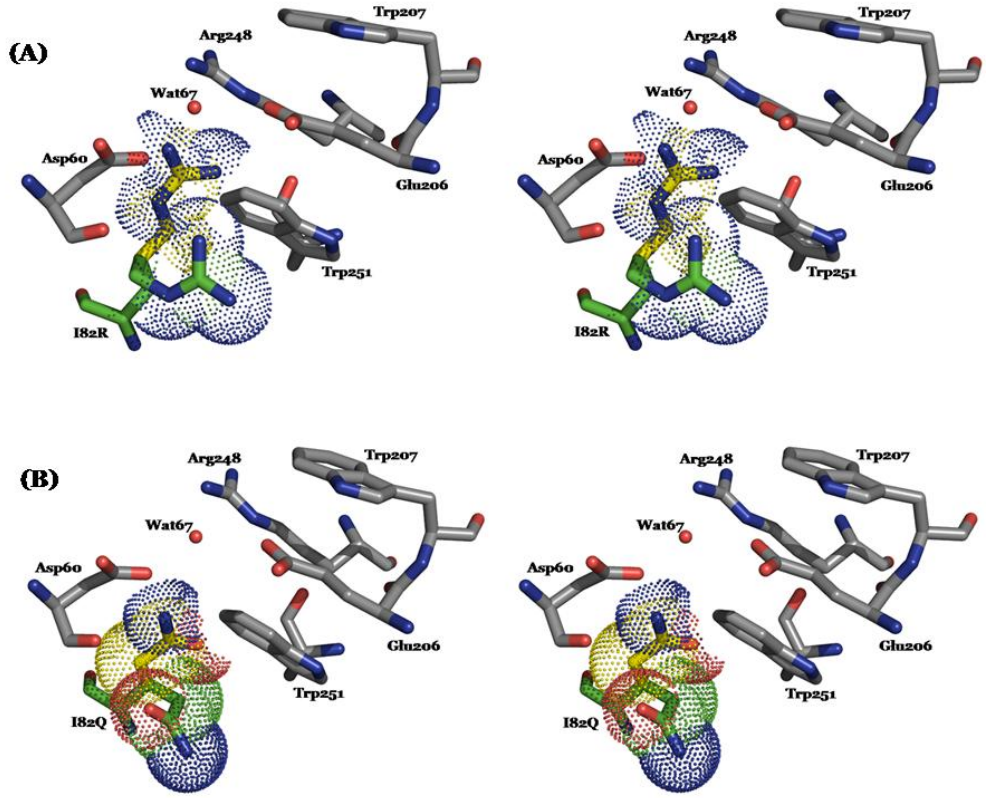


Figure 8.14: Stereo diagrams of rPNGase F with modelled mutations I82R (A) and I82Q (B).
Two possible amino acid rotamers are shown for each mutation (yellow, green). The mutations were modelled into the crystal structure model of rPNGase F (Chapter 7) using COOT (Emsley & Cowtan, 2004).

As shown in **Figure 8.14 (A)** the introduction of the long arginine side chain may cause a rearrangement or displacement of Asp60 as well as Trp251 and Wat67, leading to disturbance of residues involved in the catalytic mechanism. This is especially the case for the rotamer shown in yellow. The decrease in K_m could be explained by the formation of additional hydrogen bonds between the substrate's hydroxyl groups and the arginine's guanidinium nitrogen atoms. The second rotamer (green) is less invasive, but comes close to GOL10 (Chapter 7; not shown in **Figure 8.14 (A)**). In this scenario the arginine could provide additional binding sites for the substrate (lower K_m), but alter the way it binds so that the scissile bond is not in the correct position for catalysis (low k_{cat}). Ile82 is located at the bottom of the active site pocket and the introduction of any residue larger than isoleucine could possibly result in decreased (if no additional binding sites are introduced) or increased binding affinity, as described above. In both cases the way the substrate binds to the

enzyme is likely to change, resulting in a less optimal positioning of the scissile bond. This theory fits with the results obtained for the I82Q mutant, where binding affinity decreased (increased K_m), but catalysis still occurs with ~15% of the wildtype efficiency. The green rotamer shown in **Figure 8.14 (B)** does not directly interfere with the active site residues. It might, however, interfere with substrate binding providing fewer hydrogen bonding partners for the substrate compared to arginine. Its smaller size might also result in smaller perturbation of the substrate binding mode, allowing limited catalysis to occur (i.e. 15% relative to wildtype). The other scenario for this mutation (yellow rotamer) would result in direct interference with residues Asp60 and Trp251 although to a lesser extent than the arginine mutation. It should, however, not directly interfere with Wat67. The glutamine residue may push Asp60 slightly out of its original position and form a hydrogen bond between its NE_2 and Asp60 OD_2. In this scenario, Asp60 would not be able to form a hydrogen bond to Wat67, which would then possibly be less well positioned to act as a nucleophile. The OE_2 atom could in fact initiate a domino-effect by disrupting a number of important hydrogen bonds (Trp251 → Glu206; Arg248 → Trp207).

These explanations of the kinetic results obtained for the mutants I82R and I82Q are, however, preliminary and need to be substantiated by actual structural data. The most important information to be drawn from these mutants is that because replacing Ile82 with a polar residue did not inactivate the enzyme, a hydrophobic environment around Asp60 is not essential for activity. Replacing Ile82 with an arginine might be misleading because the introduction of the large arginine residue most likely results in a number of distortions within the active site, thus affecting substrate binding and catalysis.

8.3.6 PNGase F W207Q

The progress curves that were recorded for the mutant W207Q using a final enzyme concentration of 0.025 mg/mL are presented in **Figure 8.15** and show that product production is linear with time for ~10 minutes under the specified conditions.

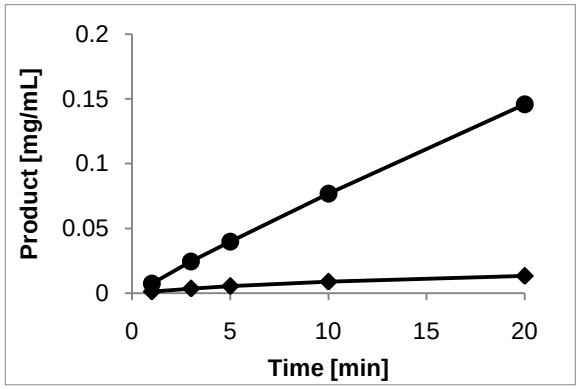


Figure 8.15: Reaction progress curve of PNGase F W207Q.
The final enzyme concentration was 0.025 mg/mL (0.691 µM). •, 0.9 mg/mL; ♦ 0.0225 mg/mL Ova-FITC.

Figure 8.16 shows the graphic analysis of the reaction velocities obtained for this PNGase F mutant using nonlinear (Michaelis-Menten model) and linear regression (Hanes-Woolf transformation).

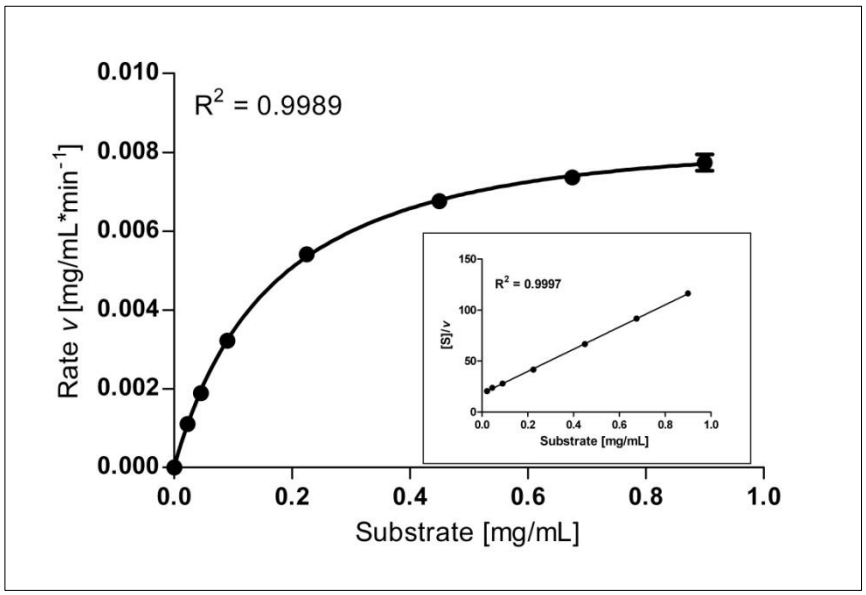


Figure 8.16: Kinetics of PNGase F W207Q.
The plot shows the rate of generation of deglycosylated Ova-FITC as a function of substrate (Ova-FITC) concentration. Each data point represents the average of three independent reactions, and error bars show the standard deviation. The inset shows the Hanes-Woolf transformation. R^2 values are given as a measure of the 'goodness of fit'. Both plots were generated using GraphPad Prism® 5.

It is apparent (**Table 8.6**), that the K_m of mutant W207Q is very similar to that of the wildtype value of 0.2 mg/mL (55.3 µM). k_{cat} on the other hand is 3000 times lower than the wildtype k_{cat}. This suggests that substrate binding

affinity was not impaired by this mutation, whereas the catalytic turnover of substrate to product was strongly affected.

Table 8.6: Kinetic parameters for PNGase F W207Q.
Values for K_m and V_{max} were determined for both the Michaelis-Menten model and the Hanes-Woolf transformation. Values in µM and µM*min^{-1} are given in parentheses. k_{cat} and k_{cat}/K_m were calculated using micromolar concentrations. Final enzyme concentration: 0.025 mg/mL (0.691 µM).

	Regression model/ Transformation		**k_{cat} [sec^{-1}]**	**k_{cat}/K_m [µM^{-1} sec^{-1}]**
	Michaelis-Menten	**Hanes-Woolf**		
K_m [mg/mL]	0.2 (55.3 µM)	0.2 (50.91 µM)		
			0.1	1.3×10^{-3}
V_{max} [mg/mL*min^{-1}]	9.7×10^{-3} (2.0 µM*min^{-1})	9.2×10^{-3} (2.8 µM*min^{-1})		

These results show that the presence of the large hydrophobic tryptophan residue in this position is important for PNGase F activity, but not for substrate binding affinity. Interestingly, Trp207 is positioned in an almost planar conformation just 'above' (slightly shifted sideways) residue Glu206 at a distance of 3.3 Å (between closest atoms Glu206 *OE_2* and Trp207 *CZ_2*; **Figure 8.14** or **Figure 8.21**). These results, combined with those obtained for the mutant W251Q, indicate that it is essential for Glu206 to be in a hydrophobic environment for the enzyme to function at optimal rates. Furthermore, it appears to be more important that Glu206 is in a hydrophobic environment than it is for Asp60.

8.3.7 PNGase F R248K

Figure 8.17 shows the reaction progress curves for two substrate concentrations for the mutant R248K. The product appearance at 4 minutes is approximately linear with time, which indicates that the chosen conditions are suitable for the determination of kinetic parameters.

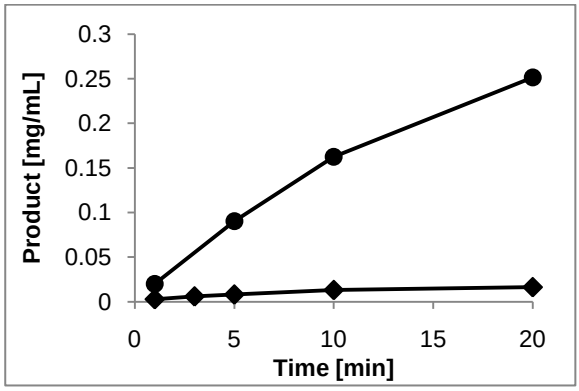


Figure 8.17: Reaction progress curve of PNGase F R248K.
The final enzyme concentration was 5.0×10^{-4} mg/mL (0.0138 μM). •, 0.9 mg/mL; ♦ 0.0225 mg/mL Ova-FITC.

Figure 8.18 shows the plots generated for the experimental data obtained for mutant R248K. As for this mutant substrate inhibition was not apparent, the Michaelis-Menten model was used for nonlinear regression.

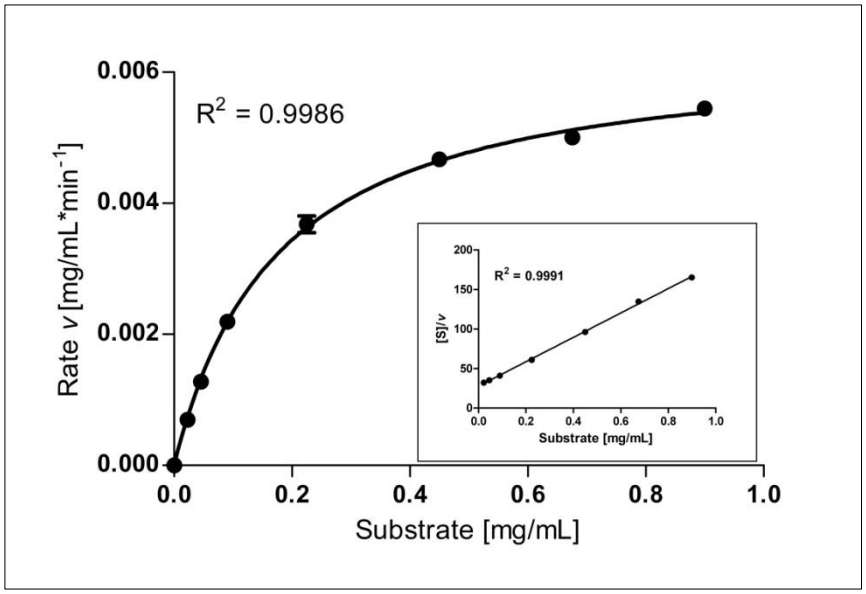


Figure 8.18: Kinetics of PNGase F R248K.
The plot shows the rate of generation of deglycosylated Ova-FITC as a function of substrate (Ova-FITC) concentration. Each data point represents the average of three independent reactions, and error bars show the standard deviation. The inset shows the Hanes-Woolf transformation. R^2 values are given as a measure of the 'goodness of fit'. Both plots were generated using GraphPad Prism® 5.

The mutation R248K seems to have minimal impact on substrate binding as the K_m is almost equal to the wildtype value. The turnover of substrate following substrate binding appears, however, to be impaired as indicated by the almost

90x decrease of k_{cat} compared to the wildtype protein. The kinetic parameters calculated for mutant R248K are summarised in **Table 8.7**.

Table 8.7: Kinetic parameters for PNGase F R248K.
Values for K_m and V_{max} were determined for both the Michaelis-Menten model and the Hanes-Woolf transformation. Values in µM and µM*min^{-1} are given in parentheses. k_{cat} and k_{cat}/K_m were calculated using the micromolar concentrations. Final enzyme concentration: 5.0×10^{-4} mg/mL (0.0138 µM).

	Regression model/ Transformation		**k_{cat} [sec^{-1}]**	**k_{cat}/K_m [µM^{-1} sec^{-1}]**
	Michaelis-Menten	**Hanes-Woolf**		
K_m [mg/mL]	0.2 (54.9 µM)	0.2 (55. 5 µM)		
			2.4	0.04
V_{max} [mg/mL*min^{-1}]	6.5×10^{-3} (2.0 µM*min^{-1})	6.5×10^{-3} (2.0 µM*min^{-1})		

8.3.8 PNGase F R248Q

The progress curves that were recorded for the mutant R248Q using a final enzyme concentration of 0.015 mg/mL are shown in **Figure 8.19**. The results demonstrate that reaction progress was approximately linear with time for ~5 minutes at the highest substrate concentration under the specified conditions and less than 15% of the initial substrate concentration was converted into product. Hence, an incubation time of 4 minutes appeared suitable for determination of kinetic constants.

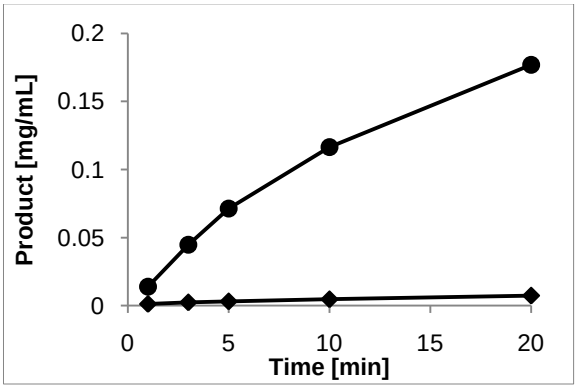


Figure 8.19: Reaction progress curve of PNGase F R248Q.
The final enzyme concentration was 0.015 mg/mL (0.414 µM). •, 0.9 mg/mL; ♦ 0.0225 mg/mL Ova-FITC.

The graphical result of the kinetic characterisation of PNGase F R248Q is shown in **Figure 8.20** and indicates that higher substrate concentrations would have been required for an accurate determination of the kinetic values. This was not possible as the availability of substrate was extremely limited and no more could be produced due to time constraints. However, estimates of the kinetic parameters can be obtained from these results. For this mutant substrate inhibition was not apparent and therefore the Michaelis-Menten model was used for nonlinear regression.

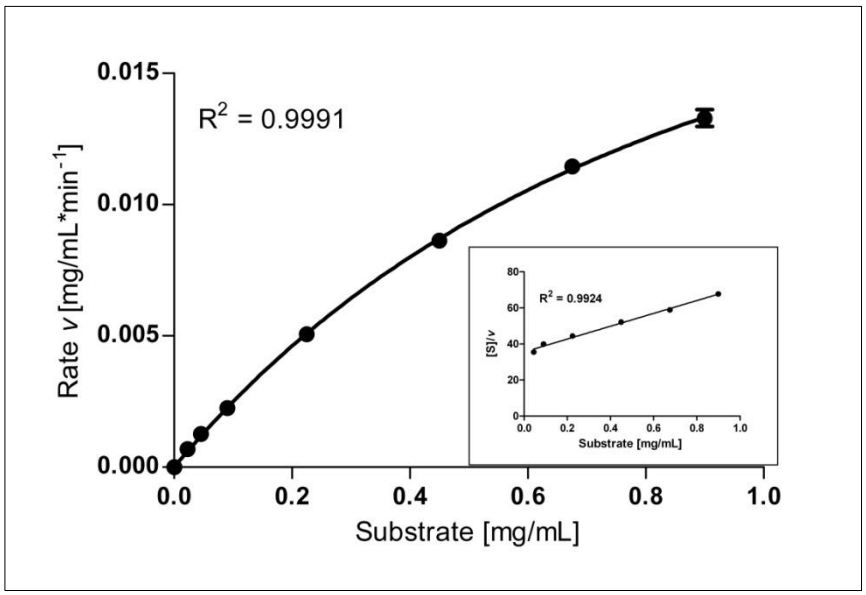


Figure 8.20: Kinetics of PNGase F R248Q.
The plot shows the rate of generation of deglycosylated Ova-FITC as a function of substrate (Ova-FITC) concentration. Each data point represents the average of three independent reactions, and error bars show the standard deviation. The inset shows the Hanes-Woolf transformation. R^2 values are given as a measure of the 'goodness of fit'. Both plots were generated using GraphPad Prism® 5.

The K_m calculated for this mutant is ~6-fold larger than that of the wildtype, indicating a decrease in substrate binding affinity. The *turnover number* k_{cat} decreased by a factor of ~550, suggesting severe distortions to the catalytic mechanism. As a result, the overall catalytic effectiveness, given by the ratio k_{cat}/K_m, decreased more than 3000-fold. The values of the kinetic parameters for this mutant are summarised in **Table 8.8**.

Table 8.8: Kinetic parameters for PNGase F R248Q.
Values for K_m and V_{max} were determined for both the Michaelis-Menten model and the Hanes-Woolf transformation. Values in µM and µM*min^{-1} are given in parentheses. k_{cat} and k_{cat}/K_m were calculated using micromolar concentrations. Final enzyme concentration: 0.015 mg/mL (0.414 µM).

	Regression model/ Transformation		**k_{cat} [sec^{-1}]**	**k_{cat}/K_m [µM^{-1} sec^{-1}]**
	Michaelis-Menten	**Hanes-Woolf**		
K_m [mg/mL]	1.1 (359.1 µM)	1.0 (307.5 µM)		
			0.4	1.1×10^{-3}
V_{max} [mg/mL*min^{-1}]	0.03 (9.7 µM*min^{-1})	0.03 (8.6 µM*min^{-1})		

Comparison of the results for mutants R248K and R248Q shows that having a positively charged residue in the position of Arg248 is probably important as the deglycosylation activity is less severely impaired when the arginine residue is substituted with a lysine. Nevertheless, the activity of R248K was still considerably lower than that of the wildtype. This might be due to the shorter side chain of lysine as compared to arginine, which is, looking at the structure, in an almost fully extended conformation (**Figure 8.21**). Therefore, in R248K the distance of the charged amino group of the lysine side chain to the carbonyl oxygen of the amide linkage might be too far. Furthermore, Arg248 is held tightly in place by a hydrogen-bond network. It is possible that a lysine residue in this position would be more flexible.

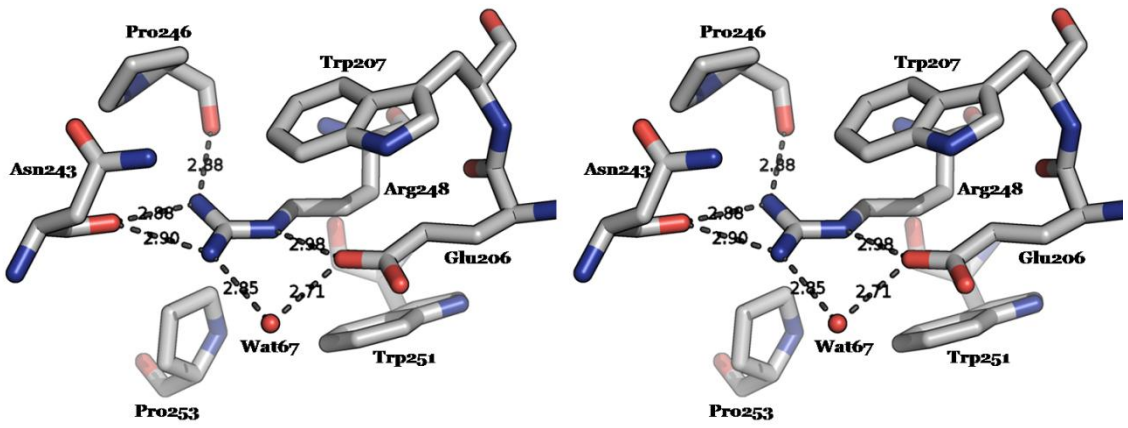


Figure 8.21: Arg248 is held tightly in place within the active site.

Mutation of arginine to the uncharged glutamine resulted in a residual catalytic efficiency of 0.03% relative to the wildtype. It is therefore likely that Arg248 plays an important role in the catalytic mechanism, one that is dependent on its charge as well as its position.

8.3.9 PNGase F W251Q

Figure 8.22 shows the reaction progress curves for two substrate concentrations for the mutant I82Q. The product appearance is approximately linear with time for 20 minutes, the maximal incubation time, and less than 15% of the initial substrate concentration was converted to product. This indicates that the conditions chosen for the reaction were suitable for the determination of kinetic parameters.

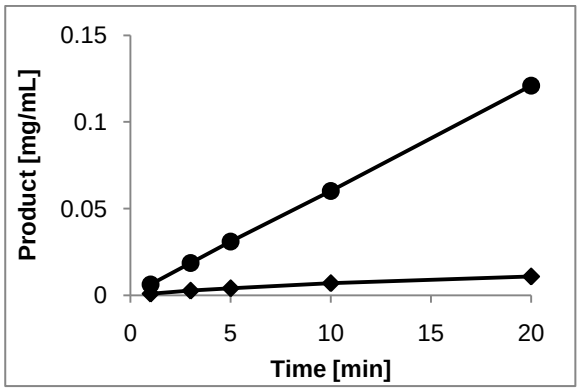


Figure 8.22: Reaction progress curve of PNGase F W251Q.
The final enzyme concentration was 7.5×10^{-4} mg/mL (0.0207 μM). •, 0.9 mg/mL; ♦ 0.0225 mg/mL Ova-FITC.

The results of the kinetic characterisation of PNGase F W251Q are shown in **Figure 8.23**.

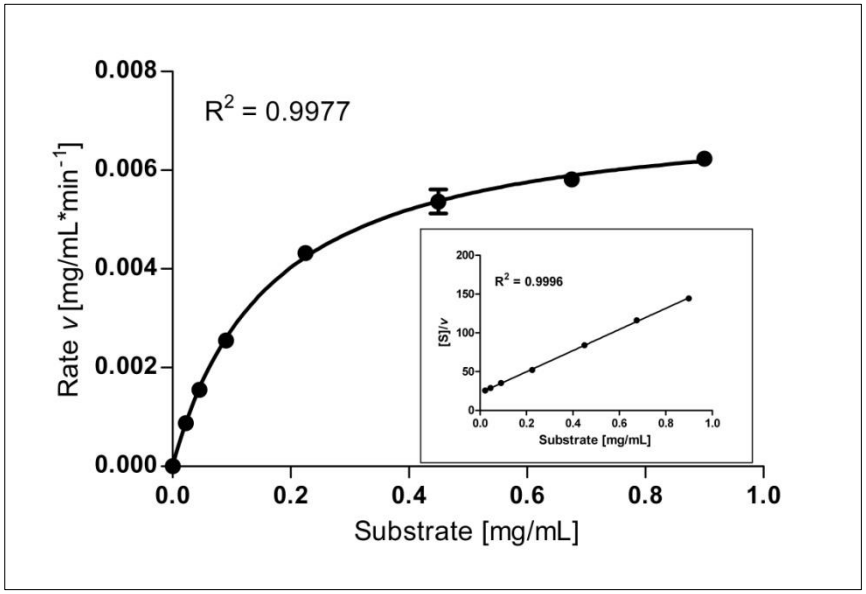


Figure 8.23: Kinetics of PNGase F W251Q.
The plot shows the rate of generation of deglycosylated Ova-FITC as a function of substrate (Ova-FITC) concentration. Each data point represents the average of three independent reactions, and error bars show the standard deviation. The inset shows the Hanes-Woolf transformation. R^2 values are given as a measure of the 'goodness of fit'. Both plots were generated using GraphPad Prism® 5.

The summary of kinetic parameters calculated for this protein is presented in **Table 8.9**. The effect of this mutation appears to be limited to processes subsequent to substrate binding. While K_m decreased slightly, k_{cat} decreased more than 100× compared to the wildtype.

Table 8.9: Kinetic parameters for PNGase F W251Q.
Values for K_m and V_{max} were determined for both the Michaelis-Menten model and the Hanes-Woolf transformation. Values in µM and µM*min^{-1} are given in parentheses. k_{cat} and k_{cat}/K_m were calculated using micromolar concentrations. Final enzyme concentration: 7.5×10^{-4} mg/mL (0.0207 µM).

	Regression model/ Transformation		**k_{cat} [sec^{-1}]**	**k_{cat}/K_m [µM^{-1} sec^{-1}]**
	Michaelis-Menten	**Hanes-Woolf**		
K_m [mg/mL]	0.2 (51.2 µM)	0.2 (51.0 µM)	1.8	0.04
V_{max} [mg/mL*min^{-1}]	7.4×10^{-3} (2.3 µM*min^{-1})	7.3×10^{-3} (2.2 µM*min^{-1})		

This mutant shows similar kinetic characteristics to the mutant W207Q (8.3.6). Interestingly, W251 is also positioned in close proximity to Glu206, almost forming a sandwich with Glu206 in the middle and Trp207 on top (**Figure 8.21**). These results reinforce the importance of tryptophan residues in the active site, not only for substrate binding but also for the generation of a hydrophobic environment. Furthermore, this result supports the hypothesis that a hydrophobic environment might be very important for the functionality of Glu206 in catalysis.

8.3.10 PNGase F V257N

Figure 8.24 shows the reaction progress curves for the mutant V251N at a final enzyme concentration of 5.0×10^{-6} mg/mL, which was identical to the concentration used for wildtype PNGase F. The product appearance at 4 minutes is linear with time, which indicates that the chosen conditions are suitable for the determination of kinetic parameters.

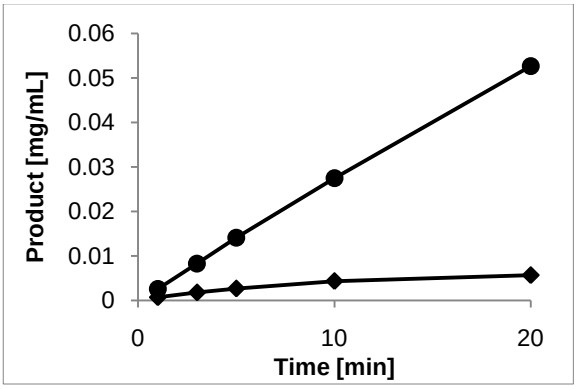


Figure 8.24: Reaction progress curve of PNGase F V257N.
The final enzyme concentration was 5.0×10^{-6} mg/mL (1.38×10^{-4} μM). •, 0.9 mg/mL; ♦ 0.0225 mg/mL Ova-FITC.

Figure 8.25 displays the graphical analysis of the velocities measured for mutant V257N.

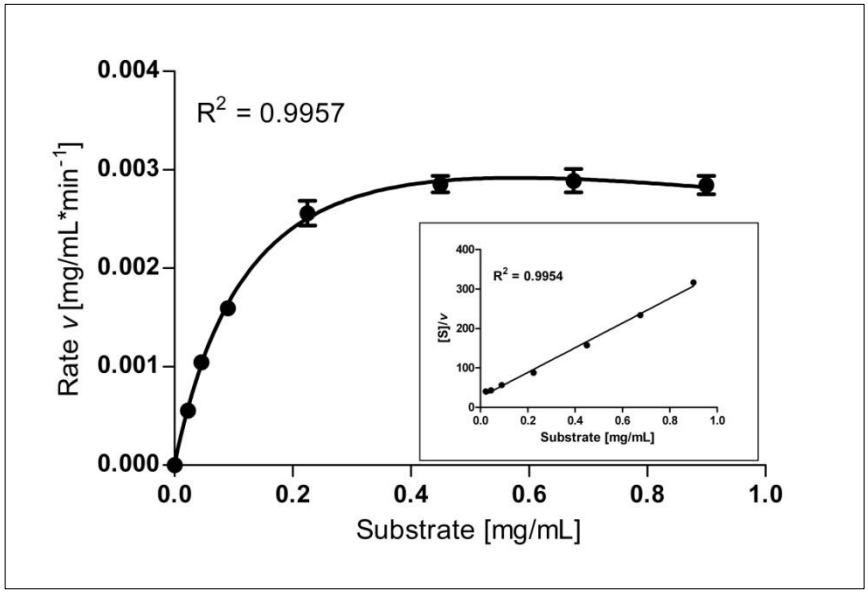


Figure 8.25: Kinetics of PNGase F V257N.
The plot shows the rate of generation of deglycosylated Ova-FITC as a function of substrate (Ova-FITC) concentration. Each data point represents the average of three independent reactions, and error bars show the standard deviation. The inset shows the Hanes-Woolf transformation. R^2 values are given as a measure of the 'goodness of fit'. Both plots were generated using GraphPad Prism® 5.

