HETEROLOGOUS PRODUCTION AND CHARACTERISATION OF A YEAST PEPTIDE:*N*-GLYCANASE

A thesis presented in partial fulfillment of the requirements for
the degree of Master of Science
in Biochemistry
at Massey University, New Zealand

Kun Hong

2007

Abstract

Peptide:*N*-glycanases (PNGases) removes *N*-linked glycans from glycoproteins. Three distinct families of PNGases have been characterised, although all of them not completely. Some of these PNGases are cytosolic, others are secreted. Cytoplasmic PNGases (Png1p) are implicated in the proteasomal degradation of newly synthesized misfolded or unfolded glycoproteins that are exported from the endoplasmic reticulum (ER). Cytoplasmic PNGases are encoded by the *PNG1* gene and have been classified as members of transglutaminase-like superfamily based on the sequence analyses. There has, however, been no report of transglutaminase activity in any PNGase. The three-dimensional structures of recombinant PNGases from yeast (*S. cerevisiae*) and mouse have been determined in complex with the XPCB domain of Rad23 and mHR23B respectively. These PNGases were both produced as insoluble proteins, and could only be refolded and crystallised in the presence of their physiological binding partners.

In this study, the gene encoding for *S. pombe* PNGase has been cloned and heterologously expressed as a soluble thioredoxin-fused protein. The proteolytic cleaved recombinant protein (rPNGase Sp) remained soluble as a monomer and retained its deglycosylating activity. It did not have, however any transglutaminase activity despite its homology to the transglutaminase family of proteins. The activity of rPNGase Sp *in vitro* is both reductant and Zn^{2+} dependent. rPNGase Sp showed apparent heterogeneity on SDS-PAGE, which was characterised by the appearance of two bands differing in their molecular weights by an ~ 2.3 kDa. This heterogeneity was eventually shown to be the result of two different local conformations that were dependent on disulfide bond and/or Zn^{2+}. The enzyme was shown to only deglycosylate the denatured glycoproteins, not their native counterparts. Moreover, it preferred to deglycosylate glycoproteins with high mannose-type glycan chains, both of which are consistent with the biological function of cytoplasmic PNGases.

Compared to bacterial PNGase F, rPNGase Sp is not very active, at least on the substrate used in this study. A higher K_m (186 μM) determined for rPNGase Sp using a FITC-labelled glycopeptide which carries a complex-type glycan as the substrate also suggests that complex glycans are not favoured substrates for these PNGases. rPNGaseSp has similar characteristics to the yeast (*S. cerevisiae*) and mouse PNGases; it has a neutral pH optimum and is strongly inhibited by Cu^{2+}, Cd^{2+} and Ni^{2+}. EDTA treatment deactivates it, and the addition of Zn^{2+} could not restore its activity. Interestingly, addition of exogenous Zn^{2+} was found to strongly inhibit rPNGase Sp.

Acknowledgements

I would like to thank the following people for their assistance in the completion of this project:

My supervisor Dr. Gillian Norris for her invaluable advice, guidance, infinite patience and encouragement through out the course of my study.

Mr. Trevor Loo for lots of technical help and useful discussions during this research.

Dr. Mark Pattchet for useful advice and providing chemicals for the reduction and alkylation experiment.

I express my gratitude to Dr. Santanu Deb Choudhury (mass spectrometry analyses), Mr. Matthew Bennett (crystal trials), Associate Prof. David Harding (chemicals for the transglutaminase assay). I also express my sincere appreciations to all the other members in the X-lab (Institute of Molecular BioScience, Massey University) for their help and support.

Thank you, Sue Flynn and Sylvia Hooker, for arranging my living and study in New Zealand. I should also thank New Zealand's International Aid and Development Agency for awarding me a NZAID scholarship.

Finally, I would particularly like to thank my wife, Yabin, for her support, encouragement and love.

Table of Contents

Chapter 2 — Materials and Methods

List of Figures

Appendix Figures

List of Tables

Abbreviations

~	approximately
Mwt	molecular weight
kb	kilo base
kDa	kilo Dalton
nm	nanometer
ERAD	endoplasmic reticulum associated degradation
DNA	deoxyribose nucleotide triphosphate
PCR	polymerase chain reaction
dNTP	deoxyribose nucleotide triphosphate
Trx	thioredoxin
E. coli	*Escherichia Coli*
Amp	ampicillin
Tet	tetracycline
Kan	kanamycine
Cam	chloramphenicol
LB	Luria Broth
IPTG	iso-propyl-beta-D-thiogalactopyranoside
DNase	deoxyribonucease
RNase	ribonuclease
SDS	sodium dodecyl sulfate
PAGE	polyacrylamide gel electrophoresis
BSA	bovine serum albumin
EDTA	ethylene diamine tetra-acetate
DTT	dithiothreitol
TCEP	Tris (2-carboxy-ethyl) phosphine hydrochloride
IMAC	immobilised metal affinity chromatography
rTEV	recombinant tobacco etch virus
SEC	size exclusion chromatography
FPLC	fast protein liquid chromatography
HPLC	high performance/pressure liquid chromatography
TCA	trichloroacetic acid

TFA	trifluoroacetic acid
MS	mass spectrometry
MALDI	matrix assisted laser desorption ionization
TOF	time-of-flight
PMF	peptide-mass fingerprinting
m/z	mass-to-charge ratio
UV	ultra-violet
HEPES	2-(4-(2-hydroxy-ethyl)-1-piperazinyl) ethane sulfonic acid
MES	2-(N-morpholino) ethane sulfonic acid
Tris	tris (hydroxymethyl) aminomethane
F.C.	final concentration
rPNGase Sp	recombinant Peptide:N-Glycanase from *S. pombe*
TGase	transglutaminase
Ova	hen egg ovalbumin-derived 11-mer glycopeptide
FITC	fluorescein isothiocyanate
FITC-Ova	FITC-dilabeled ovalbumin 11-mer glycopeptide

Chapter 1 — Introduction

1.1 *N*-linked Glycosylation of Proteins

N-glycosylation is a common post-translational modification of many proteins that contribute greatly to the their structural and functional diversity. In the endoplasmic reticulum (ER), Specific 14-saccharide "core" units are transferred on to the nascent polypeptide chain by oligosaccharyl transferase at an Asn that is part of an Asn-X-Ser/Thr sequon (where X can be any amino acid except Pro or Asp). They are then trimmed by a number of specific glucosidases before being transported to the Golgi complex where further trimming and elongation occur. The glycoproteins have to be correctly folded before being transferred to the Golgi complex [1].

The trimming and processing that *N*-linked core glycans undergo in ER is uniform and, therefore, results in no additional structural diversity of glycoproteins. The common functions of the small *N*-linked core glycans are promoting the proper folding of proteins, quality control and certain sorting events. More extensive modifications occur in the Golgi complex where the *N*-linked core glycans undergo a series of non-uniform modifications and processes. These modifications, which result in the addition of different monosaccharides in a variety of linkages to the common core, generate a huge variety of branched structures and, as a consequence, give rise to the tremendous structural and functional diversity of glycoproteins [1]. The glycans of matured glycoprotein contribute to protein stability, specificity, protein targeting and intra- / extracellular recognition [2].

1.2 De-*N*-Glycosylation

N-glycosylation of proteins is known to be a biologically important post-translational modification of proteins. However, the significance of de-*N*-glycosylation of glycoproteins is not yet clear. The ubiquitous enzyme, Peptide:*N*-glycanase (EC 3.5.1.52, PNGase, also termed glycoamidase), is known to remove the *N*-linked glycans from glycoproteins *in vivo*. It cleaves the β-aspartylglucosamine bond of *N*-linked glycans, releasing an intact oligosaccharide and in the process, converting the asparagine into an aspartic acid residue [3, 4] (Figure 1.1).

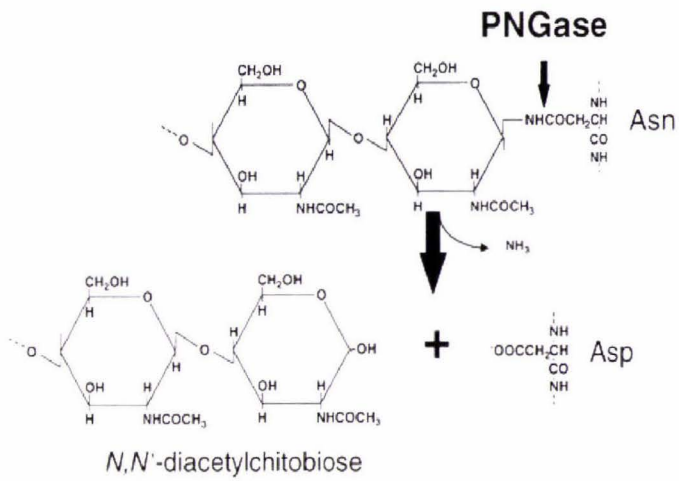


Figure 1.1 The de-*N*-glycosylation reaction catalysed by PNGase. PNGase cleaves the amide bond between the asparagine residue and the *N*-acetylglucosamine, generating an aspartic residue in the in place of the asparagine and an intact oligosaccharide having a N, N'-diacetylchitobiose at its reducing terminus. The amine linked to the oligosaccharide autolyses at physiological pH and diffuses away. Adapted from Suzuki *et al.*, 2002.

1.3 Peptide:*N*-Glycanase (PNGase) Overview

Bioinformatic analyses have shown that on the basis of sequence similarity to proteins known to have *N*-glycanase activity, there appear to be three distinct types of PNGase (Gutsche *et al.*, unpublished data, IMBS, Massey University). Although the known examples from each family catalyse the same reaction, they have absolutely no sequence similarity to each other. PNGase F secreted by the Gram-negative bacterium *Flavobacterium meningosepticum* is the sole proven example of a prokaryotic PNGase (type I). It is a single polypeptide chain made up from 314 amino acids and its 3-dimensional structure has been determined [5-8]. Type II PNGases were first found in plants (*Prunus amygdalus*, PNGase A [9]) and more recently in fungi (*Aspergillus tubigensis,* PNGase At [10]). They are usually heterodimers and can be glycosylated themselves [10, 11]. While the precise biological functions of type I and II PNGases are still not known, they are widely used as tools for studying both the glycan and protein moieties of *N*-linked glycoproteins because of their ability to efficiently release intact *N*-glycans from glycoproteins and glycopeptides under mild conditions [6, 12].

Type III PNGases are present in a variety of eukaryotes that range from yeasts to plants and mammals [4, 13]. They are monomeric and have no homology to either type I or type II PNGases, but are highly conserved in all eukaryotes [13], suggesting they play an important role. Intriguingly, they have been classified as belonging to the 'transglutaminase-like superfamily' of proteins because there is a 'transglutaminase' catalytic motif (Cys, His and Asp) in the most conserved region of the sequence [14]. However, there has, as yet, been no report or any observation of transglutaminase activity for any of these enzymes. The catalytic mechanism is not certain, although site-directed mutagenesis of a number of residues (Cys-191, His-218, Asp-235, Trp-220, Trp-231, Arg-210, and Glu-222) has produced inactive protein [15]. Recently, increasing evidence has implicated type III PNGases in the proteasomal degradation of newly synthesized misfolded or unfolded glycoproteins exported from the endoplasmic reticulum (ER) [13, 16, 17].

1.4 Potential Functions of Cytoplasmic PNGase (Type III)

1.4.1 Participating in ERAD

In eukaryotes, many newly synthesized proteins are subjected to post-translational modification before entering the secretory pathway. They are usually *N*-glycosylated by oligosaccharyl transferase in ER immediately after being synthesized. Glycoproteins that fold correctly in ER are then transported to Golgi complex, whereas misfolded or unfolded glycoproteins are retained in ER, where they undergo several cycles of chaperone mediated folding/unfolding events till the quality control system of the ER determines they are correctly folded. If they are irreversibly denatured, they are subsequently discharged from ER into cytosol and subjected to proteasomal degradation. This degradation pathway is usually called the ER-associated degradation (ERAD) [18].

Soluble PNGases (type III) have been thought to act in this degradation pathway by deglycosylating misfolded glycoproteins in the cytosol. Several lines of evidence support this hypothesis:

(1) De-*N*-glycosylated protein intermediates accumulate when proteasome inhibitors are added into mammalian cells [19].

(2) Glycopeptides transported out of the ER are deglycosylated by an enzyme having PNGase like activity [4].

(3) The appearance of intact glycoprotein (human class I myosin heavy chains (MHCs)) in the cytosol of human PNGase knockdown cells. Hirsch and co-workers [20] used small interfering RNA (siRNA) against PNGases to attenuate PNGase levels in a human cell line which had been previously specially treated. In these treated cells, major histocompatibility complex heavy chains (MHC HCs) could not assemble correctly in ER and were transferred into cytosol for proteasomal degradation. After PNGase siRNA was introduced into the cells, the intact glycosylated MHC HCs were observed in cytosol. This finding suggested that (a) PNGase prepares the glycoprotein for proteasomal degradation; (b) PNGase acts after the dislocation of misfolded protein out of the ER; (c) PNGases acts prior to proteasome in the ERAD pathway.

1.4.2 Involvement of Post-translational Remodification

Besides the involvement in the ERAD degradation pathway, soluble PNGases are also thought to be involved in the protein quality control system, by participating in the post-translational remodification of certain proteins. Because the PNGase-catalysed reaction not only releases the intact glycan but also changes the primary structure of the protein involved (converts Asn to Asp, introducing a negative charge), it can leads to changes in the physicochemical properties of the protein, promoting its maturation. An example of such remodification is observed for the maturation of hen ovalbumin. PNGase specifically hydrolyses Di-*N*-glycosylated ovalbumin to generate the mature mono-glycosylated ovalbumin found in egg white [21].

1.5 Cytoplasmic PNGases

1.5.1 Discovery of Cytoplasmic PNGases

After the finding of PNGases in plants almond and bacteria *Flavobacterium meningosepticum* [5, 9]), the first animal PNGase was observed in fish oocytes and embryos [22]. Later, the PNGase activity was observed in mouse and human cells, hen

oviduct and the budding yeast *Saccharomyces cerevisiae* [4, 21, 23]. Unlike PNGases F and A, which are secreted from cells, type III PNGases reside in the cytosol [16]. Further studies have shown that these cytoplasmic PNGases have a different substrate specificities compared to type I and II PNGases.

1.5.2 Png1p are Highly Conserved in Eukaryotes

In 2000, Suzuki and co-workers [13] isolated a PNGase defective mutant in *Saccharomyces cerevisiae*, and subsequently identified a gene (*PNG1*) encoding a PNGase in *Saccharomyces cerevisiae* (ScPng1p). By searching for sequence similarity using various DNA and protein databases, it was revealed that the *PNG1* is conserved over a wide range of higher eukaryotes [13], suggesting that the protein product (Png1p) plays a role in some fundamental biological process. Sequence alignment of *S. cerevisiae* Png1p (ScPng1p) with hPng1p (human), mPng1p (mouse), DmPng1p (*D. melanogaster*), CePng1p (*C. elegans*) and SpPng1p (*S. pombe*) showed that there is a highly conserved core domain (amino acids residues 106-362 in ScPng1p) residing in the middle of the sequence [13]. Moreover, unlike yeast Png1p, there are extended sequences at both the N- and C- termini in the mammalian homologues of Png1p, the N-terminal regions containing a PUB (Peptide:*N*-glycanase/UBA-containing protein domain) or PUG domain, which has been proposed to interact with various ubiquitin/proteasome pathway-related proteins [16, 24] (Figure 1.2). These findings suggested that Png1p might be involved in the ubiquitin/proteasome degradation pathway in eukaryotes.

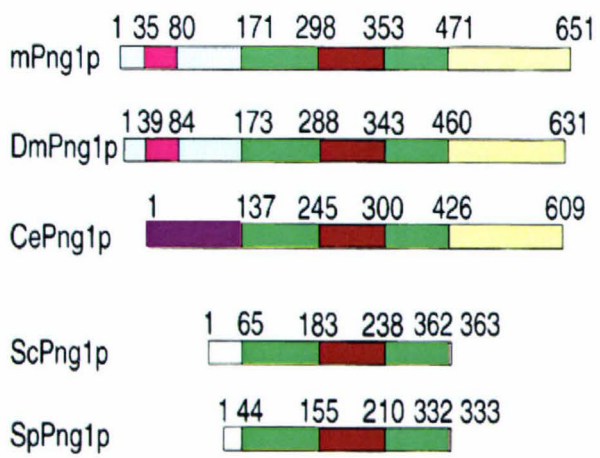


Figure 1.2 Schematic representation of the primary structure of various eukaryotic PNGases. mPng1p (mouse Png1p homologue), DmPng1p (*D. melanogaster*), CePng1p (*C. elegans*), ScPng1p (*S. cerevisiae*), and SpPng1p (*S. pombe*). The central core domain (green) is common to all Png1p homologues and contains the highly conserved transglutaminase domain (red). The higher eukaryotes (m, Dm, Ce) have extended N- and C-terminal domains (yellow and blue). The purple domain represents a unique thioredoxin-like domain so far found only in CePng1p. The PUB domain (pink) is found in proteins implicated in ubiquitin-related pathways. Adapted from Suzuki *et al.*, 2002.

1.5.3 Png1p: Members of Transglutaminase-like Superfamily

Based on sequence analyses, Png1p and its orthologues have been classified as being members of the 'transglutaminase-like superfamily [14]. Transglutaminases (TGase, EC 2.3.2.13) catalyse the post-translational modification of proteins by the formation of isopeptide bonds. This occurs either through protein cross-linking via ε-(γ-glutamyl) lysine bonds or through incorporation of primary amines at selected intra peptide glutamine residues [25] (Figure 1.3).

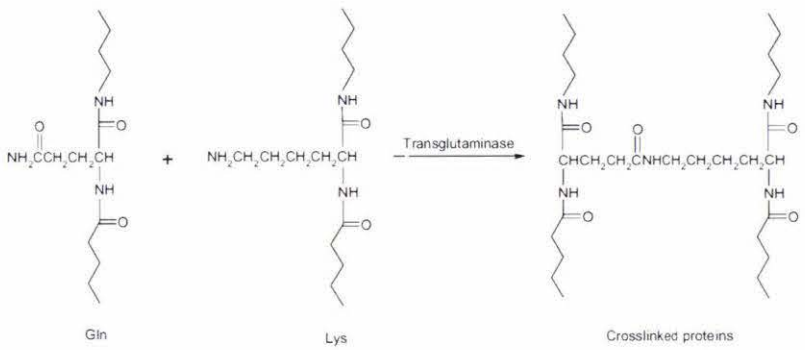


Figure 1.3 Transglutaminases cross-link proteins through an acyl-transfer reaction between the γ-carboxamide group of peptide-bound glutamine and the ε-amino group of peptide-bound lysine, resulting in ε-(γ-glutamyl)lysine isopeptide bond.

Extensively cross-linked insoluble protein polymers exhibit high resistance to proteolytic degradation and mechanical breakage, and are used by the organism to crate stable structures that act as barriers. Examples of such structues are found in blood clots (transglutaminase coagulation factor X III) as well as skin and hair (epidermal /hair follicle transglutaminase). Recent research indicates that abnormally high transglutaminase activity results in a variety of disease states including acne, cataracts, psoriasis, Huntington's and Alzhimer's disease (Reviewed by Griffin [26]).

PNGase, on the other hand, cleaves the amide bond formed between the modified asparagine residue and *N*-acetylglucosamine (GlcNAc). Proteins belonging to the transglutaminase superfamily have a conserved active site comprising cysteine (Cys), histidine (His) and aspartic acid (Asp) residues. The sequence surrounding this catalytic motif has also found to be highly conserved [26]. Site-specific mutagenesis has shown that the residues essential for Png1p activity Cys-191, His-218, and Asp-235 in ScPng1p are also essential for the activity of all known transglutaminases [15]. The hypothesis that

Png1p utilises the transglutaminase-like motif (Cys-His-Asp) as an active site for their catalytic reaction has therefore been proposed [15]. Recently, a study of the crystal structure of yPng1p in complex with one of its physiological partner (section 1.6) revealed that yPng1p and transglutaminases (e.g. coagulation factor XIII) have similar folding pattern in their core domains [27]. This further indicates there may be an evolutionary linkage between cytoplasmic PNGases and transglutaminases. Although Png1ps have been classified as being members of the 'transglutaminase-like superfamily', there has been, as yet, no report or any observation of transglutaminase activity in the type III PNGases.

Other conserved residues types such as Asp, Arg and Glu are shown to be essential for PNGase F activity through site-directed mutagenesis ([8], Norris *et al.*, personal communication). In Png1p, the aromatic residues Trp-220 and Trp-231 are thought to stabilise the structure of the protein through hydrophobic interactions, while Arg-210 and Glu-222 are predicted to maintain the protein conformation by forming a salt bridge [15]. Interestingly, these residues are reminiscent of the active site in PNGase F, where the arginine plays an essential role in making a hydrogen bond with the scissile carbonyl oxygen of Asp making it more susceptible to nucleophilic attack by a bound water molecule (Loo *et al.*, unpublished data, IMBS, Massey University, NZ).

Outside the transglutaminase catalytic motif, two conserved thioredoxin-like 'CXXC' motifs formed by four cysteines were also found to be essential for the PNGase activity of Png1p [13]. They were proposed to have a structural or a catalytic role, as reducing reagents such as dithiothreitol (DTT) are required for PNGase activity *in vitro* [4, 21, 28]. In fact, the recent crystal structure study [27] showed that these two motifs reside in a zinc-binding domain of yPng1p chelating one zinc ion. Additional residues that are conserved in type III PNGases but not in transglutaminases were also found to affect the integrity of the enzyme. These are Phe-261, Asp-268, Tyr-257 and Tyr-273 [15].

Generally, Png1p proteins differ from transglutaminases in their biological function, substrate specificity (section 1.9.3), their global three-dimensional structure (discussed in the next section), and possibly the catalytic mechanism.

1.6 Crystal structure of the Png1p

From 2005 to 2006, the crystal structures of yPng1p and mPng1p in complex with the XPC-binding (XPCB) domain of yRad23 (a DNA repair protein that provides a link between Png1p and the 26S proteasome and is involved in transport of misfolded glycoproteins from the ER to the proteasome. Section 1.8) and of mHR23B (the mouse homolog of yRad23) respectively, were determined [27, 29]. These helped to elucidate the catalytic mechanism and substrate binding specificities of cytoplasmic PNGases. The crystal structure of full-length yPng1p in complex with the XPCB binding domain ((yRad23XBD; residues 238-309) of yRad23 is shown in Figure 1.4.

Overall, yPng1p folds into α/β structure dimensions 92×55×37Å. The structure can be divided into three major domains: a Rad23 binding domain, a core and a zinc binding domain. The Rad23 binding domain, which contains an extensive exposed hydrophobic surface comprised of two N- and two C-terminal helices (H1, H2, H11 and H12). The N-terminal helix adopts an extended conformation and interacts extensively with a hydrophobic groove formed by four helices in the XPCB binding domain of yRad23. Most of the residues involved in this interface are highly conserved in both PNGase and Rad23 orthologues. The C-terminal H12, which positions against the H1 at 45°, interacts with residues from the helix 1 of yRad23 through two H-bonds and two ion pairs.

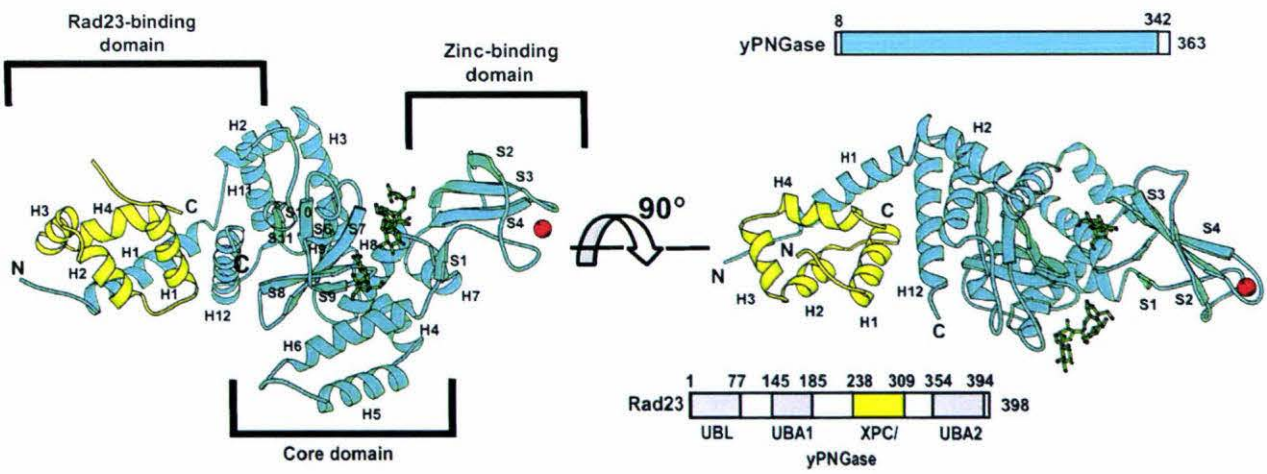


Figure 1.4 Schematic representation of the yPNGase–yRad23 complex, providing two different views of the yPNGase–yRad23 complex structure yPNGase is shown in blue and yRad23 is in yellow. yPNGase comprises three domains, an N-terminal Rad23-binding, a core, and a Zn-binding domain. Three sucrose molecules (green) are located in the deep cleft. A Zn atom (red) is coordinated by Cys-129, -132, -165, and -168 in yPNGase. Adapted from Lee *et al.*, 2005.

The zinc-binding domain located at the opposite end of the Rad23 binding domain consists of five strands (S1 to S5) and two helices (H7 and H8). One zinc ion is located between two loops that span S1&S2 and S3&S4, and contains two conserved CXXC motifs which coordinate zinc ion through the four cysteines. It has been suggested that this zinc-binding domain stabilises the enzyme conformation as exposing the enzyme to EDTA lowered its melting temperature (Tm) by 4℃, and inactivates it [27]. Mutation of any of the four cysteine residues in the two CXXC motifs to alanine resultes in complete inactivation of Png1p [15]. The precise role of zinc binding domain with respect to PNGase activity is, however, not clear, although it has been assumed to play a role in regulating substrate binding and /or the thermodynamics of the reaction. This is because the zinc-binding motif, comprised of S2, S3 and S4, shows a scaffold similar to the zinc-ribbon structures of *Bacillus stearothermophilus* adenylate kinase and transcriptional elongation factor S-II. By forming a "lid" over the active site of the enzyme, the zinc ribbon of adenylate kinase is thought to regulate the thermodynamics of catalysis by stabilizing the intermediate state and promoting product release [29]. If this assumption is right, it may explain how PNGase catalyses reactions in the reverse direction to transglutaminases, as in these enzymes, the zinc ribbon is absent. In fact, the zinc-binding domain participates the formation of a deep cleft in which the active site resides, and provides two residues (Arg-176, Arg-190) that interact with the inhibitor Z-VAD-fmk.

The core domain of yPng1p contains six central, strongly curved antiparallel β-strands (S6-S11), which are buttressed by three α-helices (H3, H5 and H6) and several surface loops. Comparison of the crystal structures of yPng1p/mPng1p with several transglutaminases including the coagulation factor XIII, arylamine *N*-acetyltransferase and avrpph B hydrolase revealed that they share a common core of two α-helices (H6 and H8) and five strands of curved antiparalleled β-sheet (S6 to S10 in yPng1p). The major structural difference between the transglutaminases and the Png1p family is that in the PNGases, a deep cleft (width 8Å, length 30Å) harboring a catalytic triad is formed between the interface of the core domain and the zinc-binding domain. H9, S8 and S9 form one side of this cleft while S1 forms the other. In contrast, in the transglutaminases, a shallow substrate-binding pocket is formed between the two domains which are much more tightly packed. Moreover, residues (Trp-123, Arg-176, Asn-217, Cys-237 and Glu-238) which are part of the large deep cleft in yPng1p are highly conserved in all the

cytoplasmic PNGases, but not in the transglutaminases. These structural differences reflect the differences seen in the catalytic mechanism and substrate specificity of PNGase and transglutaminase.

The active site of Png1p was mapped within the deep cleft because:

1) The putative transglutaminase catalytic triad is located in this cleft (Cys-191 from H8, His-218 from S7 and Asp-235 between S8 and S9);

2) When an irreversible inhibitor, the tripeptide carbobenzyloxy-Val-Ala-Asp-α-fluoromethylketone (Z-VAD-fmk, a pan-caspase inhibitor), was co-crystallised with yPng1p and mPng1p [27, 29], it was found to be covalently linked to Cys-191 through the methylene group originally linked to the α-fluoromethylketone (leaving group in covalent inhibition [30, 31]). It also formed two H-bonds and two ion bridges with arginine-176 and arginine-190 (arginine-291, arginine-305 in mPng1p) from S4 and S5 of the zinc-binding domain of the enzyme (Figure 1.5b). Although Z-VAD-fmk is not an obvious mimic of the glycosylated asparagine residue of a PNGase substrate (sequence is Val-Ala-Asp, not the Asn-X-Ser/Thr sequon), it presents a motif similar to that generated in during PNGase deglycosylation. Also, it binds exclusively to cysteine-191 residue out of the 14 cysteine residues in the yPng1p / core domain of mPng1p.

3) Three sucrose molecules (from cryo-solution when the crystals were prepared for data collection) were found in the deep cleft that bind to three specific sites as shown in Figure 1.4. The sucrose molecule at the top binding site was replaced with Z-VAD-fmk while the other two were not (Figure 1.5b). These findings suggest that there are two binding sites, one for the peptide and one the glycan of glycoproteins. Most recently, a group of haloacetamidyl-sugars were found to be strong, highly specific inhibitors of Png1p [32]. These inhibitors included a high mannose-containing oligosaccharide, $Man_9GlcNAc_2$-iodoacetoamide, as well as the shorter $GlcNAc_2$-iodoacetoamide, both of which were shown to make a covalent bond with the sulphur cysteine-191, irreversibly inhibiting Png1p. When the chitobiose (GlcNAc) molecule was docked into the proposed oligosaccharide binding site using molecular modeling software, a binding free energy of - 4.70 kcal/mol was estimated. The chitobiose moiety was also shown to be well positioned with its reducing end in close proximity to Cys-191,

forming eight H-bonds with residues that are part of the deep cleft in Png1p (Figure 1.5a). All these findings suggested the GlcNAc bound in a pocket in the deep cleft of Png1p, and that the cysteine-191 was the nucleophile in the enzymatic deglycosylation reaction, attacking the carbonyl carbon atom on one side of the β-aspartylglucosamine bond, and causing the hydrolysis of the scissile bond.

These structures containing both a carbohydrate motif ($GlcNAc_2$/sucrose) and a tripeptide inhibitor (Z-VAD-fmk) may mimic the binding of a glycopeptide substrate. When the GlcNAc moiety, glycosylated Asn, and flanking residues of some substrates (e.g., yeast carboxypeptidase Y, RNase B, ovalbumin) of Png1p were superimposed onto the sucrose moiety and the C^{α} position in the tripeptide inhibitor Z-VAD-fmk, it was found that the glycosylated Asn residue of the native/folded glycoprotein was prevented from accessing the active site by both sides of the deep cleft: the strands S2 and S3 in the Zn-binding domain, and two loops, one between H10 and H11, and the other between H9 and S10 of the core domain. Because denatured glycoproteins are, however, likely to have flexible and extended polypeptide chains, access of the glycosylated Asn residues to the active site may be possible despite the constraints imposed by the wall of the cleft. This may explain why Png1p family can only hydrolyse denatured but not native glycoproteins, a conformational selectivity that transglutaminases don't have due to the lack of a deep narrow cleft. Intriguingly, transglutaminases do not have PNGase activity.

It is noteworthy that although the global folding of yPng1p and mPng1p is similar, some notable differences exist. For example, there is an extra helix in mPng1p after H10 that was labeled as H10', and the first two helices of yPng1p (H1 and H2) are missing in the mouse protein. This results in different Png1p-Rad23 interaction patterns. The interaction between yPng1p and yRad23 primarily involves the N-terminal helix, H1 and the C-terminal helix, H12. For mouse PNGase, the XPCB domain of mHR23B interacts extensively with H12 of mPng1p and to a small degree with H11. It is thought that the difference in the interaction is not due to a conformational change but is caused by the differences in both the XPCB domain and its interactions. In fact, the yRad23-XPCB and mHR23-XPCB structures display striking differences in their molecular shapes and particularly, in the distribution of hydrophobic residues. More importantly, H12 of mPng1p is very different from yPng1p H12 in terms of its amino acid composition and hydrophobicity. This may be because, during evolution, the primary XPCB-interaction

function migrated from H1 to H12 [27, 29].

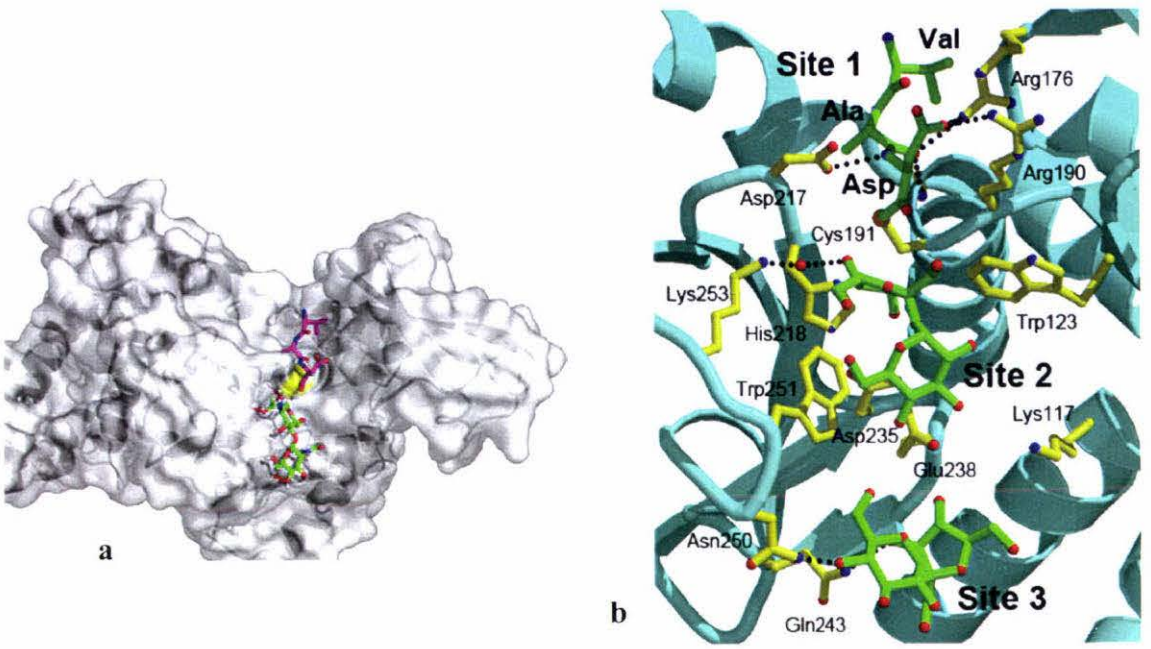


Figure 1.5 a) Close up view of $GlcNAc_2$ (green) binding to Png1p in complex with Z-VAD-fmk (pink). Png1 is shown in a partially transparent surface representation with an underlying ribbon diagram. The active site cysteine is labelled in yellow. **b) Active site of yPng1p** Two sucrose molecules (green, bound in site 2 and 3 respectively) interact with yPng1p residues through multiple H-bonds (black dotted line) and van der Waals interactions. Upon binding of Z-VAD-fmk, the thiol group of Cys-191 that was H-bonded to His–218 movs away and covalently binds to the methyl carbon adjacent to the terminal carbonyl group of the inhibitor. This Asp residue also forms ion pairs with side chains of Arg-176 and -190. The backbone carbonyl groups of Val and Ala residues of the inhibitor form H-bonds with the side chains of Arg-176 and -190. Adapted from Lee *et al.*, 2005 and Zhao *et al.* 2006.

1.7 Png1p is Residing Mainly in the Cytosol

Unlike bacterial PNGase F which is secreted, yeast Png1p and its mammalian homologues are cytoplasmic enzymes [4, 13, 21-23, 33]. Determination of the subcellular localization of Png1p is therefore important for studying its biological function. Although PNGases have been reported to be present in the endoplasmic reticulum and microsomes [21, 33], their activities have been described as mainly cytosolic or cytoplasmic, which is consistent with the hypothesis that they might be involved in a proteasome-degradation pathway in the cytosol.