Values for the kinetic parameters are summarised in **Table 8.10**.

Table 8.10: Kinetic parameters for PNGase F V257N.
Values for K_m and V_{max} were determined for both the substrate inhibition model and the Hanes-Woolf transformation. Values in µM and µM*min^{-1} are given in parentheses. k_{cat} and k_{cat}/K_m were calculated using the micromolar concentrations. Final enzyme concentration: 5.0×10^{-6} mg/mL (1.38×10^{-4} µM).

	Regression model/ Transformation		**k_{cat} [sec^{-1}]**	**k_{cat}/K_m [µM^{-1} sec^{-1}]**
	Substrate inhibition	**Hanes-Woolf**		
K_m [mg/mL]	0.2 (46.21 µM)	0.1 (25.62 µM)	164.3	3.6
V_{max} [mg/mL*min^{-1}]	4.4×10^{-3} (1.36 µM*min^{-1})	3.2×10^{-3} (0.98 µM*min^{-1})		

This mutation has little effect on the enzymatic efficiency compared to the wildtype, probably because it is too far away from the critical active site residues and the substrate binding site to have major effects on substrate binding and catalysis. It is, however, a good positive control for the site-specific mutagenesis

program, in that there are only slight differences in the kinetic parameters between this mutant and the wildtype.

8.3.11 Summary of Kinetic Parameters

The relative values of K_m and k_{cat} are presented in **Figure 8.26** to summarise the main kinetic results obtained for PNGase F and the nine mutants. The values obtained for the wildtype protein were set to 100% and results determined for the mutants were related to the wildtype.

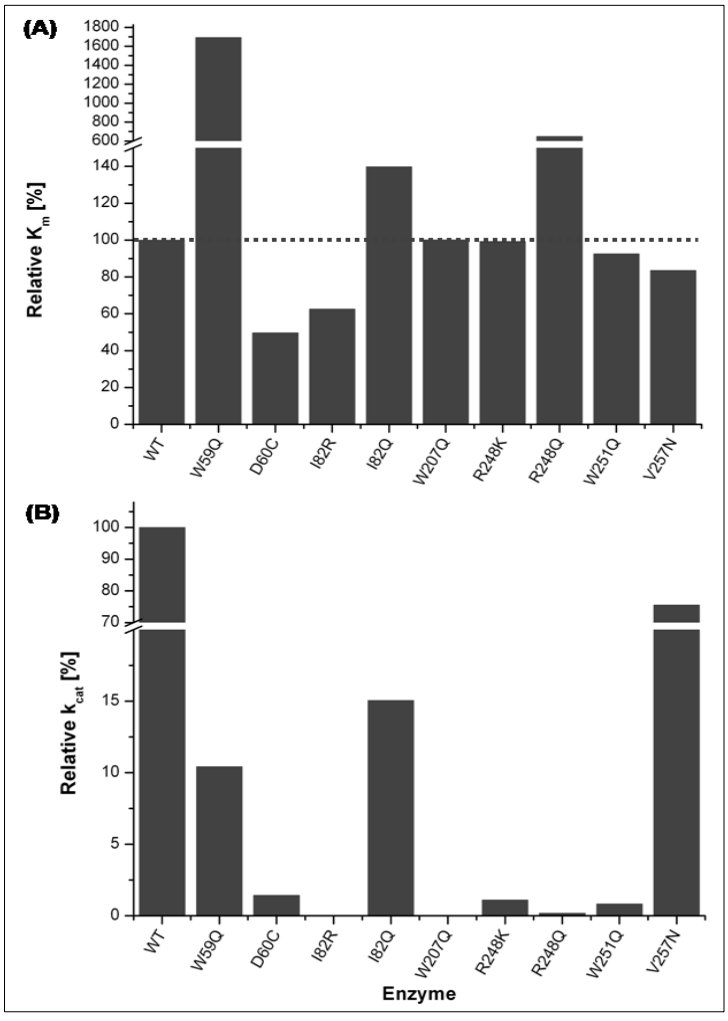


Figure 8.26: Relative kinetic parameters.
(A) Relative K_m. The dotted line indicates the 100% mark, which equals the K_m of the wildtype. **(B)** Relative k_{cat}.

The relative overall catalytic efficiency of the mutants is summarised in **Figure 8.27**.

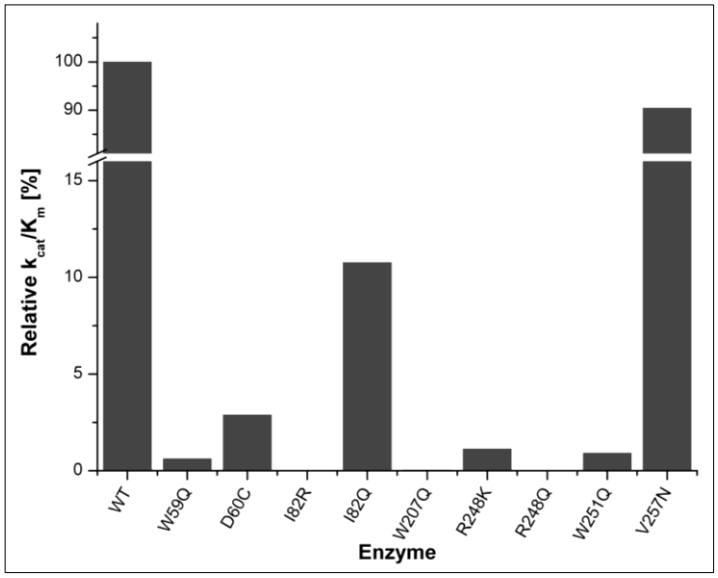


Figure 8.27: Overall catalytic efficiency k_{cat}/K_m.
The wildtype k_{cat}/K_m was set to 100%. The efficiency of the mutant enzymes is presented relative to the wildtype.

8.3.12 The Catalytic Mechanism of PNGase F

The kinetic results described in this chapter combined with the structural results presented in Chapter 7 suggest that the catalytic mechanism of PNGase F might be different to the one proposed (6.1) due to the following observations:

(i) Wat422 in structure 1PGS (Norris *et al.*, 1994b), proposed to act as the nucleophile in the cleavage of the amide bond, is replaced by a glycerol molecule in the structure presented in this work. Instead another water molecule, present in all PNGase F structures might act as the nucleophile (Wat423 in 1PGS (Norris *et al.*, 1994b); Wat67 in rPNGase F (Chapter 7), Wat338 in 1PNG (Kuhn *et al.*, 1994), Wat346 in 1PNF (numbering in publication (Kuhn *et al.*, 1995); inconsistent with PDB-file 1PNF where Wat346 is numbered as Wat652)). This bound water Wat67 is positioned between residues Asp60, Glu206 and Arg248, 2.76 Å to Asp60 OD_2, 2.71 Å to Glu206 OE_2 and 2.85 Å to Arg248 NH_1 (**Figure 7.8**). Interestingly, glycerol acts as an inhibitor and it is therefore possible that it does so by replacing Wat422. However, it appears from this work that

Asp60 is not the residue activating the water (see *(ii)*). Therefore the water proposed here, Wat67 (Wat423 in 1PGS), appears to be more reasonable as Wat422 is not close enough to Glu206, the residue most likely to abstract a proton from the water molecule (see *(iii)*).

(ii) The mutation of Asp60 to a cysteine did not inactivate the enzyme as expected. Asp60 had been proposed to play an essential role in the catalytic mechanism through activating the water to become a better nucleophile by abstraction of an H^+. This mechanism is seen in other enzymes, e.g. aspartic proteinases (Coates *et al.*, 2006). It can only do this if the pK_a of its side chain is raised to a value close to the optimal pH of the reaction, pH 8.5. This can only be achieved by the creation of a hydrophobic environment for the side chain or by the presence of residues of like charge adjacent to the side chain. If the proposed mechanism is correct, replacing Asp60 with a cysteine should inactivate the enzyme. The pK_a of the thiol side chain is normally around pH 8.3. In a hydrophobic environment such as that around Asp60 this would be increased and the thiol side chain would be protonated and therefore unable to accept a proton from the water allowing it to become nucleophilic to attack the scissile carbonyl carbon.

Although mutation of two hydrophobic residues surrounding Asp60 (W59Q, I82Q) did decrease catalytic activity, it did so to a lesser extent than expected. The W59Q and I82Q mutants showed the second and third highest catalytic efficiencies among all mutants examined here (relative k_{cat}: W59Q ~10%; I82Q ~15%; **Figure 8.26 (B)**). Decreasing the hydrophobicity should, if the proposed mechanism is correct, lead to a decrease of the side chain pK_a (< ~8.5) making it unlikely that Asp60 could abstract a proton from the active site water and result in an inactive PNGase F. In terms of adjacent residues of like charge, the only other acidic residue in close proximity to Asp60 is Tyr85. However, the pK_a of tyrosine's phenolic hydroxyl group is ~10 and it will therefore exist mainly in its protonated form at pH 8.5.

All of the results above indicate that the role of Asp60 in the catalytic mechanism of PNGase F must be different from that originally proposed. It is more likely to be involved in substrate binding or transition state stabilisation.

(iii) Kinetic results obtained for mutants W207Q and W251Q indicate that the presence of the large hydrophobic tryptophan residues in these positions is essential for catalytic activity, but not for substrate binding affinity as indicated by almost wildtype K_m values, but very low k_{cat} values. Looking at the structure, these two residues almost form a sandwich around residue Glu206, a residue that has previously been reported to be almost essential for catalytic activity (Fig 7.3 (B), **Figure 8.21**; (Kuhn *et al.*, 1995)). The importance of these hydrophobic residues, their position close to Glu206 and the importance of Glu206 itself suggest that it might be essential for Glu206 to be in a hydrophobic environment. Such an environment could possibly raise the pK_a of Glu206 to close to ~8.5, the pH optimum of the reaction. Thus, Glu206 might be the catalytic residue and accept a proton from a water molecule (most likely Wat67), activating it to become the nucleophile.

(iv) Mutation of Arg248 to an uncharged residue has a larger effect on PNGase F activity than its substitution with a charged lysine residue indicating that the proposed function for Arg248 as shown in **Figure 6.1** may be correct. The glutamine side chain in R248Q, although protonated, is unable to donate or accept protons. A lysine in this position, however, could donate a proton, but its slightly smaller side chain and possibly higher flexibility might lead to architectural alterations within the active site, disrupting the mechanism and leading to a significant decrease in catalytic rate. For an arginine to be able to fulfil the proposed function, i.e. to donate a proton, its side chain pK_a would need to be lowered. This can be achieved by two means: by positioning of other positively charged side chains in its vicinity, or by environment (Schlippe & Hedstrom, 2005). There are no positive charges in close proximity to Arg248, but it is positioned in a quite hydrophobic environment.

As shown in **Figure 8.21** Arg248 is, as Glu206, close to the tryptophan residues Trp207 and Trp251 and additionally two proline residues, Pro246 and Pro253, are in close proximity (Arg248 NH_2→Pro246 *CG*: 4.5 Å; Arg248 NH_1→Pro253 *CG*: 4.1 Å). This environment could quite possibly decrease Arg248's side chain pK_a far enough for it to be able to donate a proton to the scissile bond carbonyl oxygen, making the carbonyl carbon susceptible for the nucleophilic attack by the activated water molecule. Another feature that could have an influence on the ability of Arg248 to donate a proton is the hydrogen bond between Arg248 *NE* and Glu206 OE_2 (**Figure 8.21**). These arginine-carboxylate motifs have been observed in the active site of several enzymes where it has been suggested that they might be involved in a mechanism to activate arginine residues for acid/base chemistry (Schlippe & Hedstrom, 2005).

Another group of enzymes, the aspartic proteinases, might also give some clues as to how PNGase F functions. Both enzymes essentially cleave an amide bond, most likely using a water molecule as a nucleophile. The catalytic mechanism employed by PNGase F could be similar to the mechanism used by aspartic proteinases, which is shown in **Figure 8.28** (Coates *et al.*, 2006; Coates *et al.*, 2008; Veerapandian *et al.*, 1992). In these enzymes, a water molecule is hydrogen bonded to the carboxyl groups of two conserved aspartic acid residues. This water molecule has been implicated in catalysis by acting as the nucleophile after being polarised by one of the catalytic aspartate residues (Asp215). Following the attack on the scissile bond carbonyl group of the substrate by the activated water nucleophile, a tetrahedral transition state is observed, which is stabilised by hydrogen bonds to the other aspartate (Asp32). Fission of the scissile C-N-bond is accompanied by transfer of a proton to the leaving amino group either from Asp215 or from bulk solvent. Due to the low pH optimum of these enzymes (pH 4.5), which matches the usual pK_a of the aspartate side chain, it is likely that one aspartate is charged and the other protonated (Coates *et al.*, 2008).

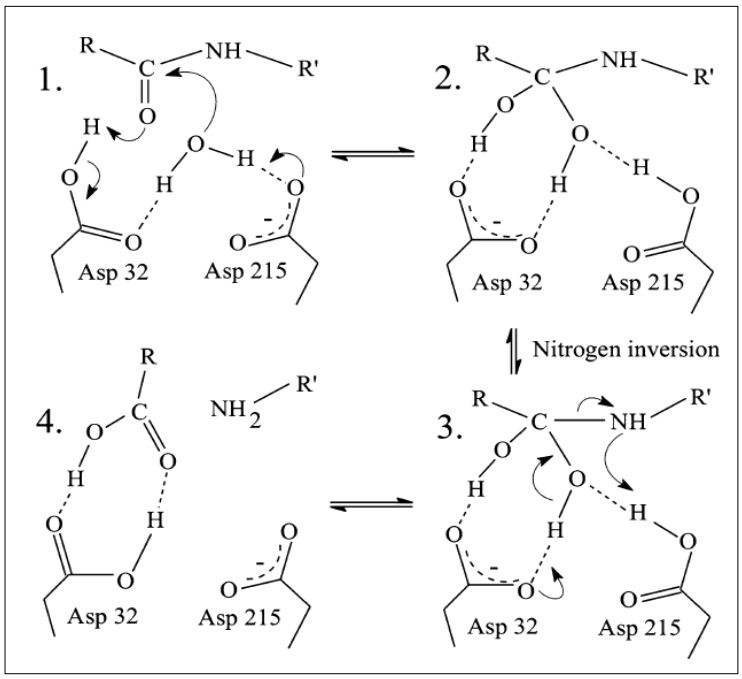


Figure 8.28: The catalytic mechanism of aspartic proteinases proposed by Veerapandian *et al.* (1992).
This mechanism is based on the X-ray structure of a difluoroketone (*gem*-diol) inhibitor bound to the aspartic proteinase endothiapepsin. This figure was taken from Coates *et al.* (2006).

According to the kinetic results described above it is unlikely that in PNGase F the two carboxylic acid side chains of Asp60 and Glu206 alone could catalyse the reaction, mainly due to the results obtained for the mutants that changed the hydrophobic environment around Asp60 (W59Q, I82Q) and D60C. If Asp60 plays either role in the mechanism shown in **Figure 8.28** it would need to be either protonated at the beginning of the reaction (=Asp32) or be able to accept a proton (=Asp215). Both cases are unlikely due to the reasons discussed above (*(ii)*).

In conclusion, while several aspects of the proposed catalytic mechanism (6.1) were verified, others were disproven. In light of the results and conclusions from the kinetic experiments (previous and present) and the possible similarity to the reaction mechanism of aspartic proteinases, a modified catalytic scheme can be proposed (**Figure 8.29**).

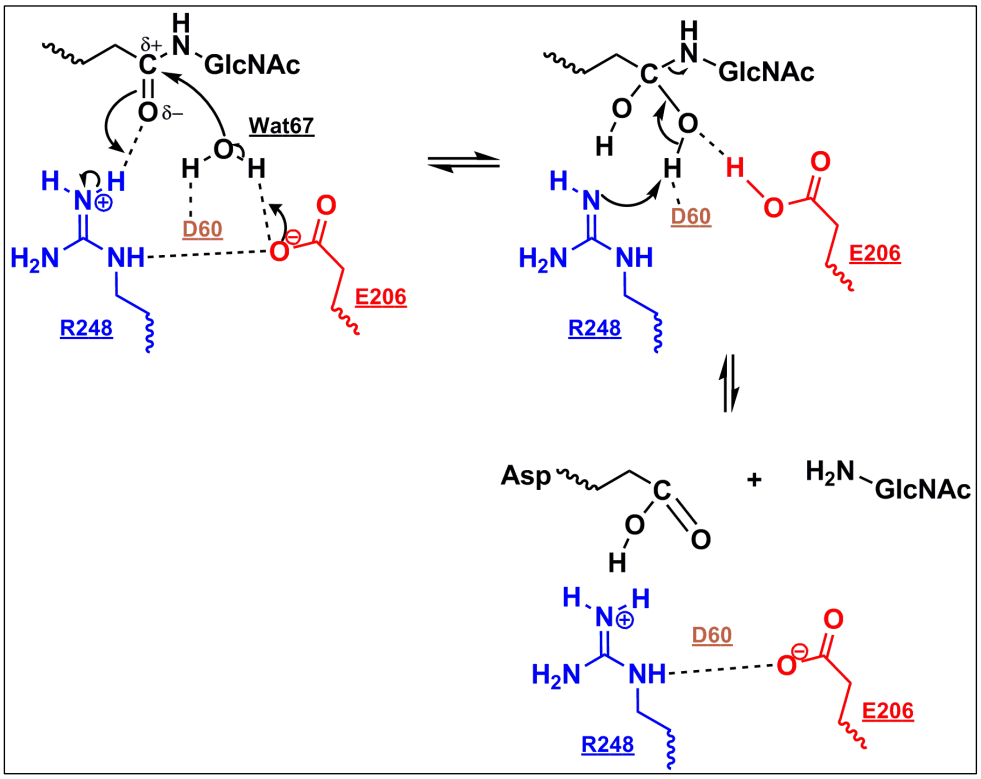


Figure 8.29: Proposed mechanism for PNGase F.

In this mechanism residue Glu206 in PNGase F could take the role of Asp215 in aspartic proteinases, i.e. accepting a proton from the active site water. The water molecule acting as the nucleophile is now proposed to be Wat67 (equivalent to Wat423 in 1PGS). This function was originally thought to be filled by Asp60, which now appears to be more important for substrate binding, coordination of Wat67 and/or possibly for stabilisation of a reaction transition state. The role proposed for residue Arg248 as shown in **Figure 6.1** could be correct and would, to some extent, correspond to the role played by Asp32 in aspartic proteinases (**Figure 8.28**). In this mechanism, Arg248 would initially donate a proton to the carbonyl oxygen of the scissile bond, priming the carbonyl carbon for the nucleophilic attack by the activated Wat67.

The next steps shown in **Figure 8.29** are hypothetical and mainly based on the suggested similarity of the reaction to the mechanism of aspartic proteinases. In this step Arg248 could be reprotonated by abstracting a proton from a tetrahedral *gem*-diol transition state, resulting in the cleavage of the scissile bond and formation of the intermediate 1-amino-*N*-acetylglucosaminyl oligosaccharide and the aspartate-containing protein/peptide. Protonation of the 1-amino group of the proximal GlcNAc could be achieved by a proton

transfer from the now protonated Glu206. The intermediate is spontaneously hydrolysed into the *N*-acetylglucosaminyl oligosaccharide and free ammonia.

However, the main problem in proposing a mechanism for PNGase F is that it is very difficult to predict how the native, very complex substrate binds to the enzyme. Will the proposed active site residues be positioned correctly for catalysis? Does PNGase F undergo some conformational changes upon substrate binding? The complex structures with *N,N'*-diacetylchitobiose (Kuhn *et al.*, 1995) and glycerols may give some indication on substrate binding, but it is hard to imagine that these rather small molecules, compared to a full glycoprotein or –peptide, will bind in the same way in the active site and have the same impact on the enzyme. Although mutagenesis has identified the main residues responsible for catalytic activity, further experimental evidence has to be obtained to verify the proposed catalytic mechanism, such as analysis of protonation states (e.g. neutron diffraction studies) and substrate binding. So far there seems to be no knowledge about the binding of the peptide part of the substrate to PNGase F, although it is known that it does affect activity (Fan & Lee, 1997). It is known that a minimum of three, but preferably more amino acids, is required for PNGase F activity. To analyse whether the general reaction mechanism is indeed similar to that of aspartic proteinases, inhibitor studies could be performed using pepstatin, a microbial peptide and typical inhibitor of aspartic proteinases. Other molecules have been used in aspartic proteinase studies to mimic the tetrahedral transition state ((Coates *et al.*, 2008) and references therein). Those could be tested on PNGase F and, if effective, be used for further crystallisation studies.

Chapter 9

Summary & Future Directions

9 Summary & Future Directions

9.1 Summary

9.1.1 Section I

Bioinformatic analyses of the amino acid sequences of PNGases (Peptide:*N*-glycanases, EC 3.5.1.52) led to the proposal of a classification scheme for these enzymes. PNGases were divided into three types based mainly on differences in amino acid sequence. Further differences between the three types include subcellular localisation, phylogenetic distribution (to date) and physiological function (if known). Also, crystal structures of members of types I and III do not show any similarity, and these types clearly employ different catalytic mechanisms. Given these findings it was concluded that the three types of PNGases are most likely the result of convergent evolution, which describes the evolution of non-homologous enzymes in different organisms to catalyse the same or very similar enzymatic reaction(s).

The putative type I PNGase from *D. radiodurans* was shown to be expressed in the native organism using RT-PCR. Subsequent recombinant expression in *E. coli* was successful in terms of obtaining soluble *Dra*PNGase, both full length and truncated. Deglycosylation activity was observed for the first preparation of full length *Dra*PNGase, but unfortunately this result could not be reproduced despite several attempts to do so. At this stage it is still not clear if the observed 'activity' was due to *Dra*PNGase activity or some other factors such as proteolytic degradation or PNGase F contamination. Misfolding of recombinant *Dra*PNGase cannot be ruled out at this stage, although misfolded proteins often are not soluble, in contrast to both full length *Dra*PNGase and truncated *Dra*PNGase. Trials to express *Dra*PNGase in insect cells using the baculovirus expression system were unsuccessful.

Gene expression analyses were successfully performed for two type II PNGases, *Sav*PNGase and *Ani*PNGase. RT-PCR products were obtained, indicating that the genes encoding these proteins are transcribed and therefore

are likely to fulfil a role in the native organism. Recombinant expression in *E. coli* was carried out for *Sav*PNGase and another putative type II PNGase from *S. solfataricus*. While this was unsuccessful for *Sso*PNGase (possibly due to the inability of *E. coli* to glycosylate *Sso*PNGase), *Sav*PNGase was expressed in a soluble form as a fusion partner of MBP. This recombinant protein, however, was very susceptible to proteolytic degradation, most likely as a result of incorrect folding. Other *E. coli* expression systems were tried, which also resulted in the production of insoluble protein, indicating that *Sav*PNGase fails to assume a native conformation when expressed in *E. coli*. To solve this problem another expression host was chosen and recombinant expression of *Sav*PNGase, *Sso*PNGase and *Ani*PNGase was attempted in insect cells using the BVES. Despite many trials this system did not lead to the production of any recombinant protein. At this stage it is still not clear why these proteins proved to be so difficult to express, especially given that PNGase *At* had been shown to be expressed in insect cells by Ftouhi-Paquin *et al.* (1998). This has to be followed up in future. Alternative expression systems using yeast may result in success, and should be trialled in the future.

9.1.2 Section II

Several mutants of PNGase F were generated, recombinantly expressed in *E. coli* and purified for subsequent structural and enzyme kinetic studies. Analysis of the results of the kinetic studies led to the proposal of a modified catalytic mechanism in which Glu206 and Arg248, and not Asp60 as previously suggested, are the main catalytic residues in this reaction. The involvement of a bound water molecule as the attacking nucleophile still persists. However, in the modified mechanism a different water molecule was identified that fulfils this role. Although at this stage no crystal structures have been obtained for any of the PNGase F mutants, circular dichroism spectroscopy showed that all mutants appeared to be folded correctly. A 1.57 Å resolution crystal structure was, however, obtained for the recombinant native PNGase F. Interestingly, in this structure three glycerol molecules were found to be bound in the active site pocket of the enzyme. While the binding of glycerol molecules may provide

some clues about the binding of the carbohydrate moiety of the natural substrate, it still is very difficult to deduce how a complex substrate such as a glycoprotein can bind to PNGase F, so that the scissile bond is in the correct orientation for catalysis to occur. One of the glycerol molecules replaced the water molecule previously assumed to be the nucleophile, leading to the proposal of the involvement of a different water molecule present in all PNGase F structures available in the protein data base. In conclusion, both the kinetic and structural studies of PNGase F and nine site-specific mutants provided valuable information towards an understanding of PNGase F's catalytic mechanism.

9.2 Future Directions

9.2.1 Section I

At this stage there are still not many PNGase F-like proteins in public databases although this is bound to change eventually due to the speed at which whole genomes can be sequenced today. However, given the relative paucity of convincing homologues it appears to be a comparatively rare enzyme, and therefore interesting for further investigations.

The putative PNGase from *D. radiodurans* could still be a good candidate for the investigation of PNGase F-like proteins due to the fact that expression of soluble recombinant protein was achieved in *E. coli* and apparent deglycosylation activity was observed in one experiment. In future, circular dichroism experiments could be carried out to determine if this soluble *Dra*PNGase (full length and truncated) is folded. A CD spectrometer was not available when the *Dra*PNGase experiments were performed during this work. Also, PNGase F can be successfully expressed and crystallised with a C-terminal His_6 tag, which does not interfere with PNGase F activity. At this stage no such construct has been generated for *Dra*PNGase. This approach could be particularly interesting for expression of only the PNGase F-like C-terminal domain of the protein. Deglycosylation activity assays should accompany

*Dra*PNGase expression experiments and both the gelshift assay and the HPLC-based assay should be used. Furthermore, it could be investigated if the presence of the histidine residue close to the active site has an effect on activity (**Figure 5.10**). A site specific mutant could be generated for *Dra*PNGase where His489 is changed to tryptophan, the residue present in the equivalent position in PNGase F (Trp191). The reverse mutation (W191H) could be introduced into PNGase F as well as the mutation E118A. Also, further attempts to crystallise *Dra*PNGase should be made. The crystallographic analysis of *Dra*PNGase would make the structural comparison with PNGase F possible and could possibly give further clues about the catalytic mechanism of these enzymes, or possibly explain the apparent inactivity of *Dra*PNGase.

For the characterisation of type II PNGases the identification of a suitable expression system has to be the main focus. It is still possible that these proteins can be successfully expressed using the BVES as this has been done before (Ftouhi Paquin *et al.*, 1998). This, unfortunately, could not be pursued further during this work due to time constraints. The use of high-yield insect cells such as High Five™ cells (Invitrogen™) could improve results. Other systems such as yeast expression systems could also be trialled. Alternatively, if these proteins continue to resist recombinant expression, purification of the target proteins from the native organisms would have to be attempted where reasonable. Should any of these attempts prove successful, crystallisation trials should be carried out as, at this stage, no crystal structure of a type II PNGase has been published. It would then be interesting to see if there are any parallels in terms of active sites between these type II PNGases and those of the other two types.

9.2.2 Section II

The characterisation of mutant enzymes presented in this work provided valuable insights into the possible catalytic mechanism of PNGase F. There are, however, many questions still to be answered. Firstly, to complement the results obtained here, crystallisation of these mutants should be pursued more intensively than was possible during this work. Although CD experiments gave some confidence that the purified mutants were folded correctly, obtaining

crystal structures of these mutants would confirm this. Furthermore, and probably even more important, crystal structures would show the structural basis for the kinetic results obtained and would provide more evidence for some aspects of the proposed modified mechanism. In addition to the mutants analysed here, more mutants need to be generated and analysed to further elucidate the catalytic mechanism. Initially it would be interesting to repeat some of the mutations performed previously by Kuhn *et al.* (1997), such as D60N and E206Q, and to analyse the mutant enzymes using the procedures established during this work. This is necessary as some data obtained here are incompatible with results obtained by this group. An interesting new mutation could be E206C, which would, at this stage, be expected to show no or very low catalytic activity. Also interesting might be the mutant E206H. The histidine could possibly take over the function of Glu206 if the proposed mechanism is correct and no major structural changes are introduced by this mutation.

Another approach to investigate the catalytic mechanism of PNGase F would be to carry out inhibitor studies. The possible parallels between the mechanisms of PNGase F and aspartic proteinases make inhibitors such as pepstatin, a characteristic inhibitor of aspartic proteinases, an attractive choice. Also, inhibitors used to imitate the *gem*-diol transition state of aspartic proteinases such as the difluoroketone-containing tripeptide CP-81,282 (Veerapandian *et al.*, 1992) may be interesting. If any of these inhibit PNGase F activity, co-crystallisation experiments could be performed. *Co*-crystallisation of PNGase F with short peptides would also be interesting as so far it is unknown where the peptide moiety of the natural PNGase F substrate(s) binds to the enzyme, and it is possible that such binding may cause some movement in the active site.

Thus, while this work has further increased our knowledge about the putative catalytic mechanism of this enzyme, it has also posed further questions. Until a suitable transition state analogue has been found that can be co-crystallised with PNGase F, the question of exactly how the enzyme effects catalysis will not be known.