Based on immunofluorescence analyses using a green fluorescent protein, a (GFP)-recombinant ScPng1p chimera was shown to be localised mainly in cytosol and nucleus of budding yeast cells [13]. This result, together with the finding that almost all the 26S proteasome subunits were localized in the nucleus of *S. cerevisiae* [34], implied that in yeast, ScPng1p might co-localise and interact with proteasomes in nucleus.

Hirsch and co-workers carried out a subcellular fractionation of cos-1 cells containing a gene expressing mouse PNGases (mPng1p). They reported that mPng1p was present only in the cytosol [17]. Katiyar and co-workers showed that human Png1p was present in the cytosol of HeLa and Cos-1 cells but was more concentrated in the ER fraction. They hypothesized that mPng1p was non-covalently associated with the ER membrane and functioned in the ER-associated degradation pathway [35]. However, because the bioinformatic analyses showed there was no membrane spanning sequence in Png1p proteins [13], this hypothesis was discounted.

1.8 Protein Interactions of Cytoplasmic PNGases — Working Modes

1.8.1 Proposed Working Model in Yeast

The finding that soluble PNGases interact strongly with proteasomal proteins suggests the involvement of PNGase in ERAD pathway. Yeast two-hybrid screening and biochemical studies showed that a DNA repair protein, yeast Rad23 (HR23B, and mHR23B in human and mouse respectively), provided a link between yeast Png1p (hPng1p and mPng1p) and the 26S proteasome (19S in mammalian cells). Rad23 binds to the proteasome subunit through its N-terminal ubiquitin-like (UbL) domain [24, 36, 37] , and to the N-terminal region of yPng1p through its XPC binding domain (yRad23XBD; residues 238-309). Interestingly, Rad23 also uses this domain to interact with Rad4/XPC, a member of the transglutaminase family involved in the DNA repair process [27, 38]. Recent evidence has shown that Rad23 also interacts with ubiquitin/multiubiquitin through its ubiquitin-associated (UBA) domain [39-41]. An *in vitro* binding experiment showed that mHR23B specifically binds to two misfolded glycoproteins, yeast carboxypeptidase Y and chicken ovalbumin, only after they were deglycosylated, indicating that Rad23 is able to act as a receptor for misfolded proteins [35]. These findings, combined with the observation that hPng1p, hHR23B and the 19S proteasome both co-localise close to the ER in HeLa cells, suggests that a complex formed by PNGase and the proteasome provides a platform for the cooperative deglycosylation and proteolysis of misfolded glycoproteins exported from the ER.

Although several glycoproteins such as denatured carboxypeptidase Y, the T cell receptor α-chain, and class I MHC are known to be substrates for PNGase *in vitro*, their turnover *in vivo* was not drastically affected in *PNG1*-knockout yeast cells [13, 20, 30]. Recently, the first *in vivo* substrate of the Png1p/Rad23-dependent pathway was identified. Experiments showed that the XPC binding domain-mediated Png1p-Rad23 interaction, the Ub chain binding activity (mediated by UBL domain of Rad23) and the deglycosylation activity of Png1p were all required for the degradation of the glycosylated ricin A chain (RTA) [38]. A working mode for yPng1p was thus proposed (Figure 1.6 a).

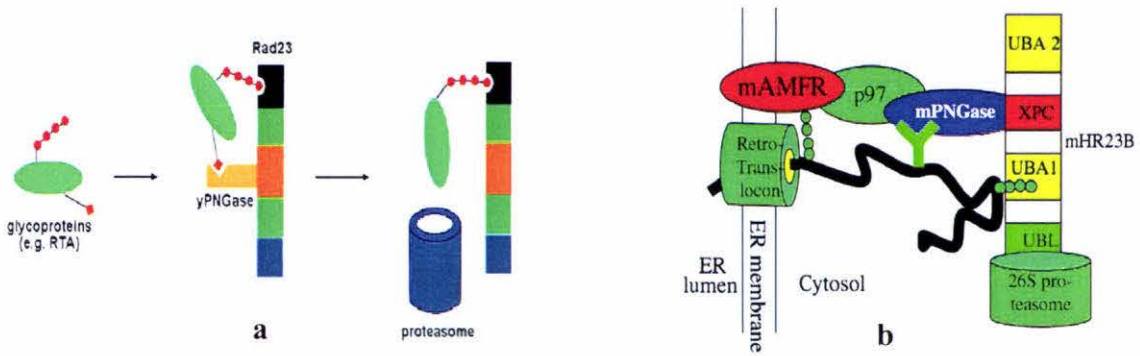


Figure 1.6 Schematic model for Rad23- mediated substrate proteolysis in a) yeast: Glycosylated ER proteins (e.g., RTA) are ubiquitylated and transported back to the cytosol. Png1 removes N-linked glycan. The XPCB domain (orange) of Rad23 Binds Png1p, which in turn facilitates the substrate recognition of Rad23. Through interactions with Ub chains and the proteasome mediated by the C-terminal UBA (black) and UBL (blue) domains in Rad23, Rad23 facilitates substrate transfer to the proteasome. The red circles depict Ub, and the orange diamonds depict sugar moiety. Adapted from Kim *et al.*, 2006. **b) mouse – mAMFR/p97/mPng1p dependent escort pathway** A protein substrate (black line) is being retrotranslocated (through the retro translocon) from the ER lumen to cytosol and recognized by the mAMFR-p97-mPNGase-mHR23B-proteasome complex in the cytosol. Poly-ubiquitin chain (green dot) is being added by mAMFR, an E3 ligase. The glycan moiety is recognized by mPNGase, and a polyubiquitin chain is bound to the ubiquitin associated (UBA) domain (yell square) of mHR23B. Adapted from Li *et al.*, 2005.

1.8.2 A model for the involvment of Png1p in the ERAD pathway in mammals

The mammalian homologues (mPng1p and hPng1p) differ from their yeast orthologues in that they have extended domains at both N- and C-termini [16], indicating that they may interact with more proteins than yPng1p. Two-hybrid library screening using mPng1p as target revealed that, besides mHR23B, mPng1p interacts with a number of proteins including mS4 (a ATPase subunit of the 19S proteasome), ubiquitin, mouse autocrine motility factor receptor (mAMFR, a putative ubiquitin ligase located on ER membrane), mY33K (a hypothetical protein containing 1 UBA and 1 UBX motif) and importin α, all of which have been implicated in the ubiquitin-dependent degradation pathway [42]. However, further study [43] using GST pull down assays revealed that only mHR23B and mS4 interact directly with mPng1p *in vitro*. Both interactions were mediated through the N-terminus of mPng1p which encompasses the PUB domain (residues 35-80), and were found to be mutually exclusive. This suggested there may be at least two pathways that promote the interaction of mPNGase with the proteasome. Interestingly, a chaperone-like

protein AAA ATPase p97 (also called VCP, Cdc48) that plays a central role in the recruitment and delivery of ubiquitylated substrates to proteasome in the ERAD pathway [44] was also found to interact directly with mPng1p [45]. A more recent study [46], in which the three-dimensional structure of the PUB domain (residues 11-109) of hPng1p was solved, found that this domain contains a conserved p97 binding site. Furthermore, a GST pull down competition experiment showed that mPng1p mediates the interaction between p97 and mHR23B forming a ternary complex [43]. Because p97 is known to bind to the ER membrane protein AMFR (an E3 ligase), a model in which p97, mPng1p, mHR23B mediate interaction of the ER with the proteasome was proposed (Figure 1.6b).

Interestingly, in two *in vivo* studies, it was found that a glycoprotein Deg-Sec62 was degraded through a Ufd2 (a Ub chain elongation factor)- and Rad23- dependent pathway, and that its degradation was unaffected in *PNG1*-deleted cells, indicating that not all glycoproteins are degraded by the Png1p / Rad23 pathway [38, 44]. In contrast, the *in vivo* degradation of RTA is Png1p-dependant but not regulated by Ufd2. These findings suggested that Rad23 might regulate the degradation of distinct ERAD substrates through its interactions with various cofactors such as Png1p and Ufd2 that appear to be involved in specific degradation pathways [38].

1.9 Common Enzymatic Properties of Png1p

The complete lack of both the sequence and structural homology between type I and III PNGases indicates that type III PNGases (Png1p) may have different catalytic mechanisms.

1.9.1 Effects of pH and Temperature

Unlike bacterial PNGase F and plant PNGase A, which work well in basic (pH 8.5) or acidic (pH 4.5) environment respectively [5, 10, 11], PNGases from budding yeast (ScPng1p) and mouse (mPng1p) show maximum activity at neutral pH, ranging from about 6.5 to 7.5 depending on the source of the enzyme and the buffer used [4, 21, 28, 47]. ScPng1p and mPng1p exhibit the maximum activity at 37℃ and 30℃ respectively, and both become unstable when exposed to temperatures above 37℃ over a long period of

time [4, 47, 48]. Due to the high sequence conservation between the various cytoplasmic PNGases, it is reasonable to expect that other type III PNGases have similar biochemical properties.

1.9.2 Effects of Metal Ions and Dithiothreitol (DTT)

Several studies [4, 7, 21, 28] have shown that both budding yeast and mammalian PNGases are not metalloproteins. However, recent crystal structural studies [27, 29] revealed that zinc is bound to yPng1p/mPng1p in 1:1 ratio and that the addition of EDTA abolishes the deglycosylation activity of yPng1p.

All type III PNGases require the presence of DTT (at least 1 mM) to be active *in vitro*. The effect of DTT addition was shown to be reversible. Moreover, both SpPng1 and mPng1p are inhibited by the thiol-modification reagent *N*-ethylmaleimide (NEM) [4, 7, 21, 28]. These properties suggest that at least one free thiol group (-SH) is needed for PNGase activity, and Cys-191 which forms part of a putative catalytic triad is thought to act as the nucleophile in the hydrolysis reaction [15]. The requirement for a reducing environment is also reminiscent of the finding that, beside the catalytic motif, there are two conserved thioredoxin-like 'Cys-X-X-Cys' motifs in the core sequence of Png1p, which are also essential for the enzyme activity. From the results of crystal structure studies, it is clear now that four Cys residues in two CXXC motifs should be maintained in a reduced state to sequester one zinc ion.

1.9.3 Substrate Specificity: PNGase Recognises Both the Structure of N-linked Glycan and the Peptide

Compared with bacterial PNGase F and plant PNGase A, cytoplasmic PNGases (from yeast and mammal) are more selective for substrates in terms of both the structure of the glycan and the peptide / protein.

Studies [4, 28, 47] showed that neither GlcNAc-peptide nor glycosylasparagine were substrates for the cytoplasmic PNGases, suggesting that the enzymes require the presence of additional structural elements on both the glycan and peptide to bind to the active site. On the one hand, a study of carbohydrate-binding properties of mPng1p indicated that the enzyme requires a trisaccharide such as Man β 1→4GlcNAc β 1→4GlcNAc β 1→peptide

as the minimum carbohydrate attached to the peptide. In contrast, PNGase F and PNGase A are active on *N, N'*-diacetyl-chitobiosyl-peptides [48]. The best substrates for mPNGase were shown to be those glycopeptides that contain a pentasaccharide $Man_3GlcANc_2$ structure [48]) (Figure 1.7). The most common types of *N*-linked glycans are "high-mannose-type", "complex-" and "hybrid-" type glycans, all of which share a common pentasaccharide $Man_3GlcANc_2$ core structure. High-mannose glycans have only mannose residues added to the core, whereas complex-type contains no mannose residues other than those in the core. The hybrid-type glycans have a mixed characteristic of high-mannose-type and complex-type. Moreover, one or more additional monosaccharide resides may branch from the core in hybrid-type *N*-glycans. Although cytoplasmic PNGases are able to deglycosylate glycopeptides bearing high-mannose-, complex- or hybrid- type of N-glycan [48 Suzuki 1994], the results of some studies suggested that the enzyme could distinguish high-mannose from complex, hybrid and sialyated glycopeptides, and preferentially deglycosylated high-mannose glycopeptide [4, 17, 48].

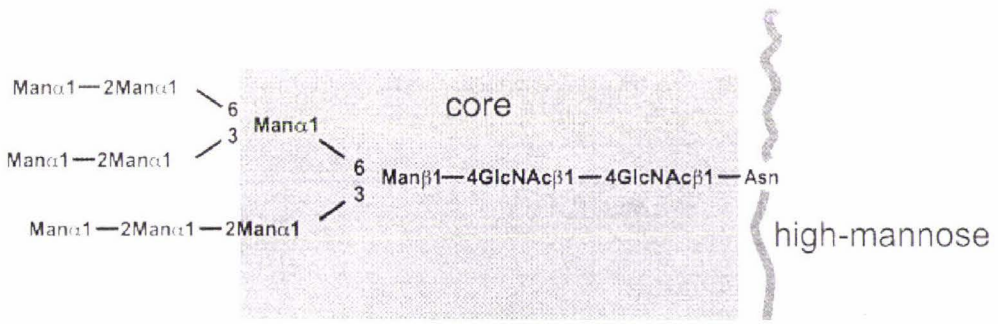


Figure 1.7 High-mannose-type glycopeptide/glycoprotein substrate of Png1p. High-mannose-type glycan chain contains nine α-mannose residue. The box encloses a common pentasaccharide $Man_3GlcANc_2$ core at the reducing termini of the glycan. Other types of glycan chain such as complex- and hybrid-types also contain this common core but are different in their sugar composition and branching patterns. In hybrid- and complex-type glycans, additional monosaccharides may be linked to the core residues. (Figure created by Rudolf Tauber, Institut fur Klinische Chemie und Pathobiochemie und)

It was also shown that cytoplasmic PNGases have different activities with substrates that have identical N-linked glycans but different amino acid sequences [47]. Moreover, the finding [21] that PNGase HO (hen oviduct) was able to specifically hydrolyse Di-*N*-glycosylated ovalbumin to generate mature mono-glycosylated ovalbumin strongly supports the idea that cytoplasmic PNGase was able to recognise structural elements of the peptide backbone.

For a long time, it was thought that cytoplasmic PNGases could only act on glycopeptides, but not full-length glycoproteins *in vitro*. Many glycopeptides, such as those derived from fetuin and hen ovalbumin peptides are well characterized substrates for both the yeast and mammalian PNGases [13, 21, 28]. These PNGases are thought to participate the proteasome degradation of misfolded proteins transported out of ER [16, 18]. Therefore, deglycosylation of glycoproteins by the PNGase *in vivo* may require the protein to be unfolded or denatured [13].

Recent studies have, in fact, demonstrated that cytoplasmic PNGases are able to act on full-length denatured glycoproteins *in vitro* [17, 49, 50]. Hirsch *et al.* observed that the α-chain of T-cell receptor (TCR α) and human class I MHC heavy chains (MHC HCs) could be deglycosylated by yeast PNGase and its mammalian homologue both *in vitro* and *in vivo* [17, 20]. Since both the TCR α and MHC HCs are known to be processed through an ER-associated degradation pathway *in vivo* [51], this finding further implicated the participation of cytoplasmic PNGases in the degradation pathway. Subsequently, it was found that yeast PNGase was able to distinguish between native and non-native (misfolded or denatured) glycoproteins, and attack only the latter [49, 50]. For example, yeast PNGase could deglycosylate denatured or misfolded bovine ribonuclease RNaseB, MHC HCs, carboxypeptidase Y (CPY) and ovalbumin, but not their native counterparts [49, 50]. This trait is thought may not be unique to yeast PNGase because other endo-glycosidases such as PNGase A are active only on denatured or partially denatured substrate.

1.10 Aims of this Investigation

1) Cytoplasmic PNGases have been classified as members of transglutaminase-like superfamily [15]. Although a PNGase homologue from *Arabidopsis thaliana* (AtPng1p) has been shown to have transglutaminase activity, it was thought this property is unique to AtPng1p because there is clear evolutionary divergence between this enzyme and other eukaryotic PNGases [52]. Are other cytoplasmic PNGases also transglutaminases? The transglutaminase activity of PNGase Sp will be investigated. If it does exist, kinetic constants will be measured using DMPDA, Z-Gln-Gly.

2) The PNGase from yeast *Schizosaccharomyces pombe* (SpPng1p) was shown to be highly homologous to ScPng1p [4, 13]. While they may have similar kinetic properties—this has yet to be established. In addition, the requirements for co-factors such as zinc or other metal ions has yet to be rigorously investigated [4, 7, 21, 27-29]. This study will determine the biochemical properties of SpPng1p which will include: determination of Km and k_{cat}/Km values, testing the effects of pH, temperature, establishing the co-factor requirement. Furthermore, if SpPng1p is also found to be an active transglutaminase, the k_{cat}/Km values of SpPng1p for the different reactions will be determined.

3) Although the cytoplasmic PNGases are thought to use the transglutaminase-like catalytic triad (Cys, His and Asp) for their activity, the exact catalytic mechanism is still not clear. It has been found that the conserved residues Trp-220, Trp-231, Arg-210 and Glu-222 are essential for PNGase activity, but exactly why has not been determined. Mutation of these conserved residues results in complete loss of activity but this may be as a result of disrupting the conformation of the enzyme. To better understand the catalytic mechanism of Png1p, one goal of this project is to produce enough active recombinant native and mutant SpPng1 for use in X-ray crystallography studies. Although the structures of two orthologues (ScPng1p and mPng1p) have been solved, they have only been solved in complex with other proteins [27, 29]. This was because both proteins could not be produced in heterologous hosts in isolation in a soluble form. It would therefore be interesting to crystallise the uncomplexed protein to see if there are structural changes that occur when the protein is complexed. It is possible that such changes, if they occur, could be involved in some sort of signaling event in the glycoprotein ERAD pathway.

To achieve the goals mentioned above, the following steps will be taken:

1) Clone and over-express soluble, recombinant SpPng1p in a bacterial expression system.
2) Determine if the gene product has PNGase activity.
3) Determine if the gene product has transglutaminase activity.
4) Properly characterise the PNGase activity of SpPng1p.
5) Purify enough recombinant SpPng1p for crystal trials.

Chapter 2 — Materials and Methods

2.1 Materials and Equipments

2.1.1 Materials

Solvent: The water used in this study was Milli Q™ deionised water. All the solvents used were analytical grade. Acetonitrile was HPLC grade and TFA was protein-sequencing grade.

Plasmids and bacterial strains: The plasmids and bacterial strains used are listed in Table 2.1 and 2.2 respectively.

Chemicals, commercial Kits:
Luria Broth base (LB) (Luria Broth was prepared as a 2.5% solution then autoclaved); 1 kb DNA ladder were from Invitrogen™ Life Technologies.

Agar bacteriological (For LB-agar plates, 2.5 g LB and 1.5 g agar were added into 100 mL water prior to autoclaving were from OXOID Ltd.

*Hind*III and *Bam*HI restriction endonuclease, *Taq* and *Pwo* polymerase, shrimp alkaline phosphatase, T4 DNA ligase, DNase I, Complete™ EDTA-free Mini protease inhibitor, High pure PCR product purification kit and high pure plasmid isolation kit were from Roche™ Diagnostics.

Hen egg albumin, trypsin, serum albumin bovine, transglutaminase (from Guinea pig liver), Tris base, HEPES, TCEP, EDTA, imidazole, Nickel (II) chloride hexahydrate; fetuin (from fetal calf, Sigma), hen egg ovalbumin, α1-acid glycoprotein (from bovine blood) and ribonuclease B (from bovine pancreas) were from Sigma-Aldrich Inc.

SDS-PAGE molecular weight makers (low range), Chelex-100 resin (A.R. grade) and P-4 resin (50-100 mesh) were from Bio-Rad® Laboratories Inc.

Crystal Screen™, Crystal Screen 2™ and Additive Screen™ were from Hampton Research. Structure Screening 1™ and 2™ screen reagents, and PACT *premier*™ screens were from Molecular Dimensions Ltd.

Recombinant tobacco etch virus (rTEV) protease was produced "in house".

2.1.2 Equipment

Thermal cycler — (Biometra™ *T* gradient)

Bench top microcentrifuge — (Eppendorf MiniSpin Plus)

Refrigerated Superspeed Centrifuge — (SORVALL Evolution™ RC)

- 80°C refrigerator — (Forma Scientific, Inc.)

Electrophoresis apparatus and power supplier — (Bio-Rad™ Mini-PROTEAN 3 Cell, and Bio-Rad MiniSub Cell GT apparatus)

ND-1000 spectrophotometer — (NanoDrop®)

UV Visual spectrometer — (Varian® Cary50)

HPLC system — (P580 binary pump, Dionex, UK)

ZMD mass spectrometer — (Micromass®, Waters®, USA)

M@ LDI mass spectrometer — (Micromass®, Waters®, USA)

Low pressure flow pump — (Bio-Rad™, Econo)

French pressure cell — (Wabash™, Aminco Instruments Co.)

FPLC system — (ÄKTA™ Explorer, GE Healthcare)

2.2 General Methods

2.2.1 Agarose Gel Electrophoresis

DNA fragments were separated on the basis of size by agarose gel electrophoresis. Typically, DNA samples containing 1X loading dye (10% glycerol and 0.15% bromophenol blue) were electrophoresed through 1% standard agarose TAE (40 mM Tris-HCl, 20 mM acetic acid, 1 mM EDTA, pH 8.0) gel. The samples were routinely electrophoresed alongside a 1 kb DNA ladder as a size marker in a Bio-Rad MiniSub Cell GT apparatus at 80 volt for ~35 minutes. The gels were stained in a 0.5 μg/mL ethidium bromide for 20 minutes and rinsed in water for 5 minutes before being visualised by UV light. Gel images were captured using a Gel Doc 2K (Bio-Rad) gel imaging system.

2.2.2 SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Protein molecules were fractionated by SDS-PAGE using a discontinuous buffer system according to the protocol of Laemmli [53]. Briefly, the discontinuous gel system consists of 4% acrylamide stacking, and 12% acrylamide separating gels containing 0.1% SDS. 5-10 μL protein samples were firstly mixed with 5-10 μL 2X sample buffer (4 mL H_2O, 1 mL 0.5 M Tris-HCl pH 6.8, 0.8 mL glycerol, 1.6 mL 10% (w/v) SDS, 0.77 g DTT, 0.5% (w/v) bromophenol blue) and boiled for 2 minutes, before being loaded onto the stacking gel. Electrophoresis was carried out at 200 volts for ~45 minutes in a Bio-Rad Mini-PROTEAN 3 Cell apparatus until the front line of dye has reached the end of the gel. The gels were visualised by being stained with 0.1% Coomassie Brilliant Blue R-250 then destained with a 25% methanol, 65% water and 10% acetic acid mixture with agitation.

2.2.3 Native Polyacrylamide Gel Electrophoresis (Native-PAGE)

Native-PAGE was carried out in a basically same way as SDS-PAGE (section 2.2.2) except that no SDS was used in both the gels and electrode reservoir buffer. Before being loaded onto the gel, the protein samples were mixed with 2X sample buffer without SDS, and the samples were not boiled (section 2.2.2).

2.3 Molecular Cloning

2.3.1 Cloning *PNG1Sp* into Different Vectors

2.3.1.1 *Polymerase Chain Reaction (PCR) Amplification of S. pombe PNG1 Gene*

Before PCR amplification, the *PNG1Sp* gene sequence (NCBI database accession No. AL031852, Appendix I) was analysed for restriction enzyme recognition sites using NEBcutter software (New England BioLabs). The full-length *PNG1Sp* gene was then amplified from the vector pProEX_HTb_*PNG* (Table 2.1) using a pair of primers carrying *Bam*HI and *Hind*III restriction sites as shown below:

PNG1.BamHI.Fwd: TTTCAGGGATCCATGGATTTTCATGCGATTTC
PNG1.HindIII.Rev: CAGCCAAGCTTACAATTATTTTCCTGCTTCC

A PCR reaction was performed using 0.7 ng of pProEX_HTb *PNG* with 1U of *Pwo* polymerase, 0.2 mM dNTPs, 5 μM of each of the forward and reverse primers, and 1× PCR buffer (Roche). The reaction was carried out in a DNA thermal cycler programmed for 32 three-step cycles of 1-minute melt at 95℃, 1-minute anneal at 55℃ and 1.5-minute elongation at 72℃. A further elongation at 72℃ was performed for 7 minutes at the end of the program. A positive PCR control using another vector carrying the *PNG1Sp* gene and two gene-specific primers was included, as was a negative control in which no template DNA was added to the reaction mixture.

2.3.1.2 *Restriction Endonuclease Digestion*

Before restriction digestion, the PCR products were analysed by agarose gel electrophoresis, and purified using the High Pure PCR Product Purification Kit (Roche) according to the manufacture's directions. For the digestion of vectors or PCR products, the following reaction was typically used: ~2 μg of vector or PCR product was double digested with 5 U *Bam*HI and *Hind*III restriction enzymes, in SuRE/Cut buffer B (Roche), in a total reaction volume of 50 μL. After incubation at 37℃ for at least 3 hours, the reaction was quenched by heating at 65℃ for 15 minutes. A single digest with either *Bam*HI or *Hind*III endonuclease and a negative control in which no restriction endonuclease was used were also included as controls for enzyme activity. Before ligation,

both the digested vectors and inserts were purified using a High Pure PCR Product Purification Kit then examined by agarose gel electrophoresis.

Three vectors were used in this study. They are: pET32a_HTBH, pSUMO_BXH and pMal_CHTBH (Table 2.1. A detailed map of pET32a_HTBH is also shown in Appendix II). All vectors were double digested with *Bam*HI and *Hind*III restriction endonuclease, and used in the following ligation reactions.

2.3.1.3 *Ligation*

For ligation, a reaction mixture with a total volume of 10 µL containing 1X ligation buffer (660 mM Tris-HCl, 50 mM $MgCl_2$, 10 mM DTT, 10 mM ATP, pH 7.5), 1 U bacteriophage T4 DNA ligase, ~ 30 ng digested PCR product and ~ 150 ng vector was incubated at room temperature for 48 hours, and then heated at 65℃ for 15 minutes to deactivate the enzyme. Because the concentration of both vector and insert was based on the intensity of electrophoretic bands on an agarose gel, three molar ratios (1:3, 1:1 and 3:1; vector : insert) were normally tried in the reaction to ensure high ligation efficiency. As controls, the ligation reaction, without insert, was incubated with or without T4 DNA ligase. The products of these reactions were used as controls in transformation to provide background for single cut (self-ligation) and uncut vector (intact) respectively.

2.3.2 Transformation, Colony Screening and Plasmid DNA Preparation

The *PNG1Sp* gene- and ampicillin resistance gene-containing vectors (in ligation reaction mixture, section 2.3.1.3) were transferred into a competent cloning strain XL-1 *E. coli* (Table 2.2) using heat shock. Briefly, up to 5 µL of the ligation reaction was mixed with 50 µL of XL-1, and incubated on ice for 15 minutes. Cells were then heat-shocked at 42℃ for 60 seconds then immediately placed on ice for further 2 minutes, after which 450 µL of SOC media (2% peptone, 0.5% yeast extract, 0.05% NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 20 mM glucose) was added to the cells and the mixture incubated at 37℃ for 1 hour to revive the cells. 100 µL of the above-said transformation culture was then plated onto LB-agar plates with ampicillin selection (F.C. 100 µg/mL) and incubated at 37℃ overnight. Three control experiments were carried out in parallel: transformation of XL-1 cells without additional DNA followed by plating onto either ampicillin + or - LB-agar media,

and cells with uncut vectors containing amp-resistance gene followed by plating onto ampicillin-containing LB-agar media.

Transformants were screened by whole cell PCR screening. Typically, seven colonies were picked from the aforementioned overnight culture and used to inoculate 5 mL of fresh liquid LB media (2.5% Luria broth base, 100 μg/mL ampicillin), followed by incubation overnight at 37℃ with shaking (200 rpm). 50 μL of the culture was then centrifuged (14,000*g*, 0.5 minute) and the pellets resuspended in 400 μL water followed by boiling for 5 minutes to break the cells. Finally, the cell lysates were re-centrifuged for 5 minutes to remove cell debris and the supernatant used as template solution in following PCR reactions: 1 μL of 10X PCR buffer (Bio Therm™), 1 μL of template solution, 0.2 U *Taq* DNA polymerase, 0.2 mM dNTPs, 0.16 μM each of a vector-specific forward and a gene-specific reverse primer, in total volume of 10 μL. The reaction conditions were identical to those used for *PNG1Sp* gene amplification (section 2.3.1.1). The vector specific primers are shown as below:

Lac_Operon fwd:	GGAATTGTGAGCGGATAACA
T7_Terminator rev:	GCTAGTTATTGCTCAGCGGT

Small-scale plasmid preparations of positive clones were carried out using a High Pure Plasmid Isolation Kit (Roche, Inc.) according to manufacturer's instructions. Normally, 3-8 μg pure plasmid DNA (in 100 μL elution buffer, 30-80 μg/mL, quantitated by measuring the UV absorbance at 230 nm using a NanoDrop® ND-1000 spectrophotometer) could be obtained from 3 mL of cell culture (depending on the cell density). Three recombinant plasmids were constructed in this study: pET32a_HTBH_*PNG1*, pSUMO_BXH_*PNG1* and pMal_CHTBH_*PNG1*. These were stored at -20℃ and used for the transformation of expression *E. coli* host cells (section 2.4.1).

2.3.3 Sequence Analyses of DNA

For DNA sequencing, 300 ng recombinant plasmid was mixed with 1 pmol of either vector-specific forward or reverse primer in a total volume of 15 μL in a 200 μL PCR tube. DNA sequencing was carried out by the Allan Wilson Centre Genome Sequencing Service (Massey University). In brief, the BigDye™ Terminator Version 3.1 Ready Reaction Cycle

Sequencing Kit was used on an ABI 3730 Genetic Analyser (Applied Biosystems Inc.). The algorithms used in the analyses of the results were:

1) Chromas 2 (version 2.31, Technelysium Pty Ltd. www.Technelysium.com.au)
2) AnnHyb (version 4.916. (c) Olivier Friard 1997-2004, http://bioinformatics.org/annhyb)
3) Protein Alignment software (http://www.ch.embnet.org/software/LALIGN_form.html)

2.4 Protein Expression in *Escherichia Coli*

2.4.1 Small-scale Expression Trials

To establish an expression system able to produce large concentrations of soluble PNGase Sp, small-scale expression trials were carried out using different *E. coli* strains with varying concentrations of the expression-inducing agent, IPTG, and varying temperatures. The Origami™ B (DE3) strain was tried first, as this strain had been used to successfully produce another construct of soluble PNGase Sp in the laboratory. While this makes no sense, as it provides a less reducing environment in the cytoplasm, and it is necessary that the cysteines are in a reduced state for this enzyme [4, 21, 28], it had been previously shown that the recombinant protein was insoluble in BL21 (DE3), but soluble when produced in Origami B (DE3) cells.

2.4.1.1 *Expressed in Origami™ B (DE3)*

Origami B (DE3) cells (Table 2.2) were transformed using heat-shock with the purified plasmid pET32a_HTBH_*PNG1* (section 2.3.2), cultured on LB-agarose plates using the methods described in section 2.3.2, except that the antibiotics used were ampicillin (100 μg/mL) plus tetracycline (12.5 μg/mL) and kanamycin (15 μg/mL). A single colony was used to inoculate 5 mL of LB broth containing the same antibiotics, and then incubated overnight at 37°C. 100 μL of the overnight culture was used to seed 5 mL of fresh LB broth, which was again incubated at 37°C until the optical density at 600 nm (OD_{600}) reached ~ 0.7, at which point expression of the recombinant PNGase Sp was induced by the addition of IPTG to a final concentration of 0.1, 0.2 or 0.5 mM respectively. Following antibiotic supplementation, the culture was incubated overnight at 16, 25, or 37 °C respectively. As a control, a culture was grown as above without the addition of IPTG.

Cells transformed with empty plasmid were also included as another negative control. The overnight culture was then analysed for protein production and solubility using SDS-PAGE (section 2.2.2).

When other *E. coli* strains such as Origami B (DE3) pLysS, Origami B (DE3) pLacI, Origami B (DE3) GroE, Rosseta™ (DE3) and BL21 (DE3) as expression hosts, the operations are basically same except that the antibiotics used for selection were different (Table 2.2).

2.4.1.2 *Total Protein and Solubility Analyses*

The protein production and solubility were analysed by SDS-PAGE. The amounts of total cell extract samples loaded onto each lane were normalized by OD_{600} measurement in order to use equal cell numbers. Briefly, a volume ($0.5/OD_{600}$ mL) of cell culture was centrifuged at 14,000*g* and the pellet was resuspended in 200 µL water, then mixed with 200 µL SDS-PAGE loading buffer, boiled for 5 min to break the cells, then centrifuged for 30 seconds before being loaded onto a 12% SDS polyacrylamide gel for electrophoretic analyses.

To check the solubility of recombinant proteins, 1 mL of cell culture was collected by centrifugation and the resulting pellet resuspended in 500 µL of lysis buffer (same as the loading buffer in section 2.5.2) plus 1× Complete™ Mini protease inhibitor (Roche), kept on ice and lysed by sonication using a VirSonic Digital 475 Cell disrupter at level 3 for 10 seconds followed by a second 10 second burst after one minute. The soluble and insoluble cell fractions were separated by centrifugation at 14,000*g* for 10 minutes at 4℃. The cell debris was resuspended in 500 µL of the same buffer. Both the supernatant (soluble fraction) and cell debris suspension (insoluble fraction) were then analysed by SDS-PAGE.

2.4.2 Large-scale Expression in Origami™ B (DE3) Strain

During large-scale expression, a single transformed Origami B (DE3) colony was used to inoculate 5 mL LB broth, followed by incubation with shaking (200 rpm) overnight at 37 ℃. 1 mL of this culture was used to further inoculate 100 mL LB broth. After incubation overnight under the same conditions, 10 mL of the culture was again used to inoculate one litre of LB broth and cultured under the same conditions. Protein expression was induced

by the addition of IPTG to a final concentration of 0.5 mM when the optical density of the culture at 600 nm (OD_{600}) was ~0.7. Before being incubated at 25℃ overnight, the culture was re-supplied with fresh antibiotics to maintain selection.

2.5 Purification of Thioredoxin-His_6-Tagged PNGase Sp

All steps in the purification protocol were carried out at 4℃ in order to minimize the denaturation and proteolytic degradation of the target enzyme.

2.5.1 Cell Lysis

Cells in one-litre overnight culture were harvested by centrifugation at 4,400*g* for 20 minutes. The pellet was resuspended in 10 mL of pre-chilled lysis buffer (same as the loading buffer used in IMAC purification, section 2.5.2) to which the appropriate amount of Complete™ Mini protease inhibitor (Roche) had been added, and then lysed by two passes through a Wabash™ French pressure cell (Aminco Instruments Co.) at 6,000 psi. A trace amount of DNase I was immediately added into the cell lysate to break down DNA molecules released from the cells. After centrifugation at 30,000*g* for 25 minutes to remove the cell debris, the supernatant was filtered through a 0.8-μM membrane before being subjected to affinity chromatography. Total protein in the cell lysate was quantitated using the Bradford method (section 2.5.6) and assayed for PNGase activity as described in section 2.6.

When lysed with lysozyme, the cell pellet from 100 mL of culture was resuspended in 1.5 mL lysis buffer (65 mM Tris-HCl buffer, pH 7.6; 10 mM $MgCl_2$; 1 mM TCEP and Complete™ Mini protease inhibitor). To this solution, 0.3 mg chicken egg white lysozyme was added and the solution left for 0.5 hour at room temperature, when it became very viscous. A trace amount of DNase I was added and the solution left for further 0.5 hours before being centrifuged at 14,000*g* for 20 minutes.

2.5.2 Immobilised Metal Affinity Chromatography (IMAC)

Recombinant PNGase Sp with a N-terminal thioredoxin-hexohistidine-tag was firstly purified by affinity chromatography method using immobilised Ni^{2+} as the affinity matrix.

The specific interaction between Ni^{2+} and the hexohistidine patch promotes preferential binding of the fusion protein over other cellular proteins. After washing with loading buffer to remove these, the fusion or tagged protein can be eluted from the resin-bound Ni^{2+} later by adding imidazole into the mobile phase. Imidazole has a similar structure to that of histidine, and thus competes with the fusion protein for access to the immobilised Ni^{2+}.