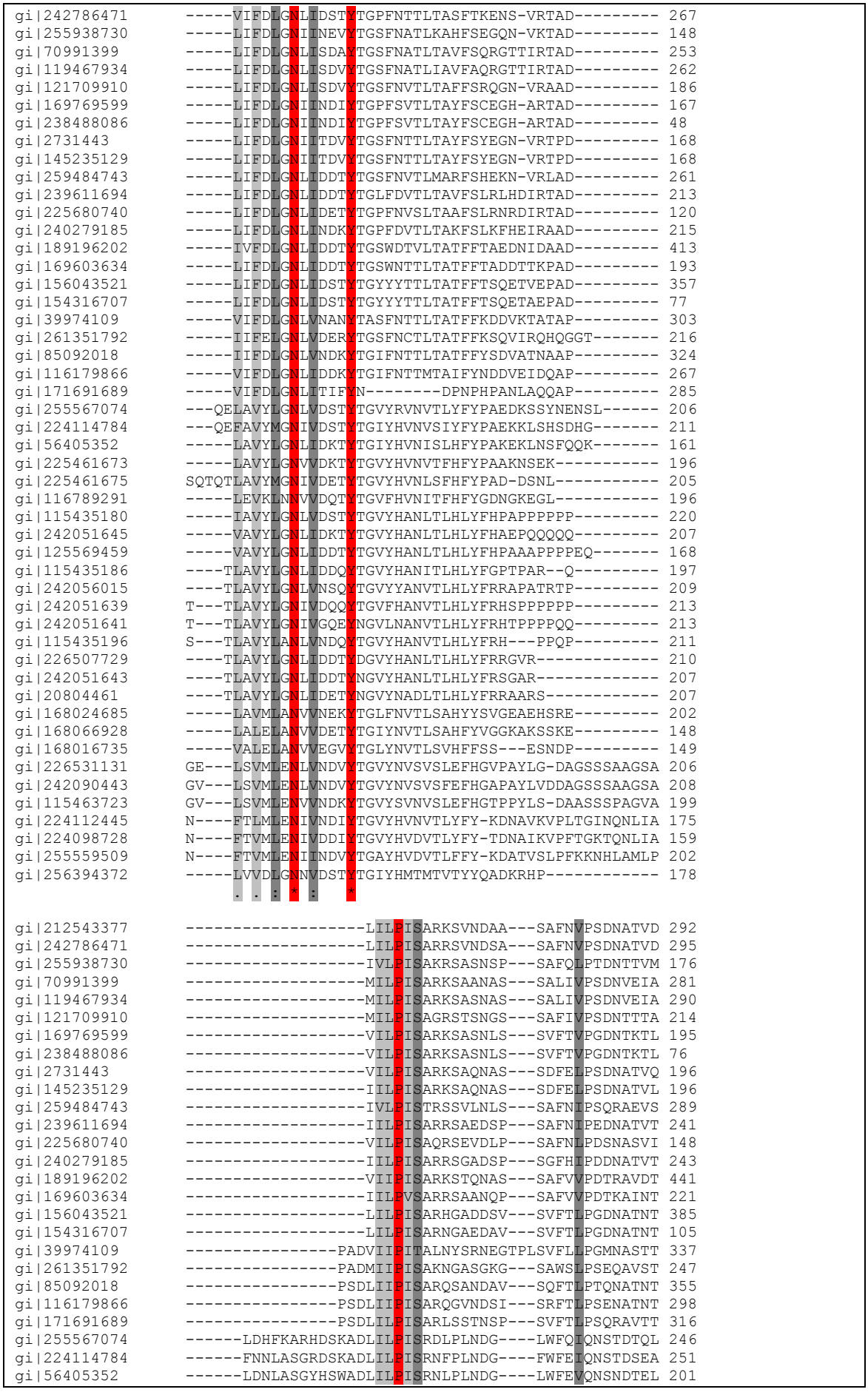
Appendices

10 Appendices

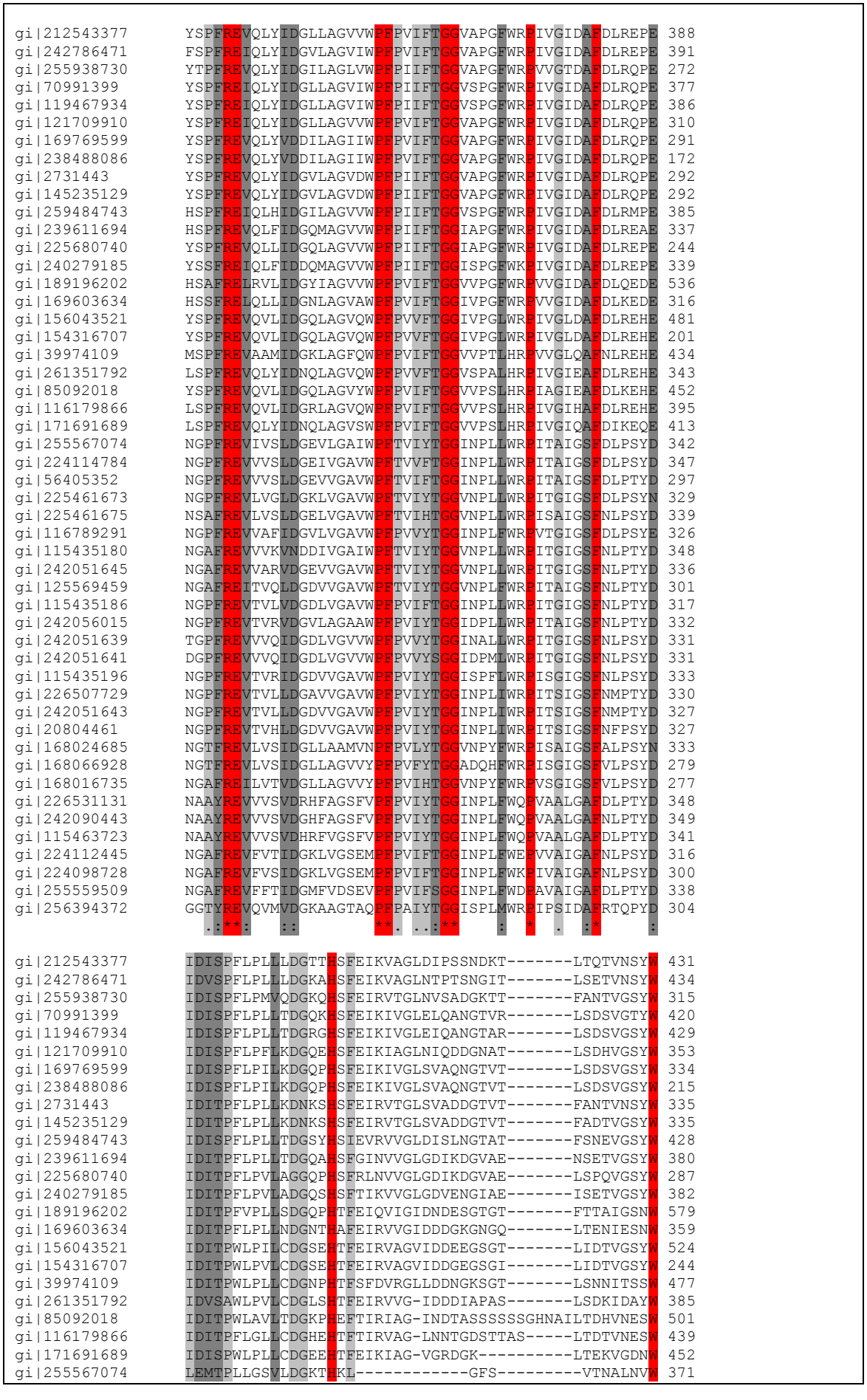
10.1 Appendix 1

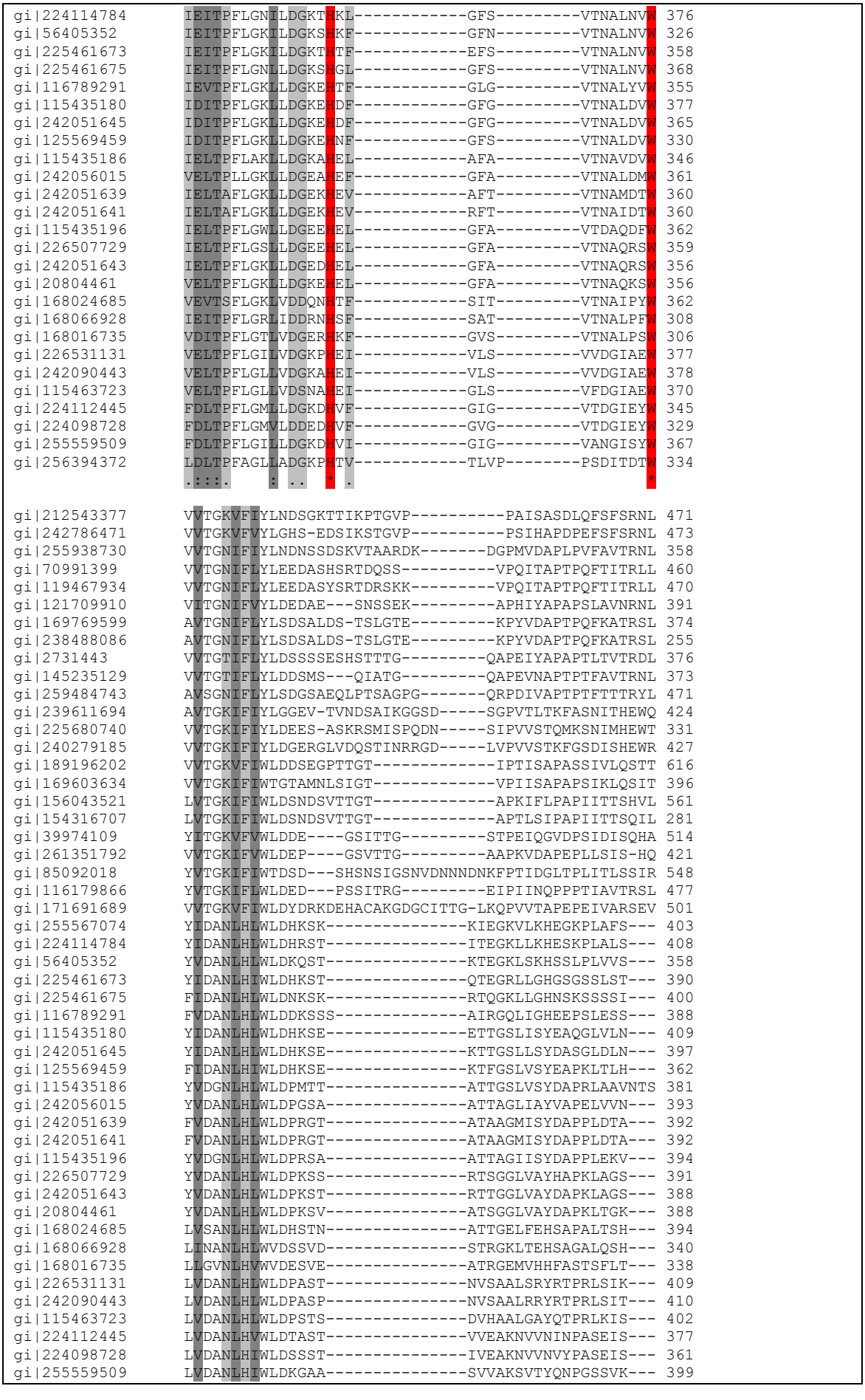
Figure 10.1 shows the multiple amino acid sequence alignment of 50 PNGase A and PNGase *At* homologues selected according to the criteria described in 3.2.2. **Table 10.1** lists details of the sequences included in the multiple sequence alignment.

```
gi|212543377      ALMYLNDTEVF----RTSTAEP--TTN-GIVWTYIKEMSQYLTLWKTPQK 229
gi|242786471      ALMYLNDTEVF----RTSTAEP--TTN-GIVWTYIKEMSQYLTLWKSPQK 232
gi|255938730      ALMYLRDNEVF----RTSTAEP--TTN-GIVWTYIKEMSQYNSLWRSPQK 113
gi|70991399       AIMYLGDAEVF----RTSTAEP--TTN-GIVWTYIKEMSHYNALWKEPQK 217
gi|119467934      AIMYLGDAEVF----RTSTAEP--TTN-GIVWTYIKEMSQYNALWKEPQK 226
gi|121709910      AIMYLGDAEVF----RTSTAEP--TSN-GIVWTYIKEMSQYNALWKEEQK 151
gi|169769599      AILFLGDTEVF----RTSTAEP--TAD-GIVWAYIKDMSQYNALWQIQQK 132
gi|238488086      ------------------------------------MSQYNALWQIQQK 13
gi|2731443        ALMYLGDTEVF----RTSTAEP--TTD-GIIWTYIKDMSQFNVLWKEKQK 133
gi|145235129      ALMYLGDTEVF----RTSTAEP--TTN-GIIWTYIKDMSQFNVLWKEKQK 133
gi|259484743      AHLWLGDIEVF----RTSTAEP--TAD-GIIWSYVKDLSQYKVLWQEPQK 226
gi|239611694      ALMFLGDIEVF----RTSTAEP--TPN-GIIWTYVKDMSLYKALWTEPQK 177
gi|225680740      ALMFLGDTEVF----RTSTAEP--TPN-GIVWTYLKDVSLYKALWMEPQK 84
gi|240279185      ALMFLGDTEVF----RTSTAEP--TAK-GIIWTYVKDMSLYKALWAKPQK 179
gi|189196202      ALMFLDDTEVF----RTSTAEP--TQD-GIIWSYVKDMSSYLALFKTTQK 377
gi|169603634      ALMYLDDIEVW----RTSTAEP--TSS-GIIWTYTKDMSAYSVLFRSPHK 157
gi|156043521      AIMYFGDSEVW----RTSTAEP--TTA-GIRWEYLKDMTEYLYFWNSPQT 321
gi|154316707      --MYFGDSEVW----RTSTAEP--TTA-GIRWEYLKDMTEYLYFWNSPQT 41
gi|39974109       ALMYLGDTEVW----RTSTAEP--TASPGIHWEYHRDMTQYLSLWKKPQT 267
gi|261351792      ALMFFNDTEVW----RTSTAEP--KPTPGIVWTYWKDMTEFLALWKSPQT 178
gi|85092018       ALMYLGDTEVW----RTSTAEP--VAPPGIRWEYLKDMTEYLSLWQQKQK 288
gi|116179866      AIMYLGDTEVW----RTSTAEP--TAPPGISWIYLKDMTHYLYFWKSPQR 231
gi|171691689      AVMYFGDTEVLNFSGRTSTAQP--TAPPGISWIYLKDMTHYMYFWKRPQK 257
gi|255567074      FGVWLGGVELL----RSCTAEP---RATGIFWSVQKDITRYYSSLLKDET 165
gi|224114784      FGVWLGGVELL----RSCTAEP---RATGIVWTVRKDITRYYSLLVKNET 171
gi|56405352       FGVWLGGVEIL----RSCTAEP---RPNGIVWTVEKDITRMYSLLKSNQT 123
gi|225461673      FGVWLGGVELL----RSCTAEP---TATGIVWTVKKDVTRYYSLLMKEET 162
gi|225461675      FGVWLDGVELL----RSCTAEP---KATGIVWTVEKDITRYSSLLLKSQT 167
gi|116789291      AGVWLSGVEIL----RTCTAEP---TAKGIEWIILKDITKYSSLLDKPQI 162
gi|115435180      FGVWLSGAELL----RSCTAEP---RATGIVWSVSRDVTRYAALLAEPGE 184
gi|242051645      FGVWLSGAELL----RSCTAEP---RPNGILWSVSRDVTRYAALLAEPGE 171
gi|125569459      FGVWLSGAELL----RSCTAEP---RATGIVWSVSRDVSRYTALLAAPGE 130
gi|115435186      FGVWLGGVELL----RSCTAEP---RPNGIVWSVSKDVTRYASLLAAGNS 162
gi|242056015      FGVWLGGAELL----RSCTAEP---RPNGIVWSVSKDITKYASLLAAGNS 172
gi|242051639      FGVWLGGAELL----RGCTAEPPIQSAGGVEWTVSKDVTKYASLLAARDS 175
gi|242051641      FGVWLGGAELL----RGCTAEPPIQSAGGVEWTVSKNVTKYASLLAARDS 175
gi|115435196      FGVWLGGVELL----RSCTAEP---RPKGVVWSVSKDVTKYASLLAARNS 176
gi|226507729      FGVWLGGAELL----RGSTAEP---RPGGVVWSVSKDVTRYAALLAAGNA 177
gi|242051643      FGVWLGGAELL----RGSTAEP---RPGGVVWSVSKDVTRYAALLAAGDA 174
gi|20804461       FGVWLGGAELM----RGSTAEP---RPGGVTWSVHKDVTKYASLLAAGNS 173
gi|168024685      SAVWLGGVEVF----RTCTAEP--TAQ-GIEWSVEKDVTPFASVFKAPQL 166
gi|168066928      SAVWLNGVEIF----RTCTAEP--TPN-GIVWTVEKDVTRFSALFKKPQM 112
gi|168016735      SAVWLGGVEIL----RTCTAEP--VQQPGIQWTVEKDVTRYSSLFSSPQP 116
gi|226531131      AAVWLDGAELL----RTTTAEP---TTDGVRWTIHKDVTRYSALLRSPRG 160
gi|242090443      AAVWLDGAELL----RTTTAEP---TPDGVRWTVRKDVTRYSALLRSPPG 161
gi|115463723      AAVWLDGAELL----RTTTAEP---TPEGVRWTVRKDVTRYSALLRSPPG 153
gi|224112445      SALWLGGSELL----RTSTAEP---EEHGIFWNVRKDITKYSSLLVQNYL 130
gi|224098728      SALWLGGSELL----RTSTAEP---GKRGIFWKVRKDITRYSSLLQQNNL 114
gi|255559509      SGLWLGGAELL----RTSTAEP---TETGIYWSIRKDITRYSSLLKQRNV 157
gi|256394372      AGVWIGGSEVF----RTSTPEP---DPAGISWHVDQDISAFIPLLRTPQP 145
                                                    :: :

gi|212543377      -----IIFDLGNLIDSKYTGPFNTTLTASFTKENN-VRMAD--------- 264
```


gi|242786471 -----VIFDLGNLIDSTYTGPFNTTLTASFTKENS-VRTAD--------- 267
gi|255938730 -----LIFDLGNIINEVYTGSFNATLKAHFSEGQN-VKTAD--------- 148
gi|70991399 -----LIFDLGNLISDAYTGSFNATLTAVFSQRGTTIRTAD--------- 253
gi|119467934 -----LIFDLGNLISDVYTGSFNATLIAVFAQRGTTIRTAD--------- 262
gi|121709910 -----LIFDLGNLISDVYTGSFNVTLTAFFSRQGN-VRAAD--------- 186
gi|169769599 -----LIFDLGNIINDIYTGPFSVTLTAYFSCEGH-ARTAD--------- 167
gi|238488086 -----LIFDLGNIINDIYTGPFSVTLTAYFSCEGH-ARTAD--------- 48
gi|2731443 -----LIFDLGNIITDVYTGSFNTTLTAYFSYEGN-VRTPD--------- 168
gi|145235129 -----LIFDLGNIITDVYTGSFNTTLTAYFSYEGN-VRTPD--------- 168
gi|259484743 -----LIFDLGNLIDDTYTGSFNVTLMARFSHEKN-VRLAD--------- 261
gi|239611694 -----LIFDLGNLIDDTYTGLFDVTLTAVFSLRLHDIRTAD--------- 213
gi|225680740 -----LIFDLGNLIDETYTGPFNVSLTAAFSLRNRDIRTAD--------- 120
gi|240279185 -----LIFDLGNLINDKYTGPFDVTLTAKFSLKFHEIRAAD--------- 215
gi|189196202 -----IVFDLGNLIDDTYTGSWDTVLTATFFTAEDNIDAAD--------- 413
gi|169603634 -----LIFDLGNLIDDTYTGSWNTTLTATFFTADDTTKPAD--------- 193
gi|156043521 -----LIFDLGNLIDSTYTGYYYTTLTATFFTSQETVEPAD--------- 357
gi|154316707 -----LIFDLGNLIDSTYTGYYYTTLTATFFTSQETAEPAD--------- 77
gi|39974109 -----VIFDLGNLVNANYTASFNTTLTATFFKDDVKTATAP--------- 303
gi|261351792 -----IIFELGNLVDERYTGSFNCTLTATFFKSQVIRQHQGGT------- 216
gi|85092018 -----IIFDLGNLVNDKYTGIFNTTLTATFFYSDVATNAAP--------- 324
gi|116179866 -----VIFDLGNLIDDKYTGIFNTTMTAIFYNDDVEIDQAP--------- 267
gi|171691689 -----VIFDLGNLITIFYN--------DPNPHPANLAQQAP--------- 285
gi|255567074 ---QELAVYLGNLVDSTYTGVYRVNVTLYFYPAEDKSSYNENSL------ 206
gi|224114784 ---QEFAVYMGNIVDSTYTGIYHVNVSIYFYPAEKKLSHSDHG------- 211
gi|56405352 -----LAVYLGNLIDKTYTGIYHVNISLHFYPAKEKLNSFQQK------- 161
gi|225461673 -----LAVYLGNVVDKTYTGVYHVNVTFHFYPAAKNSEK----------- 196
gi|225461675 SQTQTLAVYMGNIVDETYTGVYHVNLSFHFYPAD-DSNL----------- 205
gi|116789291 -----LEVKLNNVVDQTYTGVFHVNITFHFYGDNGKEGL----------- 196
gi|115435180 -----IAVYLGNLVDSTYTGVYHANLTLHLYFHPAPPPPPP--------- 220
gi|242051645 -----VAVYLGNLIDKTYTGVYHANLTLHLYFHAEPQQQQQ--------- 207
gi|125569459 -----VAVYLGNLIDDTYTGVYHANLTLHLYFHPAAAPPPPEQ------- 168
gi|115435186 ----TLAVYLGNLIDDQYTGVYHANITLHLYFGPTPAR--Q--------- 197
gi|242056015 ----TLAVYLGNLVNSQYTGVYYANVTLHLYFRRAPATRTP--------- 209
gi|242051639 T---TLAVYLGNIVDQQYTGVFHANVTLHLYFRHSPPPPPP--------- 213
gi|242051641 T---TLAVYLGNIVGQEYNGVLNANVTLHLYFRHTPPPPQQ--------- 213
gi|115435196 S---TLAVYLANLVNDQYTGVYHANVTLHLYFRH---PPQP--------- 211
gi|226507729 ----TLAVYLGNLIDDTYDGVYHANLTLHLYFRRGVR------------- 210
gi|242051643 ----TLAVYLGNLIDDTYNGVYHANLTLHLYFRSGAR------------- 207
gi|20804461 ----TLAVYLGNLIDETYNGVYNADLTLHLYFRRAARS------------ 207
gi|168024685 -----LAVMLANVVNEKYTGLFNVTLSAHYYSVGEAEHSRE--------- 202
gi|168066928 -----LALELANVVDETYTGIYNVTLSAHFYVGGKAKSSKE--------- 148
gi|168016735 -----VALELANVVEGVYTGLYNVTLSVHFFSS---ESNDP--------- 149
gi|226531131 GE---LSVMLENLVNDVYTGVYNVSVSLEFHGVPAYLG-DAGSSSAAGSA 206
gi|242090443 GV---LSVMLENLVNDVYTGVYNVSVSFEFHGAPAYLVDDAGSSSAAGSA 208
gi|115463723 GV---LSVMLENVVNDKYTGVYSVNVSLEFHGTPPYLS-DAASSSPAGVA 199
gi|224112445 N----FTLMLENIVNDIYTGVYHVNVTLYFY-KDNAVKVPLTGINQNLIA 175
gi|224098728 N----FTVMLENIVDDIYTGVYHVDVTLYFY-TDNAIKVPFTGKTQNLIA 159
gi|255559509 N----FTVMLENIINDVYTGAYHVDVTLFFY-KDATVSLPFKKNHLAMLP 202
gi|256394372 -----LVVDLGNNVDSTYTGIYHMTMTVTYYQADKRHP------------ 178
gi|212543377 -------------------LILPISARKSVNDAA---SAFNVPSDNATVD 292
gi|242786471 -------------------LILPISARRSVNDSA---SAFNVPSDNATVD 295
gi|255938730 -------------------IVLPISAKRSASNSP---SAFQLPTDNTTVM 176
gi|70991399 -------------------MILPISARKSAANAS---SALIVPSDNVEIA 281
gi|119467934 -------------------MILPISARKSASNAS---SALIVPSDNVEIA 290
gi|121709910 -------------------MILPISAGRSTSNGS---SAFIVPSDNTTTA 214
gi|169769599 -------------------VILPISARKSASNLS---SVFTVPGDNTKTL 195
gi|238488086 -------------------VILPISARKSASNLS---SVFTVPGDNTKTL 76
gi|2731443 -------------------VILPISARKSAQNAS---SDFELPSDNATVQ 196
gi|145235129 -------------------IILPISARKSAQNAS---SDFELPSDNATVL 196
gi|259484743 -------------------IVLPISTRSSVLNLS---SAFNIPSQRAEVS 289
gi|239611694 -------------------IILPISARRSAEDSP---SAFNIPEDNATVT 241
gi|225680740 -------------------VILPISAQRSEVDLP---SAFNLPDSNASVI 148
gi|240279185 -------------------IILPISARRSGADSP---SGFHIPDDNATVT 243
gi|189196202 -------------------VIIPISARKSTQNAS---SAFVVPDTRAVDT 441
gi|169603634 -------------------IILPVSARRSAANQP---SAFVVPDTKAINT 221
gi|156043521 -------------------LILPISARHGADDSV---SVFTLPGDNATNT 385
gi|154316707 -------------------LILPISARNGAEDAV---SVFTLPGDNATNT 105
gi|39974109 ----------------PADVIIPITALNYSRNEGTPLSVFLLPGMNASTT 337
gi|261351792 ----------------PADMIIPISAKNGASGKG---SAWSLPSEQAVST 247
gi|85092018 ----------------PSDLIIPISARQSANDAV---SQFTLPTQNATNT 355
gi|116179866 ----------------PSDLIIPISARQGVNDSI---SRFTLPSENATNT 298
gi|171691689 ----------------PSDLILPISARLSSTNSP---SVFTLPSQRAVTT 316
gi|255567074 ------LDHFKARHDSKADLILPISRDLPLNDG----LWFQIQNSTDTQL 246
gi|224114784 ------FNNLASGRDSKADLILPISRNFPLNDG----FWFEIQNSTDSEA 251
gi|56405352 ------LDNLASGYHSWADLILPISRNLPLNDG----LWFEVQNSNDTEL 201

```
gi|225461673      ---------LASGYGSWADLILPISRNLPLNDG----LWFEIENSTDLEV 233
gi|225461675      ---------LKSGYGSRPDLILPISRNLPLNDG----LWFEIQNQTHVEG 242
gi|116789291      ------------DNPAHAHLILPISLPSSEIGG----SWFQIENSSDVQS 230
gi|115435180      --------------PQQ-ADLIVPISRSLPLNDG----QWFAIQNSTDVQG 252
gi|242051645      -------------QQQQADLIVPISRSLPLNDG----QWFAIQNATDVQS 240
gi|125569459      ---------QQQQQQQHADLIFPISRSLPLNDG----QWFAIQNSTNVQS 205
gi|115435186      ----------PAPATAPADIIVPVSRSLPLNDG----LWFQIQNATDVES 233
gi|242056015      ----------PPPATAPADLIVPMSRGLPLNDG----LWYQIQNATDVQS 245
gi|242051639      ----------Q-PGLGPADAVVPISRSLPLNDG----LWFEIENDLDVAT 248
gi|242051641      ----------QQPGLGPADAVVPISRSLPLNDG----LWFEILNDFYDAT 249
gi|115435196      ----------PQPGLGPADVIVPISQSLPLNGG----QWFQINNNEDVES 247
gi|226507729      ----------PSAAAAAADAVVPVSRSLPLNDG----LWFVVQNDTDVQS 246
gi|242051643      ----------SSPPSAAADAVVPVSRSLPLNDG----LWFVVQNATDVQS 243
gi|208044611      ----------PTAASAPADVVVPVSRSLPLNDG----LWFVVDNTTDVES 243
gi|168024685      ------------SYGGVADLILPFAESSPLNGG----HWFQLQNESDLRT 236
gi|168066928      ------------SYGGVADVILPFAEVSPLKGG----HWFQLQNESDVQS 182
gi|168016735      ------------SFNGGADVFLPFANFS--AQE----YWFRIHNEGEAHI 181
gi|226531131      EPGGQATLKLPASYFQPADLILPISEGMGSSNG----FWFRIQNSSDPRS 252
gi|242090443      DPG-QATPKLPASYFQPADLILPISEGTGNSSG----FWFRIQNSSDSRS 253
gi|115463723      SND-PKEPMLPESYFQPADLIVPISDVAGNGKGG---FWFRIQNASDSHS 245
gi|224112445      PVLQSPLFGDKSMYDPPADLIIPISASDS-TKG----YWFIIESELDVKF 220
gi|224098728      PALELPFFGDKSMYDPPADLIIPISASDS-TKG----YWFIVEGDLDVKF 204
gi|255559509      HQIQA-----KVVYETPSDLIIPISSFHD-NRG----YWFRIEDESDVQY 242
gi|256394372      -------------QAAHSDVVVPISQSTS-APG-----WWGLTKG-QTAS 208
                                    ..*.:                :

gi|212543377      LT---FPSNAQRAVVSISACGQSEEE-FWWSGVLNQDTDDFDSTIGVLYG 338
gi|242786471      LT---FPNNVQRAVVSISACGQSEEE-FWWSSVLNQDIDDFDSTVGVLYG 341
gi|255938730      YE---IPAAASRAIVSISACGQSEEE-FWWSNVFSQDTRDFESTVGGLYG 222
gi|70991399       YR---LPSNTSRAIVSISACGQSTEE-FWWSNVFSPDTESFVNTVGELYG 327
gi|119467934      YR---LPSNTARAIVSISACGQSTEE-FWWSNVFSPDTESFVSTVGELYG 336
gi|121709910      YQ---FPDSAARAVVSISACGQSTEE-FWWSNVFSGDTESFESTVGELYG 260
gi|169769599      YQ---IPPNTSRAVVSISACGQSTEE-FWWSNVFSYDTEAFNTTMGELYG 241
gi|238488086      YQ---IPPNTSRAVVSISACGQSTEE-FWWSNVFSYDTEAFNTTMGELYG 122
gi|2731443        YQ---IPQTASRAVVSISACGQSEEE-FWWSNVLSADEYTFDNTIGELYG 242
gi|145235129      YQ---IPPTASRAVVSISACGQSEEE-FWWSNVLSADEWTFDNTIGELYG 242
gi|2594847431     YR---FDSRVSRALVSISACGQSTEE-FWWSNVFSSDTRTFDSTVGELYG 335
gi|239611694      HV---VPNDASRAVVSISACGQSTEE-FWYSNVLSSDIYTFNETTGPLYG 287
gi|225680740      HS---IPDDAYRAVVSISACGQSTEE-FWYTNVLSSDTYTFNQTTGPLYG 194
gi|240279185      HI---IPPDVSRATVSISACGQSTEE-FWYSNVLSSDVLTFNKTAGPLYG 289
gi|189196202      VT---LPQNAKKAVFSIAACGQAAEE-FWWSNVLQSDVNTFGGET-TLYG 486
gi|169603634      LE---LPKNTEKAVFTISACGQAAEE-FWWSNVFNSDTKAFGNDT-TLYG 266
gi|156043521      IA---FPRNANRAVFSISACGQSTEE-FWWGNVLQSDIEAFEDYDGTLYG 431
gi|154316707      IS---FPQNANRAVFSISACGQSAEE-FWWGNVLQSDVETFEEYDGTLYG 151
gi|39974109       VKS--FPRNANRAVLAIQANGQAAEE-FWWSNLLQSDVATFNATNGMAPG 384
gi|261351792      IY---FPQHANRAVFSLSANGQMAEE-FWWSNVLQSDVDTFNHTASSMPG 293
gi|85092018       ISN--FPLNARRAVFSVSANGQGNEE-FWWSNVLQSDTHAFSDTVGELPG 402
gi|116179866      IS---LPRNIRRAVFSVSANGQTSEEEFWWSNVLQSDVYTFNATAGKLPG 345
gi|171691689      FSAGSLPRDIRRAVVSLSTTGQASEE-FFWSNVLESDTATFEGDP--LPG 363
gi|255567074      KE-FEIPPNVYRAVLEVYVSFHENDE-FWYSNYYNEYISANNLTG--SPG 292
gi|224114784      KE-FKIPQNVYRAVLEVYVSFHENDE-FWYGNYPNEYIIANNLTG--FPG 297
gi|56405352       KE-FKIPQNAYRAVLEVYVSFHENDE-FWYSNLPNEYIAANNLSG--TPG 247
gi|225461673      KK-FKIPRNAYRAVLEVYLSFHENDE-FWYSNPPNDYISANNLTG--TPG 279
gi|225461675      KE-FIIPKNAYRAVLEVYVSFHENDE-FWYLNPPNDYIDVNNLTGS-IPG 289
gi|116789291      KS-LQLPPNAYRAVLEIFVSFHSDDE-FWYSNPPNVYIEENNLTG--TAG 276
gi|115435180      KR-LAIPSNTYRAILEVFVSFHSNDE-FWYTNPPNEYIEANNLSN--VPG 298
gi|242051645      KK-LAIPSNTYRAVLEVFVSFHSNDE-FWYTNPPDDYIQANNLSS--VPG 286
gi|125569459      KK-LVIPSNTYRAVLEVFVSFHSNDE-DWYMHPPNEYIEANNISI--LPG 251
gi|115435186      AS-IVLPSNTYRAVLEVYVSFHGDDE-FWY--THT----------P---DG 267
gi|242056015      TS-VTLPPNTYRAVLEVYASSHGDDE-FWYWYTNT----------P--GAA 282
gi|242051639      AS-VTVPANTYRAVLEVYLSYHSDDE-FWYGNT---------------AE 281
gi|242051641      AS-VTVPTNTYRAVLEVYLSYQSGDE-YWYGN----------------AD 281
gi|115435196      AS-LAVPANAYRAVLEVYLSYHGSDE-FWYTYGNP----------F---NG 283
gi|226507729      AR-VSVPRNAYRAVLEVYVSSHDADE-FWYMNTP--------------EQ 280
gi|242051643      AR-VTVPRNAYRAVLEVYVSSHDADE-FWYMNTP--------------EQ 277
gi|20804461       AR-LTVPPNAYRAVLEVYVSSHNFDE-FWYMNTP--------------DQ 277
gi|168024685      QE-IKIPRNAYKAVLEVCVSPHGSDE-FWYTNPPDDYLNANNLTEE-IPG 283
gi|168066928      RE-IQIPRNAYKAVLEICISFHGDDE-FWYANPPNDFLLSNNISDQ-VAG 229
gi|168016735      RE-IHIPRNAYKAVMEICVSFHEHDE-FWYINPPNEYLKASNVTDE-A-G 227
gi|226531131      KL-VRIPSNTYRAVLEVFVSPHSNDE-FWYSNPPDLYIRENNLTT--GRG 298
gi|242090443      KL-VSIPSNTYRAVLEVFVSPHSNDE-FWYSNPPDLYIRENNLAT--GRG 299
gi|115463723      RL-VTIPSSTYRAVLEVFVSPHSNDE-YWYSNPPDIYIRENNLTT--RRG 291
gi|224112445      EK-VRFPLNTRKVVLELYVSFHGNDE-FWYSNPSNSYIRMNNLST--PRG 266
gi|224098728      EK-VRFPLNTRKVVLELYVSFHGTDE-FWYSNPSSSYIRMNNMSN--PRG 250
gi|255559509      KK-LRFPRNTRKAVLELYVSFHGNDE-FWYSNPSNTYIRMNNLTS--LRG 288
gi|256394372      TT-VTLPRNTESADLQLYARGGGCEE-FWYSNVPDSYAASHASDG--LCG 254
                        .      . . :       :*  ::
```


gi|212543377 YSPFREVQLYIDGLLAGVVWPFPVIFTGGVAPGFWRPIVGIDAFDLREPE 388
gi|242786471 FSPFREIQLYIDGVLAGVIWPFPVIFTGGVAPGFWRPIVGIDAFDLREPE 391
gi|255938730 YTPFREVQLYIDGILAGLVWPFPIIFTGGVAPGFWRPVVGTDAFDLRQPE 272
gi|70991399 YSPFREIQLYIDGLLAGVIWPFPIIFTGGVSPGFWRPIVGIDAFDLRQPE 377
gi|119467934 YSPFREIQLYIDGLLAGVIWPFPIIFTGGVSPGFWRPIVGIDAFDLRQPE 386
gi|121709910 YSPFREIQLYIDGLLAGVVWPFPVIFTGGVAPGFWRPIVGIDAFDLRQPE 310
gi|169769599 YSPFREVQLYVDDILAGIIWPFPVIFTGGVAPGFWRPIVGIDAFDLRQPE 291
gi|238488086 YSPFREVQLYVDDILAGIIWPFPVIFTGGVAPGFWRPIVGIDAFDLRQPE 172
gi|2731443 YSPFREVQLYIDGVLAGVDWPFPIIFTGGVAPGFWRPIVGIDAFDLRQPE 292
gi|145235129 YSPFREVQLYIDGVLAGVDWPFPIIFTGGVAPGFWRPIVGIDAFDLRQPE 292
gi|259484743 HSPFREIQLHIDGILAGVVWPFPIIFTGGVSPGFWRPIVGIDAFDLRMPE 385
gi|239611694 HSPFREVQLFIDGQMAGVVWPFPIIFTGGIAPGFWRPIVGIDAFDLREAE 337
gi|225680740 YSPFREVQLLIDGQLAGVVWPFPIIFTGGIAPGFWRPIVGIDAFDLREPE 244
gi|240279185 YSSFREIQLFIDDQMAGVVWPFPIIFTGGISPGFWKPIVGIDAFDLREPE 339
gi|189196202 HSAFRELRVLIDGYIAGVVWPFPVIFTGGVVPGFWRPVVGIDAFDLQEDE 536
gi|169603634 HSSFRELQLLIDGNLAGVAWPFPVIFTGGIVPGFWRPVVGIDAFDLKEDE 316
gi|156043521 YSPFREVQVLIDGQLAGVQWPFPVVFTGGIVPGLWRPIVGLDAFDLREHE 481
gi|154316707 YSPFREVQVLIDGQLAGVQWPFPVVFTGGIVPGLWRPIVGLDAFDLREHE 201
gi|39974109 MSPFREVAAMIDGKLAGFQWPFPVIFTGGVVPTLHRPVVGLQAFNLREHE 434
gi|261351792 LSPFREVQLYIDNQLAGVQWPFPVVFTGGVSPALHRPIVGIEAFDLREHE 343
gi|85092018 YSPFREVQVLIDGQLAGVYWPFPVIFTGGVVPSLHRPIAGIEAFDLKEHE 452
gi|116179866 LSPFREVQVLIDGRLAGVQWPFPVIFTGGVVPSLHRPIVGIHAFDLREHE 395
gi|171691689 LSPFREVQLYIDNQLAGVSWPFPVIFTGGVVPSLHRPIVGIQAFDIKEQE 413
gi|255567074 NGPFREVIVSLDGEVLGAIWPFTVIYTGGINPLLWRPITAIGSFDLPSYD 342
gi|224114784 NGPFREVVVSLDGEIVGAVWPFTVVFTGGINPLLWRPITAIGSFDLPSYD 347
gi|56405352 NGPFREVVVSLDGEVVGAVWPFTVIFTGGINPLLWRPITAIGSFDLPTYD 297
gi|225461673 NGPFREVLVGLDGKLVGAVWPFTVIYTGGVNPLLWRPITGIGSFDLPSYN 329
gi|225461675 NSAFREVLVSLDGELVGAVWPFTVIHTGGVNPLLWRPISAIGSFNLPSYD 339
gi|116789291 NGPFREVVAFIDGVLVGAVWPFPVVYTGGINPLFWRPVTGIGSFDLPSYE 326
gi|115435180 NGAFREVVVKVNDDIVGAIWPFTVIYTGGVNPLLWRPITGIGSFNLPTYD 348
gi|242051645 NGAFREVVARVDGEVVGAVWPFTVIYTGGVNPLLWRPITGIGSFNLPTYD 336
gi|125569459 NGAFREITVQLDGDVVGAVWPFTVIYTGGVNPLFWRPITAIGSFNLPTYD 301
gi|115435186 NGPFREVTVLVDGDLVGAVWPFPVIFTGGINPLLWRPITGIGSFNLPTYD 317
gi|242056015 NGPFREVTVRVDGVLAGAAWPFPVIYTGGIDPLLWRPITAIGSFNLPTYD 332
gi|242051639 TGPFREVVVQIDGDLVGVVWPFPVVYTGGINALLWRPITGIGSFNLPSYD 331
gi|242051641 DGPFREVVVQIDGDLVGVVWPFPVVYSGGIDPMLWRPITGIGSFNLPSYD 331
gi|115435196 NGPFREVTVRIDGDVVGAVWPFPVIYTGGISPFLWRPISGIGSFNLPSYD 333
gi|226507729 NGPFREVTVLLDGAVVGAVWPFPVIYTGGINPLIWRPITSIGSFNMPTYD 330
gi|242051643 NGPFREVTVLLDGDVVGAVWPFPVIYTGGINPLIWRPITSIGSFNMPTYD 327
gi|20804461 NGPFREVTVHLDGDVVGAVWPFPVIYTGGINPLIWRPITSIGSFNFPSYD 327
gi|168024685 NGTFREVLVSIDGLLAAMVNPFPVLYTGGVNPYFWRPISAIGSFALPSYN 333
gi|168066928 NGTFREVLVSIDGLLAGVVYPFPVFYTGGADQHFWRPISGIGSFVLPSYD 279
gi|168016735 NGAFREILVTVDGLLAGVVYPFPVIHTGGVNPYFWRPVSGIGSFVLPSYD 277
gi|226531131 NAAYREVVVSVDRHFAGSFVPFPVIYTGGINPLFWQPVAALGAFDLPTYD 348
gi|242090443 NAAYREVVVSVDGHFAGSFVPFPVIYTGGINPLFWQPVAALGAFNLPTYD 349
gi|115463723 NAAYREVVVSVDHRFVGSFVPFPVIYTGGINPLFWQPVAALGAFDLPTYD 341
gi|224112445 NGAFREVFVTIDGKLVGSEMPFPVIFTGGINPLFWEPVVAIGAFNLPSYD 316
gi|224098728 NGAFREVFVSIDGKLVGSEMPFPVIFTGGINPLFWKPIVAIGAFNLPSYD 300
gi|255559509 NGAFREVFFTIDGMFVDSEVPFPVIFSGGINPLFWDPAVAIGAFDLPTYD 338
gi|256394372 GGTYREVQVMVDGKAAGTAQPFPAIYTGGISPLMWRPIPSIDAFRTQPYD 304
gi|212543377 IDISPFLPLLLDGTTHSFEIKVAGLDIPSSNDKT-------LTQTVNSYW 431
gi|242786471 IDVSPFLPLLLDGKAHSFEIKVAGLNTPTSNGIT-------LSETVNSYW 434
gi|255938730 IDISPFLPMVQDGKQHSFEIRVTGLNVSADGKTT-------FANTVGSYW 315
gi|70991399 IDISPFLPLLTDGQKHSFEIKIVGLELQANGTVR-------LSDSVGTYW 420
gi|119467934 IDISPFLPLLTDGRGHSFEIKIVGLEIQANGTAR-------LSDSVGSYW 429
gi|121709910 IDISPFLPFLKDGQEHSFEIKIAGLNIQDDGNAT-------LSDHVGSYW 353
gi|169769599 IDISPFLPILKDGQPHSFEIKIVGLSVAQNGTVT-------LSDSVGSYW 334
gi|238488086 IDISPFLPILKDGQPHSFEIKIVGLSVAQNGTVT-------LSDSVGSYW 215
gi|2731443 IDITPFLPLLKDNKSHSFEIRVTGLSVADDGTVT-------FANTVNSYW 335
gi|145235129 IDITPFLPLLKDNKSHSFEIRVTGLSVADDGTVT-------FADTVGSYW 335
gi|259484743 IDISPFLPLLTDGSYHSIEVRVVGLDISLNGTAT-------FSNEVGSYW 428
gi|239611694 IDITPFLPLLTDGQAHSFGINVVGLGDIKDGVAE-------NSETVGSYW 380
gi|225680740 IDITPFLPVLAGGQPHSFRLNVVGLGDIKDGVAE-------LSPQVGSYW 287
gi|240279185 IDITPFLPVLADGQSHSFTIKVVGLGDVENGIAE-------ISETVGSYW 382
gi|189196202 IDITPFVPLLSDGQPHTFEIQVIGIDNDESGTGT-------FTTAIGSNW 579
gi|169603634 IDITPFLPLLNDGNTHAFEIRVVGIDDDGKGNGQ-------LTENIESNW 359
gi|156043521 IDITPWLPILCDGSEHTFEIRVAGVIDDEEGSGT-------LIDTVGSYW 524
gi|154316707 IDITPWLPVLCDGSEHTFEIRVAGVIDDGEGSGI-------LIDTVGSYW 244
gi|39974109 IDITPWLPLLCDGNPHTFSFDVRGLLDDNGKSGT-------LSNNITSSW 477
gi|261351792 IDVSAWLPVLCDGLSHTFEIRVVG-IDDDIAPAS-------LSDKIDAYW 385
gi|85092018 IDITPWLAVLTDGKPHEFTIRIAG-INDTASSSSSSGHNAILTDHVNESW 501
gi|116179866 IDITPFLGLLCDGHEHTFTIRVAG-LNNTGDSTTAS-----LTDTVNESW 439
gi|171691689 IDISPWLPLLCDGEEHTFEIKIAG-VGRDGK----------LTEKVGDNW 452
gi|255567074 LEMTPLLGSVLDGKTHKL------------GFS---------VTNALNVW 371


gi|224114784 IEITPFLGNILDGKTHKL----------GFS---------VTNALNVW 376
gi|56405352 IEITPFLGKILDGKSHKF----------GFN---------VTNALNVW 326
gi|225461673 IEITPFLGKILDGKTHTF----------EFS---------VTNALNVW 358
gi|225461675 IEITPFLGNLLDGKSHGL----------GFS---------VTNALNVW 368
gi|116789291 IEVTPFLGKLLDGKEHTF----------GLG---------VTNALYVW 355
gi|115435180 IDITPFLGKLLDGKEHDF----------GFG---------VTNALDVW 377
gi|242051645 IDITPFLGKLLDGKEHDF----------GFG---------VTNALDVW 365
gi|125569459 IDITPFLGKLLDGKEHNF----------GFS---------VTNALDVW 330
gi|115435186 IELTPFLAKLLDGKAHEL----------AFA---------VTNAVDVW 346
gi|242056015 VELTPLLGKLLDGEAHEF----------GFA---------VTNALDMW 361
gi|242051639 IELTAFLGKLLDGEKHEV----------AFT---------VTNAMDTW 360
gi|242051641 IELTAFLGKLLDGEKHEV----------RFT---------VTNAIDTW 360
gi|115435196 IELTPFLGWLLDGEEHEL----------GFA---------VTDAQDFW 362
gi|226507729 IELTPFLGSLLDGEEHEL----------GFA---------VTNAQRSW 359
gi|242051643 IELTPFLGKLLDGEDHEL----------GFA---------VTNAQRSW 356
gi|20804461 VELTPFLGKLLDGKEHEL----------GFA---------VTNAQKSW 356
gi|168024685 VEVTSFLGKLVDDQNHTF----------SIT---------VTNAIPYW 362
gi|168066928 IEITPFLGRLIDDRNHSF----------SAT---------VTNALPFW 308
gi|168016735 VDITPFLGTLVDGERHKF----------GVS---------VTNALPSW 306
gi|226531131 VELTPFLGILVDGKPHEI----------VLS---------VVDGIAEW 377
gi|242090443 VELTPFLGLLVDGKAHEI----------VLS---------VVDGIAEW 378
gi|115463723 VELTPFLGLLVDSNAHEI----------GLS---------VFDGIAEW 370
gi|224112445 FDLTPFLGMLLDGKDHVF----------GIG---------VTDGIEYW 345
gi|224098728 FDLTPFLGMVLDDEDHVF----------GVG---------VTDGIEYW 329
gi|255559509 FDLTPFLGILLDGKDHVI----------GIG---------VANGISYW 367
gi|256394372 LDLTPFAGLLADGKPHTV----------TLVP--------PSDITDTW 334
VVTGKVFIYLNDSGKTTIKPTGVP----------PAISASDLQFSFSRNL 471
gi|242786471 VVTGKVFVYLGHS-EDSIKSTGVP----------PSIHAPDPEFSFSRNL 473
gi|255938730 VVTGNIFIYLNDNSSDSKVTAARDK-------DGPMVDAPLPVFAVTRNL 358
gi|70991399 VVTGNIFLYLEEDASHSRTDQSS----------VPQITAPTPQFTITRLL 460
gi|119467934 VVTGNIFLYLEEDASYSRTDRSKK---------VPQITAPTPQFTITRLL 470
gi|121709910 VITGNIFVYLDEDAE---SNSSEK---------APHIYAPAPSLAVNRNL 391
gi|169769599 AVTGNIFLYLSDSALDS-TSLGTE---------KPYVDAPTPQFKATRSL 374
gi|238488086 AVTGNIFLYLSDSALDS-TSLGTE---------KPYVDAPTPQFKATRSL 255
gi|2731443 VVTGTIFLYLDSSSSESHSTTTG---------QAPEIYAPAPTLTVTRDL 376
gi|145235129 VVTGTIFLYLDDSMS---QIATG---------QAPEVNAPTPTFAVTRNL 373
gi|259484743 AVSGNIFLYLSDGSAEQLPTSAGPG-------QRPDIVAPTPTFTTTRYL 471
gi|239611694 AVTGKIFIYLGGEV-TVNDSAIKGGSD-----SGPVTLTKFASNITHEWQ 424
gi|225680740 VVTGKIFIYLDEES-ASKRSMISPQDN-----SIPVVSTQMKSNIMHEWT 331
gi|240279185 VVTGKIFIYLDGERGLVDQSTINRRGD-----LVPVVSTKFGSDISHEWR 427
gi|189196202 VVTGKVFIWLDDSEGPTTGT-------------IPTISAPASSIVLQSTT 616
gi|169603634 VVTGKIFIWTGTAMNLSIGT-------------VPIISAPAPSIKLQSIT 396
gi|156043521 LVTGKIFIWLDSNDSVTTGT-------------APKIFLPAPIITTSHVL 561
gi|154316707 LVTGKIFIWLDSNDSVTTGT-------------APTLSIPAPIITTSQIL 281
gi|39974109 YITGKVFVWLDDE----GSITTG---------STPEIQGVDPSIDISQHA 514
gi|261351792 VVTGKIFVWLDEP----GSVTTG---------AAPKVDAPEPLLSIS-HQ 421
gi|85092018 YVTGKIFIWTDSD---SHSNSIGSNVDNNNDNKFPTIDGLTPLITLSSIR 548
gi|116179866 YVTGKIFLWLDED---PSSITRG---------EIPIINQPPPTIAVTRSL 477
gi|171691689 VVTGKVFIWLDYDRKDEHACAKGDGCITTG-LKQPVVTAPEPEIVARSEV 501
gi|255567074 YIDANLHLWLDHKSK---------------KIEGKVLKHEGKPLAFS--- 403
gi|224114784 YIDANLHLWLDHRST---------------ITEGKLLKHESKPLALS--- 408
gi|56405352 YVDANLHLWLDKQST---------------KTEGKLSKHSSLPLVVS--- 358
gi|225461673 YIDANLHIWLDHKST---------------QTEGRLLGHGSGSSLST--- 390
gi|225461675 FIDANLHLWLDNKSK---------------RTQGKLLGHNSKSSSSI--- 400
gi|116789291 FVDANLHLWLDDKSSS--------------AIRGQLIGHEEPSLESS--- 388
gi|115435180 YIDANLHLWLDHKSE---------------ETTGSLISYEAQGLVLN--- 409
gi|242051645 YIDANLHLWLDHKSE---------------KTTGSLLSYDASGLDLN--- 397
gi|125569459 FIDANLHIWLDHKSE---------------KTFGSLVSYEAPKLTLH--- 362
gi|115435186 YVDGNLHLWLDPMTT---------------ATTGSLVSYDAPRLAAVNTS 381
gi|242056015 YVDANLHLWLDPGSA---------------ATTAGLIAYVAPELVVN--- 393
gi|242051639 FVDANLHLWLDPRGT---------------ATAAGMISYDAPPLDTA--- 392
gi|242051641 FVDANLHLWLDPRGT---------------ATAAGMISYDAPPLDTA--- 392
gi|115435196 YVDGNLHLWLDPRSA---------------ATTAGIISYDAPPLEKV--- 394
gi|226507729 YVDANLHLWLDPKSS---------------RTSGGLVAYHAPKLAGS--- 391
gi|242051643 YVDANLHLWLDPKST---------------RTTGGLVAYDAPKLAGS--- 388
gi|20804461 YVDANLHLWLDPKSV---------------ATSGGLVAYDAPKLTGK--- 388
gi|168024685 LVSANLHLWLDHSTN---------------ATTGELFEHSAPALTSH--- 394
gi|168066928 LINANLHLWVDSSVD---------------STRGKLTEHSAGALQSH--- 340
gi|168016735 LLGVNLHVWVDESVE---------------ATRGEMVHHFASTSFLT--- 338
gi|226531131 LVDANLHLWLDPAST---------------NVSAALSRYRTPRLSIK--- 409
gi|242090443 LVDANLHLWLDPASP---------------NVSAALRRYRTPRLSIT--- 410
gi|115463723 LVDANLHLWLDPSTS---------------DVHAALGAYQTPRLKIS--- 402
gi|224112445 LVDANLHVWLDTAST---------------VVEAKNVVNINPASEIS--- 377
gi|224098728 LVDANLHIWLDSSST---------------IVEAKNVVNVYPASEIS--- 361
gi|255559509 LVDANLHIWLDKGAA---------------SVVAKSVTYQNPGSSVK--- 399

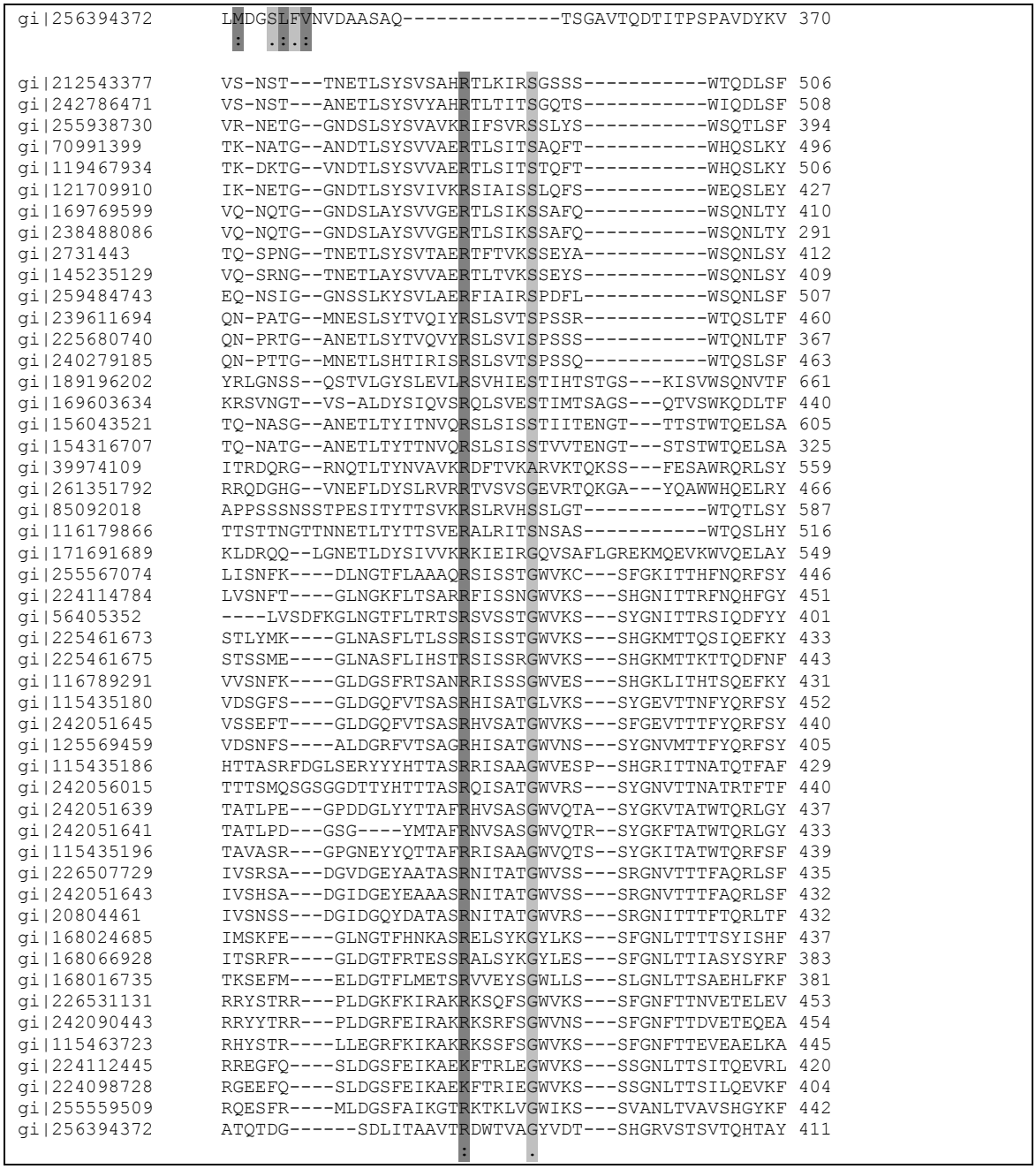


Figure 10.1: CLUSTAL W2 Multiple amino acid sequence alignment for PNGase A and PNGase *At* and related sequences.

Table 10.1: Details of sequences included in the multiple amino acid sequence alignment shown in Figure 10.1.