Normally, 10 mL of filtered cell lysate (~15-20 mg/mL) was loaded onto a Ni^{2+} charged-chelating sepharose™ column (GE Healthcare, bed volume: 5 mL), which was pre-equilibrated with loading buffer (50 mM Tris-HCl, pH 7.6, 0.5 M NaCl, 1mM TCEP, 10 mM imidazole). Once loaded, the column was extensively washed with 5 column volumes of the loading buffer (5 C.V. = 5×5 mL = 25 mL). Bound proteins were then eluted in a stepwise way using loading buffer with increasing concentrations of imidazole (4 C.V. each with 40 and 60 mM imidazole, then 2 C.V. with 250 mM imidazole). The flow rate was 1 mL/min for each step except that the initial loading was performed at 0.4 mL/min to ensure sufficient time for the binding of the fusion protein to immobilised Ni^{2+}. The flow rate was controlled using a low pressure Bio-Rad Econo pump. Fractions of one column volume (5mL) were collected, into which Complete™ Mini protease inhibitor and EDTA (Final concentration: 5 mM) were immediately added in order to inhibit any proteolytic degradation. The fractions were then analysed for the presence of rPNGase Sp by SDS-PAGE.

2.5.3 Proteolytic Removal of Fusion Tag

The N-terminal thioredoxin-hexohistidine tag was removed by recombinant tobacco etch virus (rTEV) protease digestion. The purest IMAC fractions (containing mainly rPNGase Sp) were pooled and buffer-exchanged into rTEV cleavage buffer (10 mM Tris-HCl, pH 7.6, 0.1 M NaCl, 1mM TCEP, 1 mM EDTA), then concentrated to approximately 5 mL by ultra filtration using a VivaSpin concentrator (20-mL, 10 kDa mwt cut off, VivaScience). 200 µL 1 mg/mL rTEV protease was added into the sample solution which was then incubated at 10℃ for 48 hours. The cleavage reaction was monitored by SDS-PAGE.

2.5.4 Further Purification by IMAC

The proteolytically cleaved PNGase Sp was further purified using IMAC, as both the released N-terminal tag and any un-cleaved protein bound to the Ni^{2+}-charged resin, while protein that had been cleaved contained no tag, and hence flowed through the column. All the chromatographic conditions were same as that used in the initial IMAC purification (section 2.5.2), except that the flow through fractions were collected and analysed by SDS-PAGE. The fractions containing almost pure PNGase Sp were pooled, concentrated with 10 mM HEPES buffer (pH 7.5, 1 mM TCEP, 1 mM EDTA) to approximately 15-30 mg/mL, snap-frozen in liquid nitrogen in 50 µL aliquots, and stored at - 80℃ for use later.

2.5.5 Size Exclusion Chromatography (SEC)

For crystallisation trails, the protein was further purified by SEC to remove any trace impurities from the 2nd IMAC purification step and to ensure a homogeneous population of molecules was obtained. Briefly, 200 µL of concentrated sample from the 2nd IMAC step was loaded onto a gel filtration column (Superdex™ 75 resin, GE Healthcare, 10× 300mm) pre-equilibrated with 10 mM HEPES buffer (pH 7.5, 1 mM TCEP, 1 mM EDTA) and eluted with the same buffer at a flow rate of 0.4 mL/min. The elution was monitored at 214 and 280 nm and 1 mL fractions were collected for SDS-PAGE analyses. An ÄKTA™ Explorer (Amersham, GE Healthcare) FPLC system was used for all the chromatography.

2.5.6 Protein Quantitation

Protein samples from each purification step were quantitated using the Bradford Assay [54]. The assay is based on the observation that the absorbance maximum for an acidic solution of Coomassie brilliant blue G-250 shifts from 465 nm to 595 nm when binding to positively charged residues (lysine, arginine and histidine) and aromatic residues. The high extinction coefficient of protein-color reagent complex at 595 nm results in the assay being very sensitive with a minimum detection limit of 1 µg protein. Within the linear range of the assay, the concentration of protein is directly proportional to the absorbance at 595 nm. Bovine serum albumin (BSA) was used as an external standard in this assay.

2.6 PNGase Activity Assay

PNGase activity of rPNGase Sp was measured using a discontinuous high performance liquid chromatography (HPLC)-based assay [6] using hen egg ovalbumin derived 11-mer glycopeptide as the substrate. The glycopeptide was prepared by cyanogen bromide digestion of hen egg ovalbumin followed by SEC purification as outlined in Appendix III. The 11-mer glycopeptide contains a complex, biantennary oligosaccharide linked to the asparagine residue (Figure 2.1) with nine uniformly distributed hybrid and high-mannose glycoforms (Figure 1.7) [55]. This substrate is susceptible to deglycosylation by both bacterial PNGase F and rPNGase Sp. The principle of this assay is that the natural (substrate) and deglycosylated (product) forms of the glycopeptide present in the assay reaction show different hydrophobicity, and thus can be separated on a reverse phase C-18 column. The product is more hydrophobic than substrate due to the loss of glycan chain, and thus elutes later. Both PNGase F and rPNGase Sp are eluted later, and do not interfere with the assay.

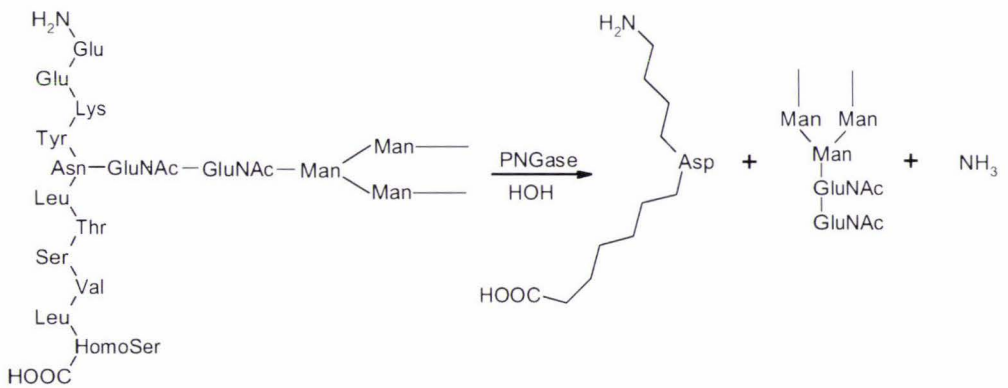


Figure 2.1 Deglycosylation reaction catalysed by PNGases. PNGase catalyses the hydrolysis of the amide bond form between the 11-mer peptide-bound Asn residue and the N-acetylglucosamine, generating an Asp residue in the in place of the asparagine and a intact oligosaccharide with a N, N'-diacetylchitobiose at its reducing terminus.

Generally, 10 μL of enzyme solution (0.5 mg/mL) was incubated with 85 μL of substrate solution (0.5 mg/mL, in activity assay buffer: 10 mM HEPES buffer, pH 7.5, 1 mM TCEP, 1 mM EDTA) at 25℃ overnight. The reaction was quenched by addition of 10 μL of formic acid and then boiled for 5 minutes. Before being loaded onto a C-18 column, the reactions were centrifuged at 14,000*g* for 15 minutes. Chromatography conditions used were: C-18 column (RP-C18 Jupiter Series, 4.6 × 250 mm, particle diameter 5 μm;

Phenomenex); Flow rate of 1 mL/min (P580 binary pump, Dionex); 15-minute gradient elution from 20% acetonitrile / 0.1% trifluoroacetic acid (TFA) to 40% acetonitrile / 0.08% trifluoroacetic acid; Detection at 214 nm.

When a fluorescein isothiocyanate (FITC)-labeled ovalbumin glycopeptide was used as substrate in kinetic investigations (section 2.9.3, 2.9.4 and 2.9.5), the elution procedure was modified as follows: (solvent A: 0.1% TFA in water; solvent B: 0.08% TFA in acetonitrile) 1) 80% A and 20% B → 60% A and 40% B over 15 minutes; 2) 60% A and 40% B → 30% A and 70% B over 10 minutes; 3) 30% A and 70% B → 80% A and 20% B over 5 minutes plus 5-minute equilibration. The substrate and product were detected using a fluorescence detector (Dionex; excitation wavelength: 495 nm, emission wavelength: 520 nm). The product was quantitated using a standard method, and the data analysed using Chromeleon™ Client software.

2.7 Transglutaminase Activity Assay

A colorimetric method known as hydroxylamine assay [56] was used to test for transglutaminase activity. Transglutaminase catalyses the formation of a covalent bond between a free amine group (e.g., protein- or peptide-bound lysine, ammonia, hydroxylamine, monoamino acid) and the γ-carboxamide group of protein- or peptide bound glutamine. The principle behind this assay is that the reaction of the γ-carboxamide group donor carbobenzyloxy-glutamyl-glycine (Z-Gln-Gly) with hydroxylamine in the presence of calcium ion forms hydroxamic acid which, in turn, reacts with ferric (III) ion to form a brown ferric hydroxamate complex (λ_{max}: 525 nm) that can be measured spectrophotometrically (Figure 2.2).

Procedure: To a solution containing 0.1 mL of 1 M Tris-acetic acid buffer (pH 6.0), 0.025 mL each of 0.1 M $CaCl_2$, 0.025 M EDTA, 2 M $NH_2OH{\cdot}HCl$ (pH 6.0, adjusted with NaOH), and 0.075 mL of carbobenzyloxy-Gln-Gly in a 1.5 mL Eppendorf tube, 0.02 mL of rPNGase Sp was added. The final volume is adjusted to 0.5 mL with water. After a 10-minute incubation at 37℃, 0.5 mL of the ferric chloride-trichloroacetic reagent (0.5 g $FeCl_3$, 1.5 g TCA, dissolved in 27.5 mL H_2O and 2.5 mL concentrated HCl) was added. A change in color from yellow to brown indicates the enzyme has transglutaminase activity.

For quantitation, the color is read immediately against a blank containing no enzyme at 525 nm. Another reaction using transglutaminase (from Guinea pig liver, Sigma) was used as a positive control.

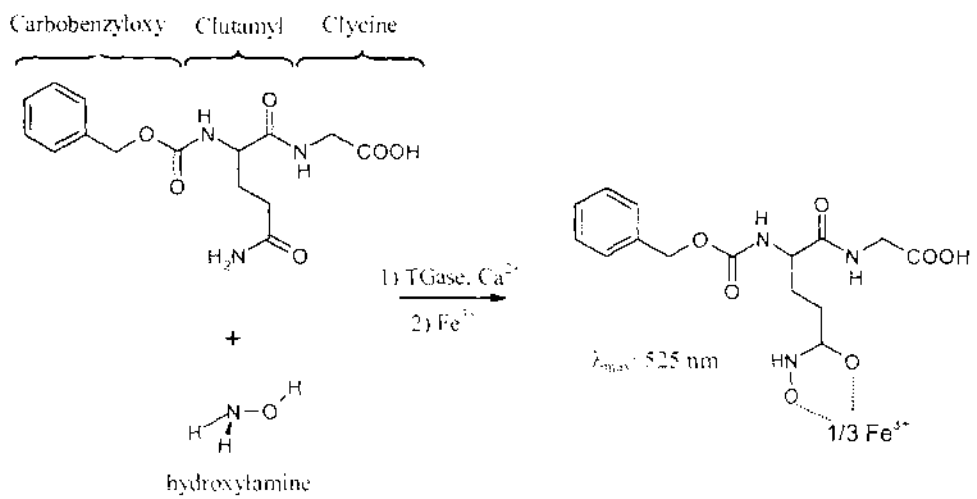


Figure 2.2 Transglutaminase activity assay. The glutamyl donor Z-Gln-Gly reacts with the free amine group on hydroxylamine to produce a hydroxamic acid which, in turn, forms a complex with ferric (III) ions that has a maximum absorbance at 525 nm. In the complex, three hydroxamic acid molecules chelate one Fe^{3+} through six coordination bonds.

2.8 Mass Spectrometry Analyses

Mass spectrometry technology was used to confirm the identity of the gene expression product in this study. Other common methods such as protein N-terminal sequencing and western blotting can also be used for the same purpose.

2.8.1 Peptide-mass Fingerprinting

Peptide-mass fingerprinting (PMF) is an analytical technique used to identify proteins by matching their constituent fragment masses (peptide masses) to the theoretical peptide masses generated from a protein or DNA database [57]. In short, an unknown protein of interest is virtually cleaved into peptides with certain protease such as trypsin. Because the digestion of every protein results in a set of peptides with unique masses, this pattern can be used to identify the protein, after accurately measuring the mass of peptides on a MALDI-TOF (matrix assisted laser desorption ionization time-of-flight) mass

spectrometer, a PMF database search is carried out (i.e. comparing the observed masses with the genome which has been previously translated into proteins and cut into peptides with the same protease using software programs). The data are statistically analysed by computer, and the identification is accomplished by matching the determined peptide masses to the theoretical peptide masses from a database. The advantage of PMF technique is that the *de novo* sequencing of proteins is not required. A disadvantage is that the protein sequence must be presented in the database. Moreover, this technique is susceptible to interference with impurities. Thus, it is ideal to purify the sample on a two-dimensional gel prior to PMF analyses.

Matrix-assisted laser desorption/ionisation is a "soft" ionisation technique that produces intact singly-protonated quasimolecular ions of biomolecules such as proteins, peptides and sugars, in mass spectrometry [58]. The ionisation is triggered by a laser beam. Generally, the analyte solution is pre-mixed with a chemical matrix solution consisting of crystallisable molecules such as cinnamic acid derivatives, and then spotted onto a sample target (a metal plate used for supporting the matrix). After the solvents vaporise, the matrix molecules re-crystallise with the analyte spreading throughout the crystals. In this case, the sample is "protected" by the matrix from being destroyed by the direct laser beam during desorption and ionisation because the matrix molecules absorb the majority of the laser light energy. The matrix also assists the irradiated sample to vaporize by forming a rapidly expanding matrix plume that aids transfer of the protein ions into the mass analyser. After being accelerated in an applied electrostatic field, the singly-protonated molecular ions drift though a field-free area of a TOF (time-of flight) mass spectrometer, and reach the detector in a time that depends on their mass to charge ratios (m/z). Based on this, the mass can be calculated from the accurately measured flight time of molecular ions.

Procedure:

1) In-gel tryptic digestion of proteins: Protein spots (from the silver-stained 12% SDS 2-D gel) or bands (Coomassie stained 12% SDS 1-D gel) were excised using a clean razor blade. These were then cut into approximately 1 × 1mm pieces on a clean glass surface, and transferred into 1.5 mL Eppendorf tubes. The de-staining and trypsin digestion were carried out according to the procedure described by David Miyamoto (Appendix IV "In-

gel tryptic digest for protein ID by mass spectrometry"). The peptides obtained were dried and stored at -20℃ until further use.

2) MALDI-TOF mass spectrometry: The rapid-evaporation method was used for sample preparation for MALDI-TOF analyses [59]. Briefly, 1 μL peptide matrix/nitrocellulose solution (2.5 mg nitrocellulose, 10 mg α-cyano-4-hydroxy-trans-cinnamic acid in 0.25 mL acetone and 0.25 mL isopropanol, prepared fresh) was applied to the sample target and left to evaporate to dryness. The dry peptide sample from the tryptic digest was dissolved in 3-5 μL 20% acetonitrile and 1% formic acid, then 1 μL of this solution was pipetted on to the matrix and evaporated to dryness at room temperature. The organic solvent in the sample was maintained below 30% to avoid the complete dissolution of the matrix layer by the sample solution. The sample target was inserted into the Micromass® M@LDI mass spectrometer with a time-of-flight analyzer (Waters®, USA) and analysed in positive-ion reflection mode. A sample of a peptide of known size, hypertensin I (DRVYIHPFHL, m.w. 1296.5 Da), was prepared in an identical way and measured to calibrate the instrument.

2.8.2 Electrospray Mass Spectrometry

Since the biochemically purified rPNGase Sp showed heterogeneity (double-banded) on SDS-PAGE gel, electrospray ionisation mass spectrometry (ESI-MS) was used to determine an accurate mass difference of the two species. Such analyses should allow identification of the truncation if there is one. Like MALDI ionisation, electrospray is another "soft" ionisation technique that produces almost exclusively multiply charged intact biomolecular ions instead of ion fragments [60]. The increase in the mass/charge ratios of the ions produced pushes them into the detectable range of most mass spectrometers, thus widening the molecular mass range for analyses.

In this technique, a solution of protein molecules suspended in a volatile solvent is introduced into a narrow capillary at 5 kV, and nebulized into small charged droplets about 10 μm diameter. Although the exact mechanism of droplet and ion formation remains unclear, it is known that further desolvation occurs, and the droplets shrink resulting in singly and multiply protonated protein ions, designated as $(M + nH)^{n+}$, being released into gas-phase. The ions then continue to the mass analyzer where they are detected. The molecular weight of protein is then calculated from observed m/z and charge

number by computer software, and used to identify the molecule. The advantage of ESI-MS is that it can be used in combination with several separation techniques such as high performance liquid chromatography (HPLC-MS) and capillary electrophoresis (CE-MS) as electrospray serves as an interface for transferring ions from the liquid-phase to the gas-phase.

Procedure: The protein sample after rTEV protease cleavage was firstly purified using RP-HPLC (Jupiter Series reverse phase C18 analytical column, 4.6 × 250 mm, Phenomenex; Flow rate: 1 mL/min. 40-minute gradient elution from 10% acetonitrile / 0.1% trifluoroacetic acid to 100% acetonitrile / 0.08% trifluoroacetic acid; Detection at 214 nm). Different elution peaks were collected manually, concentrated by Speed-Vac, and analysed by SDS-PAGE. The fraction containing almost pure expression product was quantitated by UV absorbance. When prepared samples for ESI-MS analyses, acetonitrile (final concentration: 50%, v/v) and formic acid (final concentration: 1%, v/v) were added into the fraction. The final concentration of analytes was approximately 10 pmol/μL. The sample solution was sprayed into the mass spectrometer through a capillary tube using a syringe pump (74900 series, Cole-Parmer Instrument Company), and then ionised and analysed on a Micromass® ZMD mass spectrometer with a single quadrupole analyzer (Waters®, USA).

2.9 Characterisation of Recombinant PNGase Sp

Some enzymatic properties of recombinant PNGase Sp were studied. They were substrate specificity, optimum temperature and pH for catalysis, effects of reductants and metal ions on the activity, and the Michaelis constant (K_m) and catalytic efficiency (k_{cat}/K_m) of rPNGase Sp using an ovalbumin-derived 11-mer glycopeptide as the substrate.

2.9.1 Substrate Specificity

To investigate the substrate specificity of rPNGase Sp, four well-characterised glycoproteins were used to analyse the deglycosylation activity of the enzyme. They were fetuin (from fetal calf serum and contains three sialylated triantennary *N*-linked complex sugar chains), chicken egg ovalbumin (has a single biantennary *N*-linked complex glycan),

ribonuclease B (from bovine pancreas and contains a single *N*-linked high mannose biantennary glycans), and α1-acid glycoprotein (from bovine blood and has 5 *N*-linked hybrid glycan(s) in which a fucose residue is linked to the proximal GlcNAc residue through an α 1→6 linkage). The common structural types of *N*-Glycans were briefly described in section 1.9.3. All the four substrates were used in both native and denatured forms in activity assays, and the deglycosylation reactions were detected using gel mobility shift assay.

Each substrate was made up to 1 mg/mL with activity assay buffer (50 mM MIB buffer: 12.5 mM malonate/18.75 mM imidazole/18.75 mM boric acid, pH 6.0, 1 mM TCEP, 1 mM EDTA) except for fetuin, which has multiple glycoforms, and thus a higher concentration (~ 2 mg/mL) was needed to visualise all the bands clearly by SDS-PAGE. To prepare the corresponding denatured substrates, the glycoproteins were boiled in 0.2% Thesit (a non-ionic detergent), 10 mM reducing agent dithiothreitol (DTT), 1 mM EDTA in 50 mM MIB buffer (pH 6.0) for 10 minutes. 100 μL of substrate (both natural and denatured) were incubated with 5 μL of rPNGase Sp solution (1 mg/mL) at 25 ℃ overnight, and the reaction was stopped by boiling for five minutes. Substrate deglycosylation was examined by SDS-PAGE (section 2.2.2). Both the fetuin and ovalbumin reaction mixtures were run on 12% gels, and the α 1-acid glycoprotein and ribonuclease B on 15% gels. The removal of glycans from the substrate can be visualised by a change in molecular weight on SDS polyacrylamide gels. Reactions in which either enzyme or substrate was absent were included as negative controls.

2.9.2 Effects of Reductant

It was reported that eukaryotic PNGases need the presence of reductants such as DTT (at least 1 mM) for activity *in vitro* [4, 21, 28]. To test if the activity *in vitro* is reductant dosage-dependent and hence study the relationship between the redox state of the thiol sidechains and the activity of rPNGase Sp, the thioredoxin-chimera was purified using buffers with varying concentrations of Tris (2-carboxy-ethyl) phosphine hydrochloride (TCEP), a stronger and more stable reductant compared to DTT [61], and the activity was measured and compared.

Briefly, 100 mL of *E. coli* Origami B (DE3)™ culture that overexpressed the enzyme was divided into four aliquots and each aliquot centrifuged at 4,400*g* for 20 minutes. The cell pellets were resuspended in 1-1.5 mL lysozyme lysis buffer and lysed as described in section 2.5.1. Four different concentrations of TCEP, 0.0, 0.25, 0.5 and 1.0 mM were used respectively. After lysis was complete the cell lysates were centrifuged at 30,000*g* for 25 minutes to remove the cell debris, the supernatants filtered through a 0.8 μM membrane before being loaded onto a Ni^{2+}-charged charged-chelating sepharose column and eluted as described in section 2.5.2. Again, four concentrations of TCEP were used in both the binding and elution buffers respectively. After EDTA (final concentration: 5 mM) and Complete™ protease inhibitor (1×) were added into the elution fractions, the eluents were concentrated using ultra-filtration and their protein content quantitated using the Bradford method (section 2.5.6). Activity assays were carried out according to section 2.6, except that no TCEP was added to the substrate solution to avoid perturbing the original redox states of the enzyme preparations.

2.9.3 Effects of pH

In this section, the effects of pH on initial rates of rPNGase-catalysed deglycosylation of glycopeptide were investigated. The enzyme solution was firstly dialysed against 1 mM TCEP and 1 mM EDTA water solution to remove any traces of HEPES. To negate any possible effects being due to buffer ions rather than pH, a single broad-range buffer, MIB buffer (12.5 mM malonate, 18.75 mM imidazole, 18.75 mM boric acid), was used. The assay was set up as described in section 2.6. In order measure the reaction rate (initial velocity), the assay conditions were modified as following: 10 μL of enzyme solution (0.05 mg/mL) was incubated with 85 μL of the FITC-Ova substrate (see section 2.9.4) solution (0.2 mg/mL, in 50 mM MIB buffer at pH 5.0, 5.5, 6.0, 6.5, 7.0, 7.5, 8.0, 8.5 and 9.0 respectively, 1 mM TCEP, 1 mM EDTA) for 10 minutes, then boiled for 3 minutes to stop the reaction, and subjected to HPLC analyses. All assays were carried out in triplicate.

2.9.4 Determination of the Michaelis Constant (K_m) and k_{cat}/K_m

2.9.4.1 *Fluorescein isothiocyanate (FITC)-labeling of substrate*

The initial rate values (v_0) of a catalysed reaction are used to calculate K_m and V_{max}. Since the activity of rPNGase Sp was measured in a discontinuous way, the initial rate for each substrate concentration was obtained at a point when the reaction was approximately 10%

complet [62]. Under the conditions chosen to maximise the enzyme and substrate available, the response obtained to achieve 10% conversion of substrate to product was inadequate for accurate quantitation. Therefore, to increase the sensitivity, the glycopeptide was covalently labeled with fluorescein isothiocyanate (FITC). The FITC-dilabeled ovalbumin glycopeptide (FITC-Ova) can be detected at much lower concentrations using a fluorescence detector. The excitation wavelength is 495 nm and emission wavelength is 520. Since the labeled substrate is more hydrophobic than the unlabeled one, the elution procedure in HPLC-based assay was modified accordingly (section 2.6). FTIC-Ova is schematically shown in Figure 2.3. The procedure of FITC-labeling of the glycopeptide is outlined in Appendix V.

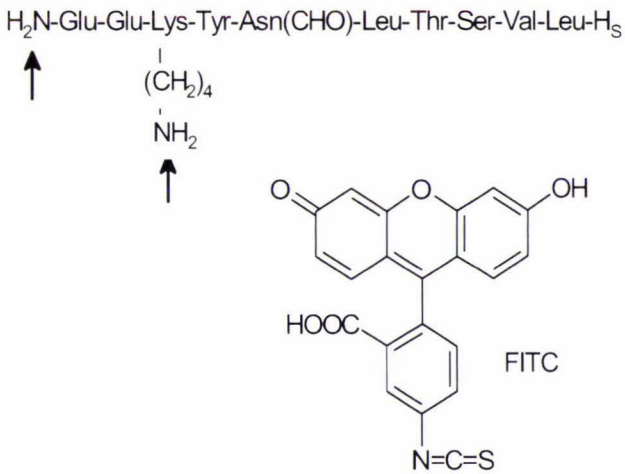


Figure 2.3 Labeling of ovalbumin glycopeptide with FITC. The ovalbumin glycopeptide (CHO indicates the carbohydrate chain on the glycosylated Asn) was labeled in the positions indicated by arrows using fluorescein isothiocyanate (FITC). Hs = Homoserine. Since the incorporated FITC molecules are at least two amino acids distant from the glycosylated asparagine residue, it was unexpected that the FITC could result in steric hindrance during the catalysis by PNGase. Adapted from Lenz 2003.

2.9.4.2 *Estimating the rough K_m of rPNGase over the substrate FITC-Ova*

An experiments were conducted under conditions that gave approximately 10% hydrolysis or less in order to measure the initial velocities (v_0) for each substrate concentration. Normally, a broad range of substrate concentrations (0.1-10 K_m) should be used to determine K_m and V_{max}. An initial experiment with a limited number of data points that spanned a broad range of substrate concentrations was carried out to obtain a rough estimate of K_m. A substrate solution with an estimated saturating concentration (e.g., 5 mg/mL) was serially diluted (10×, 100×, and 1,000×) to give a series of solutions with varying substrate concentrations. Two substrate concentration points that gave initial

velocity values that differed more than ten fold were used to estimate the rough K_m using the following equation.

$$K_{m(rough)} = [S_L] (v_H / v_L - 1)$$

K_m ----- estimated Michaelis Constant
v_H ----- initial velocity with the higher substrate concentration
v_L ----- initial velocity with the lower substrate concentration
$[S_L]$ ----- the lower substrate concentration

2.9.4.3 *Determining K_m and k_{cat}/K_m*

Based on the estimated K_m value, the substrate concentration range used was narrowed to between 0.1 and 8 K_m (e.g., 1/8, 1/4, 1/2, 1, 2, 4 and 8$\times K_m$) to obtain a larger number of data points for more accurate K_m determination. Assays were performed in triplicate. Initial rate values determined for each substrate concentration (presented in nmol/min) were then fitted to the Michaelis-Menten Equation ($v_0 = V_{max}[S] / (K_m + [S])$ using the programme GraphPad Prism® software (Version 5). From these plots values of K_m and V_{max} were obtained. The Lineweaver-Burk transform of $1/v_0$ versus $1/[S]$ was chosen to determine K_m and V_{max}.

Based on Michaelis-Menten kinetics, V_{max} data can be converted to give a k_{cat} value (turnover number) by dividing by the enzyme concentration employed in the assay solution. Subsequently, the k_{cat}/K_m ratio, a measure of the enzyme's catalytic efficiency, can be calculated.

2.9.5 Effects of Metal Ions

The effects of several metal ions on the activity of rPNGase Sp were investigated. The metal salts used were $CuCl_2$, $FeCl_3$, $MnCl_2$, $NiCl_2$, $ZnCl_2$ and $CdCl_2$. In order to eliminate the possible effects of the anion on the enzyme activity, the same anion was used where possible. To avoid the presence of trace amounts of other metal ions that might interfere the assay, Chelex-cheated (Chelex™ 100 Resin, Bio-Rad) doubly distilled water was used to make metal salt stock solutions and activity assay buffers. All glassware used was pre-washed with 50% HNO_3 and rinsed with Chelex-cheated water to remove any traces of metal ions. Lee and Zhao [27, 29] reported that the recombinant eukaryotic PNGase is a

metalloenzyme in which one zinc ion is coordinated by four cysteine thiol groups. The metal is thought to stabilise the enzyme conformation and is thus needed for the enzymatic activity. The initial velocity (v_0) was measured under the assay conditions used in kinetic studies.

To confirm if the activity of rPNGase Sp is Zn^{2+}-dependent, the enzyme solution was extensively dialysed against 50 mM MIB buffer (pH 6.0, 1mM TCEP) containing 20 mM EDTA with several changes for 72 hours to remove the enzyme-sequestered Zn^{2+}, followed by dialysis against the same buffer without EDTA to remove EDTA ions that may have bound to the protein. This apo-enzyme was assayed for deglycosylation activity, and the result was compared with the activity of the enzyme into which Zn^{2+} and other metal ions were added.

2.10 Crystallisation Trials

The goal of crystallization is to produce well-ordered protein crystals that are stable and large enough (dimensions greater than 0.1 mm along each axis) to provide a diffraction pattern when exposed to X-rays. This diffraction pattern can then be converted to an electron-density map of the diffracting protein, from which the three-dimensional structure of the protein can be modelled [63].

2.10.1 Basic Principle

To form a crystal, protein molecules assemble into a periodic lattice from super-saturated solutions. This involves starting with solution of protein with reasonable purity (95-100%) at a concentration between 0.5 and 30 mg/mL, and adding precipitating agents (e.g., ammonium sulfate, polyethylene glycol) that reduce protein solubility close to the point of precipitation. After a period of time (ranging from a few minutes to several months), often while the concentration of the precipitating agent is being slowly increased, the protein may precipitate from the solution in crystalline form in three stages: formation of nucleation sites, crystal growth and cessation of growth [64].

The two experimental formats used in this investigation to form crystals from protein solutions, hanging drop and sitting drop, are based on vapour diffusion. These methods

using micro volumes of protein solution are easy to set-up and to monitor. In both the methods, a drop of protein solution is mixed with an equal volume of precipitant solution, then placed in a sealed vessel above a reservoir of the precipitant. The protein drop and reservoir equilibrate through the vapour phase which results in a decrease in the volume of the drop — together with an increase in both the precipitant and protein concentrations resulting in a supersaturated protein solution. When optimal levels of the protein and precipitant concentration are reached, crystal nucleation may occur followed by crystal growth (Figure 2.4).

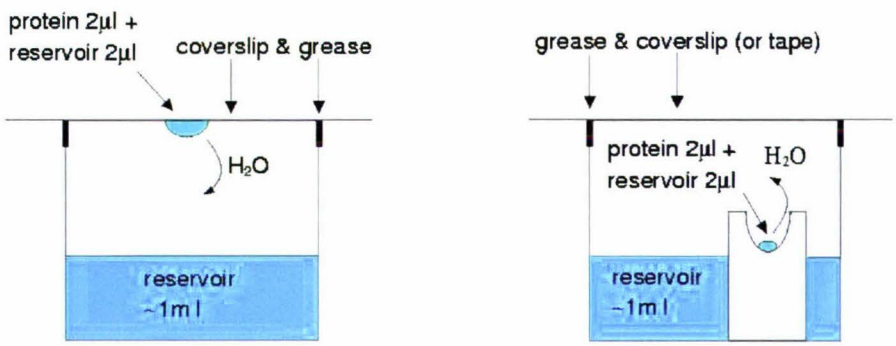


Figure 2.4 The hanging drop (left) and sitting drop (right) methods of protein crystallisation. In both the methods, a reservoir of precipitant solution was equilibrated against a drop of protein solution (1/2 protein solution plus 1/2 precipitant solution) hanging on a siliconised glass cover slip (left) or sitting on a pedestal above the reservoir solution (right) in a sealed system. Equilibrium was reached through the vapour phase. (Figure created by Airlie J McCoy, University of Cambridge)

The successful production of protein crystal is dependent on a variety of intrinsic and environmental factors because so much variation exists among proteins. Factors that may affect the protein crystallisation are protein purity and concentration, temperature, pH, presence of metal ions (e.g., Mg^{2+}, Ca^{2+}), ligands or cofactors, the addition of small amounts of miscible organic solvent (e.g., dioxan, aliphatic diols), types of precipitant, buffer type or ionic strength. Finding out an optimal combination of crystallisation conditions for an individual protein is a procedure of trial and error.

2.10.2 Setting up crystallisation trials

1) Protein sample preparing: Fresh rPNGase Sp sample purified by SEC (section 2.5.5) was diluted to ~ 10 mg/mL (containing 1×Complete™ Mini protease inhibitor), then centrifuged at the 14,000*g* in a bench-top micro-centrifuge at 4℃ for 15 minutes to remove any possible particulate matter. When the protein sample had been stored at -80℃ was used, it was heated at 37℃ for 5-10 minutes, cooled on ice, then centrifuged at 4℃ for 15 minutes before crystallisation trials were set up. The purpose of this pre-treatment was to remove any misfolded or partially damaged protein molecules and therefore to increase the homogeneity.

2) Sitting drop method: Vapour diffusion sitting drop method was used in the initial screen of crystallisation conditions, where the Structure Screen 1™ and 2™, PACT *premier*™ (Molecular Dimensions Ltd.), Crystal Screen™ and Crystal Screen 2™ (Hampton Research) kits were used to cover a number of combinations of crystallisation conditions including pH, buffers, and precipitant. The initial screening conditions (more than 200 combinations) provided by these kits are selected from known and published crystallisation conditions using sparse matrix sampling methodology [65], which allows sampling of a wide range of pH, buffers and precipitants using the minimum number of experiments.

For screening trials, 96-well plates (Greiner Bio-one) were used. 0.5 µL of protein sample was mixed with 0.5 µL of precipitant solution and placed in a small well above a 100 µL reservoir of precipitant solution. The plates were well sealed with tape (Cystalclear) and left at room temperature. The drops were monitored under a microscope and changes recorded daily during the first week and weekly afterward.

3) *Hanging drop method:* Fine screening around any conditions that provided micro-crystalline precipitates or phase separation in initial screening was carried out using the hanging drop method. In this method, 24-well (4×6) VDX plates (Hampton Research) were used. 1 µL of protein sample was mixed with 1 µL of precipitant solution on a sliconised cover slide. The slide was then sealed up side down with Vaseline over a well containing 1 mL of precipitant solution. The hanging drops were monitored in the same way as sitting drop method. Normally, in 4×6 matrix, the pH was varied from left to right

and the precipitant or salt concentration from top to bottom with certain increments. All the solutions used in the hanging drop screening were made up with double distilled water.

The formulations of the commercial screen kits can be obtained from the following links:

Structure Screen 1™:

http://www.moleculardimensions.com/us/datasheets/md1-01.htm

Structure Screen 2™:

http://www.moleculardimensions.com/us/datasheets/md1-02.htm

PACT *premier*™:

http://www.moleculardimensions.com/us/datasheets/MD1-29.pdf

Crystal Screen™:

http://www.hamptonresearch.com/assets/products/attachments/0000000001-0000000075.pdf

Crystal Screen 2™:

http://www.hamptonresearch.com/assets/products/attachments/0000000002-0000000085.pdf

Additive Screen™:

http://www.hamptonresearch.com/assets/products/attachments/0000000027-0000000467.pdf

Table 2.1 Plasmids used in this study

Plasmid	Description	Source
pProEX_HTb_*PNG*	The vector carries *Amp*^{R}, His-tag, TEV cleavage site and full length *PNG1Sp* gene (NCBI database accession No. AL031852, Appendix I).	Invitrogen
pET32a_HTBH	A modified version of pET32a vector (Novagen™) for expression of peptide sequences fused with the 109aa Trx-tag (thioredoxin), a His-tag and a TEV cleavage site at its N-terminus. It carries the ampicillin resistance gene. See appendix II for details. 'HTBH' designates His-tag, TEV site, *BamH*I and *Hind*III restriction sites respectively.	Produced by Dr. Rose Brown[a]
pSUMO_BXH	The vector is a modification of the pET32a vector (Novagen™) in which the DNA encoding for the Trx gene is replaced with the SUMO protein (93aa) amplified from the Invitrogen™ Champion™ pET SUMO vector. The protein produced with this vector carries a His-tag and SUMO at its N-terminus. The fusion can be cleaved from the SUMO protease (ULP1) cleavage site. It carries the ampicillin resistance gene. 'BXH' designates three restriction sites immediately downstream from the SUMO sequence. They are *Bam*HI, *Xba*I and *Hind*III respectively.	Produced by Mr. Trevor Loo[a]
pMal_CHTBH	A modified version of pMal-C2G vector (New England BioLabs®) for expression of peptide sequences fused with the *E. coli* maltose binding protein (MBP, encoded by *malE* gene, in which the N-terminal periplasmic signal sequence has been deleted). The protein produced with this vector carries a His-tag and a TEV cleavage site at its N-terminus. It also contains the ampicillin resistance gene. 'CHTBH' stands for cytosolic, His-tag, TEV site, *BamH*I and *Hind*III restriction sites respectively.	Produced by Mr. Trevor Loo[a]

[a] Institute of Molecular BioSciences, Massey University, Palmerston North, New Zealand

Table 2.2 *E. coli* strains used in this study

Strain	Genotype	Description	Source
XL1-Blue	*RecA endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F' *proAB lacI*^{q} *Z ΔM15 Tn10* (Tet^{r})]	High efficiency cloning strain; tetracycline resistant	Novagen™
Origami B (DE3)	F^{-} *ompT hsdS*$_B$ ($r_B^- m_B^-$) *gal dcm lacY1 ahpC gor522*::Tn*10* (Tc^{R}) *trxB*::kan (DE3)	General expression host; contains Tuner *lac* permease mutation and *trxB/gor* mutations for cytoplasmic disulfide bond formation. Tet (12.5 μg/ml), Kan (15 μg/ml)	Novagen™
Origami B (DE3) pLysS	F^{-} *ompT hsdS*$_B$ ($r_B^- m_B^-$) *gal dcm lacY1 ahpC gor522*::Tn*10* (Tc^{R}) *trxB*::kan (DE3) pLysS (Cm^{R})	High-stringency expression host; contains Tuner *lac* permease mutation and *trxB/gor* mutations for cytoplasmic disulfide bond formation; and compatible plasmids (pLysS) that provide a small amount of T7 lysozyme to control basal level of expression. Tet (12.5 μg/ml), Kan (15 μg/ml), Cam (34 μg/ml)	Novagen™
Origami B (DE3) pLacI	F^{-} *ompT hsdS*$_B$ ($r_B^- m_B^-$) *gal dcm lacY1 ahpC gor522*::Tn*10* (Tc^{R}) *trxB*::kan (DE3) pLacI (Cm^{R})	High-stringency expression host; contains Tuner *lac* permease mutation and *trxB/gor* mutations for cytoplasmic disulfide bond formation; and a compatible pLacI plasmids that supply *lac* repressor to control basal level of expression. Tet (12.5 μg/ml), Kan (15 μg/ml), Cam (34 μg/ml)	Novagen™
Origami B (DE3) pGroEL/S	F^{-} *ompT hsdS*$_B$ ($r_B^- m_B^-$) *gal dcm lacY1 ahpC gor522*::Tn*10* (Tc^{R}) *trxB*::kan (DE3) GroEL/S (Cm^{R})	Expression host; contains Tuner *lac* permease mutation and *trxB/gor* mutations for cytoplasmic disulfide bond formation; and compatible pGroEL/S plasmids that provide bacterial GroEL/GroES protein folding system. Tet (12.5 μg/ml), Kan (15 μg/ml), Cam (34 μg/ml)	Produced by Mr. Trevor Loo[a] (pGroEL/S was kind gift from Dr. Mark Patchett[a])
Rosetta (DE3)	F^{-} *ompT hsdS*$_B$ ($r_B^- m_B^-$) *gal dcm lacY1* (DE3) pRARE^{6} (Cm^{R})	General expression host; *lac* permease mutation allows control of expression level, provides rare codon tRNAs through compatible pRARE^{6} plasmids. Cam (34 μg/ml)	Novagen™
BL21 (DE3)	F^{-} *ompT hsdS*$_B$ ($r_B^- m_B^-$) *gal dcm* (DE3)	General purpose expression host. No antibiotic resistance	Novagen™

[a] Institute of Molecular BioSciences, Massey University, Palmerston North, New Zealand

Chapter 3 — Results and Discussion

3.1 Molecular Cloning

The full-length *PNG1* gene from *S. pombe* (*PNG1Sp*) had been previously cloned and expressed in a prokaryotic host (*E. coli*) in our laboratory using a pProEX™ expression vector (Invitrogen, Inc.). The gene product was however insoluble. When the gene was cloned into the pET32(+)™ vector (Novagen, Inc.) and expressed as a thioredoxin-fused protein, the product had good solubility in aqueous solution and could easily be cleaved from the hexohistidine tag to produce soluble recombinant PNGase Sp (data not shown). There were, however, about 27 extra amino acids residues on the N-terminus of the expression product due to vector design, which is not desirable for three-dimensional structure determination. In order to avoid the insertion of extra amino acids at the N-terminus of the gene product in this study, the *PNG1Sp* gene was re-cloned into a modified pET32a(+) vector, pET32a_HTBH (Figure 3.1, Table 2.1, Appendix II), in which the original unnecessary 'His tag-thrombin and S tag-enterokinase sites' were excised and replaced with a sequence encoding a His-tag and TEV site. Because the 5'-cloning site used in this study, is located immediately downstream the TEV protease cleavage site, only four extra amino acids remained on the N-terminus of the gene expression product after protease cleavage (Figure 3.1, Appendix II).

The PCR reaction using the *Bam*HI site-containing (forward) and *Hind*III site-containing (reverse) gene-specific primers together with the *PNG1Sp*-carrying plasmid pProEX_HTb_*PNG* yielded a single product of the expected size for the PNGase Sp encoding region (~1 kb; Figure 3.2 a). This was cloned into the plasmid pET32a_HTBH, which was subsequently successfully transformed into XL1-Blue cells (Table 2.2) as shown in Figure 3.2 b. DNA sequencing of the plasmids (pET32a_HTBH_*PNG1*) produced from the XL1-Blue transformants confirmed that there were no errors in the recombinant *PNG1Sp* gene sequence (data not shown).

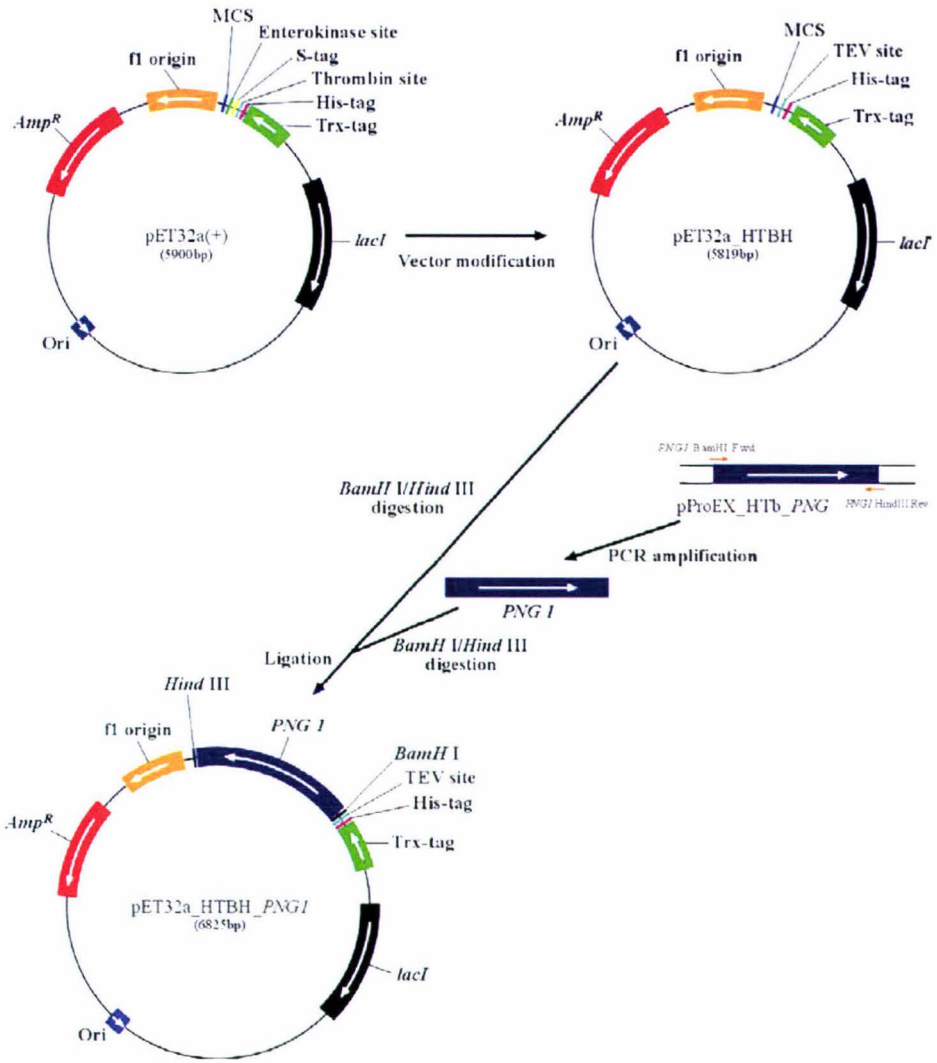


Figure 3.1 Cloning strategy The pET32a(+) vector was modified to yield pET32a_HTBH vector, in which the original sequence encoding a His-tag, thrombin site, S-tag and enterokinase site was excised and replaced with a sequence encoding a His-tag and a TEV site. The *PNG1Sp* gene was amplified from the pProEX_HTb_*PNG1* vector using *Bam*HI-containing (forward) and *Hind*III-containing (reverse) gene-specific primers and cloned into the pET32a_HTBH vector using the same restriction sites. It was heteroglogously produced in *E. coli* cells as a thioredoxin-hexohistidine-fused protein.

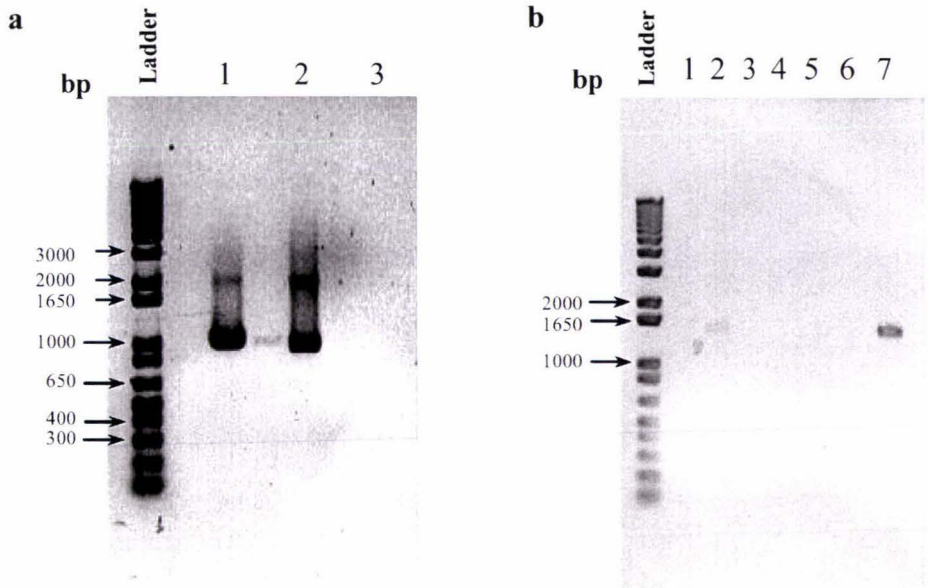


Figure 3.2. a) PCR amplification of *PNG1Sp*. The *PNG1* gene was amplified from the pProEX_HTb_*PNG1* plasmid using gene-specific forward and reverse primers resulting in a 1 kb product (lane 1) as visualized by 1% Agarose gel electrophoresis. lane 2, positive control; lane 3, negative control (without the template DNA). **b) Whole cell PCR result of colony screening.** Seven XL1-Blue colonies (corresponding to lane 1-7) were cultured overnight and then lysed. The supernatant of cell lysates containing DNA was used as template in whole cell PCR. The primers used were a plasmid-specific forward and a gene-specific reverse primers. The size of the band in lane 7 is consistent with it being the insert plus LacI and fusion tag sequence (totally 1.47 kb), thus the culture corresponding to lane 7 was used for plasmid purification.

3.2 Expression in Origami B (DE3) Strain

Small-scale expression trials showed that the Trx-His-TEV-PNGase Sp fusion protein could be overproduced in Origami B™ (DE3) cells using the pET32a_HTBH_*PNG1* vector (Figure 3.3, Lane 8). The recombinant fusion protein was soluble (> 90%) and easily recovered from the cell lysate as shown in Figure 3.4 (the detectable amount of rPNGase Sp in cell pellets was due to incomplete lysis). It was likely that the presence of the N-terminal fused Trx-tag contributed to the high solubility of the gene product. The Trx-tag's DNA sequence encodes a disulfide reductase, thioredoxin, which catalyses the formation of disulfide bonds in the cytoplasm of *E. coli* strains that carry mutations in thioredoxin reductase gene (*trxB*) and glutaredoxin reductase gene (*gor*) [66]. Considering that *PNG1Sp* encodes a eukaryotic cytosolic protein, which theoretically should not contain disulfide bonds, the solubility of the protein in Origami cells is hard to explain.

Nevertheless, rPNGase Sp was overproduced with high solubility. Interestingly, previous studies had shown that soluble PNGases require the presence of reducing reagent such as DTT for activity *in vitro* [4], indicating that free thiols are crucial for the proper folding and/or catalytic activity of the mature enzyme. Moreover, results showing the effects of the reductant TCEP on the activity of rPNGase Sp in this study (section 3.8) indicated that all the nine cysteine residues in rPNGase Sp should stay in their reduced state for maximum activity. Results of structural studies on recombinant mouse and yeast *S. cerevisiae* PNGases showed there were no disulfide bonds in the mature enzymes [27, 29]. It is likely that the N-terminal thioredoxin fusion partner enhanced the solubility of rPNGase Sp acting as a chaperone to prevent disulfide bond formation. On the other hand, the solubility of the fusion protein might be because of the solubility of thioredoxin itself. Activity assays showed however that the fusion rPNGase Sp was active (Figure 3.5), indicating that the protein was correctly folded into its native conformation in the cytoplasm of Origami™ B (DE3) cells.

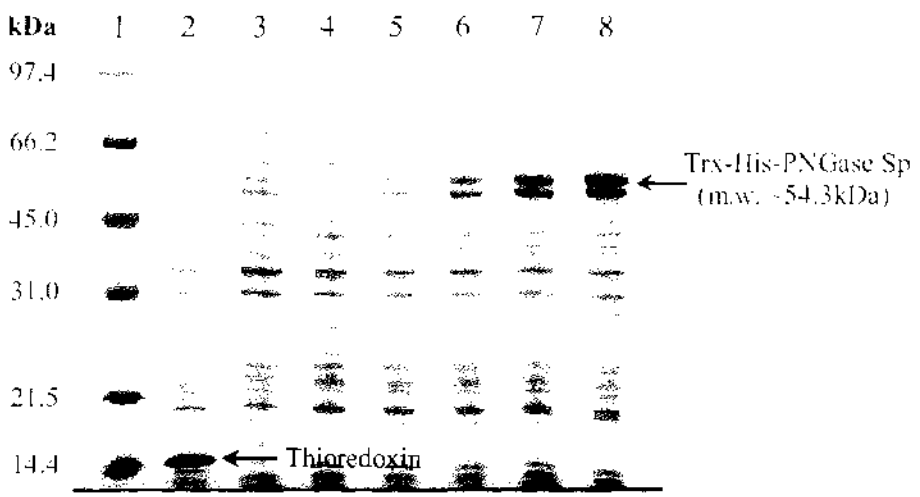


Figure 3.3 SDS-PAGE analyses of the whole cell samples of IPTG induced expression of PNGase Sp in Origami B (DE3). Lane 1, molecular weight maker (low range); Lane 2, cells carrying empty pET32a_HTBH vector; Lane 3, uninduced cells; Lane 4-8, cell samples taken at 0, 1, 3, 6 and 16 hours respectively after IPTG-induction (the final concentration of IPTG was 0.5 mM).

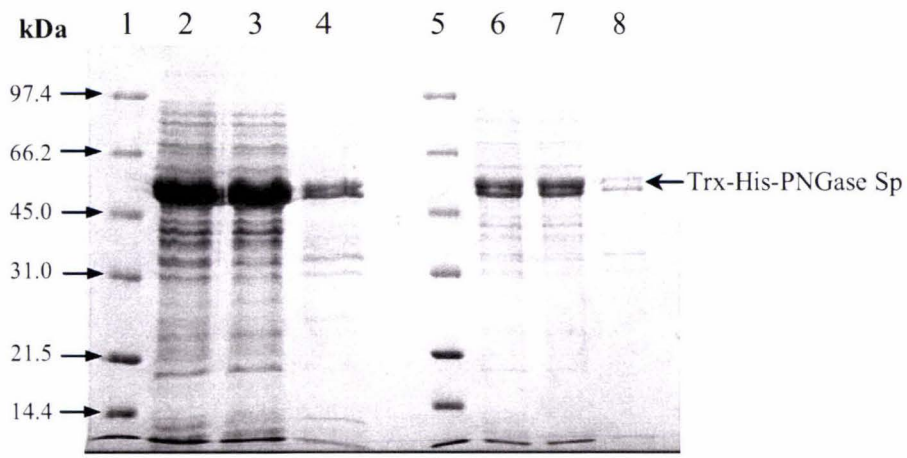


Figure 3.4 Solubility analyses of rPNGase Sp produced in Origami B (DE3) cells at 25℃ The Origami B cells shown in Figure 3.3 which strongly expressed recombinant PNGase Sp were lysed by sonication. The soluble and insoluble fractions were then separated by centrifugation before being loaded on a 12% SDS polyacrylamide gel. Lane 1, molecular weight maker; Lane 2, whole cells after lysis by sonication; Lane 3, the soluble fraction (supernatant); Lane 4, the insoluble fraction (pellet); Sample size for Lane 1-4 was 15 μL. Lane 5-8, same as Lane 1-4 except that the sample size was 3 μL so that the double band is more easily seen.

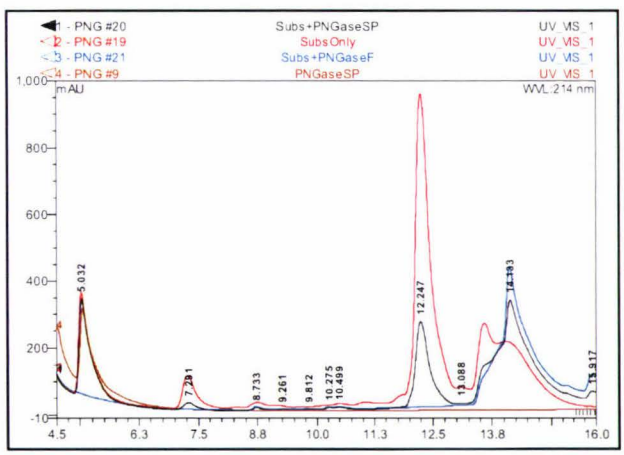


Figure 3.5 PNGase activity assay Shown is the data from a discontinuous HPLC assay. An 11-mer glycopeptide substrate (Sub) was incubated overnight at 25 ℃ with rPNGase Sp from IMAC chromatography. The reaction products were separated by HPLC using a C18 RP column. The detection wavelength was 214 nm. The peak (retention time: 14.13 minutes) in the Sub+PNGase Sp reaction (black line), that is absent in the Sub only and PNGase only controls, represents the product (deglycosylated substrate). The retention time of this peak coincides well with that of the product peak in PNGase F reaction (blue line, positive control, The product of PNGase F reaction has been previously confirmed by mass-spectrometry), indicating that identical cleavage reactions have occurred.

A large-scale expression was, therefore, carried out using Origami B (DE3) as the host strain. After induction with 0.5 mM IPTG and cultivation overnight at 25℃, SDS-PAGE analyses of the cell lysate combined with an activity assay showed that the gene product was soluble and active (data not shown), indicating that the Origami B (DE3) expression system can be easily scaled-up to one litre of total culture with a reasonable protein yield.

3.3 Purification of rPNGase Sp

The N-terminal hexahistidine-tag (His_6) allowed easy affinity purification of the fusion rPNGase Sp using immobilised metal affinity chromatography (IMAC). IMAC resulted in a significant purification of the protein as shown in Figure 3.6 a, lanes 5-8. However, the bands seen in lane 5-8 were actually comprised of two partially overlapping single bands that were calculated have a size difference of approximately 2.3 kDa using a standard curve; the higher one: ~59.2 kDa, the lower one: ~56.9 kDa. This difference, which was more obvious when a smaller amount of sample was loaded on to the SDS polyacrylamide gel, could not be confirmed by mass spectrometry (section 3.5.1). Since the calculated mass of the fusion protein is 54.3 kDa, it is more likely that the lower of the two-bands was Trx-His_6-fused rPNGase Sp. One possible reason for the double-bands is that some population of the fusion protein might bind metal ions such as Ni^{2+} ions during the IMAC purification step, affecting the charge of the protein and its migration on SDS polyacrylamide gels. Extensive dialysis of the samples against EDTA did not, however, result in a single band on SDS-PAGE. Another possible reason is that a population of the protein trapped some phosphoric acid from the buffers (originally, phosphate buffers were used, but Tris buffers were used) used in the purification and the trapped acid could not easily be removed, resulting in the apparent heterogeneity of the PNGase Sp on SDS PAGE [67]. This was discounted because: 1) the recombinant protein showed heterogeneity on SDS polyacrylamide gel at an early stage of expression before any purification (Figure 3.3, whole cell sample); 2) the protein still showed double-band when Tris-HCl buffer was used instead of phosphate buffer during lysis and IMAC purification.

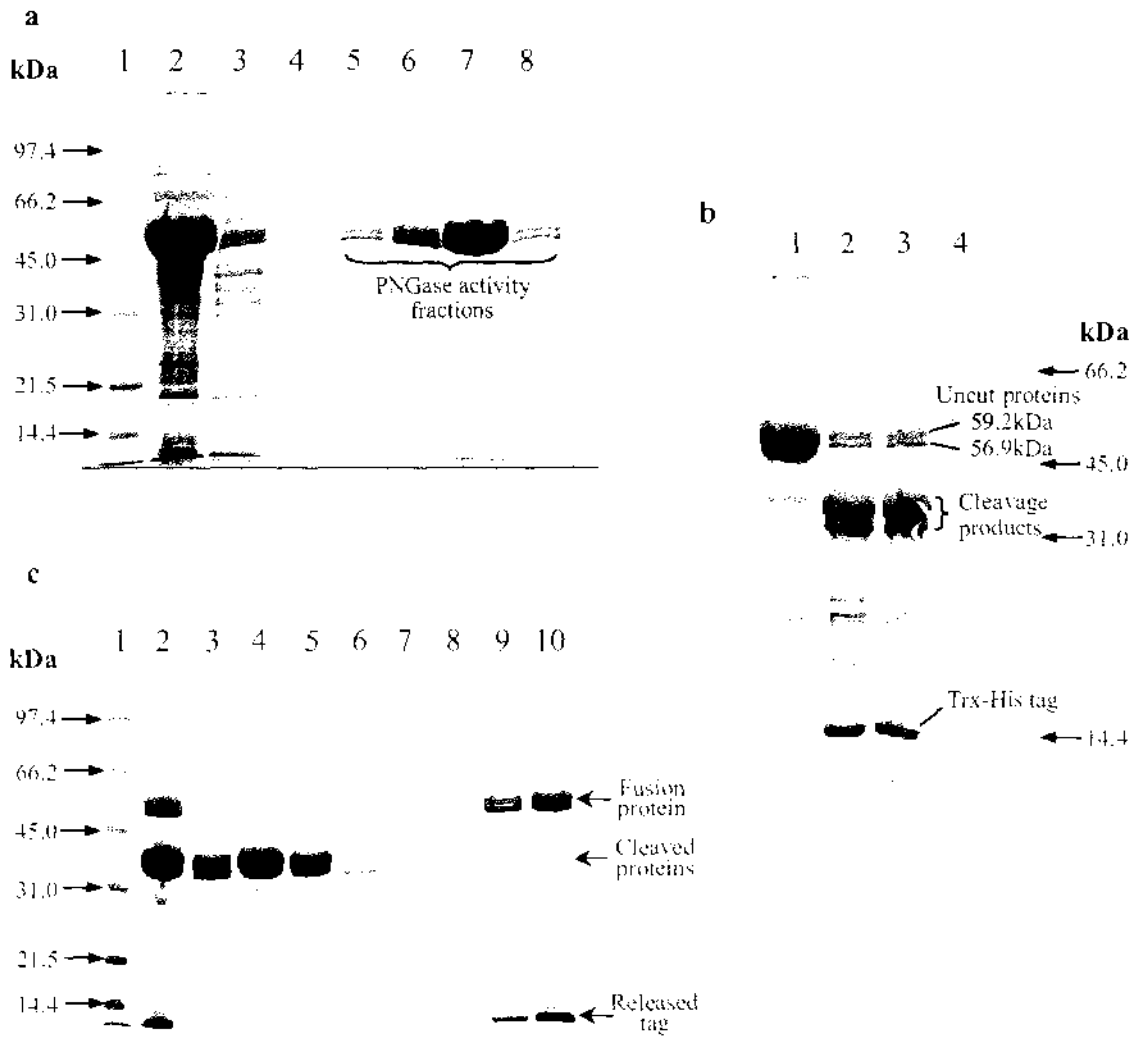


Figure 3.6 SDS-PAGE analyses of the samples taken at various purification stages a) IMAC purification of the fusion rPNGase Sp from cell lysate using a Ni^{2+} charged-chelating sepharose column. Lane 1, molecular mass marker; lane 2, material loaded on the column; lane 3, flow-through; lane 4, fraction eluted with 40 mM imidazole; Lanes 5-6, fractions eluted with 60 mM imidazole; Lanes 7-8, fractions eluted with 250 mM imidazole. **b) rTEV cleavage of the fusion protein.** Lane 1, negative control (reaction without rTev Protease); lanes 2 and 3, fusion protein plus rTev protease after overnight incubation at 25°C; lane 4, molecular mass marker. **c) 2nd IMAC purification after rTEV cleavage.** Lane 1, molecular mass marker; Lane 2, load; Lanes 3-8, washes containing rPNGase Sp; Lanes 9-10, eluted fractions containing the rPNGase Sp-Trx fusion and the released Trx-His_6 tag.

The fusion protein was then cleaved using rTEV protease, which removed the Trx-His_6-tag from the N-terminus of rPNGase Sp. The reaction was more than 90 percent completed after incubation overnight at 25°C (Figure 3.6 b). A loss of approximately 13.6 kDa in mass was consistent with the loss of the fusion partner (Trx-His_6 tag; 14.8 kDa).

The cleavage product, however, remained as two bands on SDS polyacrylamide gels, while the thioredoxin-hexahistidine-tag was clearly a single-band. These results strongly indicated that both protein components of the double band were recombinant gene products that must differ by extra sequence at their C-terminus. Such a difference might be caused by early termination during translation due to the presence of rare codons in the *PNG1Sp* gene, or by ribosome read-through the first stop codon and termination at a second one 60 bp downstream (Figure 3.9; Appendix II). Comparison of the codon usage between *E. coli* and *PNG1Sp* gene showed, however, that there are few rare codons in the gene. Moreover, the difference in apparent molecular mass between 'two' gene products is about 2.3 kDa, which is close to (although not equal to) the peptide mass (1.83 kDa) calculated from the sequence between two stop codons. Taken together, it is possible that stop codon read-through during translation resulted in dual products. Another possibility is that a very specific proteolytic activity cleaves a portion of the polypeptide chain from the C-terminus of the protein. This however, was deemed to be unlikely as there was no evidence of further proteolysis with time and the addition of Complete™ protease inhibitor to the cell lysis buffer had no effect on reducing the amount of double banding (data not shown).

Despite the apparent heterogeneity of rPNGase Sp, the cleaved protein was further purified using IMAC. As can be seen from Figure 3.6 c, the cleaved rPNGase Sp (Lanes 3-6) could be easily separated from both the released Trx-His_6 tag and un-cleaved fusion rPNGase Sp (Lanes 9-10). Trace amounts of fusion protein were observed in the wash (Lanes 3-4) and the early elution fractions (Lanes 9-10). This was probably due to non-covalent interaction between the fusion protein and cleaved rPNGase Sp. As expected, the double-band was still observed for both the cleaved (Lane 6) and fusion (Lane 9) proteins. Apart from the double band, no other bands were observed in fraction 5, indicating the enzyme was suitable for kinetic characterisation.

For crystallisation, a final purification step using size exclusion chromatography (SEC) was required. The resin used was Superdex™ 75 (GE Healthcare), which is optimized for separating proteins with molecular weight between 3-100 kDa. The buffer used was 10 mM HEPES (1 mM TCEP, 1 mM EDTA, pH 7.5). The purpose of this step was to remove any minor contaminants such as trace amounts of rTEV, and/or other proteins that may be non-specifically bound to rPNGase Sp, and improperly folded rPNGase Sp. The results of

SEC are shown in Figure 3.7. rPNGase Sp was eluted from the Superdex™ 75 column as a single peak. SDS PAGE analyses of the peak fractions showed the fractions with the highest protein concentration contained a second band of lower molecular weight (Lanes C4, C5 and C6). This band co-eluted with the target protein as it did in the 2nd IMAC (Figure 3.6 c, Lane 4). At this point, further separation was not attempted to avoid the potential effects of protein aging. Instead, further investigations into the origin of the two bands were undertaken.

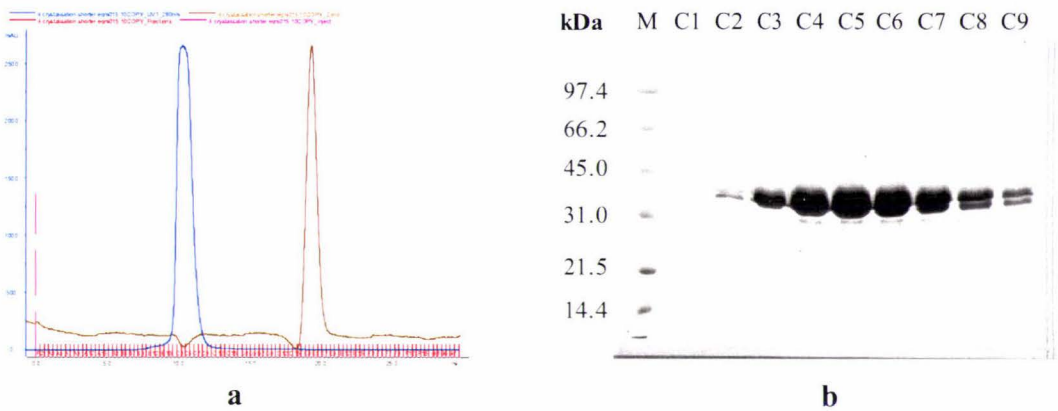


Figure 3.7 Size exclusion chromatography of rPNGase Sp. a) Chromatogram obtained during SEC run. The UV absorbance at 280 nm is shown in blue and the conductivity is shown in brown. **b) SDS-PAGE analyses of SEC fractions of interest.**

A purification table summarising the purification of rPNGase Sp to sufficient purity and quantity for enzyme kinetics and structural investigations is shown in Table 3.1.

Table 3.1 Purification Table of rPNGase Sp

Purification Step	Total Protein (mg)	Total Activity (units[a])	Specific Activity (units/mg)	Yield[b] (%)	Purification[c] (Fold)
Cell Extract	212.7	11911	56	100	1
1st IMAC	57.8	11266	195	95	3.5
2nd IMAC	35.8	10809	302	91	1.5
SEC	30.3	9265	306	78	1

[a] One unit of rPNGase Sp is defined as 1 nmol of substrate (FITC-Ova) deglycosylated per minute at 25 at pH 6.0.

[b] Yield was calculated by dividing the total activity of the current step by the total activity of the first step and expressed as percentage.

[c] Purification, expressed in fold, was calculated by dividing the specific activity of each purification step by the specific activity of the previous step.

3.4 The Use of Different Expression Hosts and Constructs to Solve the Heterogeneity problem

Several possible reasons were proposed to explain the two bands that characterised the recombinant PNGase Sp.

3.4.1 Premature Translation Termination

The possibility that the double-band resulted from an early termination of translation due to different codon usage between *E. coli* and yeast *S. pombe* was tested first. Different codon usage between the target and host DNA can have a significant impact on heterologous protein production. For example, AGA and AGG are common codons for arginine in the yeast gene, whereas they are rarely found in *E. Coli* [68]. There are eight arginines are encoded by AGA and three by AGG in the *PNG1Sp* gene. So, if these codons in the mRNA of the target gene are encountered by the ribosomes, the lack of the matching tRNAs in *E. coli* may lead to premature translation termination and, hence, truncated proteins or proteins with the wrong sequence [69, 70]. The *PNG1Sp* gene was therefore expressed in the *E. coli* expression strain Rosetta™ (DE3), which is designed to enhance the expression of target proteins that contain rare arginine, isoleucine, leucine, glycine and proline codons. Rare codon tRNAs are produced on chloramphenicol resistant plasmids (pRARE^6) that are compatible with pET32a vectors (Table 2.2). When produced in Rosetta (DE3) cells, the target protein was still double-banded on SDS PAGE both before (data not shown), and after IMAC (Figure 3.8 a), suggesting that the heterogeneity of rPNGase Sp was almost certainly not caused by premature translation termination.

3.4.2 Proteolytic Degradation

The Origami B (DE3) strain previously used to express *PNG1Sp* gene is a lysogen of bacteriophage λ DE3 and carries a chromosomal copy of the *lacI* gene encoding the lac repressor, and the gene for T7 RNA polymerase which is under the control of the IPTG-inducible *lac*UV5 promoter. IPTG induces expression by displacing the *lac* repressor from the *lac*UV5 promoter, promoting the expression of T7 RNA polymerase, which in turn binds to the bacteriophage T7 promoter upstream from the target gene (*PNG1Sp*) and starts transcription. Although the expression hosts and pET32a vectors were engineered to control basal level expression of T7 RNA polymerase (both the Origami B and Rosetta

strains are *lacYI* (lac permease) deletion mutants that enable uniform entry of IPTG into cells, and have a *lac* operator inserted downstream of the vector-encoded T7 promoter to reduce leaky transcription), the leakiness of *lac*-derived promoters such as *lac*UV5 still results in basal level production of T7 RNA polymerase and therefore, the target protein in uninduced cells [68, 71] (Figure 3.3, Lane 3). Whether or not rPNGase Sp is toxic to the host cells, premature expression may increase the level of proteolytic degradation of rPNGase Sp, resulting in heterogeneity of the target protein. This is possible because although the Origami B (DE3) and Rosetta (DE3) strains are deficient in the generic protease Lon, encoded by *lon* gene, and lack the outer membrane protease OmpT (encoded by *ompT*), other proteases may act on the target protein. The better way to reduce possible proteolytic degradation to the lowest possible level is to concomitantly control basal expression and shorten the expression period. Based on this consideration, the *PNG1Sp* gene was expressed in Origami (DE3) pLysS and Origami (DE3) pLacI (Novagen, Table 2.2), at either 25 C for 16 hours or 37 C for 4 hours. The Origami (DE3) pLysS strain contains a compatible chloramphenicol-resistant plasmid (pLysS) that provides a small amount of T7 lysozyme which can bind T7 RNA polymerase and inhibit transcription, while Origami (DE3) pLacI supplies *lac* repressors from a compatible pLacI plasmid to ensure stringent repression in the uninduced state. The results, however showed that the double banded product was not due to either the expression system or the growth conditions used (Figure 3.8 b).