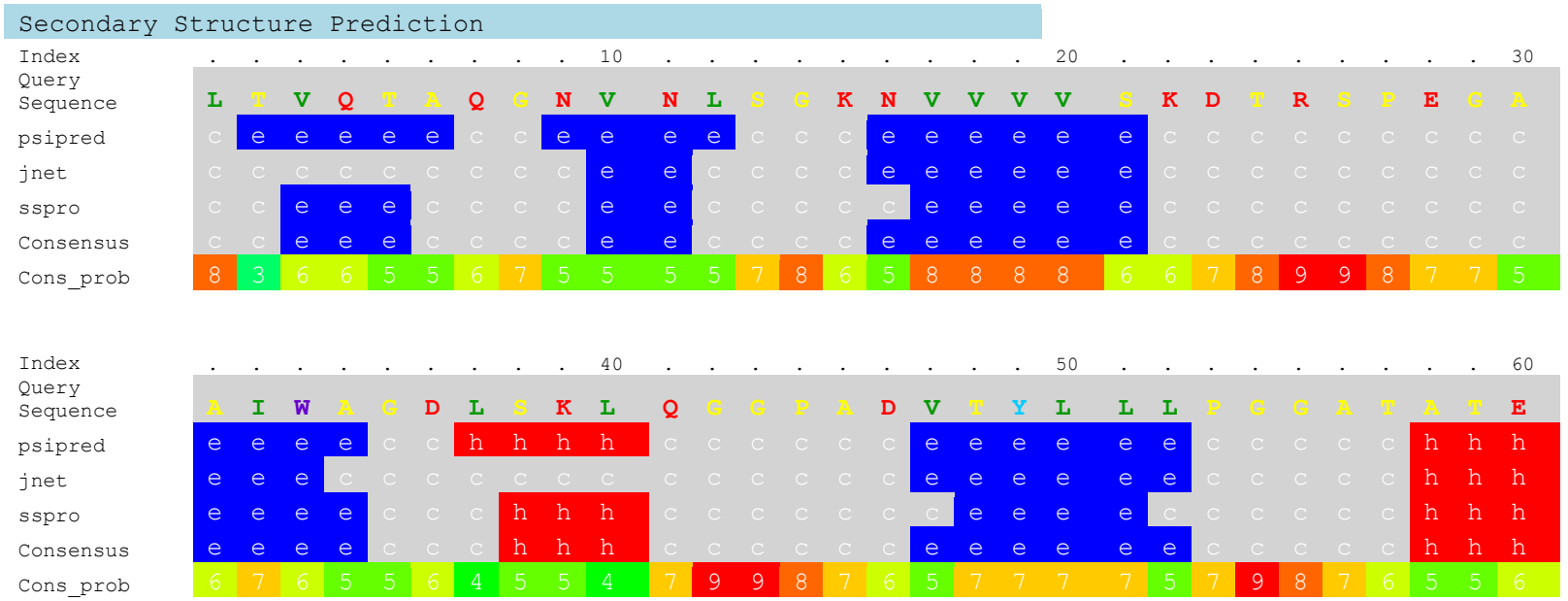
gi number	Assigned as	Organism	Taxonomy
gi|212543377	conserved hypothetical protein	*Penicillium marneffei* ATCC 18224	Fungus
gi|242786471	conserved hypothetical protein	*Talaromyces stipitatus* ATCC 10500	Fungus
gi|255938730	Pc14g01410	*Penicillium chrysogenum Wisconsin* 54-1255	Fungus
gi|70991399	conserved hypothetical protein	*Aspergillus fumigatus* Af293	Fungus
gi|119467934	hypothetical protein NFIA_052210	*Neosartorya fischeri* NRRL 181	Fungus
gi|121709910	conserved hypothetical protein	*Aspergillus clavatus* NRRL 1	Fungus
gi|169769599	hypothetical protein	*Aspergillus oryzae* RIB40	Fungus
gi|238488086	peptide-N4-(N-acetyl-beta-glucosaminyl)asparagine amidase A, putative	*Aspergillus flavus* NRRL3357	Fungus
gi|2731443	cDNA of the glycoamidase gene; PNGase *At*	*Aspergillus tubingensis*	Fungus
gi|145235129	hypothetical protein An03g03300	*Aspergillus niger*	Fungus
gi|259484743	TPA: conserved hypothetical protein	*Aspergillus nidulans* FGSC A4	Fungus
gi|239611694	peptide-N4-(N-acetyl-beta-glucosaminyl)asparagine amidase A	*Ajellomyces dermatitidis* ER-3	Fungus
gi|225680740	peptide-N4-(N-acetyl-beta-glucosaminyl)asparagine amidase A	*Paracoccidioides brasiliensis* Pb03	Fungus
gi|240279185	peptide-N4-(N-acetyl-beta-glucosaminyl) asparagine amidase A	*Ajellomyces capsulatus* H143	Fungus
gi|189196202	peptide-N4-(N-acetyl-beta-glucosaminyl)asparagine amidase A	*Pyrenophora tritici-repentis* Pt-1C-BFP	Fungus
gi|169603634	hypothetical protein SNOG_04825	*Phaeosphaeria nodorum* SN15	Fungus
gi|156043521	hypothetical protein SS1G_10764	*Sclerotinia sclerotiorum* 1980	Fungus
gi|154316707	hypothetical protein BC1G_03771	*Botryotinia fuckeliana* B05.10	Fungus
gi|39974109	hypothetical protein MGG_00799	*Magnaporthe grisea* 70-15	Fungus
gi|261351792	peptide-N4-(N-acetyl-beta-glucosaminyl)asparagine amidase A	*Verticillium albo-atrum* VaMs.102	Fungus
gi|85092018	hypothetical protein NCU04643	*Neurospora crassa* OR74A	Fungus
gi|116179866	hypothetical protein CHGG_00561	*Chaetomium globosum*	Fungus

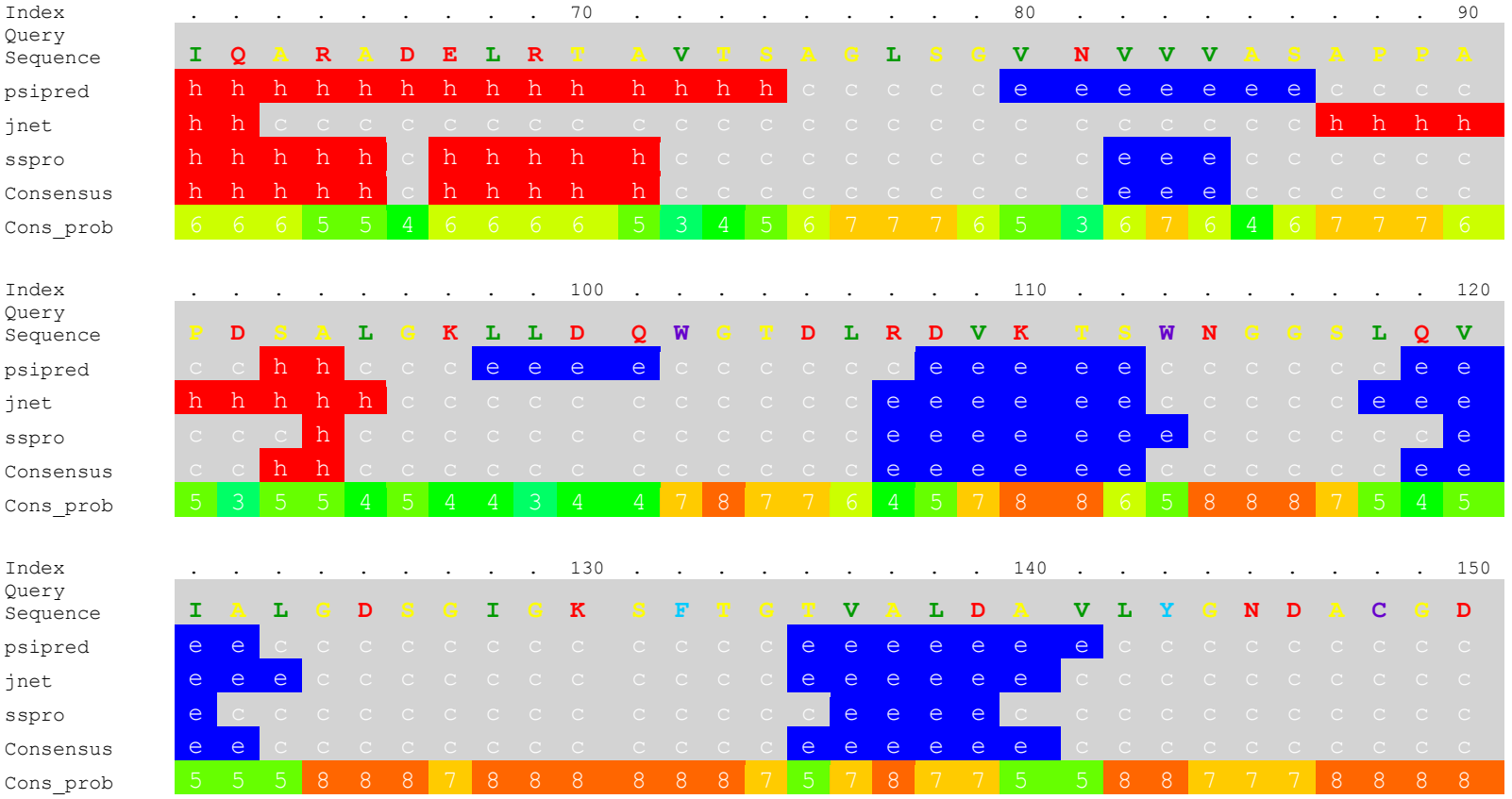
		CBS 148.51	
gi|171691689	unnamed protein product	*Podospora anserina*	Fungus
gi|255567074	Peptide-N4-(N-acetyl-beta-glucosaminyl)asparagine amidase A, putative	*Ricinus communis*	Plant
gi|224114784	predicted protein	*Populus trichocarpa*	Plant
gi|56405352	Peptide-N4-(N-acetyl-beta-glucosaminyl)asparagine amidase A; PNGase A	*Prunus dulcis*	Plant
gi|225461673	PREDICTED: hypothetical protein	*Vitis vinifera*	Plant
gi|225461675	hypothetical protein	*Vitis vinifera*	Plant
gi|116789291	unknown	*Picea sitchensis*	Plant
gi|115435180	Os01g0207200	*Oryza sativa* (japonica cultivar-group)	Plant
gi|242051645	hypothetical protein SORBIDRAFT_03g002300	*Sorghum bicolor*	Plant
gi|125569459	hypothetical protein OsJ_00817	*Oryza sativa* Japonica Group	Plant
gi|115435186	Os01g0207600	*Oryza sativa* (japonica cultivar-group)	Plant
gi|242056015	hypothetical protein SORBIDRAFT_03g002260	*Sorghum bicolor*	Plant
gi|242051639	hypothetical protein SORBIDRAFT_03g002240	*Sorghum bicolor*	Plant
gi|242051641	hypothetical protein SORBIDRAFT_03g002250	*Sorghum bicolor*	Plant
gi|115435196	Os01g0208400	*Oryza sativa* (japonica cultivar-group	Plant
gi|226507729	peptide-N4-asparagine amidase A	*Zea mays*	Plant
gi|242051643	hypothetical protein SORBIDRAFT_03g002270	*Sorghum bicolor*	Plant
gi|20804461	hypothetical protein	*Oryza sativa* Japonica Group	Plant
gi|168024685	predicted protein	*Physcomitrella patens* subsp. *patens*	Plant
gi|168066928	predicted protein	*Physcomitrella patens* subsp. *patens*	Plant
gi|168016735	predicted protein	*Physcomitrella patens* subsp. *patens*	Plant
gi|226531131	hypothetical protein LOC100274582	*Zea mays*	Plant
gi|242090443	hypothetical protein SORBIDRAFT_09g019510	*Sorghum bicolor*	Plant
gi|115463723	Os05g0395000	*Oryza sativa* (japonica cultivar-group)	Plant
gi|224112445	predicted protein	*Populus trichocarpa*	Plant
gi|224098728	predicted protein	*Populus trichocarpa*	Plant
gi|255559509	Peptide-N4-(N-acetyl-beta-glucosaminyl)asparagine amidase A, putative	*Ricinus communis*	Plant

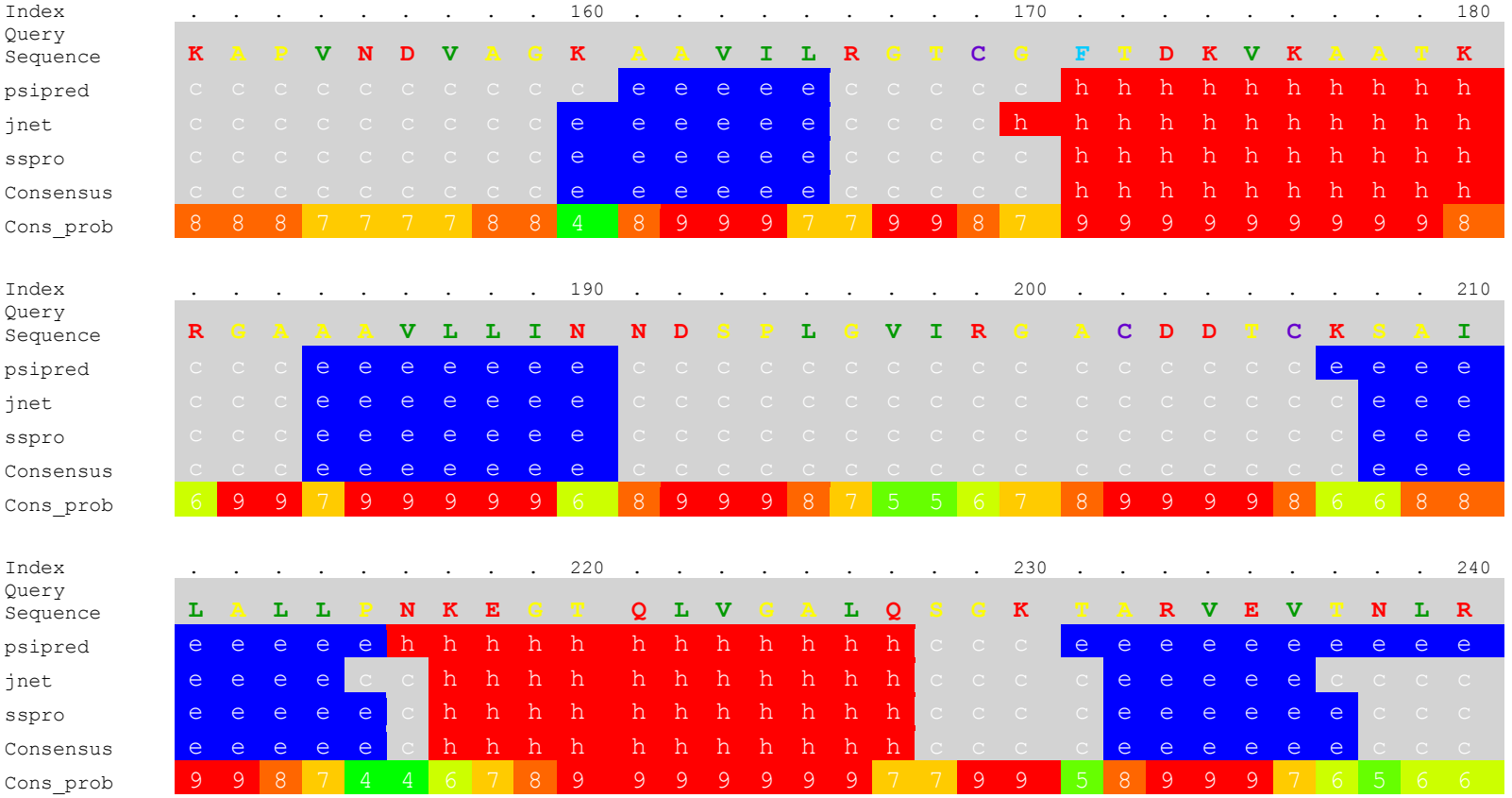
gi|256394372	peptide-N(4)-(N-acetyl-beta-glucosaminyl) asparagine amidase	*Catenulispora acidiphila* DSM 44928	Bacteria

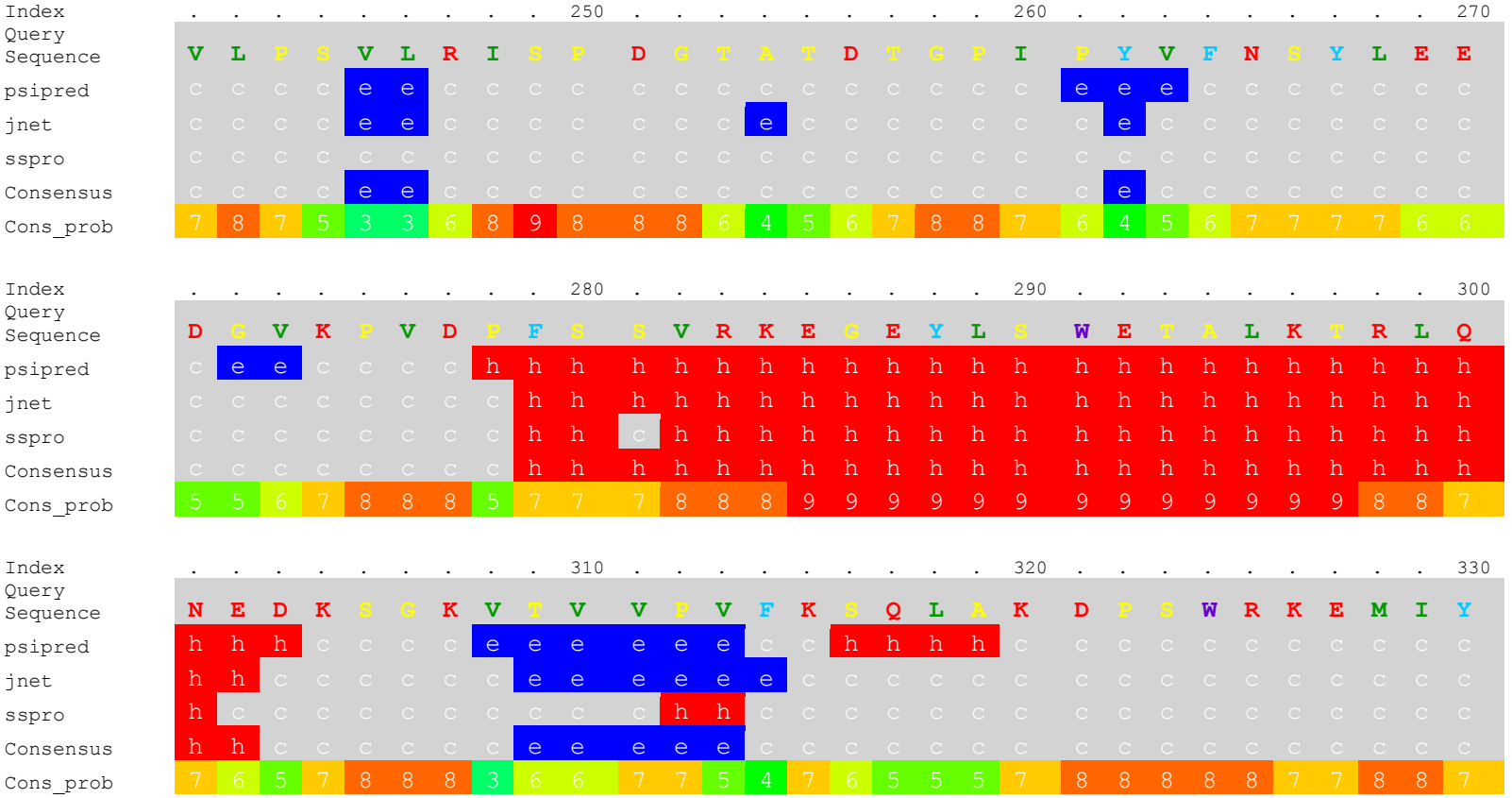
10.2 Appendix 2

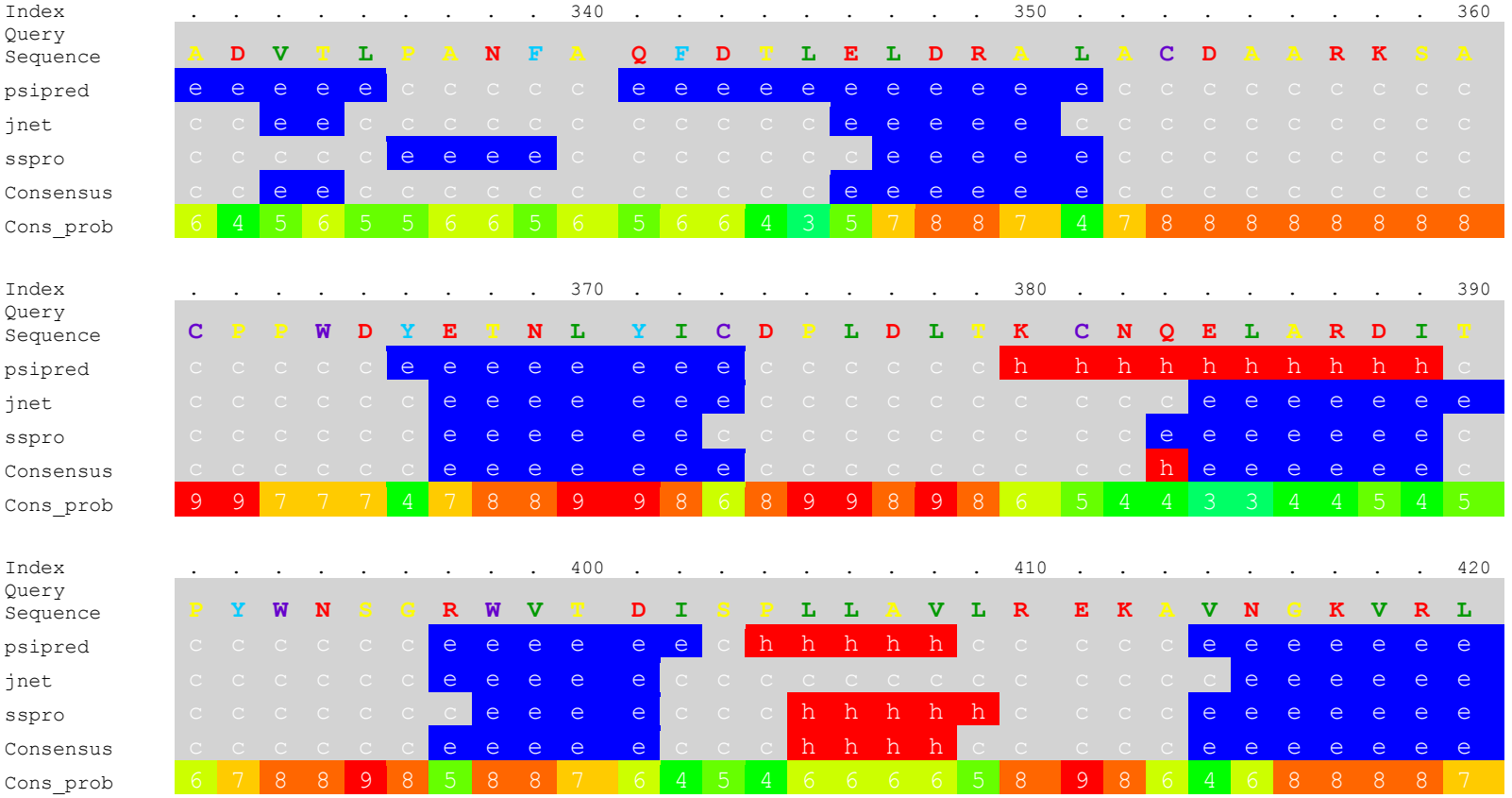
Visual representation of the secondary structure prediction for *Dra*PNGase using the Phyre server is shown in **Figure 10.2**.

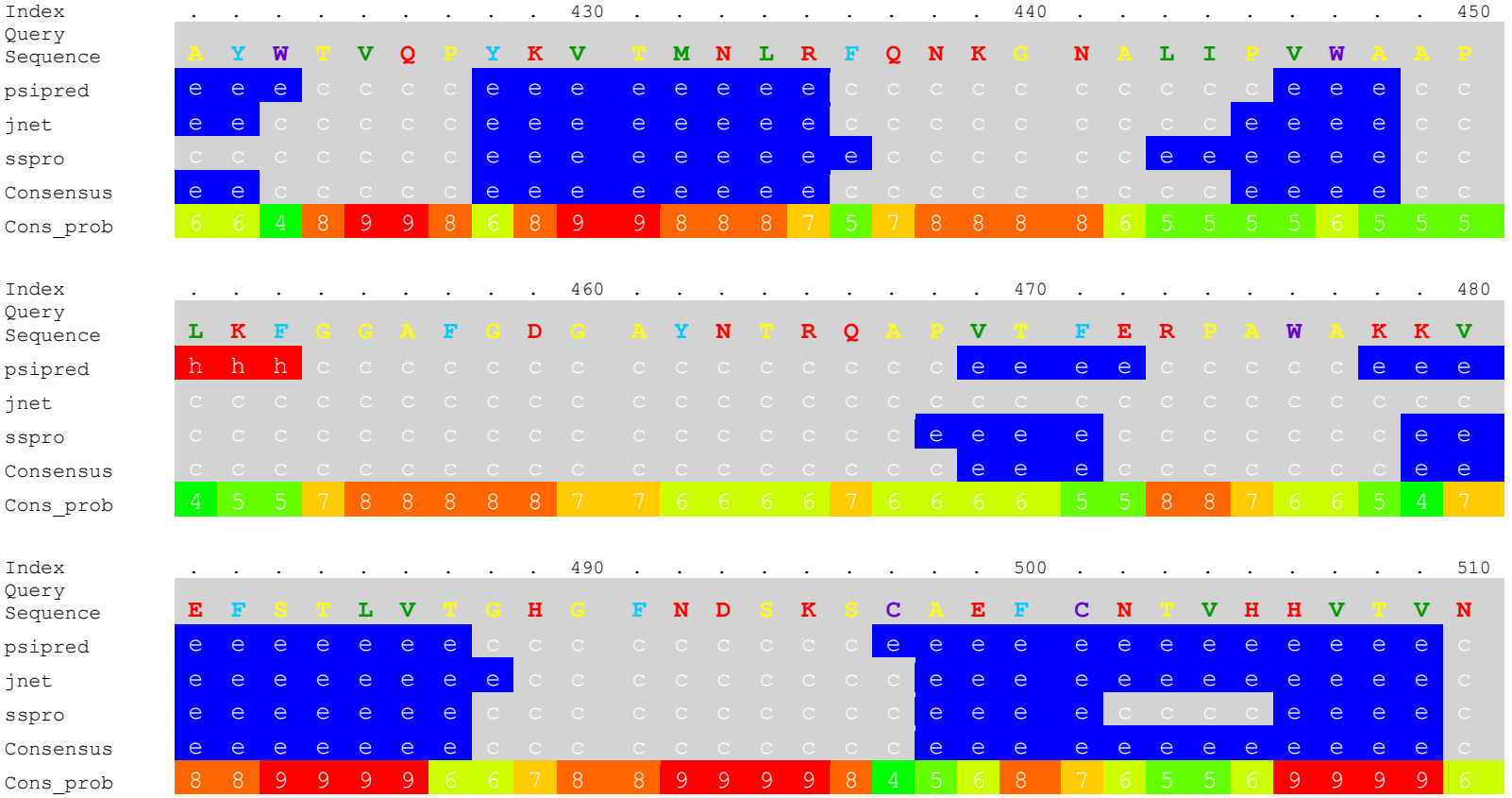


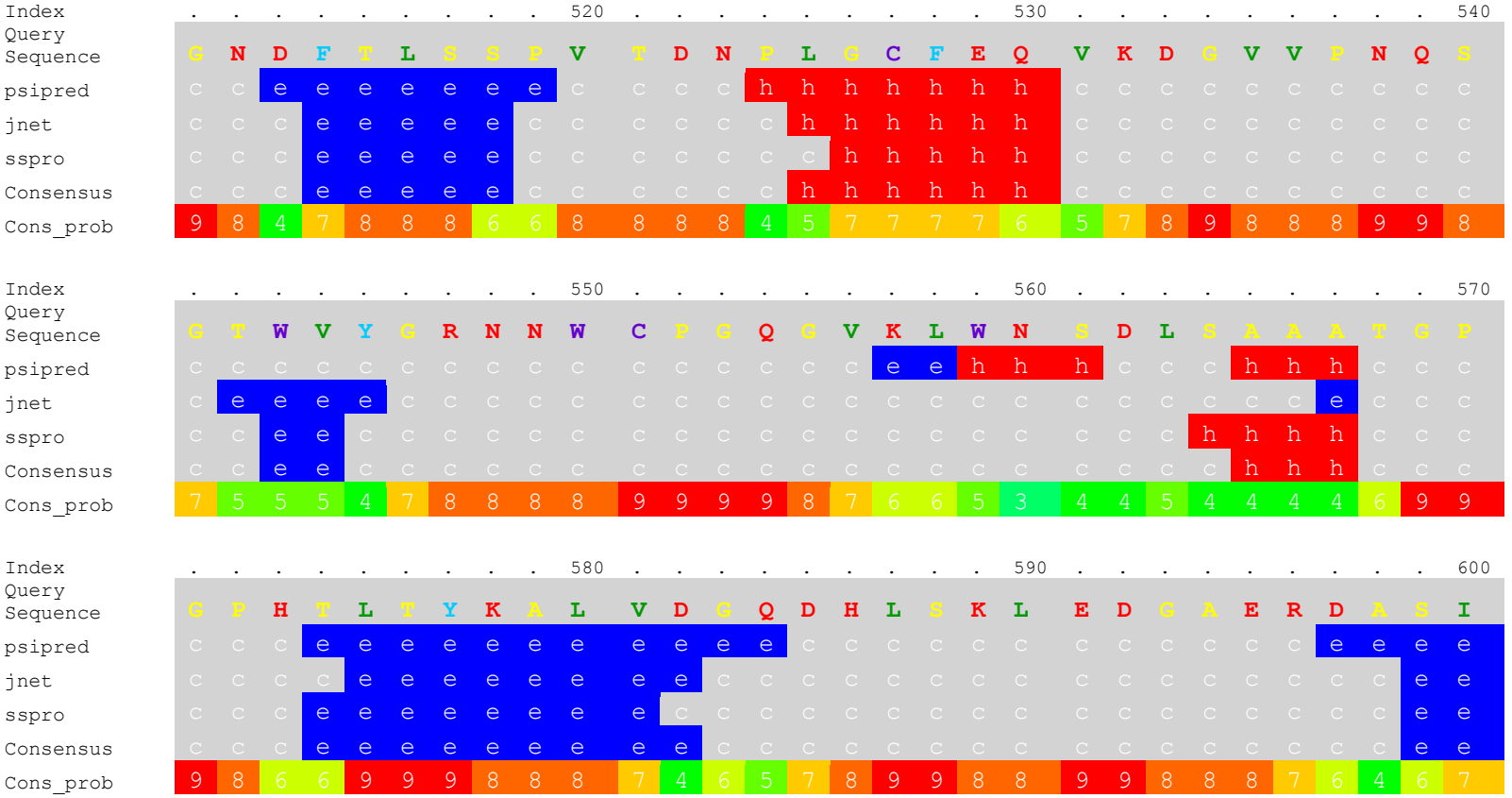

Index
Query Sequence
psipred
jnet
sspro
Consensus
Cons_prob
70
80
90
100
110
120
130
140
150


Index . . . . . 160 . . . . . 170 180
Query Sequence K A P V N D V A G K A A V I L R G T C G F T D K V K A A T K
psipred c c c c c c c c c c e e e e e c c c c c h h h h h h h h h h
jnet c c c c c c c c c e e e e e e c c c c h h h h h h h h h h h
sspro c c c c c c c c c e e e e e e c c c c c h h h h h h h h h h
Consensus c c c c c c c c c e e e e e e c c c c c h h h h h h h h h h
Cons_prob 8 8 8 7 7 7 7 8 8 4 8 9 9 9 7 7 9 9 8 7 9 9 9 9 9 9 9 9 9 8
Index . . . . . 190 . . . . . 200 210
Query Sequence R G A A A V L L I N N D S P L G V I R G A C D D T C K S A I
psipred c c c e e e e e e e c c c c c c c c c c c c c c c c c c e e e e
jnet c c c e e e e e e e c c c c c c c c c c c c c c c c c c e e e
sspro c c c e e e e e e e c c c c c c c c c c c c c c c c c c e e e
Consensus c c c e e e e e e e c c c c c c c c c c c c c c c c c c e e e
Cons_prob 6 9 9 7 9 9 9 9 9 6 8 9 9 9 8 7 5 5 6 7 8 9 9 9 9 8 6 6 8 8
Index . . . . . 220 . . . . . 230 240
Query Sequence L A L L P N K E G T Q L V G A L Q S G K T A R V E V T N L R
psipred e e e e e h h h h h h h h h h h h h c c c e e e e e e e e e e
jnet e e e e c c h h h h h h h h h h h h c c c c e e e e e c c c c
sspro e e e e e c h h h h h h h h h h h h c c c c e e e e e e c c c
Consensus e e e e e c h h h h h h h h h h h h c c c c e e e e e e c c c
Cons_prob 9 9 8 7 4 4 6 7 8 9 9 9 9 9 9 9 7 7 9 9 5 8 9 9 9 7 6 5 6 6


Index
Query
Sequence
psipred
jnet
sspro
Consensus
Cons_prob
250
260
270
280
290
300
310
320
330


Index . . . . . 340 . . . . . 350 360
Query Sequence A D V T L P A N F A Q F D T L E L D R A L A C D A A R K S A
psipred e e e e e c c c c c e e e e e e e e e e e c c c c c c c c c c
jnet c c e e c c c c c c c c c c c e e e e e c c c c c c c c c c
sspro c c c c c e e e e c c c c c c c e e e e e c c c c c c c c c c
Consensus c c e e c c c c c c c c c c c e e e e e c c c c c c c c c c
Cons_prob 6 4 5 6 5 5 6 6 5 6 5 6 6 4 3 5 7 8 8 7 4 7 8 8 8 8 8 8 8 8
Index . . . . . 370 . . . . . 380 390
Query Sequence C P P W D Y E T N L Y I C D P L D L T K C N Q E L A R D I T
psipred c c c c c e e e e e e e e c c c c c c h h h h h h h h h c
jnet c c c c c c e e e e e e e c c c c c c c c c c e e e e e e e
sspro c c c c c c e e e e e e c c c c c c c c c c e e e e e e e c
Consensus c c c c c c e e e e e e e c c c c c c c c c h e e e e e e c
Cons_prob 9 9 7 7 7 4 7 8 8 9 9 8 6 8 9 9 8 9 8 6 5 4 4 3 3 4 4 5 4 5
Index . . . . . 400 . . . . . 410 420
Query Sequence P Y W N S G R W V T D I S P L L A V L R E K A V N G K V R L
psipred c c c c c c e e e e e e c h h h h h c c c c c e e e e e e e
jnet c c c c c c e e e e e c c c c c c c c c c c c e e e e e e
sspro c c c c c c c e e e e c c c h h h h h c c c c e e e e e e e
Consensus c c c c c c e e e e e c c c h h h h c c c c c e e e e e e e
Cons_prob 6 7 8 8 9 8 5 8 8 7 6 4 5 4 6 6 6 6 5 8 9 8 6 4 6 8 8 8 8 7


Index
Query Sequence
psipred
jnet
sspro
Consensus
Cons_prob
430
440
450
460
470
480
490
500
510


Index
Query Sequence
psipred
jnet
sspro
Consensus
Cons_prob
520
530
540
550
560
570
580
590
600

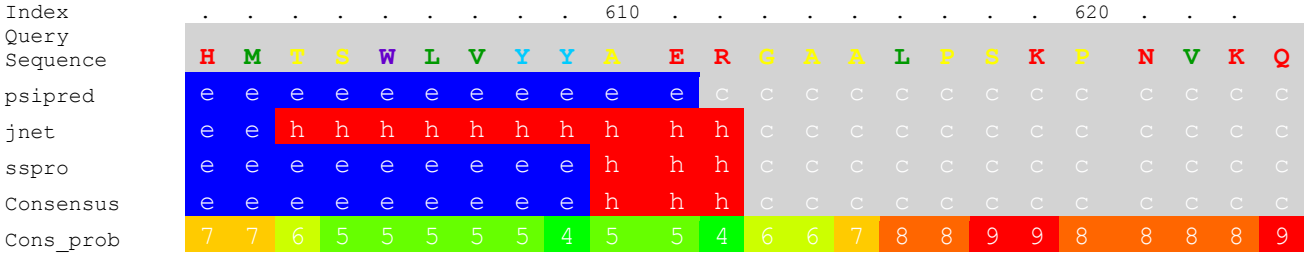
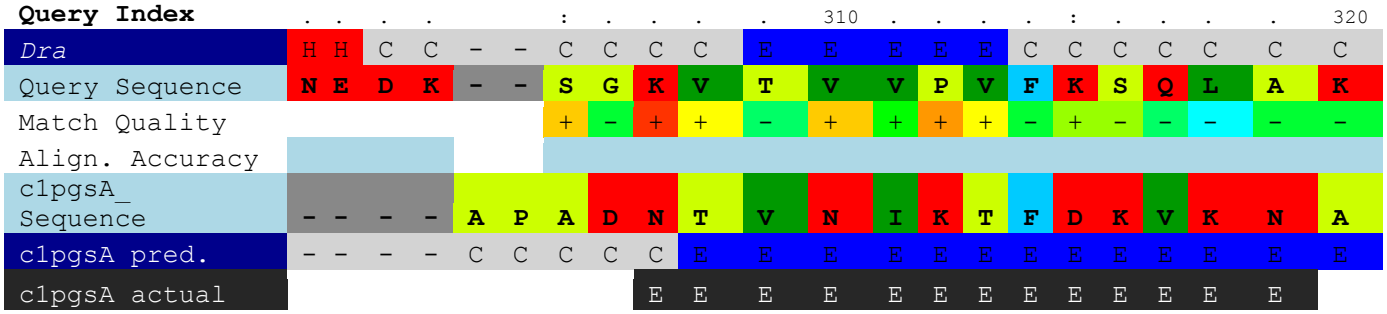
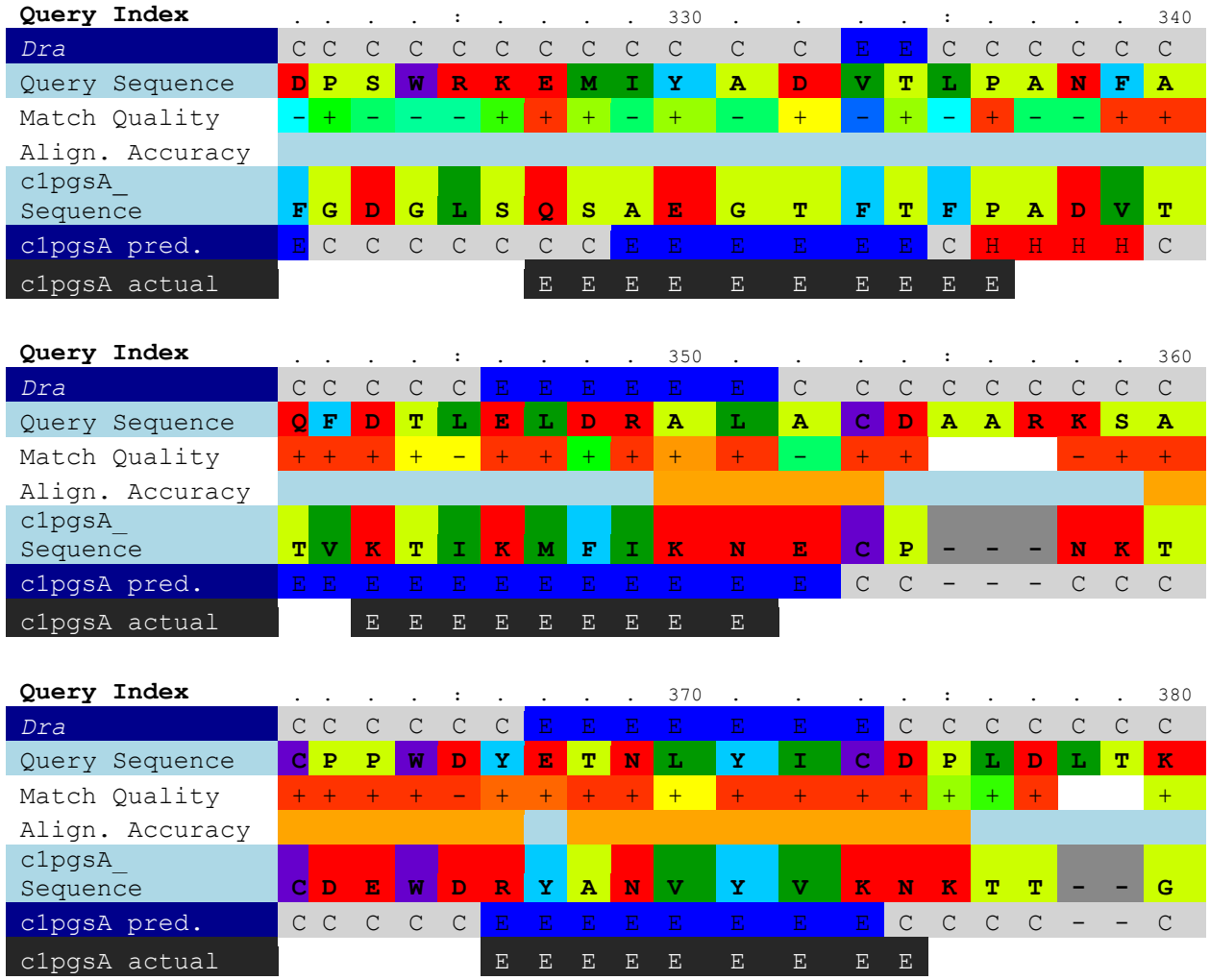
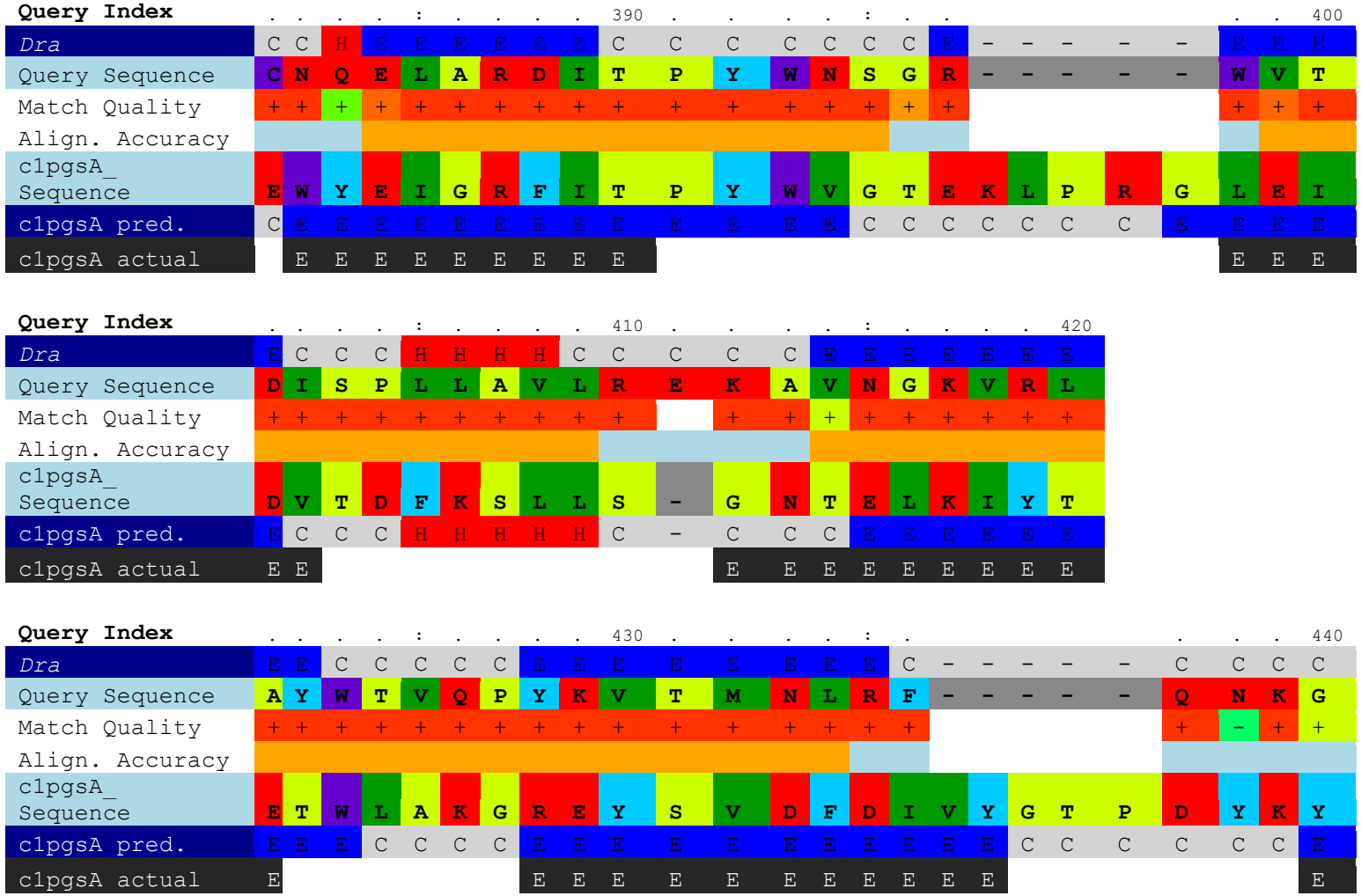


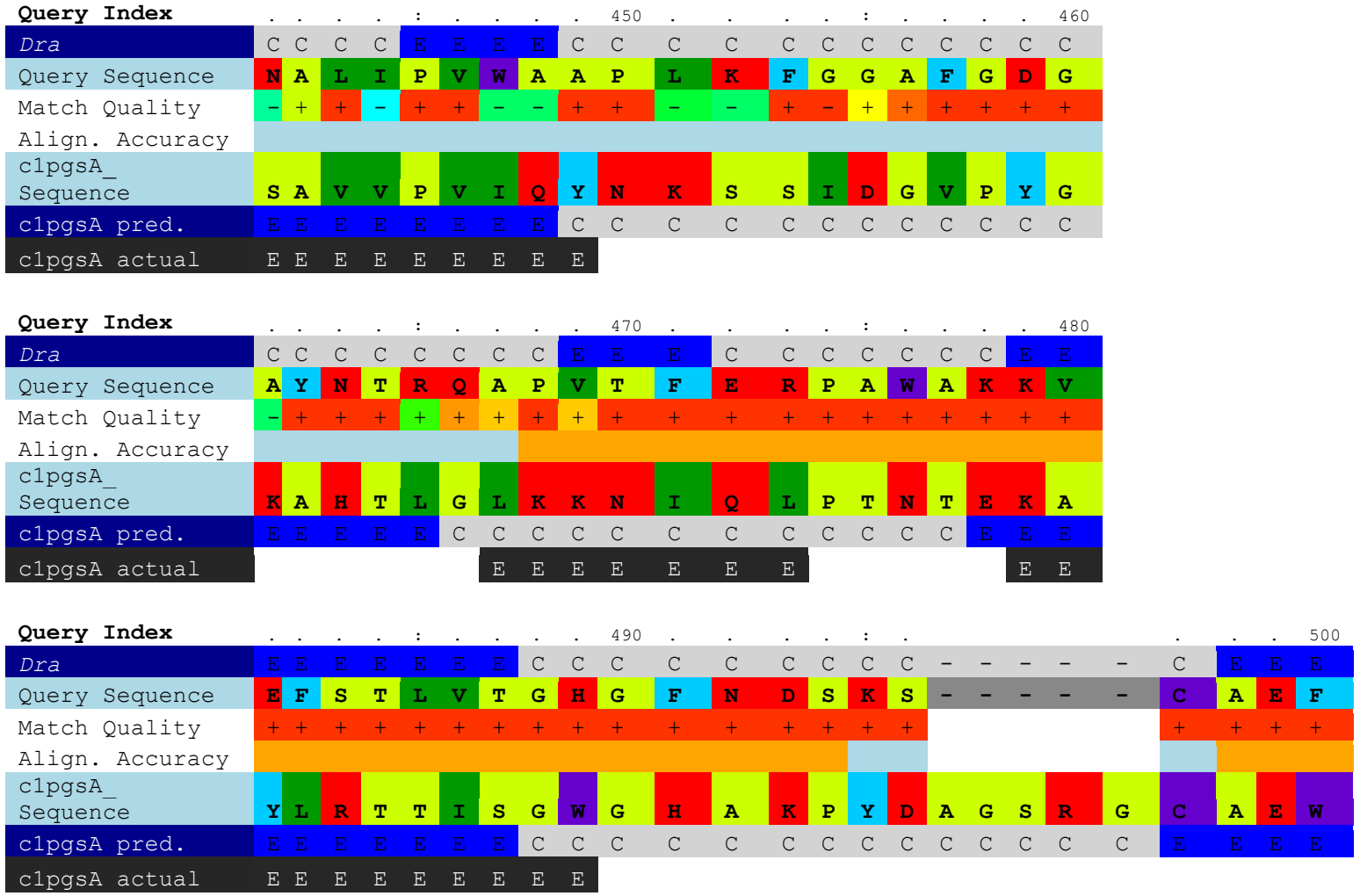
Figure 10.2: Secondary structure prediction for *Dra*PNGase using the Phyre server.
Red (h): α-helices; Blue (e): β-strands; Grey (c): coil. The Cons_prob row indicates the confidence of the prediction from 0 (low confidence) to 9 (high confidence.

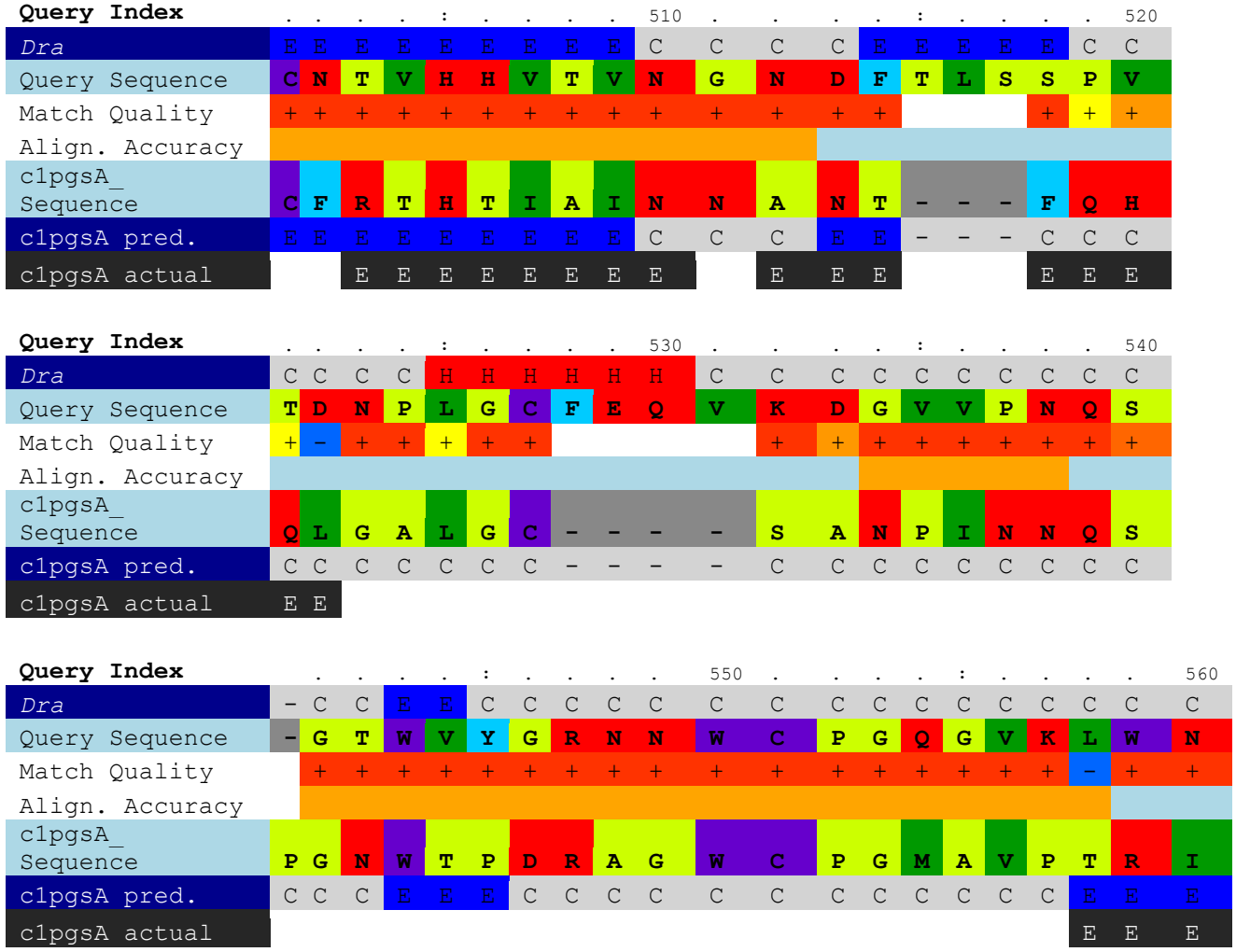
Figure 10.3 shows the alignment of *Dra*PNGase (C-terminal domain only) and PNGase F (c1pgsA) following the Phyre structure recognition scan.

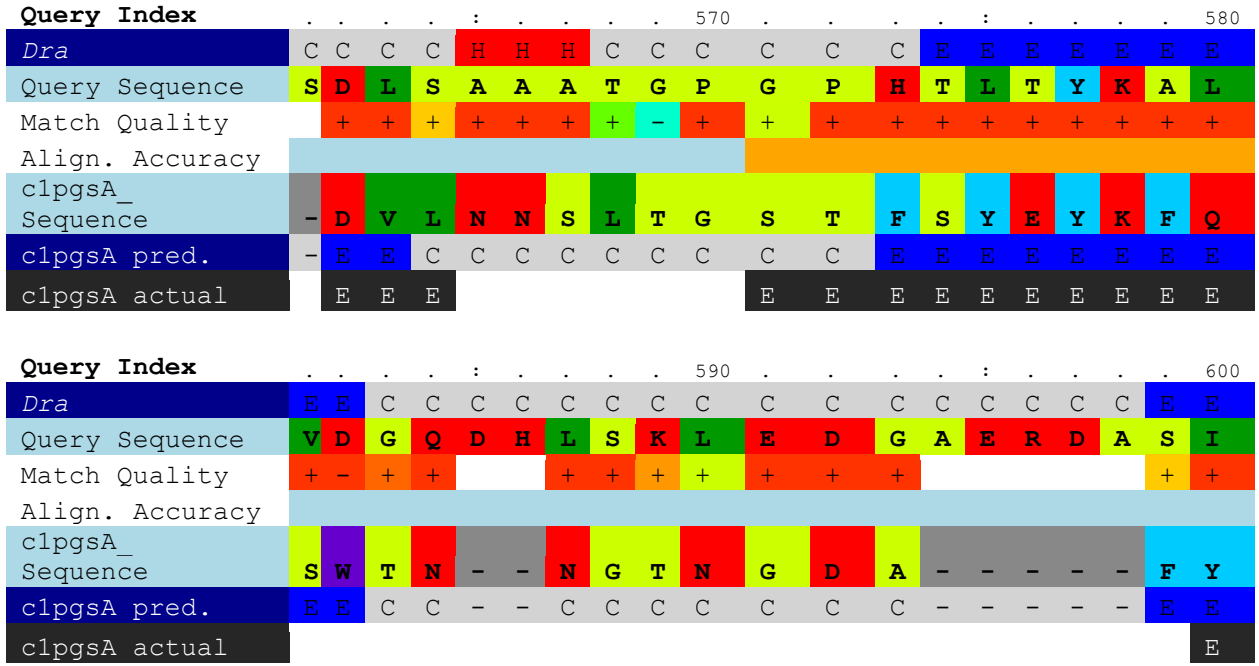



Query Index
330
340
Dra
Query Sequence
Match Quality
Align. Accuracy
c1pgsA_ Sequence
c1pgsA pred.
c1pgsA actual
Query Index
350
360
Dra
Query Sequence
Match Quality
Align. Accuracy
c1pgsA_ Sequence
c1pgsA pred.
c1pgsA actual
Query Index
370
380
Dra
Query Sequence
Match Quality
Align. Accuracy
c1pgsA_ Sequence
c1pgsA pred.
c1pgsA actual


Query Index
390
400
Dra
Query Sequence
Match Quality
Align. Accuracy
c1pgsA_ Sequence
c1pgsA pred.
c1pgsA actual
Query Index
410
420
Dra
Query Sequence
Match Quality
Align. Accuracy
c1pgsA_ Sequence
c1pgsA pred.
c1pgsA actual
Query Index
430
440
Dra
Query Sequence
Match Quality
Align. Accuracy
c1pgsA_ Sequence
c1pgsA pred.
c1pgsA actual


Query Index
450
460
Dra
Query Sequence
Match Quality
Align. Accuracy
c1pgsA_ Sequence
c1pgsA pred.
c1pgsA actual
Query Index
470
480
Dra
Query Sequence
Match Quality
Align. Accuracy
c1pgsA_ Sequence
c1pgsA pred.
c1pgsA actual
Query Index
490
500
Dra
Query Sequence
Match Quality
Align. Accuracy
c1pgsA_ Sequence
c1pgsA pred.
c1pgsA actual


Query Index
510
520
Dra
Query Sequence
Match Quality
Align. Accuracy
c1pgsA_ Sequence
c1pgsA pred.
c1pgsA actual
530
540
550
560


Query Index
570
580
Dra
C C C C H H H C C C C C C E E E E E E E
Query Sequence
S D L S A A A T G P G P H T L T Y K A L
Match Quality
+ + + + + + + - + + + + + + + + + + +
Align. Accuracy
c1pgsA_ Sequence
- D V L N N S L T G S T F S Y E Y K F Q
c1pgsA pred.
- E E C C C C C C C C E E E E E E E E
c1pgsA actual
E E E E E E E E E E
Query Index
590
600
Dra
E E C C C C C C C C C C C C C C C E E
Query Sequence
V D G Q D H L S K L E D G A E R D A S I
Match Quality
+ - + + + + + + + + + + + + +
Align. Accuracy
c1pgsA_ Sequence
S W T N - - N G T N G D A - - - - - F Y
c1pgsA pred.
E E C C - - C C C C C C C - - - - - E E
c1pgsA actual
E

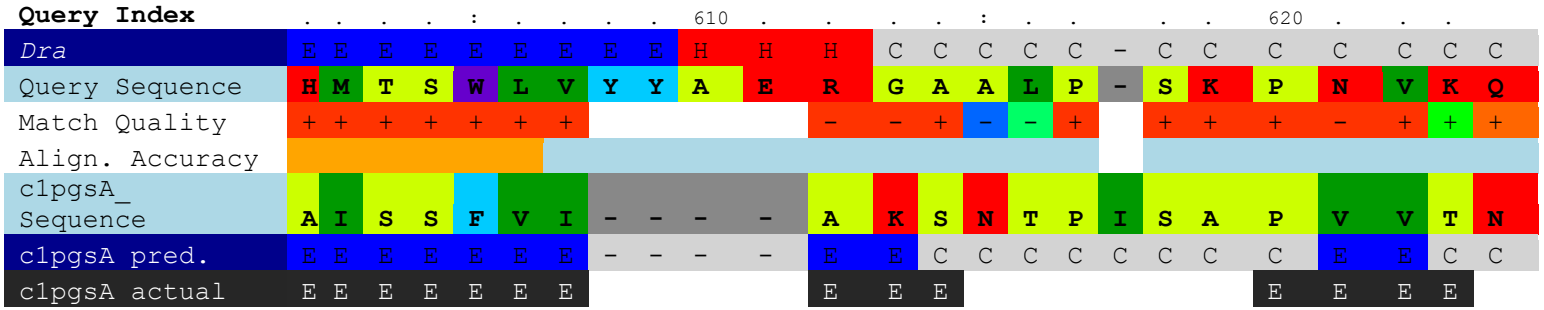


Figure 10.3: Alignment of *Dra*PNGase and PNGase F following the Phyre folding recognition scan.
The last row ('c1pgsA actual') showing the actual β-strands that were found in the PNGase F crystal structure (Norris *et al.*, 1994b) was added manually as indication of the secondary structure prediction accuracy. Colour code for 'Match Quality' row: red = high and blue = low. Colour code for 'Match Quality' row: orange: contiguous high-scoring regions; blue: low scoring or 'patchy' regions of mixed high and low scores.

10.3 Appendix 3

Result of the disorder prediction for *Dra*PNGase using PONDR®. The truncation positions are shaded grey. The sequence numbering here is based on the full length protein including the 30 amino acid long signal sequence.

```
1       MRFTLSVLSL TSVLLLSGCG LLSTPSDPNA LTVQTAQGNV NLSGKNVVVV
VLXT                     DDD DDDDDDDDDD DDDDDDDDDD DDDDDDDD

51      SKDTRSPEGA AIWAGDLSKL QGGPADVTYL LLPGGATATE IQARADELRT
VLXT                               DDDD DDDDDDDDDD DDDDDDDDDD

101     AVTSAGLSGV NVVVASAPPA PDSALGKLLD QWGTDLRDVK TSWNGGSLQV
VLXT    DDDDDDDDDD DDDDDDDDDD DDD

151     IALGDSGIGK SFTGTVALDA VLYGNDACGD KAPVNDVAGK AAVILRGTCG
VLXT

201     FTDKVKAATK RGAAAVLLIN NDSPLGVIRG ACDDTCKSAI LALLPNKEGT
VLXT

251     QLVGALQSGK TARVEVTNLR VLPSVLRISP DGTATDTGPI PYVFNSYLEE
VLXT                                DDDDD DDDD

301     DGVKPVDPFS SVRKEGEYLS WETALKTRLQ NEDKSGKVTV VPVFKSQLAK
VLXT

351     DPSWRKEMIY ADVTLPANFA QFDTLELDRA LACDAARKSA CPPWDYETNL
VLXT

401     YICDPLDLTK CNQELARDIT PYWNSGRWVT DISPLLAVLR EKAVNGKVRL
VLXT

451     AYWTVQPYKV TMNLRFQNKG NALIPVWAAP LKFGGAFGDG AYNTRQAPVT
VLXT

501     FERPAWAKKV EFSTLVTGHG FNDSKSCAEF CNTVHHVTVN GNDFTLSSPV
VLXT

551     TDNPLGCFEQ VKDGVVPNQS GTWVYGRNNW CPGQGVKLWN SDLSAAATGP
VLXT                                                DDDDDDDDD

601     GPHTLTYKAL VDGQDHLSKL EDGAERDASI HMTSWLVYYA ERGAALPSKP
VLXT    DDDD              DDD DDDD                    DDDDDDD

651     NVKQ
VLXT    DDDD
```

The VL-XT predictor integrates three feed-forward neural networks: the VL1 predictor (Romero *et al.*, 1997), the N-terminus predictor (XN), and the C-terminus predictor (XC) (Li *et al.*, 1999). VL1 was trained using 8 long

disordered regions identified from missing electron density in x-ray crystallographic studies, and 7 long disordered regions characterized by NMR. The XN and XC predictors, together called XT, were also trained using x-ray crystallographic data, where the terminal disordered regions were 5 or more amino acids in length (www.pondr.com/pondr-tut2).