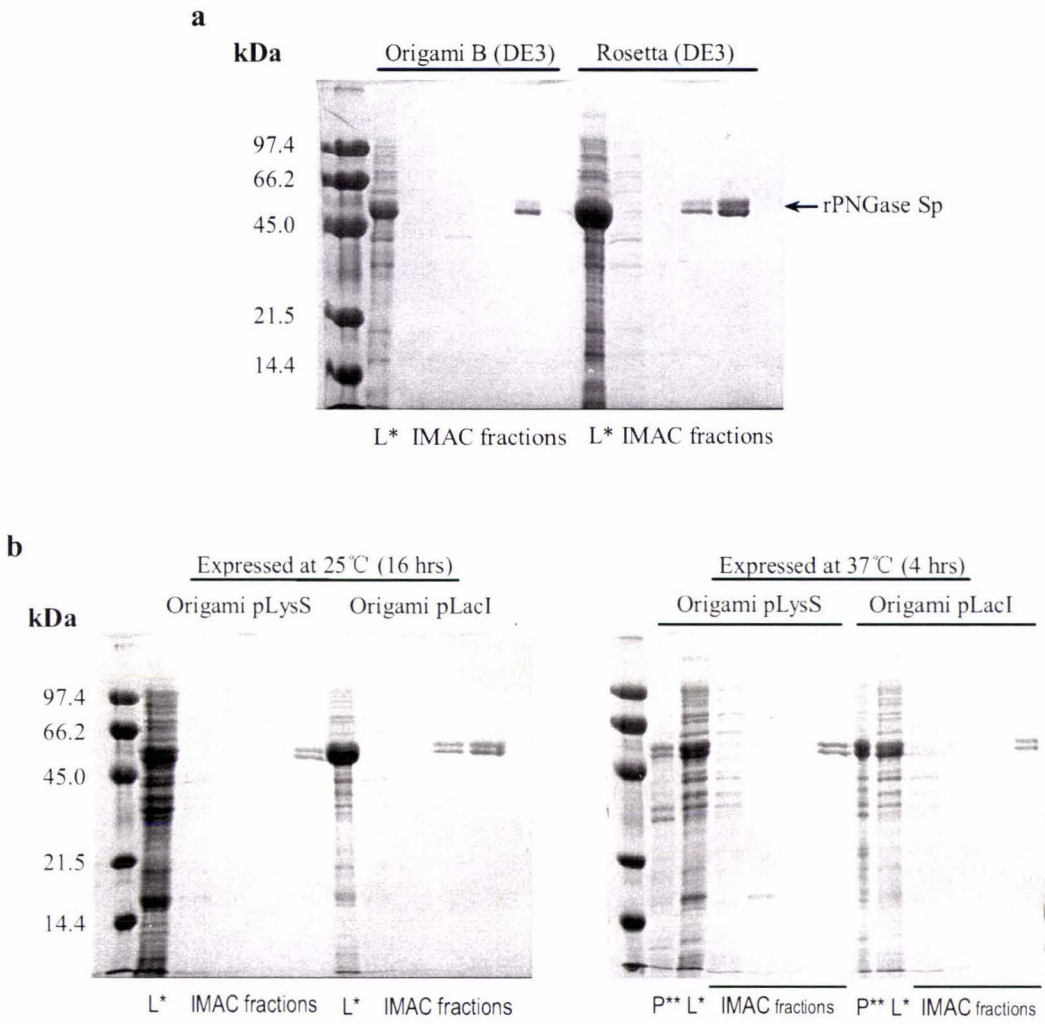


Figure 3.8 IMAC purification of the fusion rPNGase Sp produced from (a) Origami B (DE3) and Rosetta (DE3), and (b) from Origami (DE3) pLysS and Origami (DE3) pLacI at different temperatures. rPNGase was overproduced in all the four *E. coli* strains (Lane L*: material loaded on the column). However, in all cases, rPNGase showed the same double-band character on SDS-PAGE (IMAC fractions). Expression at higher temperature (37℃) for shorter time (4 hrs) showed no difference to protein produced in cells grown at 25℃ for 16 hours. The rPNGase was still highly soluble when expressed at 37℃ (comparison between soluble fraction L* and insoluble fraction P**).

3.4.3 Stop Codon Read-through

Given that the heterogeneity of the rPNGase was not a result of premature termination of translation or proteolytic degradation, it is possible that stop codon read-through during translation may have occurred resulting in two protein products, one longer than the other. Although misincorporation of amino acids at stop codons is rare event (frequency = 10^{-4} in intact cells), read-through does occur in both prokaryotes and eukaryotes [72], and a

number of factors are known to influence the efficiency of translation termination. These are the stop codon identity and the nucleotide context surrounding the stop codon, the identity of the last two amino acids incorporated into the polypeptide chain, and the presence of stimulatory elements downstream etc [73].

Firstly, the nature of stop codons themselves can result in different degrees of fidelity. The UAA sequence has the highest fidelity, followed by UAG, and then UGA which is the least reliable having the highest probability of read-through [72]. In this study, although the natural stop codon of *PNG1Sp* (UGA) has been changed to UAA and there are two stop codons in the vector pET32a_HTBH_*PNG* (Figure 3.9, in caption) supposedly with high fidelity for translational termination, the results indicated that read-through might still be occurring, resulting in two gene products 2.3 kDa different in their molecular mass (Figure 3.6 b). Therefore, one extra stop codon (T/UGA) was inserted between the first and second stop codons to make another construct (pET32a_HTBH_*PNG*_STOP, Figure 3.9). Expression of *PNG1Sp* using this new construct in Origami B (DE3), did not, however, result in a single banded protein (Figure 3.10).

It has been recognised that the 3'- nucleotide context of stop codons also influences the fidelity of stop-codon recognition [74]. One explanation for this effect is that release factors (RFs) recognise a mRNA : rRNA base-paired complex rather than single-stranded mRNA, and that the base pairing extends outside the triplet stop codon [75]. For example, the six nucleotide sequence CA(A/G)N(U/C/G)A immediately downstream of the termination codon can base-pair perfectly with certain regions of the rRNA which may destabilise secondary structure in the ribosome, affecting the binding of release factor to the stop codon and inducing read-through [76]. From Figure 3.9, it can be seen that there is no such a 3' hexanucleotide read-through determinant presented downstream from the stop codon (TAA). In *E. coli*, a wide range of genes show obvious bias in the first base immediately downstream from the stop codon (which is referred as the fourth base), and the binding affinity of protein release factors (RFs) to stop codons is affected by this fourth base [74]. When Sang Hyeon Kang and coworkers [77] expressed human erythropoietin (hEPO) in *E. coli* BL21 (DE3) using the pET-series vectors, two proteins 2 kDa different in molecular weight were produced when the fourth base was G or C. When one more stop codon was introduced as UGAUGA to examine the effect of U as the fourth base, read-through still occurred, although the ratio of correct translational termination to

read-through decreased from 2:1 (G or C) to 4:1 (U). When the fourth base was A, only single protein of the expected size was obtained, suggesting that the effect of the fourth base on termination efficiency is in the order of A >U >G ≈C for UGA. For the UAA stop codon, a hierarchy of stop signals with the decreasing order of efficiency UAAU > UAAG > UAAA/C, has been found in prokaryotes [75]. Thus, the tetranucleotide termination signal (TAAT) located at the end of *PNG1Sp* gene (Figure 3.9) should have relatively high termination efficiency, making read-through unlikely.

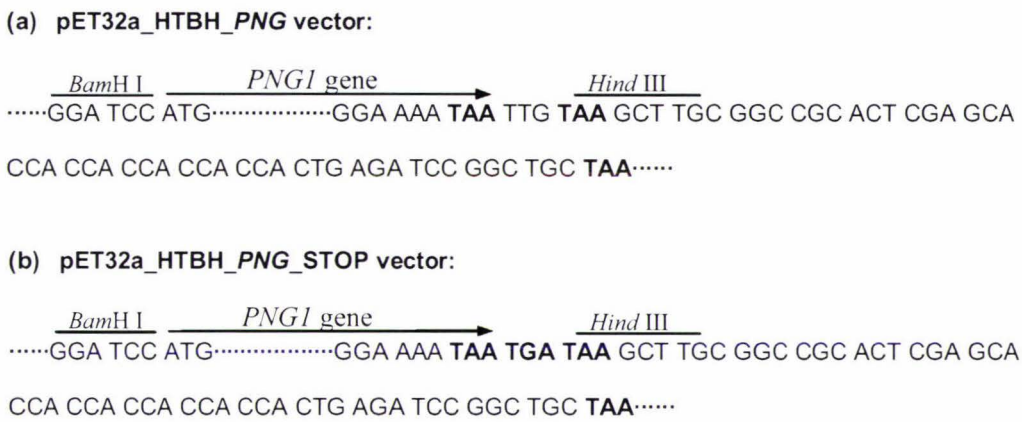


Figure 3.9 Part sequences surrounding the stop codons in the original and newly constructed recombinant plasmids. (a) Original pET32a_HTBH_*PNG* vector; **(b)** Newly constructed pET32a_HTBH_*PNG*_STOP vector. The *PNG1* gene was cloned into BamHI and HindIII restriction sites. The TAA stop codon located ~60 bp downstream of *PNG1* came from the vector pET32a_HTBH.

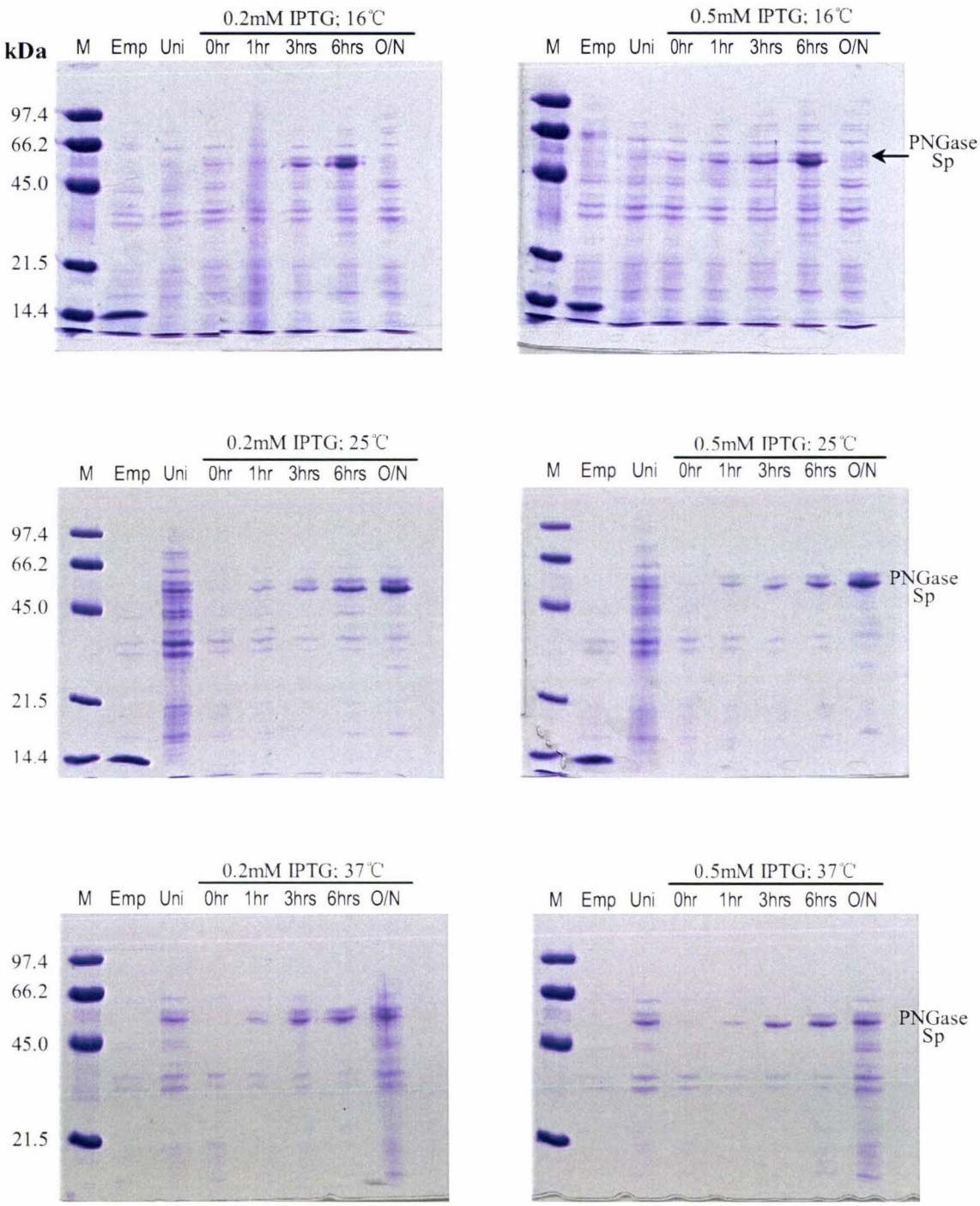


Figure 3.10 Expression trials using Origami B (DE3) with the pET32a_HTBH_*PNG*_STOP vector. The expression of *PNG1* gene was induced with either 0.2 mM or 0.5 mM of IPTG, and incubated at 16, 25 and 37℃ respectively. Whole cell samples were taken at 0, 1, 3, 6 and 16 hours (overnight) after induction. M: molecular weight marker; Emp: cells transformed with empty vectors; Uni: uninduced cells; O/N: overnight.

The 5′ nucleotide context of the mRNA is also thought to affect the translation termination efficiency both in eukaryotes and prokaryotes, and may mediate read-through via direct interaction between the P site tRNA and release factor 1 (RF1) or indirect interaction the ribosome [75, 78]. Interestingly, Sanaa Tork and coworkers [79] reported that the presence of two adenines immediately upstream of the stop codon plays a major role in translational read-through in *S. cerevisiae*. They proposed that the presence of two adenines upstream of the stop codon induces structural modification in the ribosomal P site, possibly through their high stacking potential, which is then transmitted to the ribosomal A site. Consequently, competition between the RF1 and/or RF3 and the natural suppressor tRNA is displaced in favor of read-through. A similar mechanism is thought to be used in *E. coli*. If this theory is applicable, the two adenines immediately upstream of the first (TAA) and the second (TGA) stop codons in the *PNG1Sp* gene (Figure 3.9) might contribute to possible read-through. How the third stop codon was also skipped to give such a high read-through ratio (two bands that are comparable in their intensity on SDS polyacrylamide gel) is, however, hard to explain.

Lastly, if a bacterial selenocysteine insertion sequence (bSECIS), which can form localised specific mRNA hairpin secondary structure, happened to exist downstream the UGA codon, the stop codon may be decoded as selenocysteine (the 21st amino acid) [73]. This possibility was not examined, because there were two UAA stop codons flanking the UGA codon (Figure 3.9), making it unlikely. Taken together, the possibility that stop codon read-through caused dual products during expression of *PNG1Sp* was quite low.

3.4.4 The Use of Alternative Vectors and Fusion Partners

Two other constructs: pSUMO_BXH_*PNG1* and pMal_CHTBH_*PNG1*, were also made in this study in an effort to remove the heterogeneity of the protein product.

The pSUMO_BXH vector (Table 2.1) adds a hexahistidine tag and the SUMO protein at the N terminus of the recombinant proteins. SUMO (small ubiquitin-related modifier, 93 amino acids) has been shown to modulate protein structure and function by covalently binding to the lysine side chains of the target proteins in eukaryotes [80]. It was reported that a number of 'difficult-to-express' proteins have been heterologously produced in *E.*

coli yields and remarkably enhanced solubility as SUMO-fused proteins in *E. coli* compared to using conventional tags such as glutathione S-transferase (GST), maltose-binding protein (MBP) and thioredoxin (Trx) [81, 82]. How SUMO fusion enhances expression is not known, but it has been proposed that the tightly packed and rapidly folding structure of SUMO enables it to act as a nucleation site ('priming' effect) for the folding of the C-terminally fused target protein, promoting proper folding and enhancing solubility. A distinct advantage in using the SUMO fusion system is that the SUMO can be cleaved at its C-terminus with remarkable fidelity and efficiency by ULP1, a natural SUMO protease that recognises the tertiary structure of SUMO, not a linear amino acid sequence like the other more commonly used proteases. This has two advantages: 1) the protease will not erroneously cleave within the target protein; 2) protease cleavage sites do not need to be engineered between SUMO and target protein, allowing the generation of recombinant protein with a native N-terminal sequence, which is desirable for structural analyses [81].

The pMal_CHTBH vector (Table 2.1) carries an *E. coli* maltose binding protein (MBP, encoded by the *malE* gene), a His-tag and a TEV cleavage site at its N-terminus. The fusion protein can be purified using either amylose or Ni^{2+} affinity chromatography. As mentioned in last section, it is possible that the local mRNA secondary structure downstream of the stop codon may perturb the normal decoding of the stop codon, resulting in read-through or a frame-shift [83]. Because it was not easy to test whether this specific secondary structure was formed in the target mRNA produced from pET32a_HTBH_*PNG1*, a modified version of pMal_C2G vector (New England BioLabs®), pMal_CHTBH, was used. This vector has a different sequence downstream from the multiple cloning sites, so should be unable to promote the localised secondary structural elements discussed above.

Although the exact reasons for heterogeneity of rPNGase Sp had not been identified, these two different constructs were tried to see if a homogeneous product would result. Unfortunately, both the SUMO- and MBP- fused recombinant PNGase Sp were still double-banded on SDS-PAGE (Figure 3.11).

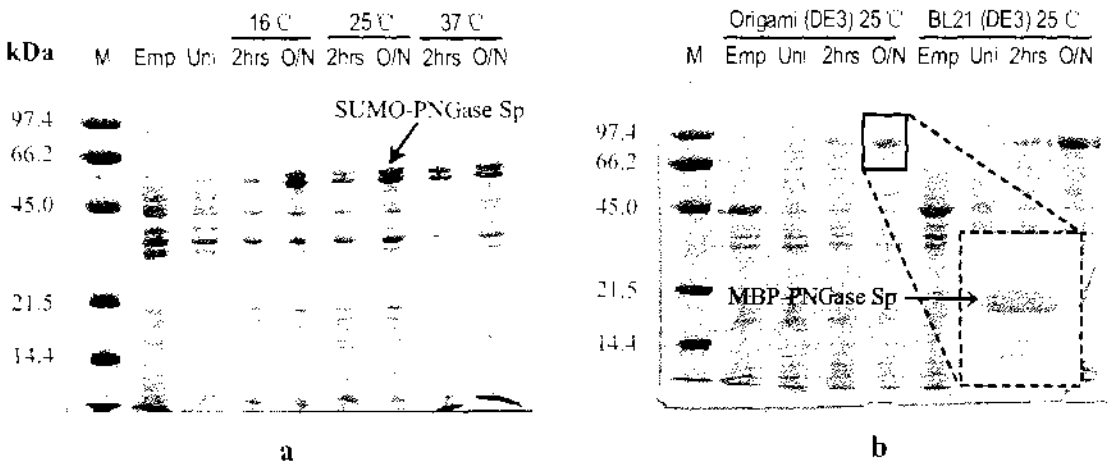


Figure 3.11 a) Expression of the *PNG1* gene in *E. coli* BL21(DE3) strain using the pSUMO_BXH_*PNG*_STOP plasmid. Expression was induced with 0.5 mM IPTG followed by incubation with shaking at different temperatures for either two hours or overnight. **b) Expression of the *PNG1* gene in Origami(DE3) and BL21(DE3) strains using the pMal_CHTBH_*PNG*_STOP plasmid.** Expression was induced with 0.5 mM IPTG followed by incubation with shaking at 25 °C for either 2 hours or overnight. M: molecular mass marker; Emp: cells transformed with empty plasmid; Uni: uninduced cells; inset: a close-up view of the double band. Both the SUMO- and MBP- fused rPNGase Sp showed heterogeneity (double-band) on SDS-PAGE (12% acrylamide).

3.5 Anomalous SDS-PAGE Behaviour of rPNGase Sp

Since all the measures mentioned in previous sections failed to solve the problem of the 'double-banded protein', we started to suspect the identity of the expression product even though the enzyme preparation was shown to have PNGase activity. HPLC purification followed by mass spectrometry (MS) analyses was therefore performed to determine the exact mass(es) of the double-banded protein(s). Peptide-mass fingerprinting was also carried out to confirm the identity of the expression product. The results of these analyses demonstrated that the protein preparation was homogeneous and the gene product was indeed rPNGase Sp. Moreover, it was found that the rPNGase Sp migrates anomalously on SDS-PAGE, which may be related to the redox state of its cysteine residues.

3.5.1 Mass spectrometry analyses of rPNGase Sp

If the rPNGase Sp preparation was indeed composed of two proteins corresponding to the double bands on SDS polyacrylamide gel, two mass peaks with different mass to charge ratio (m/z) should be observed in the mass spectra. The rTEV cleavage mixture was therefore further purified using HPLC and the eluted fractions containing either double-banded protein (Figure 3.12 a; Lane 1) or the released fusion tag (Lane 6) were collected for ESI-MS analyses. The results of mass spectrometry showed that the sample containing the double-banded protein was homogeneous, as there was only one peak produced (Figure 3.12 c), with an observed mass of 39442.0 Da, exactly the same as the calculated mass of rPNGase Sp (39441.9 Da). Considering the relative ratio between the intensities of two bands on SDS polyacrylamide gel as well as the high intensity/abundance of the mass peak (39442.0 Da) in the mass spectra, it is unlikely that the concentration of the top band was too low to be detected, confirming that the sample was homogeneous. Meanwhile, the mass (14.7 kDa, Figure3.12 b) of the fusion tag also matched well with the calculated one (14.8 kDa). These results demonstrated that the two apparently mass-different protein species visualised on SDS-PAGE had the same molecular weight, and strongly suggested they were the same protein. However, what caused the double-band pattern of rPNGase Sp on SDS-PAGE needs to be examined.

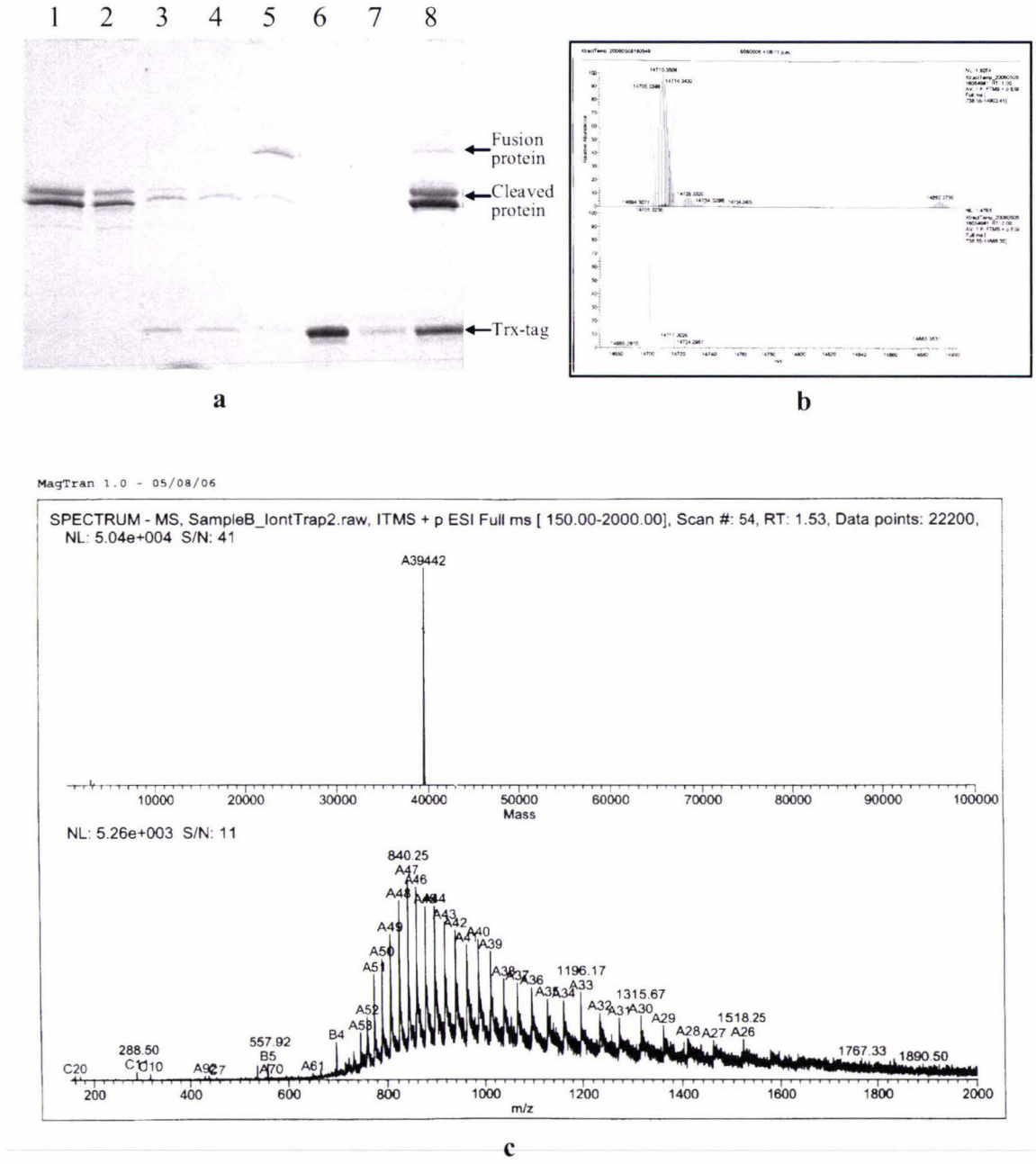


Figure 3.12 a) HPLC separation of rTEV cleavage mixture. The rTEV cleavage mixture was separated using HPLC (C-18 column, 250 × 4.6mm; A 40-minute gradient elution from 10% acetonitrile / 0.1% TFA to 100% acetonitrile / 0.08% TFA; Flow rate of 1mL/min; Detection at 214 nm). Lanes 1-7: fractions eluted from the HPLC C-18 column; both the fraction corresponding to Lane 1 which contained double-banded protein and the fraction corresponding to Lane 6 which contained the released fusion tag were collected for mass-spectrometry analyses. Lane 8: the material (rTEV cleavage mixture) loaded onto the column. **b) ESI-mass spectra of the sample containing released fusion tag.** The determined mass for the released tag was 14.7 kDa that was close to the calculated value, 14.8 kDa. **c) ESI-mass spectra of the sample containing the double-banded protein.** A single peak was observed in the deconvoluted spectrum at 39.442 kDa, which agrees with the calculated value.

3.5.2 Anomalous SDS-PAGE Behaviour of rPNGase Sp Was Related to the Redox State of Its Cysteine Residue(s)

The heterogeneity of recombinant proteins can result from posttranslational modifications that may be characteristic of the expression system used. Possible modifications to the polypeptide chain include glycosylation, phosphorylation, proteolytic cleavage and disulfide bond formation, or may be the result of external factors that occur during the culture period or as a result of the subsequent downstream processing [67, 84]. Since the glycosylation machinery is absent in *E. coli*, and the results of mass spectrometry analyses ruled out the possibility of proteolysis, we speculated that the apparent heterogeneity was related to the disulfide bond formation, hence folding of the rPNGase Sp. When Michael Fountoulakis [67] studied the heterogeneity of the *E. coli*-derived recombinant interferon γ receptors, he found that, upon disruption of the cells, both native and non-native disulfide bridges were formed, which caused apparent heterogeneity of the protein on a non-reducing SDS polyacrylamide gel (Figure 3.13). Mispaired disulfides resulted in generation of both dimeric and monomeric non-native interferon γ receptors. Only correctly folded (the receptor protein includes eight cysteine residues forming four sequential disulfides) monomers show ligand binding activity. The different forms of the receptor migrated differently on non-reducing SDS-gel, and the use of reducing agents during extraction did not decrease the percentage of non-native oligomeric and monomeric forms.

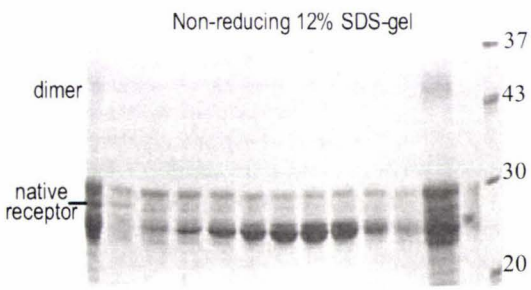


Figure 3.13 Size exclusion purified recombinant interferon γ receptor was further purified using Polybuffer Exchange. Protein was eluted by applying a NaCl-gradient. Lane 1, material loaded on the column. Lanes 2-11, fractions eluted with 0.35 M NaCl. Lane 12, proteins eluted with 0.50 M NaCl. This fraction mainly includes non-native forms. The native monomers migrated to 27 kDa, and the monomers containing either mispaired or non-formed disulfides migrated slightly slower, between 27 and 30 kDa. Adapted from Michael Fountoulakis *et al.*, 1996.

This is an example of heterogeneity that is molecular weight-independent but disulfide state-related, and reminiscent to what is observed for rPNGase Sp. But the major difference was that the rPNGase Sp showed heterogeneity on reducing instead of non-reducing SDS-PAGE. Normally, all disulfide bonds, if any, in proteins should be reduced

during sample preparation for SDS-PAGE. As previously mentioned (section 3.2), cytosolic PNGases need the presence of reducing agents (at least 1 mM DTT) for their activity *in vitro*, and it is likely that all the nine cysteine residues stay in their reduced state in the native form of rPNGase Sp (section 3.8) [27, 29]. It is therefore logical to speculate that, during expression or/and upon cell lysis, one or more non-native disulfides were formed in the rPNGase Sp molecules. This is likely, because: (1) the cytoplasm of Origami (DE3) has relatively low reducing potential due to the lack of thioredoxin reductase and glutathione reductase; (2) the reducing agents (5 mM DTT or 1 mM TCEP) present in the buffers might not be sufficiently concentrated enough to keep all the nine cysteines reduced, although the Cys residues that are crucial to PNGase activity (e.g. Cys-191; four Cys in two CXXC motifs) may remain reduced. We also hypothesise that the 'mis-formed' disulfide bond(s) help to form a strong, regional hydrophobic folding area, which remains impervious to SDS and reducing agents. A similar resistance to SDS and reducing agents is seen in the prion protein PrP^{Sc}, and amyloid proteins which have been shown to aggregate through hydrophobic interactions [114]. If only some populations of the rPNGase Sp molecules were totally denatured while others remained partially folded when boiled in SDS sample buffer, an apparent heterogeneity would be observed on SDS polyacrylamide gels. This is because these two different conformations will have different numbers of SDS molecules attached, resulting in different charge/mass ratios. This is illustrated schematically in Figure 3.14.

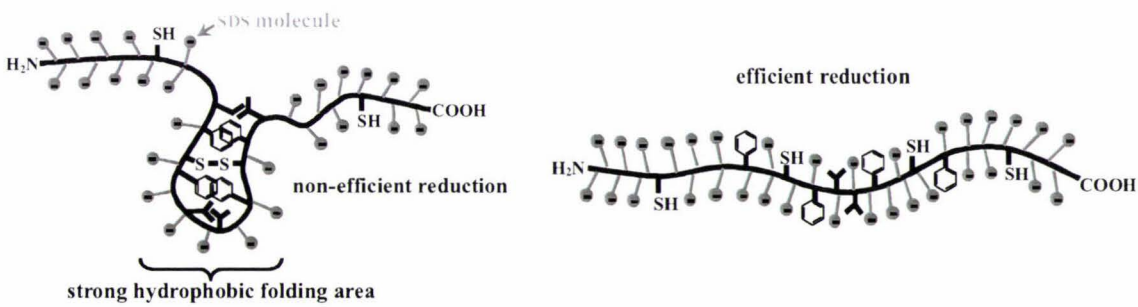


Figure 3.14 Schematic diagram showing possible differences in the modification of the same protein in the presence of SDS and reducing agents. (Left) The tightly packed hydrophobic area prevents the access of anion detergent SDS and reducing agent DTT to the inside of the hydrophobic folding area and the disulfide bond (s). **(Right)** The totally denatured protein.

According to this hypothesis, if all the thiol groups in the protein are reduced and alkylated so they can no longer form disulfides, the apparent heterogeneity observed on SDS-PAGE gels should be eliminated. To test this assumption, both fusion and cleaved rPNGase Sp samples that showed obvious double-bands on SDS PAGE were firstly denatured and reduced with 6 M guanidine hydrochloride and 10 mM DTT (100 μL of ~ 15mg/mL protein in 50mM Tris-HCl buffer (pH 8.0) plus 300 μL of 8 M guanidine hydrochloride and 22 μL of 0.2 M DTT aqueous solutions) for one hour at room temperature. The reduced thiols were then alkylated by the addition of iodoacetamide to a final concentration of 55 mM and incubation at room temperature for 45 minutes. After desalting by C-18 RP HPLC, samples of each were analysed by SDS-PAGE using the same conditions as before. As predicted, both the fusion protein and the cleaved rPNGase Sp were presented in a single band (Figure 3.15 a), which migrated slightly more slowly than the original double-bands, which was attributed to the change in the proteins as a result of these procedures.

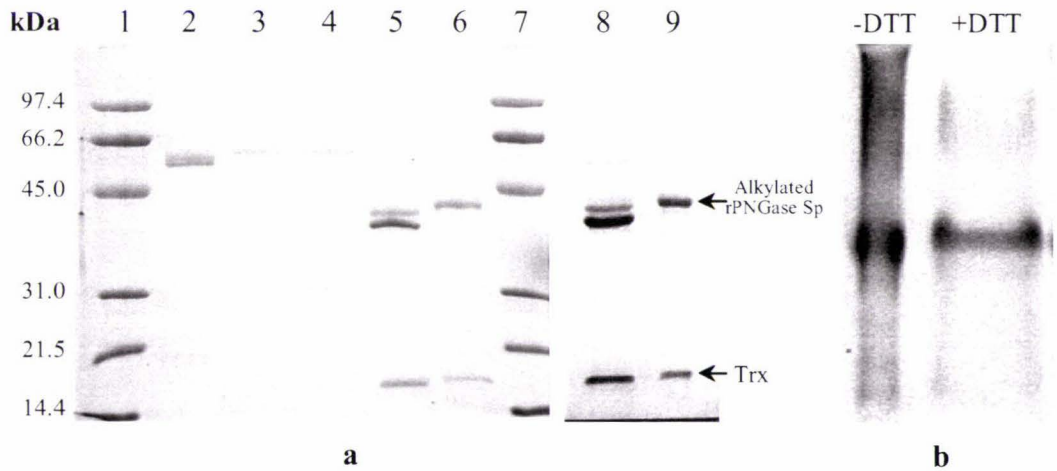


Figure 3.15 a) SDS gels of reduced and alkylated rPNGase Sp. Both fusion and rTEV cleaved rPNGase Sp were treated with 6 M guanidine hydrochloride and 10 mM DTT at room temperature for one hour, then incubated with 55 mM iodoacetamide at room temperature for 45 minutes. After desalting by C-18 RP HPLC, samples were analysed on a 12% SDS polyacrylamide gel. Lanes 1 and 7, marker; Lane 2, untreated fusion rPNGase Sp; Lane 3, reduced and alkylated fusion rPNGase Sp; Lane 4, reduced and alkylated fusion rPNGase Sp after treatment with rTEV protease. Lanes 5 and 8, untreated rPNGase Sp after cleavage with rTEV protease; Lanes 6 and 9, reduced and alkylated rPNGase Sp after treatment with rTEV protease. Both the fusion and cleaved rPNGase Sp migrated as a single band after being reduced and alkylated. rTEV protease was not able to cleave the reduced and alkylated fusion rPNGase Sp (Lane 4). The band migrated to ~31 kDa in Lane 4 was rTEV. **b) Native-PAGE of the fusion rPNGase Sp.** The fusion rPNGase Sp was run on native polyacrylamide gel after being mixed with the sample buffers in which either DTT was absent (lane on the left) or presented (lane on the right). The protein showed single band on native gel.

The results presented here not only support the results of the MS analyses but also indicate that the anomalous migration of rPNGase Sp is related to the redox states of its cysteine residues. Therefore, the two bands seen on SDS gels are clearly not a result of an impurity in the preparation. Rather they represent two conformations of the single polypeptide chain, that are dictated by the cysteine residues in the protein. rPNGase Sp contains nine cysteine residues, four of which are likely to be involved in zinc (II) ion (Zn^{2+}) binding, and therefore are susceptible to oxidation. If, during the preparation, or growth of the cells, this protein is exposed to any oxidising agent, it is likely, especially in the presence of denaturants, to form disulfide bonds. Another characteristic of the protein is the large proportion of hydrophobic residues (aliphatic index: 57.09, gravy value: 0.966, 11.3% aromatics, 9.9% Leu/Ile). If disulfide bond formation brings two hydrophobic secondary structural elements together, it is possible that an extremely stable structure will occur, reminiscent to the prion structure [114]. Of course, this will make part of the polypeptide chain inaccessible to SDS which will result in an anomalous band in SDS PAGE. The results suggest that there are two favoured conformations and that the populations inhabiting each of these conformations can vary (Figures 3.6 b, 3.7 b and 3.15 a). The native gel (Figure 3.15 b) supports this theory. The sample without DTT is smeared along the length of the lane, suggesting a none-homogeneous population of molecules, whereas when DTT is added (without SDS) — a single band is observed indicating that in denaturing gels, the difference is related to the number of SDS molecule that are able to associate with protein. Because the appearance of the double bands does not change with the amount of reducing agent added, or the length of the time the samples are boiled in the presence of reducing agent and SDS (Figure 3.16), it is likely that these bonds form in the Origami B (DE3) cells, which have lower reducing environment.

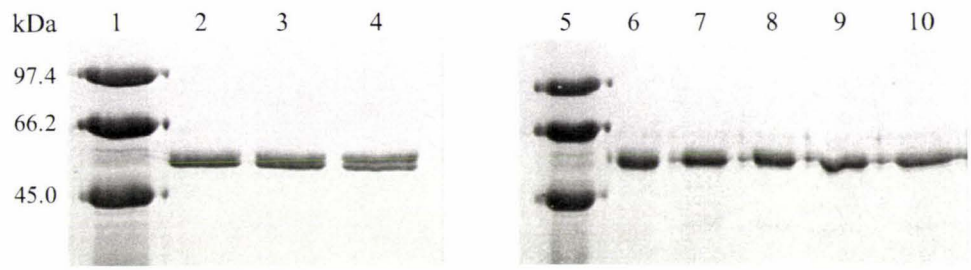


Figure 3.16 SDS-PAGE analyses of fusion rPNGase Sp samples boiled for varying time (left gel) or with varying amount of DTT (right gel). Lanes 1 and 5, molecular weight marker. Lanes 2-4, rPNGase boiled in SDS loading buffer containing 2 × (1.25 M) DTT for 2, 4 and 8 minutes respectively. Lanes 6-10, the same sample boiled in the loading buffers containing 2×, 3×, 4×, 5× and 6× DTT respectively for 2 minutes.

3.5.3 Expression Product was rPNGase Sp

To finally confirm the protein identity, the Coomassie stained band corresponding to the alkylated rPNGase Sp (Figure 3.15 a, Lane 9) was excised from the gel, and tryptic peptides generated from this band were analysed using MALDI-TOF mass spectrometry as described in section 2.8.1. MALDI-TOF analyses resulted in a series of *m/z* values for the peptides (Figure 3.17) which were queried against the NCBI database. For identification of the protein, the monoisotopic masses of peptides generated from the band were analysed using the search programme MASCOT (Matrix Science) with the NCBI database. The search criteria allowed for two missed cleavages and fixed modification of carbamidomethylation of cysteine residues. The results confirmed that the protein sample was PNGase from yeast *S. pombe* (Match to: T39642 Score: 99). Among 17 mass values used in searching, there were 11 matched the calculated masses of the tryptic peptides of PNGase Sp (Table 3.2), and the matched peptides covered 34% of the PNGase Sp sequence.

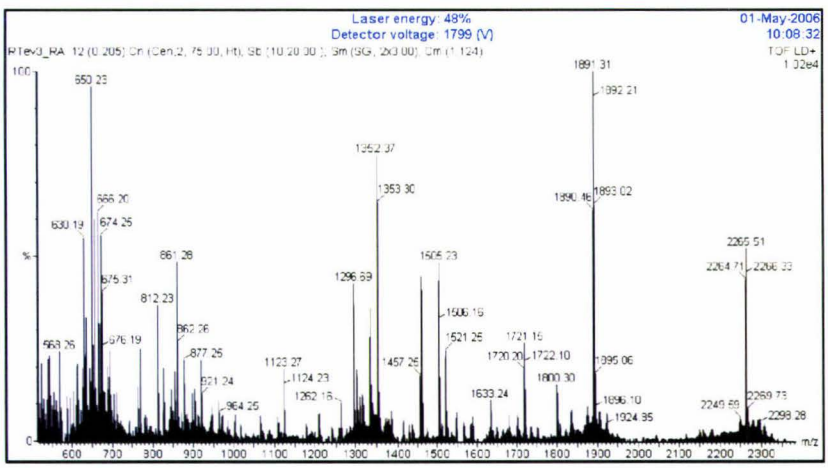


Figure 3.17 MALDI-TOF mass spectra of a tryptic digest of the band corresponding to alkylated rPNGase Sp sample from Figure 3.15 a (Lane 9)

Table 3.2 Observed and calculated masses of the tryptic peptides from the band corresponding to alkylated rPNGase Sp from Figure 3.15 a, Lane 9

Observed mass (Da)	Calculated mass (Da)	ΔDa	Missed cleavage	Position of matched peptides
812.2300	811.3720	-0.1493	0	F^{10}–R^{15}
2264.7100	2263.0294	0.6734	0	N^{18}–R^{36}
2557.0000	2555.0917	0.9010	1	C^{108}–R^{130}
1890.4600	1889.8260	-0.3733	0	N^{131}–R^{145}
674.2500	673.3759	-0.1331	0	A^{154}–R^{159}
1720.2000	1719.6956	-0.5028	0	C^{163}–R^{175}
1632.3000	1631.7799	-0.4872	1	F^{226}–R^{239}
1504.2800	1503.6850	-0.4122	0	M^{227}–R^{239}
1351.0000	1350.7156	-0.7229	1	Y^{245}–R^{255}
920.2700	918.4671	0.7956	0	H^{248}–R^{255}
963.3000	962.4781	-0.1854	1	S^{276}–R^{283}

3.6 rPNGase Sp Appears To Have No Transglutaminase Activity

All known cytosolic PNGases have been found to possess a putative transglutaminase catalytic triad in their highly conserved core region and the residues surrounding this catalytic triad are also conserved in both transglutaminases and cytosolic PNGases. Mutation of any of residues in the transglutaminase catalytic triad (Cys-163, His-190 and Asp-207 in PNGase Sp) was shown to abolish the deglycosylation activity of cytosolic PNGases [15]. Other evidence suggested that cytosolic PNGases also use this catalytic triad for their deglycosylating activity: 1) When the conserved residues Trp-220 or Trp-231 in PNGase Sc were mutated to Ala, deglycosylation activity was completely lost [15]. The residues that are equivalent to Trp-220 and Trp-231 in coagulation factor XIII (Trp-375 and Trp-392 respectively) have been shown to stabilise the orientation of His-373 which is part of the catalytic triad, and Arg-333, an important structural residue, through hydrophobic interaction [85]. 2) Recent structural studies of recombinant PNGases from *S. cerevisiae* and mouse revealed that these PNGases and some transglutaminases such as coagulation factor XIII, arylamine *N*-acetyltransferase and avrpphB hydrolase, share a common core made up of two α-helices and five strands of curved antiparalleled β-sheet. The catalytic triad (Cys-His-Asp) resides within a cleft of the core domain. The rms deviations between these proteins are 2.6-3.4 Å for the 77–121 equivalent C^{α} atoms [27, 29]. The fact that the central helix and the catalytic Cys and His residues aligned well in the superimposed structures of these homologues suggests a close evolutionary linkage between the cytosolic PNGases and the transglutaminases. 3) An irreversible inhibitor tripeptide carbobenzyloxy-Val-Ala-Asp-α-fluoromethylketone (Z-VAD-fmk) was shown to inhibit the rPNGases from *S. cerevisiae* and mouse by forming a covalent bond to S^{γ} of Cys-191. This sulfur was shown to form a hydrogen bond with N^{δ} of His-218 in the transglutaminase catalytic triad. A more specific inhibitor, $Man_9GlcNAc_2$-iodoacetoamide, which contains a high mannose-oligosaccharide, was also shown to make a covalent bond with the S^{γ} of Cys-191. Both the inhibitors bound only to Cys-191 out of the 14 cysteines available in PNGase Sc. All these findings suggested that the cytosolic PNGases utilise the transglutaminase catalytic triad to carry out the deglycosylation reaction, and that Cys-191 acts as the nucleophile, attacking the scissile carbonyl carbon atom from one side of the β-aspartylglucosamine bond during the deglycosylation reaction. The proposed mechanism for the hydrolysis reactions carried out by both the transglutaminases and cytosolic PNGases are shown in Figure 3.18.

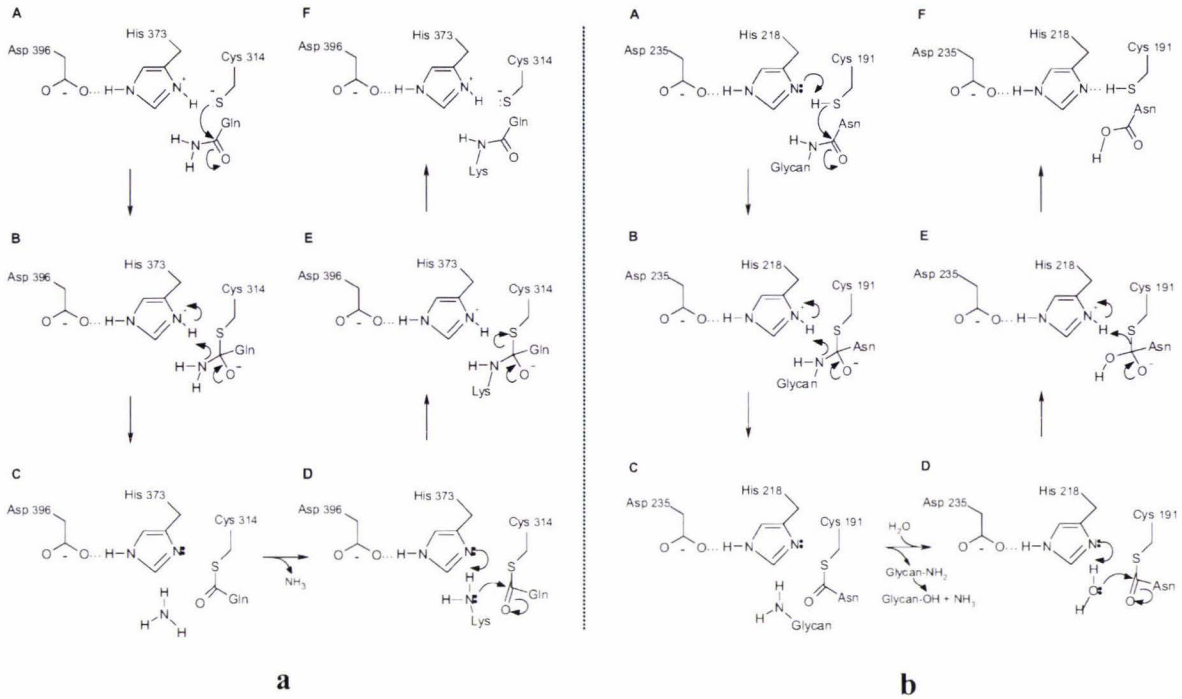


Figure 3.18 a) Proposed mechanism for transamidation catalysed by transglutaminases The residues are numbered according to the sequence of the transglutaminase, factor XIIIa. **b) Proposed mechanism for deglycosylation of glycopetides/glycoproteins by cytosolic PNGases** The residues are numbered according to the sequence of PNGase Sc. Adapted from Pedersen *et al.*, 1994 and Katiyar *et al.* 2002 with some modifications.

As previously mentioned, the reaction catalysed by transglutaminases (synthesis of an amide bond) is the reverse of that catalysed by PNGases (hydrolysis of an amide bond), suggesting similar mechanisms may be applied. Transglutaminases are members of the papain-like superfamily of cysteine proteases and have a Ping Pong-based mechanism [14]. The spatial arrangement of the amino acids in the catalytic centre, Cys-His-Asp, form a charge-relay system that is thought to activate the S^{γ} of the cysteine to attack the scissile carbon of either a peptidylglutamine moiety in transglutaminase-catalysed reactions, or a peptidylasparagine moiety in PNGase-catalysed reactions (Figure 3.18 a, b, Step B) [15, 86]. The imidazole ring of His-373/218 is prone to take up the liberated proton to form an imidazolium ion, a process which is aided by the polarizing effect of the carboxylate sidechain of Asp-396/235 (Figure 3.18, a, b, Step B and C). After the release of ammonia (transamidation reactions) or glycosylamine (deglycosylation reactions), the thioester intermediate undergoes a nucleophilic attack by a primary amine (transamidation) or water (deglycosylation) to release the aspartic acid at the site of cleavage.

Based on the fact that cytosolic PNGases and transglutaminases share both structural and catalytic homology, it is plausible that, besides their deglycosylating activity, cytosolic PNGases may also act as transglutaminases. To test this possibility, a colorimetric assay using carbobenzyloxy-Gln-Gly and hydroxylamine as substrates was carried out as described in section 2.7. The preliminary results showed however that the recombinant PNGase Sp did not have any transglutaminase activity (Figure 3.19).

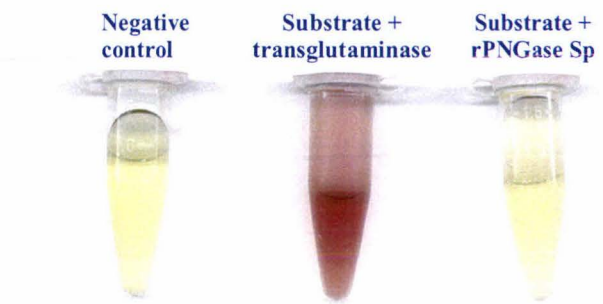


Figure 3.19. Transglutaminase activity assay of the recombinant PNGase Sp Substrates (75 µL 0.2 M carbobenzyloxy-Gln-Gly and 25 µL 2 M $NH_2OH{\cdot}HCl$) were added to 100 µL of 1 M Tris-acetic acid buffer (pH 6.0), followed by 25 µL of each of 0.1 M $CaCl_2$ and 0.02 M EDTA, and 230µL of H_2O. The reaction solution was preincubated for 3 minutes at 37°C, prior to the addition of 20 µL (0.2 U) of guinea pig liver transglutaminase (Sigma) or 20 µL (3.9 mg/mL) of rPNGase Sp to a final volume of 500 µL. After incubation for 10 minutes at 37°C and the subsequent addition of 0.5 mL of stop solution (500 mg $FeCl_3$, 1.5 g TCA, 27.5 mL H_2O, and 2.5 mL concentrated HCl), a brown complex was formed between the product of transglutaminase-catalysed reaction (hydroxamic acid) and Fe^{3+}. The reaction without enzyme served as a negative control.

This result is not surprising because transglutaminase activity has not been reported for any of the cytosolic PNGases. It has, however, been reported that PNGase At from the plant *Arabidopsis thaliana* does have transglutaminase activity [52]. PNGase At, shows obvious evolutionary divergence from the cytosolic PNGases, exemplified by the N- and C-terminal domains which extend out of the core domain [16, 24, 52]. Knowledge of the exact catalytic mechanism(s) of cytosolic PNGases will be essential in order to understand the molecular factors that are responsible for substrate specificity. Structural analyses of the two types of enzymes might provide the answers to this question.

Structural studies [86] of the transglutaminase coagulation factor XIII suggested that both the tetrahedral oxyanion intermediates in Step B and E (Figure 3.18) were stabilised by

NH···O^- hydrogen bonds, where the mainchain NH of the cysteine nucleophile (as in cysteine proteinases) and the atom Nε1 of tryptophan-279 act as the hydrogen bond donors. The cytosolic PNGases do not have residue in this position that can act as a hydrogen bond donor. Although similar oxyanion intermediates were formed in the PNGase-catalysed deglycosylation reactions, they differ from those formed in transamidation reaction in terms of their conformation. These two different enzymes use the same catalytic triad to catalyse what is essentially the same reaction, except that for the PNGases, the reverse reaction is dominant. Furthermore, the residues that surround the catalytic triad and are involved in stabilising the transition state intermediates may be different both in terms of amino acid composition and conformation. It was found that the zinc-binding domain of cytosolic PNGases has a scaffold that is similar to the zinc-ribbon structures of *Bacillus stearothermophilus* adenylate kinase and transcriptional elongation factor S-II [29]. By forming a "lid" over the active site of the enzyme, the zinc ribbon of adenylate kinase is thought to regulate the thermodynamics of catalysis by stabilising the intermediate state and promoting product release [29, 87]. It is possible that the zinc-ribbon-like structure seen in the cytosolic PNGases may play a similar role by either modulating glycopeptide/glycoprotein substrate binding or controlling the thermodynamics of the reaction. Interestingly, this domain is absent in the transglutaminases, so it is tempting to attribute its presence in the PNGases to their ability to catalyse the hydrolysis rather than the transferase reaction. The zinc-binding domain participates in the formation of a deep cleft which is unique to the cytosolic PNGase subset of the transglutaminase superfamily and which contains the active site. Within this cleft, two residues Arg-176 and Arg-190 that are part of this zinc-binding domain are conserved in cytosolic PNGases but not in transglutaminase factor XIII, and make two specific interactions with the inhibitor Z-VAD-fmk (section 1.6). These are two salt bridges formed between the side chain of the aspartic acid in Z-VAD-fmk and the side chains of Arg-176 and Arg-190. The backbone carbonyl groups of the Val and Ala residues in this inhibitor also form H-bonds with the side chains of Arg-176 and Arg-190. It is therefore thought that these two conserved Arg residues may be essential for substrate binding by specifically interacting with the glycan-linked Asn residue, and promote the release of the products of the reaction through electrostatic interactions with the Asp residue which is generated as a result of the hydrolysis [29].

It is possible that cytosolic PNGases can also form a thioester intermediate with the asparagine residue, mimicking the TGase-peptidyl glutamine intermediate in the transamidation reaction (Figure 3.18, Step C). However, because water molecule can access the active site of the PNGases, the deacylation reaction involves a nucleophilic attack by water, converting the intermediate to the free aspartic acid. In the transglutaminases, the deacylation reaction involves a nucleophilic attack by a primary amine, rather than water. Although the thio-acyl enzyme intermediate formed in transamidation reaction catalysed by transglutaminases can also undergo hydrolysis, it is thought to be unfavourable for the enzymes [26].

Considering that the result presented here might be due to the substrate used and/or the assay conditions chosen (pH, temperature and ionic strength), the possibility that rPNGase Sp and/or other cytosolic PNGases may have residual transglutaminase activity cannot be ruled out. For example, calcium ions (Ca^{2+}) are required for the activation of most transglutaminases, because the binding of Ca^{2+} to the enzyme induces a conformational change that allows access of substrate to the active site which is buried in the 'resting' state of the enzyme. It is possible that the concentration of Ca^{2+} used in the assay may have had an inhibitory effect on rPNGase Sp. The assay conditions therefore need to be examined carefully, and such factors eliminated as the cause of the inactivity of the enzyme.

3.7 Substrate Specificity of rPNGase Sp

As mentioned in section 1.9.3, unlike PNGase F from Flavobacterium *meningosepticum* and PNGase A from the plant *Prunus amygdalus*, cytosolic ScPng1p and mPng1p show high substrate specificity in terms of both the structure of the glycan moiety and of the peptide backbone. Generally, it is thought that the cytosolic PNGases preferentially deglycosylate high-mannose glycopeptide compared to complex, hybrid and sialyated glycopeptides, and can only act on denatured/misfolded glycoproteins *in vitro*.

To investigate the substrate specificity of rPNGase Sp, its deglycosylating activity towards a variety of glycoprotein substrates was analysed. Four glycoprotein substrates were used: 1) Fetuin from fetal calf serum contains three sialylated triantennary *N*-linked complex sugar chains [109]; 2) hen egg ovalbumin has a single biantennary *N*-linked complex glycan [55]; 3) ribonuclease B from bovine pancreas contains a single *N*-linked high mannose biantennary glycans [110]; and 4) α1-acid glycoprotein from bovine blood has a mixture of 5 *N*-linked hybrid and complex glycan(s) in which a fucose residue is linked to the proximal GlcNAc residue through an α 1→ 6 linkage. All the four substrates were used in both native and denatured forms in the assays, and the deglycosylation reaction were monitored by gel mobility shifts.

It can be seen from the results that the rPNGase Sp was not able to deglycosylate any of the four substrates in their native forms (Figure 3.20 a, b; Lanes 3, 4, 7 and 8). For the denatured glycoproteins, rPNGase Sp was able to remove the glycans from α1-acid glycoprotein and ribonuclease B, as there was significant increase in the abundance of protein bands that migrated faster than fully glycosylated substrates (Lane 6 and Lane 10, Figure 3.20). Moreover, the enzyme appeared to show much higher activity towards the denatured ribonuclease B, which has a high mannose glycan, than the denatured α1-acid glycoprotein, which has some hybrid glycan(s), because the deglycosylation of ribonuclease B was almost 100% completed while this was not the case for α1-acid glycoprotein after incubation for the same period (overnight). This result is consistent with the concept that cytosolic PNGases are part of the ERAD or GERAD pathway [4, 17, 48]. In the case of denatured α1-acid glycoprotein, two bands appeared at ~27 and ~31 kDa respectively, which may represent two deglycosylated products with different numbers of glycans removed.

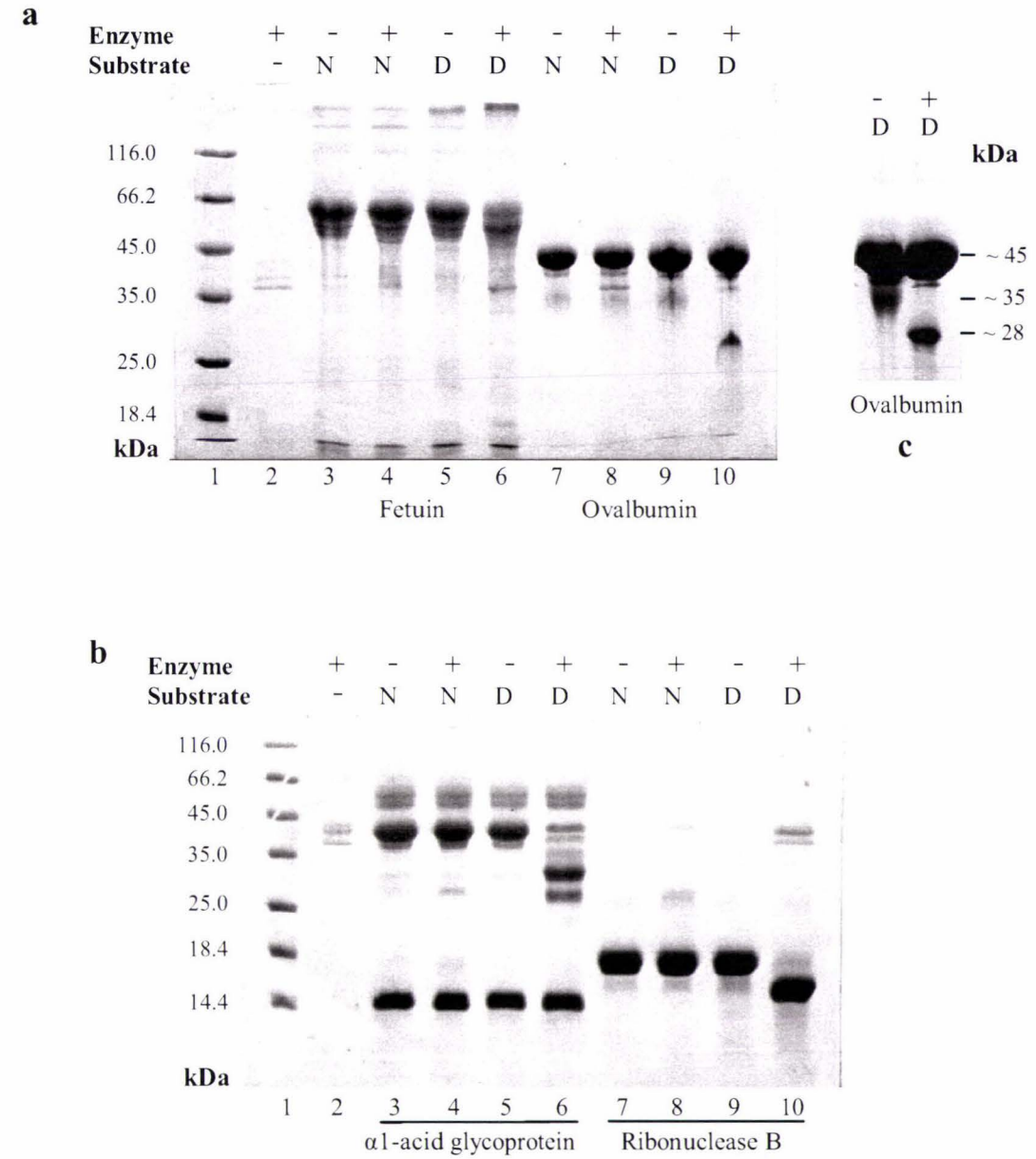


Figure 3.20 Gel shift assay for substrate specificity of rPNGase Sp Four glycoprotein substrates in either native (N) or denatured (D) form were incubated overnight at room temperature with (+) or without (-) rPNGase Sp and the overnight reaction mixtures run on either 12% (**a**) or 15% (**b**) SDS polyacrylamide gels. Gels were stained with Coomassie blue R-250. Deglycosylation of a substrate is represented by a change in its apparent molecular weight after incubation with the enzyme. **a**) Lanes 3-6, reactions using fetuin as substrate; Lanes 7-10, reactions using ovalbumin as substrate. **b)** Lanes 3-6, reactions using α1-acid glycoprotein as substrate; Lanes 7-10, reactions using ribonuclease B as substrate. In both **a** and **b,** lane 1 contains protein molecular weight standards and Lane 2 contains the reaction mixture with no added substrate. 15 μL of each reaction was loaded. In **c)**, the same samples corresponding to Lanes 9 and 10 in picture **a** were loaded, but the sample size is five times as that in picture **a**.

It seems as though there was very limited degradation of the denatured fetuin by rPNGase Sp (Lane 6, Figure 3.20 a). Comparing the negative control (Lane 5, which shows different glyco-forms clustering between 66 and 45 kDa) with the reacted sample, it can be seen that the intensity of the strongest band in substrate has decreased, and there is a concomitant increase in the intensity of a band with a lower molecular weight, which could be a deglycosylated product. Two other minor bands that migrated slightly faster than the major band in the substrate (Lane 5) also appeared to shift downwards after rPNGase Sp treatment, suggesting that the enzyme is able to hydrolyse some of the glycans of denatured fetuin but not others. This has not been previously reported. The experiment was repeated at least twice to ensure the results were consistent, but further investigations into the structure of the glycans removed needs to be carried out to confirm the results.

It appears that the denatured ovalbumin was not deglycosylated, as the intensity of ovalbumin band remained basically same after enzyme treatment (comparing Lane 9 and 10, Figure 3.20 a). The band migrating to ~ 28 kDa in Lane 10 is absent in Lane 9, the control could represent the deglycosylated product of the minor band in the denatured substrate at ~ 35 kDa. This may be a degradation product of the denatured protein. The change was more obvious when the samples of higher concentration were loaded onto the gel (Figure 3.20 c). As the 11-mer glycopeptide used in the HPLC-based activity assays in this study is derived from hen egg ovalbumin and thus bears the same glycan moiety, this seems logical. It is possible that the denaturing process cleaves a small portion of ovalbumin molecules, and that the 35 kDa band seen in the denatured sample represents a glycosylated breakdown product. It is obvious that the enzyme is capable of removing complex glycans as shown by the assay, but only from short peptide rather than longer one. The fact that the ovalbumin irreversibly precipitated when boiled in the denaturing buffer in contrast to the other substrates suggest the protein became aggregated, which might prevent the enzyme accessing the scissile bond. The solution was thus incubated at 60-70 ℃ for 30 minutes with cooling down after every 10 minutes when the solution became milky (to avoid precipitation).

However, the results presented here confirm that cytosolic PNGases can distinguish between native and non-native (misfolded or denatured) glycoproteins attacking only the

latter [17, 49, 50], a property consistent with their proposed biological function – participating in ER-associated degradation (ERAD) of misfolded / unfolded glycoproteins that exported from ER to cytosol. In this pathway, cytosolic PNGases are thought to remove the glycan chains from unfolded glycoproteins before their final degradation in the proteasome [16, 38, 43]. The molecular basis for substrate recognition by cytosolic PNGases is supported by the structural features of PNGase Sc. It is proposed that the deep and narrow cleft which harbors the active site imposes a steric hindrance to tightly folded polypeptides but allows the access by an unstructured and extended polypeptide [27].

With respect to the recognition of the glycan moiety in glycoprotein substrates, the results presented in this study are consistent with the finding that the cytosolic PNGases preferentially deglycosylate high mannose glycoproteins, such as ribonuclease B rather than those bearing complex- or hybrid- type of *N*-glycans. This is again consistent with function, as glycoproteins that are misfolded exit the ER without passing through the glycosylation machinery of the Golgi complex. As such, they will be high-mannose, as complex glycans are only formed during procession through the Golgi complex.

3.8 Effects of Reductant TCEP on PNGase Activity

In order to test whether the activity of rPNGase Sp *in vitro* was dependant on reducing agents, the enzyme was purified using buffers containing varying concentrations of the reducing agent Tris (2-carboxy-ethyl) phosphine hydrochloride (TCEP) at final concentrations: 0, 0.25, 0.5 and 1.0 mM respectively). This reducing agent was chosen because (1) it could be used with IMAC, (2) it is much more resistant to oxidation by air in comparison to DTT or BME, (3) it is a much more efficient reducing agent at room temperature, and at nearly any pH. The thioredoxin-fused rPNGase Sp obtained after the 1st IMAC purification step was used in the assays. As expected, the results showed that the *in vitro* activity of rPNGase required the presence of reducing agent, and that the activity was reducing agent-dosage dependent. That is, the activity increased with increasing concentrations of TCEP between 0 and 1 mM (Figure 3.21). The enzyme was almost completely inactivated when TCEP was not included in the buffer.

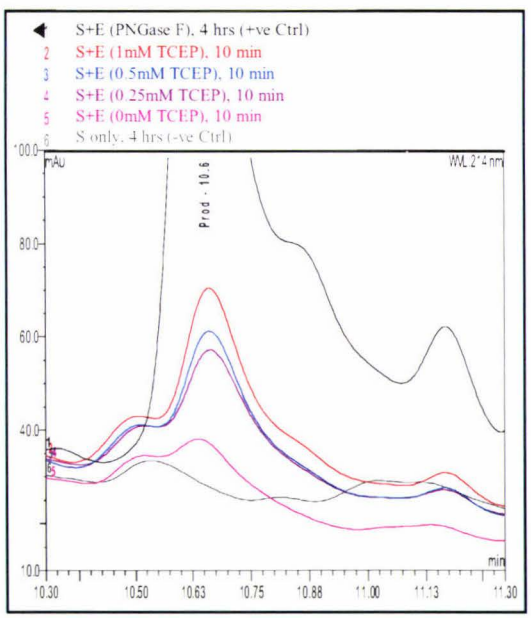


Figure 3.21 Effects of TCEP on deglycosylation activity of rPNGase Sp Trx-rPNGase Sp was purified using buffers with varying concentrations of TCEP, then assayed for the activity using the hen egg 11-mer glycopeptide as substrate. peaks representing the deglycosylated product of the reaction are shown in different colours. Green: -ve control (incubated for 4 hrs); black: +ve control (the glycopeptide was completely deglycosylated using PNGase F); pink, purple, blue and red correspond to reactions of Trx-rPNGase Sp from preparations carried out in the presence of 0 (pink), 0.25 (purple), 0.5 (blue) and 1 (orange) mM TCEP. All the buffers used to prepare the substrate also contained the matching concentration of TCEP.

The result suggested that all nine cysteine residues in rPNGase Sp need to be reduced for the enzyme to be active. According to the results of structural studies of ScPNGase and mPNGase in complexes with Rad23 [27, 29], the four Cys residues (Cys-105, -108, -137 and -140 in PNGase Sp) residing in two CXXC motifs tetrahedrally coordinate one zinc ion to form a zinc-ribbon-like structure which is thought to be important for the activity. One Cys residue (Cys-163 in PNGase Sp) in the transglutaminase-like catalytic triad (Cys-

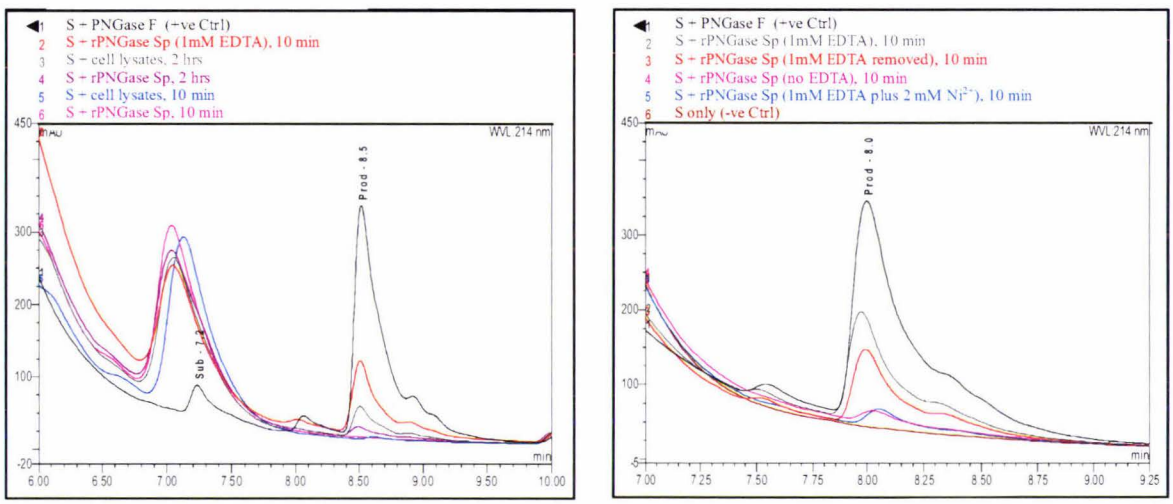
His-Asp) was found to covalently interact with the aspartic acid residue of a tripeptide inhibitor, indicating that this cysteine is most likely to be the nucleophile in the deglycosylation mechanism. It appears that the remaining four Cys residues, Cys-169 and Cys-174, which reside in the highly conserved transglutaminase domain, are conserved in SpPNGase, mPNGase and hPNGase but not in ScPNGase, while Cys-230 and Cys-258 in the central core are not conserved but also need to be reduced.

3.9 rPNGase Sp Appears to be Sensitive to Nickel Ions

Previously, when purifying rPNGase Sp, at least 1 mM EDTA was included in all the buffers used after the IMAC purification step in order to inhibit metalloprotease activity. After it was reported that the ScPNGase and mPNGase were zinc-binding proteins [27, 29], EDTA was excluded from the buffers to avoid removal of the intrinsic zinc ions. Intriguingly, this appeared to inactivate rPNGase Sp, as the activity of the 1st IMAC elution fractions was lower than that of the cell extracts (purple and green product peaks respectively in Figure 3.22 a, two-hour incubation). Addition of 1 mM EDTA was found to restore the activity of the fusion rPNGase Sp after IMAC as shown in Figure 3.22 a, where the red peak shows the products of a 10-minute incubation.