10.4 Appendix 4

The following tables give the rates obtained for the determination of kinetic parameters of PNGase F and its site-specific mutants analysed in this study. Rates were calculated for each integrated product peak area using the appropriate standard curve. Rate means and standard deviations were calculated automatically in GraphPad Prism® 5.

Table 10.2: Rates for PNGase F wildtype

Substrate [mg/mL]	Rate *v* 1	Rate *v* 2	Rate *v* 3
	[µg/mL*min^{-1}]		
0	0	0	0
0.023	0.70	0.69	0.71
0.045	1.18	1.21	1.17
0.09	1.88	1.80	1.87
0.225	3.03	2.64	2.84
0.45	3.47	3.04	3.00
0.675	2.97	2.78	2.94
0.9	2.71	2.61	2.73

Table 10.3: Rates for PNGase F D60C

Substrate [mg/mL]	Rate *v* 1	Rate *v* 2	Rate *v* 3
	[µg/mL*min^{-1}]		
0	0	0	0
0.023	0.76	0.76	0.73
0.045	1.13	1.11	1.06
0.09	1.54	1.55	1.56
0.225	2.18	2.11	2.10
0.45	2.15	2.04	2.04
0.675	1.92	1.93	1.95
0.9	1.80	1.68	1.64

Table 10.4: Rates for PNGase F W59Q

Substrate [mg/mL]	Rate *v* 1	Rate *v* 2	Rate *v* 3
	[µg/mL*min^{-1}]		
0	0	0	0
0.023	0.695	0.703	0.699
0.045	1.354	1.361	1.391
0.09	2.659	2.604	2.689
0.225	6.511	6.386	6.775
0.45	11.239	11.384	-
0.675	17.240	16.685	16.475
0.9	21.930	20.350	20.458

Table 10.5: Rates for PNGase F I82Q

Substrate [mg/mL]	Rate *v* 1	Rate *v* 2	Rate *v* 3
	[µg/mL*min^{-1}]		
0	0	0	0
0.023	0.757	0.739	0.734
0.045	1.375	1.339	1.370
0.09	2.385	2.314	2.217
0.225	4.305	4.066	4.054
0.45	5.826	5.525	5.652
0.675	6.464	6.350	6.278
0.9	6.960	6.830	6.708

Table 10.6: Rates for PNGase F I82R

Substrate [mg/mL]	Rate *v* 1	Rate *v* 2	Rate *v* 3
	[µg/mL*min^{-1}]		
0	0	0	0
0.023	0.683	0.677	0.683
0.045	1.148	1.072	1.124
0.09	1.732	1.667	1.690
0.225	2.655	2.617	2.503
0.45	3.077	3.025	3.011
0.675	3.378	3.204	3.172
0.9	-	3.590	3.553

Table 10.7: Rates for PNGase F W207Q

Substrate [mg/mL]	Rate *v* 1	Rate *v* 2	Rate *v* 3
	[µg/mL*min^{-1}]		
0	0	0	0
0.023	1.131	1.080	1.097
0.045	1.956	1.821	1.885
0.09	3.200	3.252	3.198
0.225	5.577	5.369	5.289
0.45	6.843	6.716	6.738
0.675	7.503	7.335	7.238
0.9	7.966	7.561	7.689

Table 10.8: Rates for PNGase F R248K

Substrate [mg/mL]	Rate *v* 1	Rate *v* 2	Rate *v* 3
	[µg/mL*min^{-1}]		
0	0	0	0
0.023	0.717	0.699	0.677
0.045	1.286	1.249	1.294
0.09	2.205	2.217	2.152
0.225	3.750	3.533	3.759
0.45	4.733	4.584	4.707
0.675	5.044	4.980	5.002
0.9	5.509	5.462	5.364

Table 10.9: Rates for PNGase F R248Q

Substrate [mg/mL]	Rate *v* 1	Rate *v* 2	Rate *v* 3
	[µg/mL*min^{-1}]		
0	0	0	0
0.023	0.718	0.658	0.703
0.045	1.258	1.331	1.216
0.09	2.262	2.291	2.214
0.225	5.118	5.021	5.061
0.45	8.599	8.641	8.650
0.675	11.671	11.566	11.136
0.9	13.619	12.978	13.291

Table 10.10: Rates for PNGase F W251Q

Substrate [mg/mL]	Rate *v* 1	Rate *v* 2	Rate *v* 3
	[µg/mL*min^{-1}]		
0	0	0	0
0.023	0.874	0.901	0.858
0.045	1.507	1.624	1.528
0.09	2.423	2.571	2.664
0.225	4.344	4.335	4.288
0.45	5.646	5.250	5.199
0.675	5.921	5.680	5.841
0.9	6.300	6.321	6.081

Table 10.11: Rates for PNGase F V257N

Substrate [mg/mL]	Rate *v* 1	Rate *v* 2	Rate *v* 3
	[µg/mL*min^{-1}]		
0	0	0	0
0.023	0.55	0.562	0.558
0.045	1.063	1.029	1.042
0.09	1.527	1.638	1.623
0.225	2.643	2.619	2.416
0.45	2.938	2.773	2.854
0.675	3.021	2.795	2.851
0.9	2.932	2.744	2.859

References

References

Abu-Qarn, M. & Eichler, J. (2006). Protein N-glycosylation in Archaea: defining *Haloferax volcanii* genes involved in S-layer glycoprotein glycosylation. *Mol Microbiol* **61**, 511-525.

Abu-Qarn, M., Yurist-Doutsch, S., Giordano, A., Trauner, A., Morris, H. R., Hitchen, P., Medalia, O., Dell, A. & Eichler, J. (2007). *Haloferax volcanii* AglB and AglD are involved in N-glycosylation of the S-layer glycoprotein and proper assembly of the surface layer. *J Mol Biol* **374**, 1224-1236.

Abu-Qarn, M., Eichler, J. & Sharon, N. (2008a). Not just for Eukarya anymore: protein glycosylation in Bacteria and Archaea. *Curr Opin Struct Biol* **18**, 544-550.

Abu-Qarn, M., Giordano, A., Battaglia, F., Trauner, A., Hitchen, P. G., Morris, H. R., Dell, A. & Eichler, J. (2008b). Identification of AglE, a second glycosyltransferase involved in N-glycosylation of the *Haloferax volcanii* S-layer glycoprotein. *J Bacteriol* **190**, 3140-3146.

Adams, P. D., Afonine, P. V., Grosse-Kunstleve, R. W., Read, R. J., Richardson, J. S., Richardson, D. C. & Terwilliger, T. C. (2009). Recent developments in phasing and structure refinement for macromolecular crystallography. *Curr Opin Struct Biol* **19**, 566-572.

Albers, S. V. & Driessen, A. J. M. (2002). Signal peptides of secreted proteins of the archaeon *Sulfolobus solfataricus*: a genomic survey. *Arch Microbiol* **177**, 209-216.

Altmann, F., Schweiszer, S. & Weber, C. (1995). Kinetic comparison of Peptide N-Glycosidase F and N-Glycosidase A reveals several differences in substrate specificity. *Glycoconjugate J* **12**, 84-93.

Altmann, F., Paschinger, K., Dalik, T. & Vorauer, K. (1998). Characterisation of peptide-N-4-(N-acetyl-beta-glucosaminyl)asparagine amidase A and its N-glycans. *Eur J Biochem* **252**, 118-123.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic Local Alignment Search Tool. *J Mol Biol* **215**, 403-410.

Apweiler, R., Hermjakob, H. & Sharon, N. (1999). On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim Biophys Acta-Gen Subj* **1473**, 4-8.

Baker, E. N. & Hubbard, R. E. (1984). Hydrogen bonding in globular proteins. *Prog Biophys Mol Bio* **44**, 97-179.

Banerjee, A., Wang, R., Supernavage, S. L., Ghosh, S. K., Parker, J., Ganesh, N. F., Wang, P. G., Gulati, S. & Rice, P. A. (2002). Implications of phase variation of a gene (*pgtA*) encoding a pilin galactosyl transferase in gonococcal pathogenesis. *J Exp Med* **196**, 147-162.

Baneyx, F. & Mujacic, M. (2004). Recombinant protein folding and misfolding in *Escherichia coli*. *Nat Biotechnol* **22**, 1399-1408.

Barsomian, G. D., Johnson, T. L., Borowski, M., Denman, J., Ollington, J. F., Hirani, S., McNeilly, D. S. & Rasmussen, J. R. (1990). Cloning and expression of Peptide-N-4-(N-acetyl-beta-D-glucosaminyl)asparagine amidase F in *Escherichia coli*. *J Biol Chem* **265**, 6967-6972.

Baumeister, W. & Pouch, M. N. (1998). Proteasome and protein degradation. *Biofutur*, 62-66.

Bendtsen, J. D., Nielsen, H., von Heijne, G. & Brunak, S. (2004). Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* **340**, 783-795.

Bennett-Lovsey, R. M., Herbert, A. D., Sternberg, M. J. E. & Kelley, L. A. (2008). Exploring the extremes of sequence/structure space with ensemble fold recognition in the program Phyre. *Proteins* **70**, 611-625.

Berger, S., Menudier, A., Julien, R. & Karamanos, Y. (1995). Endo-N-acetyl-beta-D-glucosaminidase and Peptide-N-4-(N-acetyl-glucosaminyl)asparagine amidase activities during germination of *Raphanus sativus*. *Phytochemistry* **39**, 481-487.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Res* **28**, 235-242.

Birnboim, H. C. & Doly, J. (1979). Rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res* **7**, 1513-1523.

Boraston, A. B., McLean, B. W., Kormos, J. M., Alam, M., Gilkes, N. R., Haynes, C. A., Tomme, P., Kilburn, D. G. & Warren, R. A. J. (1999). Carbohydrate-binding modules: diversity of structure and function. *Roy Soc Ch*, 202-211.

Boraston, A. B., Bolam, D. N., Gilbert, H. J. & Davies, G. J. (2004). Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem J* **382**, 769-781.

Bradford, M. M. (1976). Rapid and sensitive method for quantitation of microgram quantities of protein utilising principle of protein-dye binding. *Anal Biochem* **72**, 248-254.

Brooks, B. W. & Murray, R. G. E. (1981). Nomenclature for *Micrococcus radiodurans* and other radiation-resistant cocci - *Deinococcacae* Fam Nov and *Deinococcus* Gen-Nov, Including 5 Species. *Int J Syst Bacteriol* **31**, 353-360.

Brunger, A. T. (1992). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472-475.

Brunger, A. T. (1997). Free R value: Cross-validation in crystallography. In *Macromolecular Crystallography, Pt B*, pp. 366-396.

Bugg, T. D. H. & Brandish, P. E. (1994). From peptidoglycan to glycoproteins - common features of lipid-linked oligosaccharide biosynthesis. *Fems Microbiol Lett* **119**, 255-262.

Burda, P. & Aebi, M. (1999). The dolichol pathway of N-linked glycosylation. *Biochim Biophys Acta-Gen Subj* **1426**, 239-257.

Burg, R. W., Miller, B. M., Baker, E. E. & other authors (1979). Avermectins, new family of potent anthelmintic agents - producing organism and fermentation. *Antimicrob Agents Ch* **15**, 361-367.

Bussink, H. J. D., Buxton, F. P. & Visser, J. (1991). Expression and sequence comparison of the *Aspergillus niger* and *Aspergillus tubigensis* genes encoding Polygalacturonase-II. *Curr Genet* **19**, 467-474.

Chaban, B., Voisin, S., Kelly, J., Logan, S. M. & Jarrell, K. F. (2006). Identification of genes involved in the biosynthesis and attachment of *Methanococcus voltae* N-linked glycans: insight into N-linked glycosylation pathways in Archaea. *Mol Microbiol* **61**, 259-268.

Chang, T., Kuo, M. C., Khoo, K. H., Inoue, S. & Inoue, Y. (2000). Developmentally regulated expression of a peptide : N-Glycanase during germination of rice seeds (*Oryza sativa*) and its purification and characterization. *J Biol Chem* **275**, 129-134.

Cheng, C. H. & Shuman, S. (2000). Recombinogenic flap ligation pathway for intrinsic repair of topoisomerase IB-induced double-strand breaks. *Mol Cell Biol* **20**, 8059-8068.

Chu, F. K. (1986). Requirements of cleavage of high mannose oligosaccharides in glycoproteins by Peptide N-Glycosidase F. *J Biol Chem* **261**, 172-177.

Coates, L., Erskine, P. T., Mall, S., Gill, R., Wood, S. P., Myles, D. A. A. & Cooper, J. B. (2006). X-ray, neutron and NMR studies of the catalytic mechanism of aspartic proteinases. *Eur Biophys J Biophys Lett* **35**, 559-566.

Coates, L., Tuan, H. F., Tomanicek, S., Kovalevsky, A., Mustyakimov, M., Erskine, P. & Cooper, J. (2008). The catalytic mechanism of an aspartic proteinase explored with neutron and X-ray diffraction. *J Am Chem Soc* **130**, 7235-7237.

Cole, C., Barber, J. D. & Barton, G. J. (2008). The Jpred 3 secondary structure prediction server. *Nucleic Acids Res* **36**, W197-W201.

Compton, L. A. & Johnson, W. C. (1986). Analysis of protein circular dichroism spectra for secondary structure using a simple matrix multiplication. *Anal Biochem* **155**, 155-167.

Copeland, R. A. (2000). *Enzymes: a practical introduction to structure, mechanism, and data analysis, 2nd edition*: John Wiley & Sons Inc.

Cserzo, M., Wallin, E., Simon, I., vonHeijne, G. & Elofsson, A. (1997). Prediction of transmembrane alpha-helices in prokaryotic membrane proteins: the dense alignment surface method. *Protein Eng* **10**, 673-676.

Davis, I. W., Leaver-Fay, A., Chen, V. B. & other authors (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* **35**, W375-W383.

de Beer, T., Vliegenthart, J. F. G., Loffler, A. & Hofsteenge, J. (1995). The hexopyranosyl residue that is C-glycosidically linked to the side chain of tryptophan-7 in human Rnase U-S Is alpha-marmopyranose. *Biochemistry-US* **34**, 11785-11789.

de Graaff, L. H., Vandenbroeck, H. C., Vanooijen, A. J. J. & Visser, J. (1994). Regulation of the Xylanase-encoding *xlnA* gene of *Aspergillus tubigensis*. *Mol Microbiol* **12**, 479-490.

de Peredo, A. G., Klein, D., Macek, B., Hess, D., Peter-Katalinic, J. & Hofsteenge, J. (2002). C-mannosylation and O-fucosylation of thrombospondin type 1 repeats. *Mol Cell Proteomics* **1**, 11-18.

DeLano, W. L. (2002). The PyMOL Molecular Graphics System. *DeLano Scientific, San Carlos, USA* http://wwwpymolorg.

Della Mea, M., Caparros-Ruiz, D., Claparols, I., Serafini-Fracassini, D. & Rigau, J. (2004). AtPng1p. The first plant transglutaminase. *Plant Physiol* **135**, 2046-2054.

Demain, A. L. (1999). Pharmaceutically active secondary metabolites of microorganisms. *Appl Microbiol Biotechnol* **52**, 455-463.

Dempski, R. E. & Imperiali, B. (2002). Oligosaccharyl transferase: gatekeeper to the secretory pathway. *Curr Opin Chem Biol* **6**, 844-850.

Deras, I. L., Takegawa, K., Kondo, A., Kato, I. & Lee, Y. C. (1998). Synthesis of a high-mannose-type glycopeptide analog containing a glucose-asparagine linkage. *Bioorgan Med Chem* **8**, 1763-1766.

Diepold, A., Li, G., Lennarz, W. J., Nurnberger, T. & Brunner, F. (2007). The Arabidopsis AtPNG1 gene encodes a peptide: N-glycanase. *Plant J* **52**, 94-104.

Doucey, M. A., Hess, D., Cacan, R. & Hofsteenge, J. (1998). Protein C-mannosylation is enzyme-catalysed and uses dolichyl-phosahate-mannose as a precursor. *Mol Biol Cell* **9**, 291-300.

Doucey, M. A., Hess, D., Blommers, M. J. J. & Hofsteenge, J. (1999). Recombinant human interleukin-12 is the second example of a C-mannosylated protein. *Glycobiology* **9**, 435-441.

Dubois, M., Gilles, K. A., Hamilton, J. K., Rebers, P. A. & Smith, F. (1956). Colorimetric method for determination of sugars and related substances. *Anal Chem* **28**, 350-356.

Eichler, J. & Adams, M. W. W. (2005). Posttranslational protein modification in Archaea. *Microbiol Mol Biol R* **69**, 393-425.

Eisenhaber, B., Bork, P. & Eisenhaber, F. (1998). Sequence properties of GPI-anchored proteins near the omega-site: constraints for the polypeptide binding site of the putative transamidase. *Protein Eng* **11**, 1155-1161.

Eisenthal, R., Danson, M. J. & Hough, D. W. (2007). Catalytic efficiency and k_{cat}/K_M: a useful comparator? *Trends Biotechnol* **25**, 247-249.

Elder, J. H. & Alexander, S. (1982). Endo-Beta-N-Acetylglucosaminidase-F - Endoglycosidase from *Flavobacterium meningosepticum* that cleaves both high-mannose and complex glycoproteins. *PNAS* **79**, 4540-4544.

Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr D* **60**, 2126-2132.

Evans, P. (2006). Scaling and assessment of data quality. *Acta Crystallogr D* **62**, 72-82.

Fan, J.-Q. & Lee, Y. C. (1997). Detailed studies on substrate structure requirements of glycoamidases A and F. *J Biol Chem* **272**, 27058-27064.

Farrell, P. & Iatrou, K. (2004). Transfected insect cells in suspension culture rapidly yield moderate quantities of recombinant proteins in protein-free culture medium. *Prot Expres Purif* **36**, 177-185.

Faye, L., Johnson, K. D., Sturm, A. & Chrispeels, M. J. (1989). Structure, biosynthesis and function of asparagine-linked glycans on plant glycoproteins. *Physiol Plantarum* **75**, 309-314.

Fernandez, M. D., Canada, F. J., Jimenez-Barbero, J. & Cuevas, G. (2005). Molecular recognition of saccharides by proteins. Insights on the origin of the carbohydrate-aromatic interactions. *J Am Chem Soc* **127**, 7379-7386.

Ferro, V., Weiler, L., Withers, S. G. & Ziltener, H. (1998). N-Glycosyl phosphonamidates: potential transition-state analogue inhibitors of glycopeptidases. *Can J Chem-Rev Can Chim* **76**, 313-318.

Freeze, H. H. & Westphal, V. (2001). Balancing N-linked glycosylation to avoid disease. *Biochimie* **83**, 791-799.

French, S. & Wilson, K. (1978). Treatment of negative intensity observations. *Acta Crystallogr A* **34**, 517-525.

Ftouhi-Paquin, N., Hauer, C. R., Stack, R. F., Tarentino, A. L. & Plummer, T. H., Jr. (1997). Molecular cloning, primary structure, and properties of a new glycoamidase from the fungus *Aspergillus tubigensis*. *J Biol Chem* **272**, 22960-22965.

Ftouhi Paquin, N., Tarentino, A. L. & Plummer, T. H., Jr. (1998). Overexpression of PNGase at from baculovirus-infected insect cells. *Protein Expr Purif* **14**, 302-308.

Fu, Z. B., Ng, K. L., Lam, T. L. & Wong, W. K. R. (2005). Cell death caused by hyper-expression of a secretory exoglucanase in *Escherichia coli*. *Prot Expres Purif* **42**, 67-77.

Gasser, B., Saloheimo, M., Rinas, U. & other authors (2008). Protein folding and conformational stress in microbial cells producing recombinant proteins: a host comparative overview. *Microb Cell Fact* **7**, 18.

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D. & Bairoch, A. (2003). ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res* **31**, 3784-3788.

Gherardini, P. F., Wass, M. N., Helmer-Citterich, M. & Sternberg, M. J. E. (2007). Convergent evolution of enzyme active sites is not a rare phenomenon. *J Mol Biol* **372**, 817-845.

Gielkens, M. M. C., Visser, J. & deGraaff, L. H. (1997). Arabinoxylan degradation by fungi: Characterization of the arabinoxylan-arabinofuranohydrolase encoding genes from *Aspergillus niger* and *Aspergillus tubingensis*. *Curr Genet* **31**, 22-29.

Glockner, F. O., Fuchs, B. M. & Amann, R. (1999). Bacterioplankton compositions of lakes and oceans: a first comparison based on fluorescence in situ hybridization. *Appl Environ Microbiol* **65**, 3721-3726.

Greenfield, N. J. (2006). Using circular dichroism spectra to estimate protein secondary structure. *Nat Protoc* **1**, 2876-2890.

Gregoret, L. M., Rader, S. D., Fletterick, R. J. & Cohen, F. E. (1991). Hydrogen bonds involving sulfur-atoms in proteins. *Proteins* **9**, 99-107.

Gupta, R., Jung, E. & Brunak, S. (2004). Prediction of N-glycosylation sites in human proteins. *In preparation*.

Gustafson, G. L. & Milner, L. A. (1980). Occurrence of N-acetylglucosamine-1-phosphate in proteinase-I from *Dictyostelium discoideum*. *J Biol Chem* **255**, 7208-7210.

Hall, B. G., Yokoyama, S. & Calhoun, D. H. (1983). Role of cryptic genes in microbial evolution. *Mol Biol Evol* **1**, 109-124.

Hammond, C. & Helenius, A. (1994). Folding of Vsv G-protein - sequential interaction with Bip and calnexin. *Science* **266**, 456-458.

Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. *J Mol Biol* **166**, 557-580.

Haneda, K., Inazu, T., Mizuno, M. & other authors (2001). Chemo-enzymatic synthesis of a bioactive peptide containing a glutamine-linked oligosaccharide and its characterization. *Biochim Biophys Acta-Gen Subj* **1526**, 242-248.

Harrison, P. M. & Gerstein, M. (2002). Studying genomes through the aeons: Protein families, pseudogenes and proteome evolution. *J Mol Biol* **318**, 1155-1174.

Hart, G. W. (1997). Dynamic O-linked glycosylation of nuclear and cytoskeletal proteins. *Annu Rev Biochem* **66**, 315-335.

Hartley, J. L., Temple, G. F. & Brasch, M. A. (2000). DNA cloning using *in vitro* site-specific recombination. *Genome Res* **10**, 1788-1795.

Hartmann, S. & Hofsteenge, J. (2000). Properdin, the positive regulator of complement, is highly C-mannosylated. *J Biol Chem* **275**, 28569-28574.

Hashimoto, H. (2006). Recent structural studies of carbohydrate-binding modules. *Cell Mol Life Sci* **63**, 2954-2967.

Haynes, P. A. (1998). Phosphoglycosylation: a new structural class of glycosylation? *Glycobiology* **8**, 1-5.

Helenius, A. & Aebi, M. (2001). Intracellular functions of N-linked glycans. *Science* **291**, 2364-2369.

Helenius, A. & Aebi, M. (2004). Roles of N-linked glycans in the endoplasmic reticulum. *Annu Rev Biochem* **73**, 1019-1049.

Helenius, J. & Aebi, M. (2002). Transmembrane movement of dolichol linked carbohydrates during N-glycoprotein biosynthesis in the endoplasmic reticulum. *Semin Cell Dev Biol* **13**, 171-178.

Hentz, N. G., Richardson, J. M., Sportsman, J. R., Daijo, J. & Sittampalam, G. S. (1997). Synthesis and characterization of insulin-fluorescein derivatives for bioanalytical applications. *Anal Chem* **69**, 4994-5000.

Hirsch, C., Blom, D. & Ploegh, H. L. (2003). A role for N-glycanase in the cytosolic turnover of glycoproteins. *Embo J* **22**, 1036-1046.

Hirsch, C., Misaghi, S., Blom, D., Pacold, M. E. & Ploegh, H. L. (2004). Yeast N-glycanase distinguishes between native and non-native glycoproteins. *Embo Rep* **5**, 201-206.

Hofsteenge, J., Blommers, M., Hess, D., Furmanek, A. & Miroshnichenko, O. (1999). The four terminal components of the complement system are C-mannosylated on multiple tryptophan residues. *J Biol Chem* **274**, 32786-32794.

Hofsteenge, J., Huwiler, K. G., Macek, B., Hess, D., Lawler, J., Mosher, D. F. & Peter-Katalinic, J. (2001). C-mannosylation and O-fucosylation of the thrombospondin type 1 module. *J Biol Chem* **276**, 6485-6498.

Holliday, G. L., Bartlett, G. J., Almonacid, D. E., O'Boyle, N. M., Murray-Rust, P., Thornton, J. M. & Mitchell, J. B. O. (2005). MACiE: a database of enzyme reaction mechanisms. *Bioinformatics* **21**, 4315-4316.

Holliday, G. L., Mitchell, J. B. O. & Thornton, J. M. (2009). Understanding the functional roles of amino acid residues in enzyme catalysis. *J Mol Biol* **390**, 560-577.

Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H., Shiba, T., Sakaki, Y., Hattori, M. & Omura, S. (2003). Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. *Nat Biotechnol* **21**, 526-531.

Ilg, T., Overath, P., Ferguson, M. A. J., Rutherford, T., Campbell, D. G. & McConville, M. J. (1994a). O-glycosylation and N-glycosylation of the *Leishmania mexicana* secreted Acid-phosphatase - Characterization of a new class of phosphoserine-linked glycans. *J Biol Chem* **269**, 24073-24081.

Ilg, T., Stierhof, Y. D., Wiese, M., McConville, M. J. & Overath, P. (1994b). Characterization of phosphoglycan-containing secretory products of *Leishmania*. *Parasitology* **108**, S63-S71.

Ilg, T. (2000). Proteophosphoglycans of *Leishmania*. *Parasitology Today* **16**, 489-497.

Imperiali, B. & O'Connor, S. E. (1999). Effect of N-linked glycosylation on glycopeptide and glycoprotein structure. *Curr Opin Chem Biol* **3**, 643-649.

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

Johansen, P. G., Neuberger, A. & Marshall, R. D. (1961). Carbohydrates in protein .3. Preparation and some of properties of a glycopeptide from hens-egg albumin. *Biochem J* **78**, 518-&.

Joshi, S., Katiyar, S. & Lennarz, W. J. (2005). Misfolding of glycoproteins is a prerequisite for peptide: N-glycanase mediated deglycosylation. *FEBS Lett* **579**, 823-826.

Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J Appl Crystallogr* **26**, 795-800.

Kantardjieff, K. A. & Rupp, B. (2003). Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. *Protein Sci* **12**, 1865-1871.

Katiyar, S., Suzuki, T., Balgobin, B. J. & Lennarz, W. J. (2002). Site-directed mutagenesis study of yeast Peptide: N-Glycanase. Insight into the reaction mechanism of deglycosylation. *J Biol Chem* **277**, 12953-12959.

Kato, T., Kawahara, A., Ashida, H. & Yamamoto, K. (2007). Unique peptide : N-glycanase of *Caenorhabditis elegans* has activity of protein disulphide reductase as well as of deglycosylation. *J Biochem* **142**, 175-181.

Keegan, R. M. & Winn, M. D. (2007). Automated search-model discovery and preparation for structure solution by molecular replacement. *Acta Crystallogr D* **63**, 447-457.

Kelley, L. A. & Sternberg, M. J. E. (2009). Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc* **4**, 363-371.

Kelly, J., Jarrell, H., Millar, L., Tessier, L., Fiori, L. M., Lau, P. C., Allan, B. & Szymanski, C. M. (2006). Biosynthesis of the N-linked glycan in *Campylobacter jejuni* and addition onto protein through Block transfer. *J Bacteriol* **188**, 2427-2434.

Kelly, S. M., Jess, T. J. & Price, N. C. (2005). How to study proteins by circular dichroism. *BBA-Proteins Proteom* **1751**, 119-139.

Kim, S. B. & Goodfellow, M. (2002). *Streptomyces avermitilis* sp nov., nom. rev., a taxonomic home for the avermectin-producing streptomycetes. *Int J Syst Evol Micr* **52**, 2011-2014.

Kimura, Y. & Ohno, A. (1998). A new peptide-N-4-(acetyl-beta-glucosaminyl)asparagine amidase from soybean (*Glycine max*) seeds: Purification and substrate specificity. *Biosci Biotech Bioch* **62**, 412-418.

Kishimoto, N., Kosako, Y. & Tano, T. (1991). *Acidobacterium Capsulatum* Gen-Nov, Sp-Nov - An acidophilic chemoorganotrophic bacterium containing menaquinone from acidic mineral environment. *Curr Microbiol* **22**, 1-7.

Kitajima, K., Suzuki, T., Kouchi, Z., Inoue, S. & Inoue, Y. (1995). Identification and distribution of Peptide - N-Glycanase (Pngase) in mouse organs. *Arch Biochem Biophys* **319**, 393-401.

Kobata, A. (2000). A journey to the world of glycobiology. *Glycoconjugate J* **17**, 443-464.

Kobayashi, T., Nishizaki, R. & Ikezawa, H. (1997). The presence of GPI-linked protein(s) in an archaeobacterium, *Sulfolobus acidocaldarius*, closely related to eukaryotes. *Biochim Biophys Acta-Gen Subj* **1334**, 1-4.

Kornfeld, R. & Kornfeld, S. (1985). Assembly of asparagine-linked oligosaccharides. *Annu Rev Biochem* **54**, 631-664.

Kowarik, M., Numao, S., Feldman, M. F., Schulz, B. L., Callewaert, N., Kiermaier, E., Catrein, I. & Aebi, M. (2006a). N-linked glycosylation of folded proteins by the bacterial oligosaccharyltransferase. *Science* **314**, 1148-1150.

Kowarik, M., Young, N. M., Numao, S. & other authors (2006b). Definition of the bacterial N-glycosylation site consensus sequence. *Embo J* **25**, 1957-1966.

Kraut, J. (1977). Serine Proteases - Structure and mechanism of catalysis. *Annu Rev Biochem* **46**, 331-358.

Krieg, J., Hartmann, S., Vicentini, A., Glasner, W., Hess, D. & Hofsteenge, J. (1998). Recognition signal for C-mannosylation of Trp-7 in RNase 2 consists of sequence Trp-x-x-Trp. *Mol Biol Cell* **9**, 301-309.

Krissinel, E. & Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D* **60**, 2256-2268.

Krittanai, C. & Johnson, W. C. (1997). Correcting the circular dichroism spectra of peptides for contributions of absorbing side chains. *Anal Biochem* **253**, 57-64.