This finding indicated that rPNGase Sp is inhibited by metal ions during downstream processing, most likely by nickel ions (Ni^{2+}) leaching from the metal ion affinity column. It was unlikely that the EDTA molecule itself was activating the enzyme because the cell lysate was more active than the IMAC-purified enzyme. To confirm that the drop in the activity was due to Ni^{2+} inhibition, the following tests were carried out using rPNGase Sp that had the tag removed: EDTA was added into the enzyme solution to a final concentration of 1 mM. After incubation for three hours, excess EDTA, and EDTA-metal ion complexes were removed by ultra-centrifugation followed by several washes in the buffer that had been exposed to Chelex resin to render it metal ion free. The activities of the original, EDTA-treated and metal ion/EDTA-free samples were measured and their activities compared. The results showed that, similar to what was found for the Trx-PNGase Sp, the enzyme was activated by addition of 1 mM EDTA (pink and green peaks in Figure 3.22 b). After EDTA was removed from the enzyme solution, the activity remained basically the same within the accuracy of the assay.

When $NiCl_2$ was added to a final concentration of 2 mM into the enzyme, the activity dropped (green and blue peaks, Figure 3.22 b). It would therefore appear that rPNGase Sp is sensitive to Ni^{2+} ions. This is not surprising when the number of free thiols in the protein is considered. It is highly likely that excess metal ions bind to the free thiols, which may or may not include the active site thiol, causing conformational change and/or preventing catalysis.



a **b**

Figure 3.22 The effects of Ni^{2+} ions on rPNGase Sp **a)** Fusion rPNGase Sp obtained after the 1st IMAC purification step was less active than the cell lysate, but could be reactivated by the addition of EDTA. The product peaks (retention time: 8.5 min) are shown in different colours, and the corresponding reactions are indicated above the chromatograms. **b)** Incubation of PNGase Sp that was treated with EDTA (green), that was extensively dialysed against metal-free chelexed buffer to remove the EDTA (red), that had not been treated with EDTA (pink) and that had been treated with EDTA, then 2 mM Ni^{2+} added back (blue). All the incubations were 10 minutes.

3.10 Determination of Optimum pH

As mentioned in section 2.9.4.1, to ensure the response obtained on conversion of 10% of the substrate to product was high enough for accurate quantitation in the HPLC-based activity assay, the 11-mer glycopeptide was covalently labeled with FITC, which significantly increased the sensitivity of detection. The enzyme was assayed between pH 5 and 9 using a wide range buffer, 50 mM malonyl-imidazole-borate (MIB), 1mM EDTA, 5mM DTT. The results are shown in Figure 3.23 and the experimental data in Data Sheet #1 (Appendix VI)

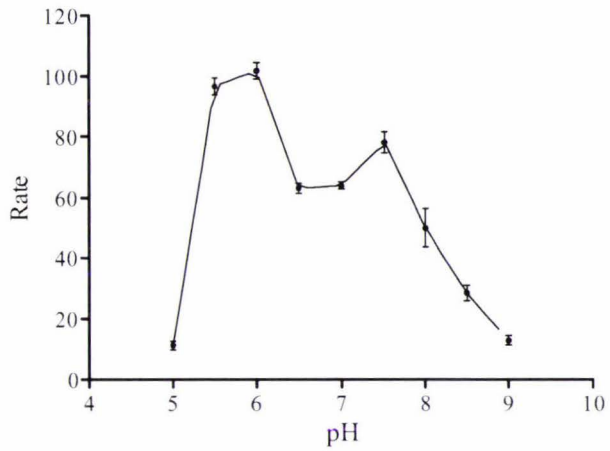


Figure 3.23 Influence of pH on rPNGase Sp activity The activity of rPNGase Sp using FITC-Ova as substrate were measured between pH 5.0 and 9.0 at 25℃. Each data point represents the average of three independent experiments, and error bars correspond to the standard deviation for each determination.

The results indicated that rPNGase Sp exhibits relatively high activity in a weak acidic environment, with maximum activity around pH 6.0. This was not surprising because a neutral pH ranging from 6.5 to 7.5 has been reported to be optimal for other cytosolic PNGases such as ScPng1p and mPng1p [4, 21, 28, 47]. Intriguingly, under the assay conditions used here, the activity of rPNGase Sp reached another apex around pH 7.5. It was unlikely that the drop in the activity between the two peaks was due to the effect of isoelectric point (pI) precipitation, because a drop in enzymatic activity caused by insolubility around the isoelectric point normally occurs within a very narrow pH range. Furthermore, the calculated pI for rPNGase Sp is 5.6, making this unlikely. The results presented here were reproducible under the assay conditions used for the unlabeled 11-mer glycopeptide substrate as shown in Figure 3.24.

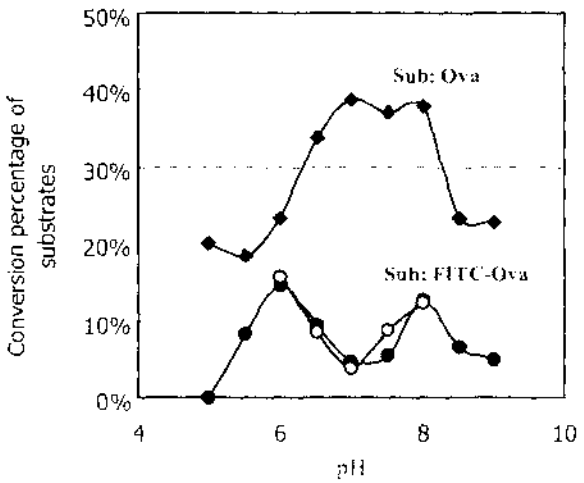


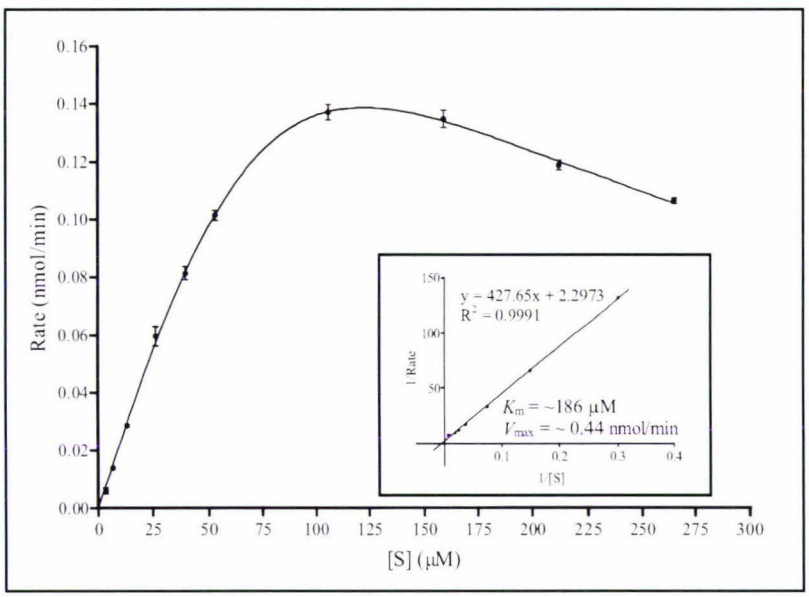
Figure 3.24 Primary estimation of optimal pH for rPNGase Sp using both the FITC-Ova and unlabeled substrates *Filled diamonds*: data for the assay using the unlabeled 11-mer glycopeptide (Ova) as the substrate. *Filled circles*: data for the assay using FITC-Ova as the substrate. *Open circles*: repeated data point using FITC-Ova as the substrate.

When the unlabeled substrate was used, the optimum pH range appears to shift one pH unit becoming more basic compared to that for FITC-Ova. Although the data presented in Figure 3.21 for the unlabelled substrate did not represent the initial velocity of the catalytic reaction (conversion to product was >15%), the pH profile is similar to that for the labelled substrate. While this bimodal pH effect could be due to the buffer used. The reproducibility of the results suggests that the profile observed is real. To test this, the results should be corroborated by repeating the experiments using different buffers. Kinetic parameters were determined using 50 mM MIB buffer, pH 6.0.

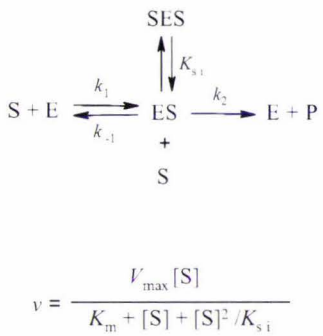
3.11 Kinetic Studies of the rPNGase Sp-catalysed Deglycosylation

3.11.1 Determination of Michaelis Constant (K_m) for FITC-Ova

A K_m value of 0.13 mg/mL (~ 37 μM) had been previously determined for recombinant PNGase F using the FITC-dilabeled ovalbumin 11-mer glycopeptide as substrate [112]. To evaluate the affinity of rPNGase Sp for the same substrate and to compare it with that of PNGase F, the K_m of rPNGase Sp for the FITC-Ova was estimated as described in section 2.9.5. The results are shown in Figure 3.25 and the experimental data in Data Sheet #2 (Appendix VI).



a



b

Figure 3.25 (a) The kinetic curve of rPNGase Sp-catalysed deglycosylation of FITC-Ova generated by GraphPad Prism® software (Version 5). The reaction exhibited substrate inhibition, when [S] was greater than ~ 100 μM. Each data point represents the average of three independent experiments, and error bars correspond to the standard deviation for each determination. The insert shows the Lineweaver-Burke plot of the eight data points at low [S]. The rate data used in transformation was the predicted rate value calculated based on the fitting results. **(b) Schematic shown of the Substrate inhibition model.** Adapted from Alejandro G. Marangoni, 2003

It can be seen from Figure 3.25 (a) and experimental data sheet #2 (Appendix VI) that the enzyme reaction was subjected to substrate or product inhibition as shown by the decline in the reaction rate (v) at substrate concentrations ([S]) greater than ~ 100 μM. To estimate the K_m, the original data points were fitted using a substrate inhibition model (Figure 3.25 b) with GraphPad Prism® software (Version 5). A series of predicted v was generated and

the eight data points generated at low [S] were transformed to produce a Lineweaver-Burke plot to estimate the K_m (inset, Figure 3.25 a). K_m and V_{max} were calculated as ~ 186 μM and ~ 0.44 nmol/min respectively. Because the ovalbumin glycopeptide is actually a mixture of different hybrid- and oligomannose glycans [55], the calculation of the exact molecular mass of the substrate is impossible. Therefore, an average mass of 2.59 kDa as determined by mass spectrometry, was used for the 11-mer glycopeptide (Appendix VI), resulting in a molecular mass of 3.37 kDa for the di-labelled substrate (FITC-Ova). The K_m (~ 186 μM) determined for rPNGase Sp is about five times greater than that of the PNGase F (37 μM), suggesting rPNGase Sp has a much lower affinity for this complex glycoconjugate compared to PNGase F. However, as the enzyme clearly prefer high mannose substrates, this is hardly surprising.

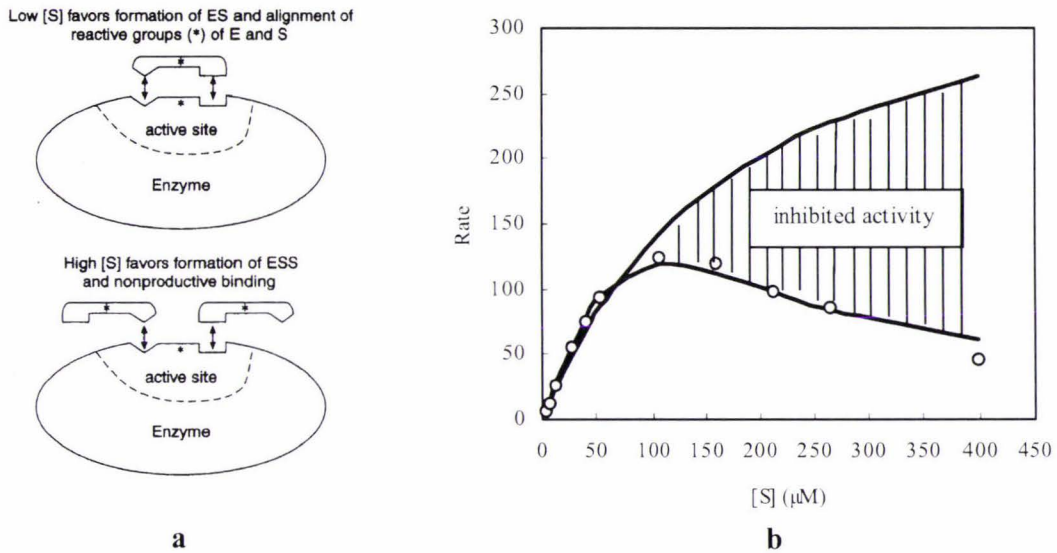


Figure 3.26 a) Visualization of model derived for substrate inhibition of enzyme in Figure 3.22 b. Adapted from Alejandro G. Marangoni, 2003. **b) Assessment of the type of inhibition.** Observed reaction rates (presented in the fluorescence absorbance peak area of products) contrasted with the predicted behavior (upper curve) of an uninhibited enzyme with the V_{max} and K_m values derived from Figure 3.22 (inset).

How can excess substrate inhibit an enzyme? The following model has been suggested as each of two substrate molecules bind to different subsites of the enzyme's active site, resulting in nonalignment of reactive groups (designated as "*") on enzyme and substrate [88] (Figure 3.26 a). In this model, at high concentrations of FITC-Ova, two substrate molecules bind to one enzyme molecule. One binds to the glycan-binding site, while the other to the peptide-binding site, forming a non-productive ternary complex. This is

possible because:

(1) as mentioned in section 1.9.3, early studies on the substrate specificities showed that cytosolic PNGases recognise both the structures of *N*-linked glycan and the peptide their substrate. Moreover, free substrate-derived glycan chains have been shown to inhibit mPNGase [89]. Recent structural studies of mPNGase have shown PNGases possess separate binding sites for both peptide and glycan chain moieties respectively [27, 29, 90] (section 1.6). In a proposed substrate binding model of mouse PNGase (Figure 3.27 a) [90], a continuous cleft composed of the core and the C-terminal domains of mPNGase is observed, with one end accommodating the peptide moiety, represented by an inhibitor Z-VAD-fmk, the center accommodating the chitobiose moiety, and the C-terminal domain, which is though to be a oligomannose carbohydrate-binding domain, accommodating the mannotetraose. These results show that a broad surface area of the enzyme is involved in the normal substrate binding. Although such an extended C-terminal domain is absent in yeast PNGases, the presence of an equivalent carbohydrate binding area on the core domain has been suggested by Lee and co-workers [27, 29]. In their model (Figure 3.27 b), the peptide and the chitobiose moieties are docked in the deep cleft and in close proximity to the catalytic triad, while the three mannose residues in the $Man_3GlcNAc_2$ core (Figure 1.7) in glycan moieties common to all *N*-linked glyco-peptides/proteins are bound at the end of the cleft. Moreover, several conserved regions that extend outside the cleft are thought be involved in recognition and binding of the extended or branched carbohydrate moieties, although the precise interactions remain unclear.

(2) The active site of cytosolic PNGases is thought to be located in a deep and narrow cleft on the core domain of the enzymes [27, 29] which may impose steric hindrance to large globular glycoprotein substrates. The substrate used in this study was a small 11-amino acid glycopeptide which by definition is likely to have a somewhat flexible structure. This lack of steric constraint may allow competitive binding to occur with the formation of a non-productive substrate-enzyme-substrate complex.

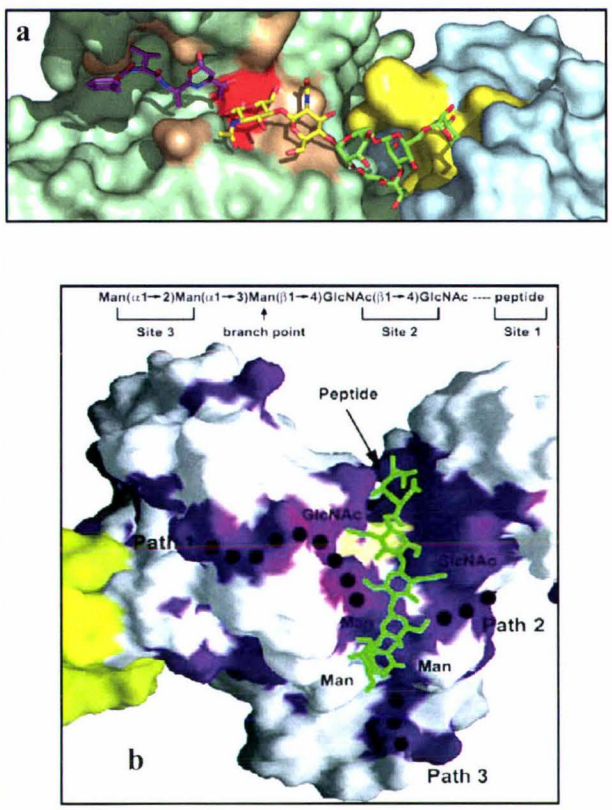


Figure 3.27 a) Hypothetical model of the relative arrangements of the core and C-terminal domains of mPNGase. In the core domain (green), residues of the catalytic triad are colored in red, and other conserved residues in the binding pocket are colored in pale orange. The mannose-binding residues in the C-terminal domain (cyan) are displayed in yellow. The carbon atoms of the chitobiose are colored in yellow, and those of the mannotetraose are colored in green. The carbon atoms of thePNGase inhibitor Z-VAD-fmk are shown in magenta. Adapted from Zhou *et al.* 2006. **b) Molecular model of the yPNGase–substrate complex.** A glycan moiety containing two GlcNAc residues and three mannose molecules was modeled on the active site of the cleft. The dotted lines indicate the three conserved regions where the additional carbohydrate molecules can bind. The common structure of a glycan motif is shown at the top. Adapted from Lee *et al.* 2005 and Zhao *et al.* 2006.

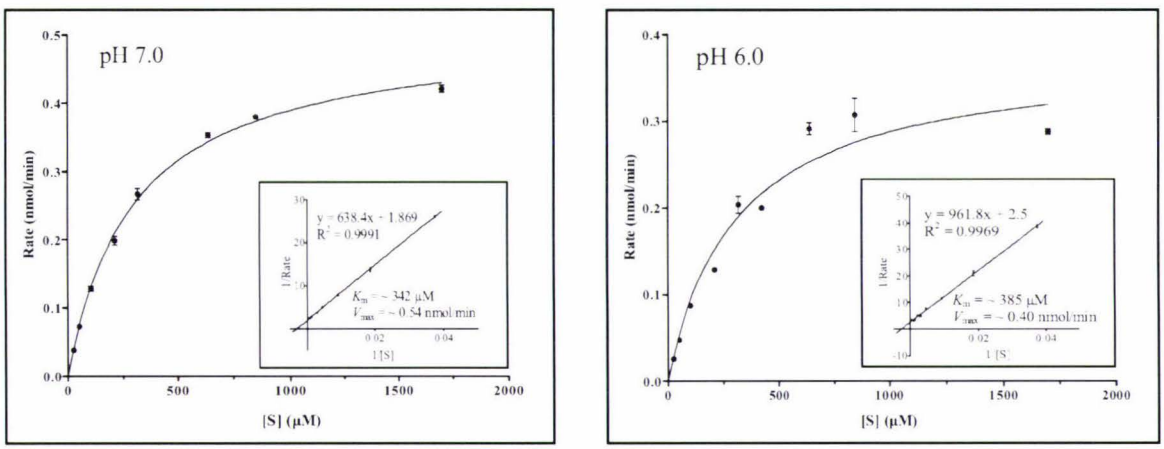
To make an assessment of the type of inhibition, the curve for the reaction rate versus [S] observed (inhibited) was compared to the predicted curve for an uninhibited reaction using Michaelis kinetics with estimated values for K_m and V_{max} to yield the upper curve in Figure 3.26 b. From this analysis, it can be seen that rPNGase Sp was inhibited at the FITC-Ova concentrations greater than ~ 50 μM.

3.11.2 A comparison study on the kinetics of unlabeled and FITC-labeled substrates

As the rPNGase Sp was found to be inhibited by the labeled substrate, experiments were carried out to test whether the unlabeled substrate, the hen egg ovalbumin derived 11-mer glycopeptide (Ova), also had the same effect. To this end, and to primarily compare the affinity of the rPNGase Sp towards FITC-Ova relative to that towards Ova, the kinetic parameters of rPNGase-catalysed deglycosylation of Ova were also estimated using the same assay conditions. Because the results presented in Figure 3.24 (section 3.10, pH-dependency) indicated that the reaction rate was higher at pH 7.0 than at pH 6.0 for this

substrate, the assays were carried out at both pH 7.0 and 6.0. The results are shown in Figure 3.28 and the experimental data tabulated in Data Sheets 3 and 4 (Appendix VI).

Firstly, The results confirmed that, pH 7.0 is the optimum pH for the unlabelled glycopeptide substrate, which is in good agreement with the optimal pH reported for some cytosolic PNGases [4, 21, 28, 47]. Secondly, in contrast to what was seen for the labelled substrate, there was no sign of substrate inhibition of the enzyme within the tested [Ova] range (from 26 to 1700 μM) (Figure 3.28). Although it had been suggested that recombinant PNGase F exhibited substrate inhibition at a very low [Ova] [113], the active site structure and catalytic residues of PNGase F are quite different from those of the cytosolic PNGases [6, 27, 29].



a **b**

Figure 3.28 K_m estimation of rPNGase Sp for Ova substrate at either pH 7.0 (a) or pH 6.0 (b) The kinetic curves were generated using GraphPad Prism® software (Version 5) based on the Michaelis-Menten model. Each data point represents the average of two to three independent experiments, and error bars correspond to the standard deviation for each determination. The inserts show the Lineweaver-Burke plot of the original data points.

A K_m of either ~ 385 μM at pH 6.0 or ~ 342 μM at pH 7.0 was estimated for the Ova substrate, while this was ~ 180 μM for the FITC-Ova at pH 6.0, indicating a higher affinity of rPNGase Sp for the FITC-Ova than Ova substrate. As mentioned in section 2.9.4.1, during labeling of the 11-mer glycopeptide substrate, two fluorescein isothiocyanate (FITC) molecules were covalently linked to the N-terminal amino group and the ε-amino group of the lysine residue that is one amino acid apart from the glycan-bound asparagine residue (reactive group) in the glycopeptide substrate (Figure 2.3). This

modification was originally expected to lower the affinity between the substrate and the enzyme due to possible steric hindrance imposed by the proximity of the FITC group to the reaction center. How the covalent labeling of the substrate positively affected its binding to the active site of rPNGase Sp is not clear.

Using the determined V_{max} and K_m values, the catalytic constant (k_{cat}, also called turnover number) and catalytic efficiency (k_{cat}/K_m) of rPNGase Sp for two substrates were calculated, and the results listed in Table 3.3.

Table 3.3 Values of K_m, k_{cat}, and k_{cat}/K_m for rPNGase SP and Ova and FITC-Ova

Reaction systems	V_{max} (μmol·min^{-1})	$[E_t]^a$ (μM)	K_m (μM)	$k_{cat}{}^b$ (min^{-1})	k_{cat}/K_m (min^{-1}·μM^{-1})
Ova, pH 7.0	5.68	1.33	342	4.27	0.0125
Ova, pH 6.0	4.21	1.33	385	3.17	0.00823
FITC-Ova pH 6.0	4.63	1.33	186	3.48	0.0187

[a] total enzyme concentration ([E]+[ES]) in the reaction system [b] catalytic constant ($k_{cat} = V_{max}/[E_t]$)

It can be seen from the table that the turnover number (k_{cat}) at pH 7.0 is higher than that at pH 6.0. This is logical because, when the amino acid composition in the active site of ScPNGase and mPNGase (Figure 1.5 b and Figure 3.29 a) [27, 29] is examined, the only residue whose ionisation state changes as the pH varies between 6 and 7 is most likely to be the histidine residue in the catalytic triad, although as pK_a of this histidine can be influenced by its environment, its is not certain what its pKa is in this enzyme. According to the proposed catalytic mechanism (Figure 3.29 b) [15], the catalytic residues (Cys-His-Asp) spatially arrange into a charge-relay system, in which the polarising effect of the carboxylate group of the aspartic acid promotes the imidazole ring to accept the liberated proton from the free thiol group of the catalytic cysteine making it a better nucleophile (Figure 3.18 b and Figure 3.29 b).

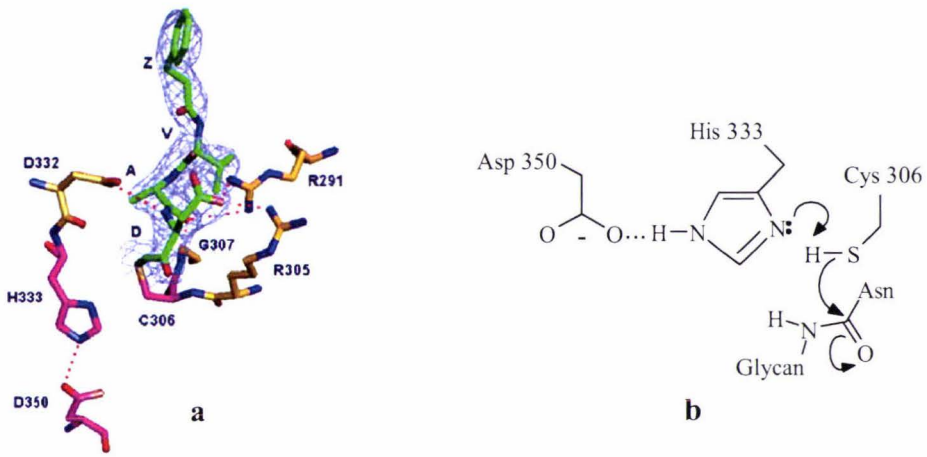


Figure 3.29 (a) active site of the mouse PNGase in complex with a tripeptide inhibitor carbobenzyloxy-Val-Ala-Asp-α-fluoromethylketone (Z-VAD-fmk) The catalytic triad is shown in pink while the inhibitor in green. Upon binding of the Z-VAD-fmk, the thiol group of Cys-306 that was H-bonded to His-333 moved away and covalently bound to the methylene carbon in the inhibitor that was originally linked to the α-fluoromethylketone, a leaving group during covalent inhibition. The inhibitor is bound to the enzyme via two hydrogen bonds and ionic interactions involving residues Arg-291 and Arg-305 of PNGase and the aspartate side chain of the inhibitor. Hydrogen bonds and ionic interactions are shown as red dashed lines. Adapted from Zhao *et al.*, 2006. **(b) Proposed charge-relay system consisting of the transglutaminase-like catalytic triad** The thiol group of Cys-306 makes a nucleophilic attack on the carbonyl carbon atom of the amide bond in glycopeptide substrate. This process is aided by the polarising effect of Asp-350 through His-333. Adapted from Katiyar *et al.* 2002.

To make a direct comparison between two substrates in terms of reaction rate, reactions at four different substrate concentrations (from 53 to 425 μM) were tested for both labelled and unlabelled substrates at pH 6.0 and 7.0 (Figure 3.30). Some unexpected results were obtained. Lowering the pH from 7.0 to 6.0 did not affect the binding of the unlabelled substrate to the enzyme, as the K_m values were approximately the same. It did, however, affect V_{max} indicating that the ionisation state of the histidine in the catalytic triad is important (Figure 3.30 d). However, comparing the labelled with unlabelled substrates at pH 6.0 shows there are significant changes. Addition of the label significantly reduced K_m and increased the catalytic efficiency, suggesting that the affinity of the enzyme for the labelled substrate is markedly increased. V_{max} remains approximately the same, however, indicating that the catalytic mechanism remains unaffected. This indicates that the label is providing some sort of specific interaction in the substrate binding site. As the function of these enzymes is to remove the glycan from improperly folded proteins, it is possible that the FITC group is mimicking the hydrophobic resides that might be exposed in an unfolded protein, and may be recognised by the binding site.

It can be seen from Figure 3.30 a, that at pH 6.0, the rates of reaction with the labelled substrate were significantly higher than those of the unlabelled substrate when [S]<200μM. Above this substrate concentration however, the rates for the labelled substrate dropped quickly to be lower than those of the unlabelled substrate due to substrate inhibition. At pH 7.0 (Figure 3.30 b) it was a different story, the rates for both labelled and unlabelled substrates followed the same pattern, with those for the unlabelled substrate being slightly higher over the whole range of substrate concentrations used. Figures 3.30 c and d show the rates of reaction for the labelled substrate at pH 6.0 and pH 7.0, and the unlabelled substrate at pH 6.0 and pH 7.0, highlighting these differences. The results clearly show there are two effects — one due to the presence of the label, and one due to the pH.

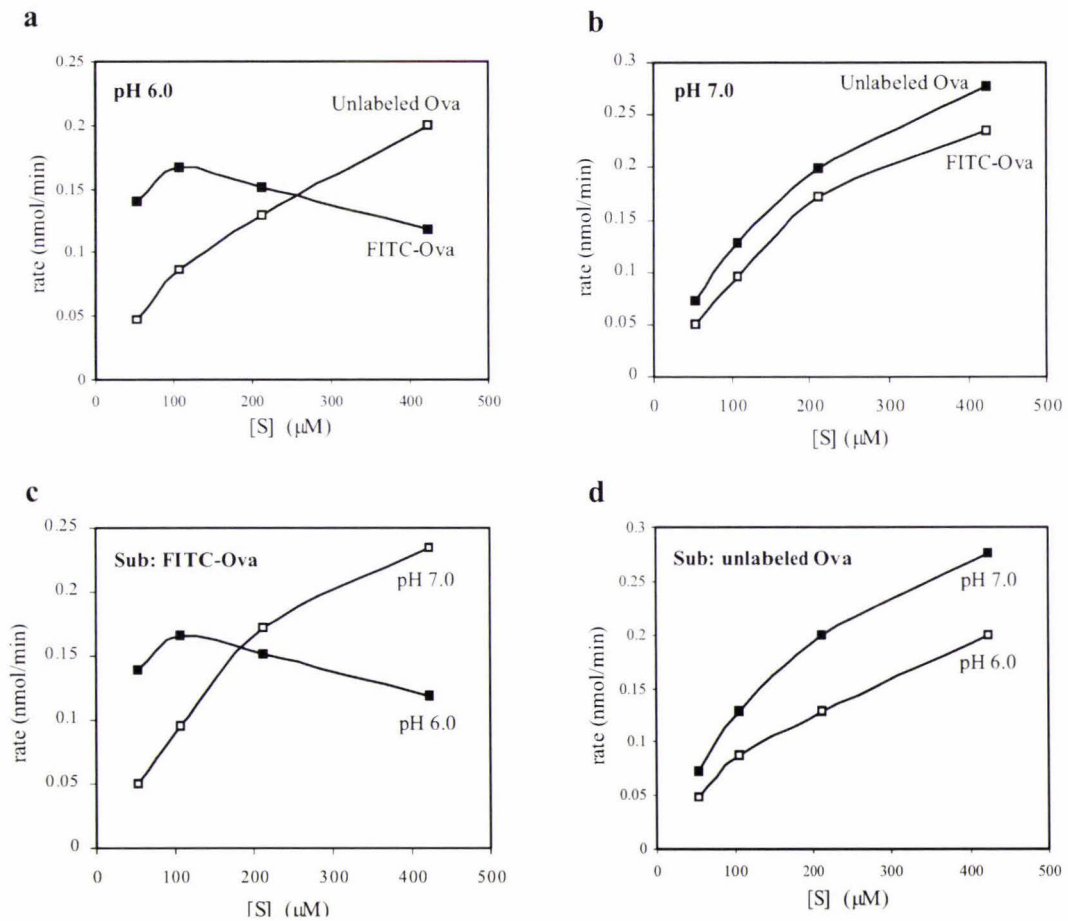


Figure 3.30 **(a)** Comparison of the reaction rates between two substrates at pH 6.0; **(b)** Comparison of the reaction rates between two substrates at pH 7.0; **(c)** Comparison of the reaction rates for FICT-Ova at pH 7.0 and pH 6.0; **(d)** Comparison of the reaction rates for Ova at pH 7.0 and pH 6.0. Four [S] ranging from 53 to 425 μM were tested. Each data point represents the average of three independent experiments.

Considering the effect of the label: FITC, is a relatively large hydrophobic molecule that will be attached to the glycopeptide at the N-terminus and to the primary amide of the lysine sidechain through a thioester linkage. This will result in the removal of two positive charges from the peptide at neutral pH. However, because the fluorescein group itself exhibits a number of different ionisation states, depending on pH, these will be replaced by one to two negative charges (Figure 3.31) [91]. At pH 6.0, the fluorescein moiety will have mostly one negative charge, while at pH 7.0, a significant number of the molecules will have two negative charges, the second residing on the phenolic oxygen of the xanthene ring system. The three dimensional structures of both mPng1p and ScPng1p are homologous in the active site region [27, 29]. As SpPng1p has a high sequence homology to ScPng1p, especially in the active site region, it is not unreasonable to assume that their three dimensional structures will be very similar, especially in this region. The structures show that the active site cleft is quite wide and deep, certainly large enough to accommodate the extra bulk of the fluorescein label. Furthermore, it is lined with a number of basic, acidic, and hydrophobic residues. The results show that at pH 7.0, both the labelled and the unlabelled substrates are processed by the enzyme although the rate is significantly lower at all substrate concentrations for the labelled substrate (Figure 3.30 b). This is not surprising and can be attributed to the presence of the label lowering the affinity of the enzyme for the substrate, probably because of less favourable interactions between enzyme and substrate. At pH 6.0 however, this pattern is reversed, especially at low substrate concentrations. The labelled substrate exhibits a higher rate of reaction than that of either the labelled or unlabelled substrate at pH 7.0, and higher than the unlabelled substrate at pH 6.0. When the substrate concentration reaches ~ 50 μM, however, the rate begins to fall, probably due to substrate inhibition. Interestingly, the labelled substrate does not seem to cause inhibition at pH 7.0. It would therefore appear that the differences observed for the labelled substrate at pH 6.0 and 7.0 must be due to the substrate itself rather than to any effect that is dependent on the enzyme itself. At pH 6.0, most of the fluorescein groups will have only a single negative charge, and the xanthene rings will be neutral. It is possible that one or both of these make a favourable interaction with the enzyme that results in the substrate assuming a more transition like conformation. As substrate inhibition is observed, and only at this pH, it is possible that one or the other of these labels can interact in a non-productive way to the same site, blocking proper access to the active site. At pH 7.0, most of the fluorescein will have a second charge as shown below, which disrupts this interaction, resulting in more normal behaviour of the enzyme.

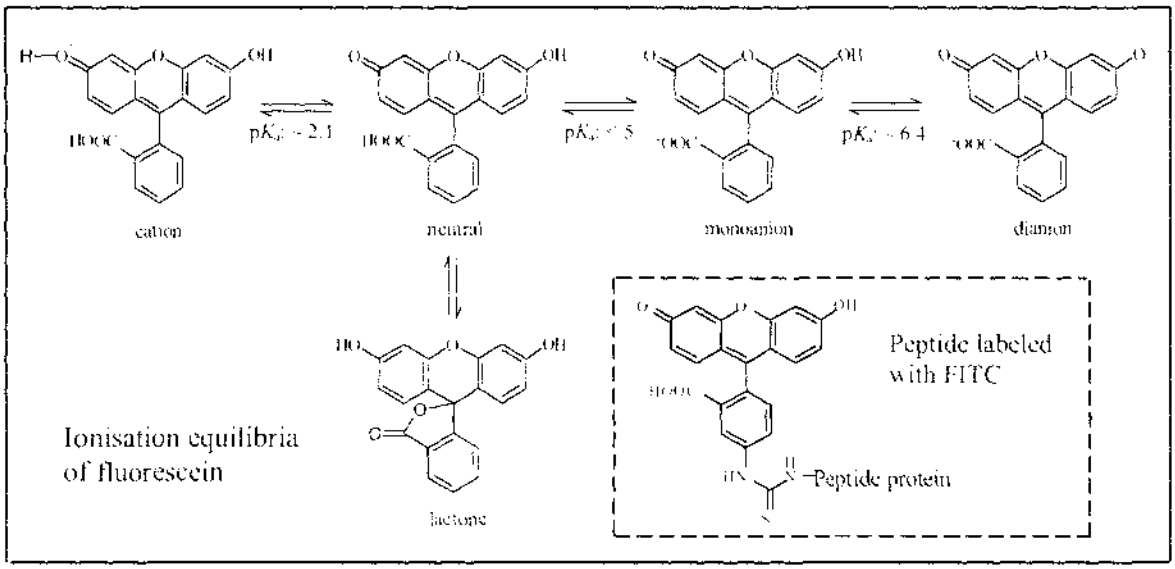


Figure 3.31 Multiple ionisation of fluorescein, the parent compound of FITC. Inset: common structure of FITC-peptide/protein conjugates FITC reacts with primary amino groups of proteins peptides to yield fluorescein-5'-thiocarbamoyl (FTC) conjugates. Adapted from Volker Schnaible *et al.*, 1999.