Kubelka, V., Altmann, F., Kornfeld, G. & Marz, L. (1994). Structures of the N-linked oligosaccharides of the membrane-glycoproteins from 3 lepidopteran cell-lines (Sf-21, Izd-Mb-0503, Bm-N). *Arch Biochem Biophys* **308**, 148-157.

Kuhn, P., Tarentino, A. L., Plummer, T. H., Jr. & Van Roey, P. (1994). Crystal structure of peptide-N4-(N-acetyl-beta-D-glucosaminyl)asparagine amidase F at 2.2-A resolution. *Biochemistry-US* **33**, 11699-11706.

Kuhn, P., Guan, C., Cui, T., Tarentino, A. L., Plummer, T. H., Jr. & Van Roey, P. (1995). Active site and oligosaccharide recognition residues of peptide-N4-(N-acetyl-beta-D-glucosaminyl)asparagine amidase F. *J Biol Chem* **270**, 29493-29497.

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

Landy, A. (1989). Dynamic, structural and regulatory aspects of Lambda-site-specific recombination. *Annu Rev Biochem* **58**, 913-949.

Lang, L., Couso, R. & Kornfeld, S. (1986). Glycoprotein phosphorylation in simple eukaryotic organisms - Identification of Udp-Glcnac-glycoprotein N-acetylglucosamine-1-phosphotransferase activity and analysis of substrate specificity. *J Biol Chem* **261**, 6320-6325.

Larkin, M. A., Blackshields, G., Brown, N. P. & other authors (2007). Clustal W and clustal X version 2.0. *Bioinformatics* **23**, 2947-2948.

Lee, J. H., Choi, J. M., Lee, C. W., Yi, K. J. & Cho, Y. J. (2005). Structure of a peptide : N-glycanase-Rad23 complex: Insight into the deglycosylation for denatured glycoproteins. *PNAS* **102**, 9144-9149.

Lee, S. G., Pancholi, V. & Fischetti, V. A. (2002). Characterization of a unique glycosylated anchor endopeptidase that cleaves the LPXTG sequence motif of cell surface proteins of gram-positive bacteria. *J Biol Chem* **277**, 46912-46922.

Lemp, D., Haselbeck, A. & Klebl, F. (1990). Molecular cloning and heterologous expression of N-Glycosidase F from *Flavobacterium meningosepticum*. *J Biol Chem* **265**, 15606-15610.

Lenz, D. H. (2003). *N*-linked glycopeptide mimetics as tools in kinetic, mechanistic and structural studies of Peptide-*N*:Glycanase F. In *Institute of Molecular BioSciences*. Palmerston North: Massey University. PhD Thesis.

Lhernould, S., Karamanos, Y., Bourgerie, S., Strecker, G., Julien, R. & Morvan, H. (1992). Peptide-N(4)-(N-acetylglucosaminyl)asparagine amidase (Pngase) activity could explain the occurrence of extracellular xylomannosides in a plant-cell suspension. *Glycoconjugate J* **9**, 191-197.

Li, G., Zhou, X., Zhao, G., Schindelin, H. & Lennarz, W. J. (2005). Multiple modes of interaction of the deglycosylation enzyme, mouse peptide N-glycanase, with the proteasome. *PNAS* **102**, 15809-15814.

Li, G., Zhao, G., Zhou, X., Schindelin, H. & Lennarz, W. J. (2006). The AAA ATPase p97 links peptide N-glycanase to the endoplasmic reticulum-associated E3 ligase autocrine motility factor receptor. *PNAS* **103**, 8348-8353.

Li, X., Romero, P., Rani, M., Dunker, A. K. & Obradovic, Z. (1999). Predicting protein disorder for N-, C-, and integral regions. *Genome Inform* **10**, 30-40.

Lilley, B. N. & Ploegh, H. L. (2004). A membrane protein required for dislocation of misfolded proteins from the ER. *Nature* **429**, 834-840.

Linton, D., Dorrell, N., Hitchen, P. G. & other authors (2005). Functional analysis of the *Campylobacter jejuni* N-linked protein glycosylation pathway. *Mol Microbiol* **55**, 1695-1703.

Loo, T., Patchett, M. L., Norris, G. E. & Lott, J. S. (2002). Using secretion to solve a solubility problem: high-yield expression in *Escherichia coli* and purification of the bacterial glycoamidase PNGase F. *Protein Expr Purif* **24**, 90-98.

Loo, T. S. (2000). Expression, purification and characterisation of recombinant Peptide:*N*-Glycosidase F. *MSc thesis, Massey University*.

Ludwig, W., Bauer, S. H., Bauer, M. & other authors (1997). Detection and in situ identification of representatives of a widely distributed new bacterial phylum. *Fems Microbiol Lett* **153**, 181-190.

Makarova, K. S., Aravind, L. & Koonin, E. V. (1999). A superfamily of archaeal, bacterial, and eukaryotic proteins homologous to animal transglutaminases. *Protein Sci* **8**, 1714-1719.

Maley, F., Trimble, R. B., Tarentino, A. L. & Plummer, T. H. (1989). Characterization of glycoproteins and their associated oligosaccharides through the use of endoglycosidases. *Anal Biochem* **180**, 195-204.

Marchler-Bauer, A., Anderson, J. B., Chitsaz, F. & other authors (2009). CDD: specific functional annotation with the Conserved Domain Database. *Nucleic Acids Res* **37**, D205-D210.

Matsudaira, P. (1987). Sequence from picomole quantities of proteins electroblotted onto polyvinylidene difluoride membranes. *J Biol Chem* **262**, 10035-10038.

Matthews, B. W. (1968). Solvent content of protein crystals. *J Mol Biol* **33**, 491-497.

Matthews, D. A., Alden, R. A., Birktoft, J. J., Freer, S. T. & Kraut, J. (1977). Re-examination of charge relay system in subtilisin and comparison with other serine proteases. *J Biol Chem* **252**, 8875-8883.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674.

McDonald, I. K. & Thornton, J. M. (1994). Satisfying hydrogen-bonding potential in proteins. *J Mol Biol* **238**, 777-793.

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

Megnegneau, B., Debets, F. & Hoekstra, R. F. (1993). Genetic variability and relatedness in the complex group of black Aspergilli based on random amplification of polymorphic DNA. *Curr Genet* **23**, 323-329.

Merello, S., Parodi, A. J. & Couso, R. (1995). Characterization and partial purification of a novel enzymatic activity - Udp-GlcNAc-Ser-Protein N-acetylglucosamine-1-phosphotransferase from the cellular slime mold *Dictyostelium discoideum*. *J Biol Chem* **270**, 7281-7287.

Mescher, M. F. & Strominger, J. L. (1976). Purification and characterization of a prokaryotic glycoprotein from cell-envelope of *Halobacterium salinarium*. *J Biol Chem* **251**, 2005-2014.

Messner, P. (1997). Bacterial glycoproteins. *Glycoconjugate J* **14**, 3-11.

Morris, A. L., Macarthur, M. W., Hutchinson, E. G. & Thornton, J. M. (1992). Stereochemical quality of protein structure coordinates. *Proteins* **12**, 345-364.

Moyer, T. R. & Hunnicutt, D. W. (2007). Susceptibility of zebra fish *Danio rerio* to infection by *Flavobacterium columnare* and *F. johnsoniae*. *Dis Aquat Org* **76**, 39-44.

Munte, C. E., Gade, G., Domogalla, B., Kremer, W., Kellner, R. & Kalbitzer, H. R. (2008). C-mannosylation in the hypertrehalosaemic hormone from the stick insect *Carausius morosus*. *Febs J* **275**, 1163-1173.

Muramats, T. (1971). Demonstration of an Endo-glycosidase acting on a glycoprotein. *J Biol Chem* **246**, 5534-&.

Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D* **53**, 240-255.

Murzin, A. G., Brenner, S. E., Hubbard, T. & Chothia, C. (1995). SCOP - a Structural Classification of Proteins database for the investigation of sequences and structures. *J Mol Biol* **247**, 536-540.

Nakamoto, H. & Bardwell, J. C. A. (2004). Catalysis of disulfide bond formation and isomenization in the *Escherichia coli* periplasm. *Biochim Biophys Acta-Mol Cell Res* **1694**, 111-119.

Nauseef, W. M., McCormick, S. J. & Clark, R. A. (1995). Calreticulin functions as a molecular chaperone in the biosynthesis of myeloperoxidase. *J Biol Chem* **270**, 4741-4747.

Navaza, J. (1994). AMORE - an automated package for molecular replacement. *Acta Crystallogr A* **50**, 157-163.

Nematollahi, A., Decostere, A., Pasmans, F. & Haesebrouck, F. (2003). *Flavobacterium psychrophilum* infections in salmonid fish. *J Fish Dis* **26**, 563-574.

Norris, G. E., Flaus, A. J., Moore, C. H. & Baker, E. N. (1994a). Purification and crystallization of the endoglycosidase PNGase F, a peptide:N-glycosidase from *Flavobacterium meningosepticum*. *J Mol Biol* **241**, 624-626.

Norris, G. E., Stillman, T. J., Anderson, B. F. & Baker, E. N. (1994b). The three-dimensional structure of PNGase F, a glycosylasparaginase from *Flavobacterium meningosepticum*. *Structure* **2**, 1049-1059.

Olden, K., Parent, J. B. & White, S. L. (1982). Carbohydrate moieties of glycoproteins - a re-evaluation of their function. *Biochem Biophys Acta* **650**, 209-232.

Ord, T., Adessi, C., Wang, L. & Freeze, H. H. (1996). Two cysteine proteinase genes *cprF* and *cprG* from *Dictyostelium discoldeum* contain unusual serine-rich domains where GlcNAc-1-P residues are added. *Glycobiology* **6**, 313-313.

Orlean, P. & Menon, A. K. (2007). GPI anchoring of protein in yeast and mammalian cells, or: how we learned to stop worrying and love glycophospholipids. *J Lipid Res* **48**, 993-1011.

Parenicova, L., Benen, J. A. E., Samson, R. A. & Visser, J. (1997). Evaluation of RFLP analysis of the classification of selected black aspergilli. *Mycol Res* **101**, 810-814.

Parenicova, L., Skouboe, P., Frisvad, J., Samson, R. A., Rossen, L., ten Hoor-Suykerbuyk, M. & Visser, J. (2001). Combined molecular and biochemical approach identifies *Aspergillus japonicus* and *Aspergillus aculeatus* as two species. *Appl Environ Microbiol* **67**, 521-527.

Park, H., Suzuki, T. & Lennarz, W. J. (2001). Identification of proteins that interact with mammalian peptide : N-glycanase and implicate this hydrolase in the proteasome-dependent pathway for protein degradation. *PNAS* **98**, 11163-11168.

Patanjali, S. R., Swamy, M. J., Anantharam, V., Khan, M. I. & Surolia, A. (1984). Chemical modification studies on abrus agglutinin - Involvement of tryptophan residues in sugar binding. *Biochem J* **217**, 773-781.

Pel, H. J., de Winde, J. H., Archer, D. B. & other authors (2007). Genome sequencing and analysis of the versatile cell factory *Aspergillus niger* CBS 513.88. *Nat Biotechnol* **25**, 221-231.

Plummer, T., Jr & Tarentino, A. (1981). Facile cleavage of complex oligosaccharides from glycopeptides by almond emulsin peptide: N-glycosidase. *J Biol Chem* **256**, 10243-10246.

Plummer, T. H., Jr., Elder, J. H., Alexander, S., Phelan, A. W. & Tarentino, A. L. (1984). Demonstration of peptide:N-glycosidase F activity in endo-beta-N-acetylglucosaminidase F preparations. *J Biol Chem* **259**, 10700-10704.

Plummer, T. H., Jr., Phelan, A. W. & Tarentino, A. L. (1987). Detection and quantification of peptide-N4-(N-acetyl-beta-glucosaminyl)asparagine amidases. *Eur J Biochem* **163**, 167-173.

Pollastri, G., Przybylski, D., Rost, B. & Baldi, P. (2002). Improving the prediction of protein secondary structure in three and eight classes using recurrent neural networks and profiles. *Proteins* **47**, 228-235.

Porter, C. T., Bartlett, G. J. & Thornton, J. M. (2004). The Catalytic Site Atlas: a resource of catalytic sites and residues identified in enzymes using structural data. *Nucleic Acids Res* **32**, D129-D133.

Potterton, E., Briggs, P., Turkenburg, M. & Dodson, E. (2003). A graphical user interface to the CCP4 program suite. *Acta Crystallogr D* **59**, 1131-1137.

Provencher, S. W. & Glockner, J. (1981). Estimation of globular protein secondary structure from circular dichroism. *Biochemistry-US* **20**, 33-37.

Quaiser, A., Ochsenreiter, T., Lanz, C., Schuster, S. C., Treusch, A. H., Eck, J. & Schleper, C. (2003). *Acidobacteria* form a coherent but highly diverse group within the bacterial domain: evidence from environmental genomics. *Mol Microbiol* **50**, 563-575.

Risley, J. M. & Vanetten, R. L. (1985). H-1-NMR evidence that almond Peptide-N-glycosidase is an amidase - Kinetic data and trapping of the intermediate. *J Biol Chem* **260**, 5488-5492.

Romero, P., Obradovic, Z. & Dunker, A. K. (1997). Sequence data analysis for long disordered regions prediction in the calcineurin family. *Genome Inform* **8**, 110-124.

Rossmann, M. G., Abadzapatero, C., Murthy, M. R. N., Liljas, L., Jones, T. A. & Strandberg, B. (1983). Structural comparisons of some small spherical plant viruses. *J Mol Biol* **165**, 711-736.

Sahalan, A. Z. & Dixon, R. A. (2008). Role of the cell envelope in the antibacterial activities of polymyxin B and polymyxin B nonapeptide against *Escherichia coli*. *Int J Antimicrob Agents* **31**, 224-227.

Schachter, H. (2001). Congenital disorders involving defective N-glycosylation of proteins. *Cell Mol Life Sci* **58**, 1085-1104.

Schlapschy, M., Grimm, S. & Skerra, A. (2006). A system for concomitant overexpression of four periplasmic folding catalysts to improve secretory protein production in *Escherichia coli*. *Protein Eng Des Sel* **19**, 385-390.

Schlippe, Y. V. G. & Hedstrom, L. (2005). A twisted base? The role of arginine in enzyme-catalyzed proton abstractions. *Arch Biochem Biophys* **433**, 266-278.

Schmidt, M. A., Riley, L. W. & Benz, I. (2003). Sweet new world: glycoproteins in bacterial pathogens. *Trends Microbiol* **11**, 554-561.

Schuster, E., Dunn-Coleman, N., Frisvad, J. C. & van Dijck, P. W. M. (2002). On the safety of *Aspergillus niger* - a review. *Appl Microbiol Biotechnol* **59**, 426-435.

Seko, A., Kitajima, K., Inoue, Y. & Inoue, S. (1991). Peptide:N-glycosidase activity found in the early embryos of *Oryzias latipes* (Medaka fish). The first demonstration of the occurrence of peptide:N-glycosidase in animal cells and its implication for the presence of a de-N-glycosylation system in living organisms. *J Biol Chem* **266**, 22110-22114.

Seko, A., Kitajima, K., Iwamatsu, T., Inoue, Y. & Inoue, S. (1999). Identification of two discrete peptide: N-glycanases in *Oryzias latipes* during embryogenesis. *Glycobiology* **9**, 887-895.

Sharon, N. L. (1982). Glycoproteins. In *The Proteins*, pp. 1-144. Edited by R. L. H. H. Neurath. New York: Academic.

She, Q., Singh, R. K., Confalonieri, F. & other authors (2001). The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. *PNAS* **98**, 7835-7840.

Shevchenko, A., Wilm, M., Vorm, O. & Mann, M. (1996). Mass spectrometric sequencing of proteins from silver stained polyacrylamide gels. *Anal Chem* **68**, 850-858.

Shuman, S. (1991). Recombination mediated by vaccinia virus-DNA Topoisomerase-I in Escherichia coli is sequence specific. *PNAS* **88**, 10104-10108.

Shuman, S. (1994). Novel approach to molecular cloning and polynucleotide synthesis using vaccinia DNA Topoisomerase. *J Biol Chem* **269**, 32678-32684.

Silberstein, S. & Gilmore, R. (1996). Biochemistry, molecular biology, and genetics of the oligosaccharyltransferase. *Faseb J* **10**, 849-858.

Souza, G. M., Hirai, J., Mehta, D. P. & Freeze, H. H. (1995). Identification of 2 novel *Dictyostelium discoideum* cysteine proteinases that carry N-acetylglucosamine-1-P modification. *J Biol Chem* **270**, 28938-28945.

Speed, M. A., Wang, D. I. C. & King, J. (1996). Specific aggregation of partially folded polypeptide chains: The molecular basis of inclusion body composition. *Nat Biotechnol* **14**, 1283-1287.

Spiro, R. G. (2002). Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology* **12**, 43R-56R.

Spiwok, V., Lipovova, P., Skalova, T., Buchtelova, E., Hasek, J. & Kralova, B. (2004). Role of CH/pi interactions in substrate binding by *Escherichia coli* beta-galactosidase. *Carbohydr Res* **339**, 2275-2280.

Stewart, E. J., Aslund, F. & Beckwith, J. (1998). Disulfide bond formation in the *Escherichia coli* cytoplasm: an in vivo role reversal for the thioredoxins. *Embo J* **17**, 5543-5550.

Sugiyama, K., Ishihara, H., Tejima, S. & Takahashi, N. (1983). Demonstration of a new Glycopeptidase, from jack-bean meal, acting on aspartylglucosylamine linkages. *Biochem Bioph Res Co* **112**, 155-160.

Suzuki, T., Seko, A., Kitajima, K., Inoue, S. & Inoue, Y. (1993a). Demonstration of the presence of Peptide-N-glycanase activities in mammalian-derived cultured cells - a possible occurrence of N-glycosylation de-N-glycosylation system in a wide variety of living organisms as the universal biologic processes. *Glycoconjugate J* **10**, 223-223.

Suzuki, T., Seko, A., Kitajima, K., Inoue, Y. & Inoue, S. (1993b). Identification of Peptide-N-glycanase activity in mammalian derived cultured cells. *Biochem Bioph Res Co* **194**, 1124-1130.

Suzuki, T., Kitajima, K., Inoue, S. & Inoue, Y. (1994a). Does an animal peptide: N-glycanase have the dual role as an enzyme and a carbohydrate-binding protein? *Glycoconjugate J* **11**, 469-476.

Suzuki, T., Kitajima, K., Inoue, S. & Inoue, Y. (1994b). Occurrence and biological roles of 'proximal glycanases' in animal cells. *Glycobiology* **4**, 777-789.

Suzuki, T., Seko, A., Kitajima, K., Inoue, Y. & Inoue, S. (1994c). Purification and enzymatic properties of Peptide-N-glycanase from C3h mouse-derived L-929 fibroblast cells - Possible widespread occurrence of posttranslational remodification of proteins by N-deglycosylation. *J Biol Chem* **269**, 17611-17618.

Suzuki, T., Kitajima, K., Emori, Y., Inoue, Y. & Inoue, S. (1997). Site-specific de-N-glycosylation of diglycosylated ovalbumin in hen oviduct by endogenous peptide: N-glycanase as a quality control system for newly synthesized proteins. *PNAS* **94**, 6244-6249.

Suzuki, T., Park, H., Kitajima, K. & Lennarz, W. J. (1998). Peptides glycosylated in the endoplasmic reticulum of yeast are subsequently deglycosylated by a soluble peptide: N-glycanase activity. *J Biol Chem* **273**, 21526-21530.

Suzuki, T., Park, H., Hollingsworth, N. M., Sternglanz, R. & Lennarz, W. J. (2000). PNG1, a yeast gene encoding a highly conserved peptide : N-glycanase. *J Cell Biol* **149**, 1039-1051.

Suzuki, T., Park, H., Kwofie, M. A. & Lennarz, W. J. (2001a). Rad23 provides a link between the Png1 deglycosylating enzyme and the 26 S proteasome in yeast. *J Biol Chem* **276**, 21601-21607.

Suzuki, T., Park, H., Till, E. A. & Lennarz, W. J. (2001b). The PUB domain: A putative protein-protein interaction domain implicated in the ubiquitin-proteasome pathway. *Biochem Bioph Res Co* **287**, 1083-1087.

Suzuki, T., Park, H. & Lennarz, W. J. (2002). Cytoplasmic peptide: N-glycanase (PNGase) in eukaryotic cells: occurrence, primary structure, and potential functions. *Faseb J* **16**.

Suzuki, T., Hara, I., Nakano, M. & other authors (2006). Site-specific labeling of cytoplasmic Peptide:N-glycanase by N,N'-diacetylchitobiose-related compounds. *J Biol Chem* **281**, 22152-22160.

Suzuki, T., Tanabe, K., Hara, I., Taniguchi, N. & Colavita, A. (2007). Dual enzymatic properties of the cytoplasmic peptide:N-glycanase in *C. elegans*. *Biochem Bioph Res Co* **358**, 837-841.

Szymanski, C. M., Yao, R. J., Ewing, C. P., Trust, T. J. & Guerry, P. (1999). Evidence for a system of general protein glycosylation in *Campylobacter jejuni*. *Mol Microbiol* **32**, 1022-1030.

Szymanski, C. M., Burr, D. H. & Guerry, P. (2002). *Campylobacter* protein glycosylation affects host cell interactions. *Infect Immun* **70**, 2242-2244.

Szymanski, C. M. & Wren, B. W. (2005). Protein glycosylation in bacterial mucosal pathogens. *Nature Reviews Microbiology* **3**, 225-237.

Taga, E. M., Waheed, A. & Vanetten, R. L. (1984). Structural and chemical characterization of a homogeneous Peptide N-glycosidase from almond. *Biochemistry-US* **23**, 815-822.

Takahashi, N. (1977). Demonstration of a new amidase acting on glycopeptides. *Biochem Bioph Res Co* **76**, 1194-1201.

Takasaki, S., Mizuochi, T. & Kobata, A. (1982). Hydrazinolysis of asparagine-linked sugar chains to produce free oligosaccharides. *Method Enzymol* **83**, 263-268.

Tarentino, A., Quinones, G., Trumble, A., Changchien, L., Duceman, B., Maley, F. & Plummer, T., Jr (1990). Molecular cloning and amino acid sequence of peptide-N4-(N-acetyl-beta- D-glucosaminyl)asparagine amidase from *Flavobacterium meningosepticum* [published erratum appears in J Biol Chem 1990 Jul 5;265(19):11405]. *J Biol Chem* **265**, 6961-6966.

Tarentino, A. L. & Plummer, T. H. (1982). Oligosaccharide accessibility to Peptide-N-Glycosidase is promoted by protein-unfolding reagents. *J Biol Chem* **257**, 776-780.

Tarentino, A. L., Gomez, C. M. & Plummer, T. H., Jr. (1985). Deglycosylation of asparagine-linked glycans by peptide:N-glycosidase F. *Biochemistry-US* **24**, 4665-4671.

Tarrago-Trani, M. T. & Storrie, B. (2004). A method for the purification of Shiga-like toxin 1 subunit B using a commercially available galabiose-agarose resin. *Prot Expres Purif* **38**, 170-176.

Ten Hagen, K. G., Bedi, G. S., Tetaert, D., Kingsley, P. D., Hagen, F. K., Balys, M. M., Beres, T. M., Degand, P. & Tabak, L. A. (2001). Cloning and characterization of a ninth member of the UDP-GalNAc : polypeptide N-acetylgalactosaminyltransferase family, ppGaNTase-T9. *J Biol Chem* **276**, 17395-17404.

Towbin, H., Staehelin, T. & Gordon, J. (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets - Procedure and some applications. *PNAS* **76**, 4350-4354.

Tretter, V., Altmann, F. & Marz, L. (1991). Peptide-N4-(N-acetyl-beta-glucosaminyl)asparagine amidase F cannot release glycans with fucose attached alpha-1,3 to the asparagine-linked N-acetylglucosamine residue. *Eur J Biochem* **199**, 647-652.

Udenfriend, S. & Kodukula, K. (1995). How glycosyl-phosphatidylinositol-anchored membrane-proteins are made. *Annu Rev Biochem* **64**, 563-591.

Vagin, A. & Teplyakov, A. (1997). MOLREP: an automated program for molecular replacement. *J Appl Crystallogr* **30**, 1022-1025.

van Stokkum, I. H. M., Spoelder, H. J. W., Bloemendal, M., van Grondelle, R. & Groen, F. C. A. (1990). Estimation of protein secondary structure and error analysis from circular dichroism spectra. *Anal Biochem* **191**, 110-118.

Varga, J., Kevei, F., Fekete, C., Coenen, A., Kozakiewicz, Z. & Croft, J. H. (1993). Restriction-Fragment-Length-Polymorphisms in the mitochondrial DNAs of the *Aspergillus niger* aggregate. *Mycol Res* **97**, 1207-1212.

Varki, A. (1993). Biological roles of oligosaccharides - All of the theories are correct. *Glycobiology* **3**, 97-130.

Veerapandian, B., Cooper, J. B., Sali, A., Blundell, T. L., Rosati, R. L., Dominy, B. W., Damon, D. B. & Hoover, D. J. (1992). Direct observation by X-ray-analysis of the tetrahedral intermediate of aspartic proteinases. *Protein Sci* **1**, 322-328.

Villaverde, A. & Carrio, M. M. (2003). Protein aggregation in recombinant bacteria: biological role of inclusion bodies. *Biotechnol Lett* **25**, 1385-1395.

Vogelstein, B. & Gillespie, D. (1979). Preparative and analytical purification of DNA from agarose. *PNAS* **76**, 615-619.

Wacker, M., Linton, D., Hitchen, P. G. & other authors (2002). N-linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli*. *Science* **298**, 1790-1793.

Wallace, A. C., Laskowski, R. A. & Thornton, J. M. (1995). LIGPLOT - A Program to generate schematic diagrams of protein ligand interactions. *Protein Eng* **8**, 127-134.

Wang, L. X., Tang, M., Suzuki, T., Kitajima, K., Inoue, Y., Inoue, S., Fan, J. Q. & Lee, Y. C. (1997). Combined chemical and enzymatic synthesis of a C-glycopeptide and its inhibitory activity toward glycoamidases. *J Am Chem Soc* **119**, 11137-11146.

Weerapana, E. & Imperiali, B. (2006). Asparagine-linked protein glycosylation: from eukaryotic to prokaryotic systems. *Glycobiology* **16**, 91R-101R.

White, O., Eisen, J. A., Heidelberg, J. F. & other authors (1999). Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1. *Science* **286**, 1571-1577.

Whitmore, L. & Wallace, B. A. (2004). DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res* **32**, W668-W673.

Wiertz, E., Jones, T. R., Sun, L., Bogyo, M., Geuze, H. J. & Ploegh, H. L. (1996). The human cytomegalovirus US11 gene product dislocates MHC class I heavy chains from the endoplasmic reticulum to the cytosol. *Cell* **84**, 769-779.

Wiese, M., Ilg, T., Lottspeich, F. & Overath, P. (1995). Ser/Thr-rich repetitive motifs as targets for phosphoglycan modifications in *Leishmania mexicana* secreted Acid-phosphatase. *Embo J* **14**, 1067-1074.

Winter, G. (2010). xia2: an expert system for macromolecular crystallography data reduction. *J Appl Crystallogr* **43**, 186-190.

Woodman, P. G. (2003). p97, a protein coping with multiple identities. *J Cell Sci* **116**, 4283-4290.

Woody, R. W. (1994). Contributions of tryptophan side chains to the far-ultraviolett circular dichroism of proteins. *Eur Biophys J Biophys Lett* **23**, 253-262.

Xin, F. X., Wang, S. J., Song, L., Liang, Q. F. & Qi, Q. S. (2008). Molecular identification and characterization of peptide: N-glycanase from *Schizosaccharomyces pombe*. *Biochem Bioph Res Co* **368**, 907-912.

Yanagida, H., Matsuura, T. & Yomo, T. (2008). Compensatory evolution of a WW domain variant lacking the strictly conserved trp residue. *J Mol Evol* **66**, 61-71.

Yet, M. G. & Wold, F. (1988). Purification and characterization of 2 Glycopeptide hydrolases from jack beans. *J Biol Chem* **263**, 118-122.

Young, N. M., Brisson, J. R., Kelly, J. & other authors (2002). Structure of the N-linked glycan present on multiple glycoproteins in the gram-negative bacterium, *Campylobacter jejuni*. *J Biol Chem* **277**, 42530-42539.

Yurist-Doutsch, S., Chaban, B., VanDyke, D. J., Jarrell, K. F. & Eichler, J. (2008). Sweet to the extreme: Protein glycosylation in Archaea. *Mol Microbiol* **68**, 1079-1084.

Yurist-Doutsch, S. & Eichler, J. (2009). Manual annotation, transcriptional analysis, and protein expression studies reveal novel genes in the *agl* cluster responsible for N-glycosylation in the halophilic archaeon *Haloferax volcanii*. *J Bacteriol* **191**, 3068-3075.

Zhao, G., Zhou, X., Wang, L., Li, G., Kisker, C., Lennarz, W. J. & Schindelin, H. (2006). Structure of the mouse Peptide N-glycanase-HR23 complex suggests co-evolution of the endoplasmic reticulum-associated degradation and DNA repair pathways. *J Biol Chem* **281**, 13751-13761.

Zhao, G., Li, G. T., Zhou, X. K., Matsuo, I., Ito, Y., Suzuki, T., Lennarz, W. J. & Schindelin, H. (2009). Structural and mutational studies on the importance of oligosaccharide binding for the activity of yeast PNGase. *Glycobiology* **19**, 118-125.

Zhou, P., Tian, F. F., Lv, F. L. & Shang, Z. C. (2009). Geometric characteristics of hydrogen bonds involving sulfur atoms in proteins. *Proteins* **76**, 151-163.

Zhou, X. K., Zhao, G., Truglio, J. J., Wang, L. Q., Li, G. T., Lennarz, W. J. & Schindelin, H. (2006). Structural and biochemical studies of the C-terminal domain of mouse peptide-N-glycanase identify it as a mannose-binding module. *PNAS* **103**, 17214-17219.

Zillig, W., Stetter, K. O., Wunderl, S., Schulz, W., Priess, H. & Scholz, I. (1980). The *Sulfolobus caldariella* group - Taxonomy on the basis of DNA-dependent RNA-polymerases. *Arch Microbiol* **125**, 259-269.