The rate of the reaction using FITC-Ova (pH 7.0) was measured at only four substrate concentrations. The data was therefore insufficient to determine the kinetic parameters for this reaction system, but a K_m was roughly estimated as ~ 658 µM, which is significantly higher than that determined at pH 6.0 (Table 3.3).

3.12 Effects of Metal Ions

3.12.1 Effects of Bivalent Metal Cations

The effects of different metal cations on the activity of rPNGase Sp were studied in this section. The results are presented in Figure 3.32 and tabulated in Data Sheet #5 (Appendix VI). The activities are expressed as percentage of the activity without addition of any metal ions, which was taken as 100%. Originally, the metal ion effects were tested at two final concentrations, 5 μM and 5 mM respectively. However, once it was found that the enzyme was inhibited by trace amounts of Ni^{2+} ions released during IMAC, and that this bound Ni^{2+} was not easily removed during subsequent purification steps, 1 mM EDTA was included in all the buffers to keep the enzyme active (section 3.9). In addition, considering that rPNGase Sp is a zinc-protein, dialysis against EDTA before testing for metal ion effects was avoided. Taken together, these limitations made it difficult to test the effect at very low concentrations of metal ions.

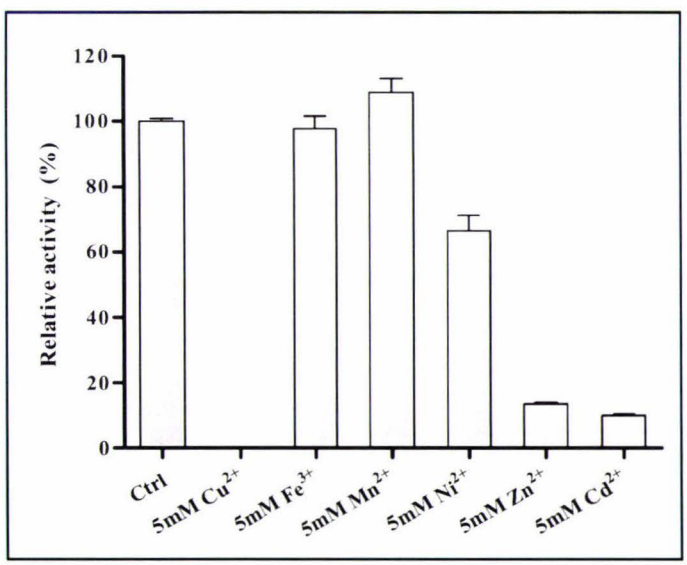


Figure 3.32 The effects of metal ions on rPNGase Sp activity all the assays were carried out under the standard assay conditions described in section 2.9.4. Assay buffer: 50 mM MIB, 1 mM TCEP, 1 mM EDTA, pH 6.0; substrate: FITC-Ova.

The results showed that the presence of 5 mM Fe^{3+} had little effect on the enzyme activity (98% activity remained). It seems that 5 mM Mn^{2+} activated the enzyme compared to the control incubation (increased to 108%). This effect was however barely significant. An interesting observation made during observed during the purification process was that the concentrated enzyme solution showed (not always) a pink colour after the 1st IMAC step in some preparations, indicating that other metal ions, may be such as Mn^{2+} or Co^{2+} may have become coordinated to the enzyme molecule during expression/purification. The

difference in the colour between different preparations might just reflect the varying contents of metal ions in the media or the water used, although this did not seem to affect the activity of the enzyme.

The presence of 5 mM Cu^{2+} completely inhibited rPNGase Sp. Cu^{2+} is known to react with deprotonated cysteine thiolates, and deprotonated histidine residues, both of which are part of the catalytic triad in PNGase Sp [92]. Although Cys-SH and His-imidazole groups have pK_a values of ~ 8.5 and ~ 6.0 respectively as isolated amino acids, the actual pK_a value can vary from 5 to 10 and 4 to 10 respectively depending on their local chemical environments. At pH 6.0, the Cys and His residues are still likely to react with Cu^{2+}. Considering that the activity was completely lost in the presence of Cu^{2+} ions, it is likely that Cu^{2+} binds to one of the catalytic residues, either Cys-163 or His-190, or both.

As expected, Ni^{2+} ions also inhibited the enzyme (67% activity left). Ni^{2+} ions preferentially react with cysteine and histidine residues. However, because the activity was not completely destroyed, it is likely that the metal ion is binding to residue involved in maintaining conformation rather than to those of the catalytic triad.

Cd^{2+}, which has even higher affinity than Cu^{2+} for cysteine (stability/formation constant of Cd^{2+}-Cys and Cu^{2+}-Cys complexes are 43,500 L/mol and 13,900 L/mol respectively [93]), was also found to strongly inhibit rPNGase Sp (10% activity remained). Again, the observed effect was probably due to the Cd^{2+} ion binding to the catalytic Cys-163, although the effect was not as pronounced as for Cu^{2+} ions, it may have preferentially bound to the other thiol groups in the protein, having an effect on conformation. It is also possible that the Cd^{2+} may have replaced the tetrathiolate-coordinated Zn^{2+} in the zinc-binding domain, inducing a local (or global) conformational change in the zinc-binding domain. This domain forms one wall of the substrate binding cleft, so its construction is important for maintenance of activity. Zn^{2+} and Cd^{2+} have similar coordination properties, preferring a tetrahedral coordination geometries, and commonly ligate to four cysteine residues [94]. Isostructural or non-isostructural replacement of Zn^{2+} with Cd^{2+} has been shown to occur in other enzymes. The replacement sometimes has an effect on activity [95-97]. In the crystal structures of ScPNGase and mPNGase [27, 29], the zinc-chelating site is located near the outer surface of the zinc-binding domain (Figure 1.4), and appears to be readily accessible to solvents and exogenous metal ions.

Zn^{2+} has been shown to be essential for the functional integrity of the cytosolic PNGases [27, 29]. In this work, the addition of even low concentrations of zinc ion was found to strongly inhibit rPNGase Sp (Data Sheet #6, Appendix VI). A number of both the zinc-dependent and zinc-independent enzymes are known be inhibited by the addition of Zn^{2+} ions and the mechanisms of inhibition are various. For example, carboxypeptidase A (regarded as a prototype of a zinc-protease family) contains one catalytic Zn^{2+} that is strongly coordinated (by His-69, His-196, Glu-72 and one H_2O) in the active site. The addition of millimolar concentrations or less of Zn^{2+} inhibits the enzyme because a second Zn^{2+} binds to the carboxylate oxygen of Glu-270, which is close to the active site, interfering with the proper positioning of the substrate [98]. A number of other zinc-proteases such as bacterial thermolysin [99], neutral metallo-endopeptidase [100] and human neutrophil collagenase [101] are also known to be inhibited by excess Zn^{2+}. Normally, the apoprotein forms of these enzymes (obtained by EDTA treatment) are readily reactivated by addition of the stoichiometric amounts of Zn^{2+} but strongly inhibited by an excess of Zn^{2+}. Likewise, a number of non-metalloenzymes are also inhibited by Zn^{2+}, and appear to be much more sensitive to this cation (IC_{50}: from less than 10 to 200 nM) than metalloenzymes that require Zn^{2+} for activity. These are epoxide hydrolase (zinc inhibits epoxide hydrolase by binding to a cysteine thiol group on the surface of the enzyme) [102], glyceraldehyde and glycerol phosphate dehydrogenases [103], caspase-3 and a tyrosine phosphatase *etc.* [104]. It appears that Zn^{2+} inhibits these enzymes in a noncompetitive way. In most cases, the inhibition of both the zinc-enzymes and non-metalloenzymes by Zn^{2+} can be reversed by thionein which preferential complexes Zn^{2+} from binding sites that are not involved in catalysis or in the maintenance of structure probably because they are more stable and/or are not so accessible to solvent [105, 106]. In the case of rPNGase Sp, whether Zn^{2+} inhibited the enzyme through binding to certain active site residue(s) or to a residue site that plays an essential structural role is not clear.

In summary, the results are basically in agreement with those from earlier studies on the native mouse PNGase [33, 47]. In those studies, 5 mM of Zn^{2+} and Cu^{2+} were found to decrease the mPNGase activity to 7.2% and 4.6% of the control respectively. Moreover, Co^{2+} was also shown to be inhibitory. In contrast to the observation in this study that Fe^{3+} had little effect on rPNGase Sp (~98% activity remained), 5 mM of Fe^{3+} caused a significant loss of activity (reduced to 14%) of mPNGase [47]. This difference might be because mPNGase carries a potential binding site for Fe^{3+} ions that is absent in PNGase

Sp. It is possible that this binding site may reside in the extended C-terminal domain, absent in lower eukaryotic PNGases but which participates in the formation of the full substrate binding site in mPNGase [90]. Also, effects observed on the addition of Mn^{2+} to rPNGase Sp do not agree with what has been found for other eukaryotic PNGases. For example, 108% activity in this study (rPNGase Sp + 5 mM Mn^{2+}); 90% activity in [47] (mPNGase + 5 mM Mn^{2+}); and 33% activity in [33] (mPNGase + 2 mM Mn^{2+}).

3.12.2 Active Enzyme Could not be Reconstituted from EDTA-treated rPNGae Sp

To confirm that rPNGase Sp is a zinc-binding protein and the zinc ion is essential for its deglycosylating activity, the enzyme was extensively dialysed against the assay buffer containing 20 mM of EDTA, followed by dialysis against the same buffer without EDTA. The so-treated enzyme, designated as apo-rPNGase Sp, was tested for its activity. Varying amount of Zn^{2+} was then added into the enzyme solution and incubated for one hour prior to the assays. The results are presented in Figure 3.33 and experimental data listed in Data Sheet #6 (Appendix VI).

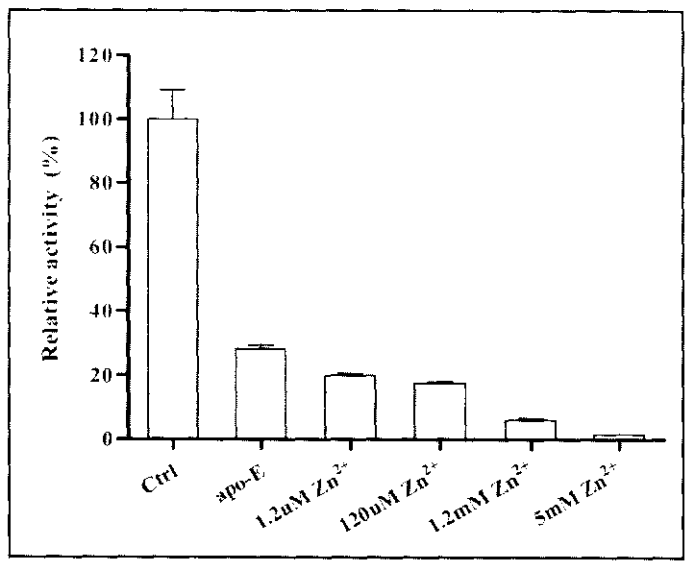


Figure 3.33 The effects of EDTA and Zn^{2+} on rPNGase Sp activity
All the assays were carried out under the standard assay conditions described in section 2.9.4. Assay buffer: 50 mM MIB, 1 mM TCEP, pH 6.0; substrate: FITC-Ova.

The results showed that the treatment with chelating agent EDTA caused a remarkable decrease in the PNGase activity (apo-E: 28% activity left), but that addition of Zn^{2+} with concentrations ranging from 1.2 μM to 5 mM could not restore the enzymic activity. In addition, the inhibition by exogenous Zn^{2+} ions became more pronounced as the Zn^{2+} concentration increased. The results suggested at least two things: (1) The presence of

Zn^{2+} in the zinc-binding domain is essential for the activity; (2) Addition of excess Zn^{2+} failed to restore activity, and in fact inhibited the enzyme further.

Obviously, zinc plays an important role in the maintenance of the tertiary structure of this enzyme. EDTA is a very potent zinc-chelating agent and can effectively remove zinc from zinc-binding proteins including those in which Zn^{2+} is most tightly coordinated such as transcriptional factors (ligands: Cys_2-His_2 or Cys_4; e.g. K_d of Zn^{2+}-Cys_2His_2 complex is ~ 10^{-10}M, while that of Zn^{2+}-EDTA complex is ~ 10^{-16}M), and irreversibly denature these zinc-proteins [107].

It appeared that the deactivation or denaturation of rPNGase Sp by EDTA was irreversible because addition of Zn^{2+} could not reactivate or induce the refolding of rPNGase Sp. This observation should however be examined carefully because, as already mentioned, a number of proteases are extremely sensitive to Zn^{2+} inhibition (IC_{50} from <10 to ~ 200 nM). The total concentration of rPNGase Sp presented in assay was 1.3 μM, which was approximately stoichiometrically equivalent to the lowest concentration (1.2 μM) of Zn^{2+} tested, but whether Zn^{2+} at this concentrate is already in excess, or whether the Zn^{2+} preferentially binds to other sites on the protein, is not known. It cannot therefore be concluded for certain that the deactivation of rPNGase Sp by EDTA is irreversible. In fact, some zinc-finger proteins including Sp1 and certain nuclear hormone receptors can be refolded following the addition of micromolar concentrations of Zn^{2+} [107].

An alternative way to study the correlation between the catalytic activity and the zinc content of rPNGase Sp might be heterologous expression of the protein in the presence of EDTA as suggested by Saur and Thauer [108]. Methyltransferase is a metalloenzyme that contains a single Zn^{2+} probably coordinated by a sulfur ligand and three oxygen or nitrogen ligands. The apoenzyme obtained by incubating the holoenzyme with EDTA was irreversibly inactivated, while the native apoenzyme expressed in *E. coli* in the presence of 2 mM EDTA could be reactivated via incubation with Zn^{2+} or Co^{2+}. In the case of rPNGase Sp, if Zn^{2+}-binding plays a crucial role in the folding process and is responsible for generating the correct tertiary structure, such an experiment would fail to produce active enzyme.

These results produce a clue as to why the chimera with thioredoxin produced soluble

rPNGase Sp, when other expression vectors failed: The binding of a structural zinc to the tetrathiolate coordination site in the enzyme is likely to be required for the initiation of the proper folding. In common with many metal ions, Zn^{2+} is sequestered by proteins such as metallothionein and is thus not freely available in the cell (cellular $[Zn^{2+}_{free}]$ is in the picomolar range). Zinc-proteins must acquire their Zn^{2+} from metallothionein [103], a process that may not be completed quickly. Thioredoxin therefore probably helps to keep the four cysteine residues in two CXXC motifs reduced until the zinc ion is properly inserted, avoiding the quick formation of misfolded or aggregated rPNGase Sp.

Estimation of the secondary structure of the native, EDTA-treated and Zn^{2+}-resupplied rPNGase Sp was performed by circular dichroism (CD) spectroscopy. All measurements were made at room temperature using a Jasco 715 Circular Dichroism Spectropolarimeter with a 1 mm path length quartz cell. A spectrum of the buffer (5 mM phosphate, pH 6.0, 1mM TCEP) was subtracted from each sample spectrum. The data was analysed using the CDPro software (http://lamar.colostate.edu/~sreeram/CDPro/main.html [114, 115]) and results shown in Figure 3.34. Intriguingly, treatment with EDTA did not appear to affect the content of secondary structures in rPNGase Sp despite the decrease in activity (Figure 3.33). Both the untreated and EDTA-treated rPNGase Sp had approximately 77.8% helix, 1.04% sheet and 19.3% random coil, while the addition of 5 mM Zn^{2+} to the treated enzyme resulted in a decrease in the content of helix from 77.8% to 58.2% and increased the content of sheet and random coil to 12.9% and 27.3% respectively, indicating that the binding of Zn^{2+} induced a significant conformational change in the rPNGase molecule, which is consistent with the inhibitory effect of Zn^{2+} on the enzyme (Figures 3.32 and 3.33).

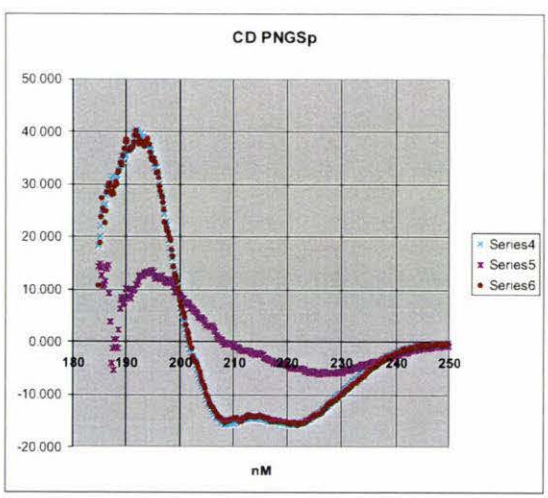


Figure 3.34 Far UV CD spectra of untreated and EDTA-treated rPNGase Sp, together with rPNGase Sp that had been treated with EDTA then exposed to 5 mM Zn^{2+} Series 4: untreated rPNGase Sp; Series 5: EDTA-treated rPNGase Sp; Series 6: Zn^{2+}-exposed rPNGase Sp sample. The secondary structure composition of the untreated and EDTA-treated rPNGase Sp are basically the same: 77.8% helix, 1.04% sheet and 19.3% random coil. After the addition of 5 mM Zn^{2+}, the content of helix decreased to 58.2%, while that of sheet and random coil increased to 12.9% and 27.3% respectively.

3.12.3 Mn^{2+} Also Appears to Inhibit rPNGase Sp

As mentioned in section 3.12.1, the concentrated rPNGase Sp samples (> 10 mg/mL) from some preparations showed pink colour, indicating that certain metal ions such as Mn^{2+} or Co^{2+} might compete with Zn^{2+} for the tetrathiolate-coordination site in rPNGase Sp. In addition, Mn^{2+} appeared to activate the native rPNGase Sp (Figure 3.32). To test whether Mn^{2+} has an effect on the apo-rPNGase Sp, varying concentrations of Mn^{2+} were added into the EDTA-treated and subsequently Zn^{2+}-resupplemented rPNGase Sp samples, and the activity assayed. The results are shown in Figure 3.34 and data in Data Sheet #7 and #8 (Appendix VI).

It can be seen from Figure 3.35 (a) that the addition of Mn^{2+} had little effect on the activity of EDTA-treated rPNGase Sp sample (apo-E), although there was a trend towards increased inhibition with increasing concentrations of Mn^{2+}. When 1.2 μM Zn^{2+} plus varying concentrations of Mn^{2+} were added into the EDTA-treated rPNGase Sp sample and assayed, the results were consistent with those shown in Figure 3.33 and 3.35 (a) (Figure 3.35 b). That is, Zn^{2+} started to inhibit rPNGase Sp at micromolar concentrations, and as the concentration of Mn^{2+} was increased, the trend was to slightly decrease the activity of the enzyme.

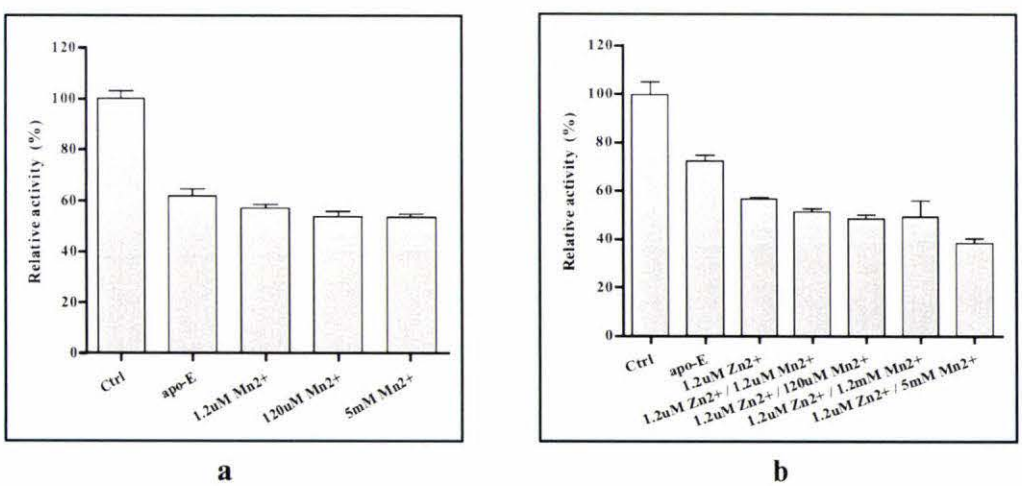


Figure 3.35 The effects of Mn^{2+} on the activity of EDTA-treated rPNGase Sp **(a)** Varying concentrations (from 1.2 μM to 5 mM) of Mn^{2+} were added into EDTA-treated rPNGase Sp and the activities were assayed. **(b)** 1.2 μM Zn^{2+} plus varying concentrations of Mn^{2+} were added into the EDTA-treated rPNGase Sp sample and assayed

It can be seen from Figures 3.33 and 3.35 that, compared to the controls, the activities of apo-enzymes dropped to different extent (to ~ 28% and ~ 60% respectively). This was due to the intrinsic Zn^{2+} ions were removed with different efficiency in the two experiments.

3.13 Crystallisation Trials

3.13.1 Initial Crystallisation Screens

Initial crystallisation screens were performed at room temperature using the sitting drop method with commercially available "sparse matrix" crystallisation screens. These were Structure Screens 1™ and 2™, PACT *premier*™, and Hampton Crystal Screens 1™& 2™, each of which provides 96 crystallisation conditions. Out of 288 conditions (some of them were same from different screen kits) tested, only one condition from Hampton Crystal Screen 1™ resulted in the production of small square crystals of rPNGase Sp. The condition was 4.0 M sodium formate, and the protein concentration used was 5.5 mg/mL. The crystals (shown in Figure 3.35) did not seem to be salt crystals, but did not diffract. Some other conditions that gave micro-crystals of rPNGase Sp also appeared to be optimisable, but fine screening based on those conditions failed to produce larger crystals.

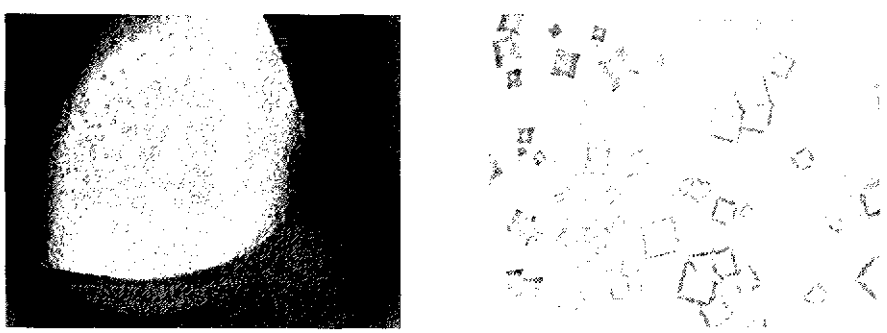


Figure 3.36 rPNGase Sp crystals obtained from initial crystallisation screens at room temperature **(left)**, and a close up view **(right)**.

3.13.2 Optimisation Screens

Based on the condition that gave small square rPNGase Sp crystals, two fine screens using hanging drop method were carried out at room temperature. In one of these, the protein concentration (2, 5, 10 and 15 mg/mL) and sodium formate concentration (2, 3, 4 and 5 M) were varied. In the second screen, the pH (4.6, 5.5, 6.5 and 7.6) and sodium formate concentrations (2, 3, 4 and 5 M) were varied at two different protein concentrations (5 and 10 mg/mL). A 96-well Additive Screen™ was also carried out to try to improve the quality and the size of rPNGase Sp crystals. Unfortunately, none of the conditions tested resulted in rPNGase Sp crystals, although most of them produced either micro-crystals or precipitates.

As mentioned in section 2.10.1 (Material and Methods), many variables may affect protein crystal growth, including the purity of protein which is crucial. The difficulty in obtaining rPNGase Sp crystals was most likely due to the conformational heterogeneity of rPNGase Sp as discussed in section 3.5.2. This could have been due to the formation of non-native disulfide bond(s) and/or caused by the partial loss of the intrinsic structural-Zn^{2+} during the purification process. In order to produce rPNGase Sp that is suitable for crystallisation, three possible changes to the protocols used in this study could be make: (1) Producing the protein in *E. coli* BL-21 (DE3)™ cells, which have a more reducing cytoplasmic environment than the Origami B (DE3)™ cells; (2) Using higher concentrations (5mM or even up to 10 mM) of the reducing agent TCEP in all downstream processes; (3) Avoiding using EDTA, a highly potent Zn^{2+}-chelating agent [107], in any buffers when purify rPNGase Sp for crystallisation trials.

The following crystallisation conditions are also worth trying: (1) Growing the crystal at a lower or higher temperature such as 4°C or 30°C. Because a number of conditions tried in both the initial and fine screens produced micro-crystals, changing the temperature may help to produce single, large rPNGase Sp crystals. (2) Trying detergent screens. Because the crystal structure of the ScPNGase-Rad23 complex showed that the ScPNGase interacted with Rad23 through its Rad23-binding domain which contains a high content of exposed hydrophobic residues [27]. This might also be why the author (and other researchers) could not produce the reasonable amount of soluble rPNGse Sc in the absence of its natural binding partner yRad23. Although rPNGase Sp appeared to fold into its native conformation and showed relatively high solubility in this study, the protein might still be unstable in the absence of its native binding partner, if any. Including certain detergents in the crystallisation solutions may therefore help to stabilise rPNGase Sp molecules and prevent them from non-specifically aggregating via hydrophobic interactions.

Chapter 4 — Summary

The aims of this study were to heterologously produce soluble and active recombinant peptide:*N*-glycanase from yeast *S. pombe* (rPNGase Sp), to biochemically characterise the enzyme, and to crystallise the enzyme for x-ray diffraction analyses. All but the last target were failed.

By using an N-terminal thioredoxin (Trx) fusion, the full-length rPNGase Sp was successfully overproduced in an *E. coli* strain Origami B™ cells in a soluble and active form. Thioredoxin fusion appeared to enhance the solubility of rPNGase Sp by keeping the nine cysteine residues, especially those participating in the formation of the zinc-ribbon-like domain of the cytosolic PNGases reduced. As zinc-binding may play a important role in stabilising or even generating the native conformation of these cytosolic PNGases [27, 29], it is essential the cysteine residues are maintained in a reduced state in the cell. The recombinant PNGase Sp produced from this system remained soluble and active in the presence of reducing agents (DTT or TCEP) even after the thioredoxin-His_6 fusion partner had been removed by specific proteolytic cleavage.

All the eukaryotic PNGases possess a transglutaminase-like catalytic triad (Cys-His-Asp) in their highly conserved core domain and these residues have been shown to be essential for the deglycosylation activity of the cytosolic PNGases. In this study, it was shown that despite this homology with transglutaminase, rPNGase Sp did not possess tranglutaminase activity. Thus, although the cytosolic PNGases and transglutaminases use the same catalytic machinery to catalyse what is essentially the same reaction, the differences in both their amino acid composition and the conformation around their active sites have resulted in two enzymes that are specific for totally different substrates, and carry out different chemical reactions – albeit that one is almost the reverse of the other: breaking of an amide bond (PNGase) or formation of an amide bond (TGase).

rPNGase Sp showed apparent heterogeneity on SDS-PAGE that was characterised by the appearance of two bands with an apparent ~ 2.3 kDa difference in their molecular weights. Mass spectrometry combined with other biochemical analyses revealed that the two bands seen on SDS-PAGE resulted from the anomalous electrophoretic migration of rPNGase Sp,

and probably represented two conformationally different populations of the enzyme molecules able to bind different amounts of SDS.

This study also investigated the substrate specificity of rPNGase Sp. The results showed that, like other cytosolic PNGases [17, 49, 50], rPNGase Sp only exhibits the deglycosylating activity towards the denatured glycoproteins but not their native counterparts. Moreover, the high activity rPNGase showed towards the denatured ribonuclease B among other substrates tested indicates the enzyme prefer to deglycosylate the glycoproteins with high mannose-type glycan chains than those bear other types. This is consistent with the recognised biological function of eukaryotic PNGases: Participating in the endoplasmic reticulum-associated degradation of misfolded/unfolded glycoproteins (ERAD pathway) [16, 38, 43], because the misfolded/unfolded glycoproteins retrotranslocated from ER uniformly carry a 'core glycan' containing nine mannose resides. The results also suggested that rPNGase is able to deglycosylate glycoproteins/peptides that carry complex- and hybrid-type glycan chains such as denatured fetuin, which has not been previously reported.

Kinetics study using both an 11-mer glycopeptide (Ova) derived from hen egg ovalbumin and its FITC-dilabeled derivative (FITC-Ova) showed that, compared to rPNGase F from bacteria which has a K_m of ~ 2.1 μM (pH 8.0) for Ova [113] and ~ 37 μM (pH 8.5) for FITC-Ova [112], rPNGase Sp has much lower affinity to both the substrates. That is a K_m of ~ 386 μM for Ova (pH 6.0) and ~ 186 μM for FITC-Ova (pH 6.0). These results are reasonable because cytosolic PNGases are known to be more stringent than PNGase F in terms of their substrate specificity, and the substrate used for the HPLC assays in this study bore a complex-type glycan chain, which is not a favoured substrate for cytosolic PNGases.

The purpose of labeling the substrate with FITC was to increase the detection sensitivity in the HPLC-based assay. This was, however, unexpectedly resulted in a higher affinity of the substrate to rPNGase Sp compared to unlabeled one, and also caused substrate inhibition at high [S]. This might suggest a non-specific and non-productive binding between the enzyme and the FITC-Ova, a process that was possibly promoted by the electrostatic attraction between the charged FITC group and certain amino acid residues surrounding the active site.

rPNGase Sp from yeast *S. pombe* was strongly inhibited by millimolar concentration of copper (II) and zinc (II) ions which had also been found for mouse PNGase [33, 47]. Manganese (II) appeared to have slightly inhibitory effect on rPNGase Sp, which agreed well with the result from Suzuki *et al.* [47], who also found that iron (III) ions strongly inhibited the mPNGase. rPNGase Sp was basically unaffected by iron (II) at the concentration used (5 mM) for mPNGase, suggesting a potential iron (III) binding site in mPNGases is absent from yPNGase. Cadmium (II) and nickel (II) ions were also found to severely inhibit PNGase Sp in this study. A number of possible interactions were proposed, all of which could result in loss of activity from the enzyme.

Dialysis of rPNGase Sp against the metal ion chelating agent EDTA resulted in a remarkable decrease in the enzymic activity, and the subsequent addition of zinc (II) ions could not restore the activity, strongly suggesting that the zinc ion is required for the activity of rPNGase Sp, as well as playing a role in maintaining the conformation of rPNGase Sp. The evidence that the presence of thioredoxin fusion significantly increased the solubility of rPNGase Sp also supports that zinc binding is crucial for maintaining the conformational integrity of rPNGase Sp, because the premise of the zinc binding is that those cysteines coordinating the zinc ion should stay in reduced state. Since the activation of non-EDTA-treated enzyme by addition of TCEP was a reversible process, the function of reducing agents presented in the buffers appears to keep the catalytic nucleophile Cys-163 reduced.

Although rPNGase Sp needs zinc ions for its structural integrity, it is inhibited by exogenous zinc ions at micromolar concentrations. The free zinc ions might bind to either the catalytic residues such as Cys-163, His-190 and Asp-207 and inhibit the enzyme in a competitive way or an inhibitory site on the protein surface, thus function in a noncompetitive way, or both. An inhibitory kinetic study should help to elucidate the mechanism of zinc inhibition.

Although small crystals of rPNGase Sp were obtained using 4.0 M sodium formate aqueous as the mother liquid and a protein concentration of ~ 5.5 mg/mL, they did not diffract. Further screening around this condition failed to produce diffracting crystals.

Appendix I

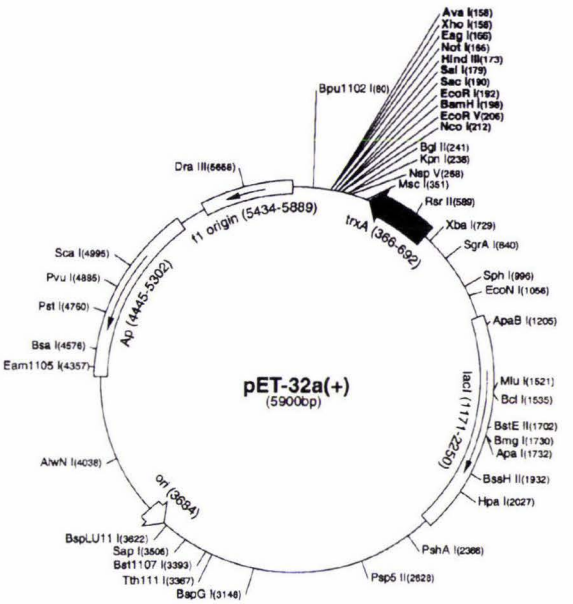
PNG1Sp Sequence

(NCBI database accession No. AL031852)

```
atggattttcatgcgatttctcaacgtttcattgatatgatgagaagcaagaattcccag  60
 M  D  F  H  A  I  S  Q  R  F  I  D  M  M  R  S  K  N  S  Q

aacgcttctcagcctccagagacatatccctttttatcatgaagtacgtcaaatgtcacaa  120
 N  A  S  Q  P  P  E  T  Y  P  F  Y  H  E  V  R  Q  M  S  Q

cacccgtggatgtatgaagacccagagttgcaggattatgctcttagcattttaccactt  180
 H  P  W  M  Y  E  D  P  E  L  Q  D  Y  A  L  S  I  L  P  L

gacaaactatttcaagatgcttctgaattagaaaaagagggagatggatcttggggttac  240
 D  K  L  F  Q  D  A  S  E  L  E  K  E  G  D  G  S  W  G  Y

caagattatgtgattcaagccttgttaaagtggttcaagcgagaattctttgtttgggtt  300
 Q  D  Y  V  I  Q  A  L  L  K  W  F  K  R  E  F  F  V  W  V

aatcaaccaccttgcgaaaaatgtggaggtgaaactcatatgacaggcaacggtcccccc  360
 N  Q  P  P  C  E  K  C  G  G  E  T  H  M  T  G  N  G  P  P

aatgaggaagaaaaatggaatggagtccgcaacgtagagctttatcaatgcaatgtatgc  420
 N  E  E  E  K  W  N  G  V  R  N  V  E  L  Y  Q  C  N  V  C

gggcataatcagagatttcccaggtataatcgcattcgagcattgcttgattcaaggaag  480
 G  H  N  Q  R  F  P  R  Y  N  R  I  R  A  L  L  D  S  R  K

ggaagatgtggagagtgggccaactgtttcactttcctatgcagagcacttggatctaga  540
 G  R  C  G  E  W  A  N  C  F  T  F  L  C  R  A  L  G  S  R

gctaggtggatttggaatgctgaggaccatgtgtggacggaagtctatagtaataaacag  600
 A  R  W  I  W  N  A  E  D  H  V  W  T  E  V  Y  S  N  K  Q

caacgctgggtgcatgtcgatagtggtgaagaatcattcgatgaacctttgatatacgaa  660
 Q  R  W  V  H  V  D  S  G  E  E  S  F  D  E  P  L  I  Y  E

caaggttggggaaagaagatgtcatattgtttgggttttgggatcgatagtgttcgcgat  720
 Q  G  W  G  K  K  M  S  Y  C  L  G  F  G  I  D  S  V  R  D

gtatcgcatagatatattcgtcaccctgagaatggtcttcctcgtgatcgatgccccgaa  780
 V  S  H  R  Y  I  R  H  P  E  N  G  L  P  R  D  R  C  P  E

tctgtcctacaacaggctttgcatgagattaatattgagtttcggtcccgactaactgat  840
 S  V  L  Q  Q  A  L  H  E  I  N  I  E  F  R  S  R  L  T  D

tctgaacgtaaggctctagaggaggaagacaaacgtgaaaaagatgaacttgatggttat  900
 S  E  R  K  A  L  E  E  E  D  K  R  E  K  D  E  L  D  G  Y

atgcgtcctgtttcccaggccacgcccactaacactgacctgccagcaagacaaacgggg  960
 M  R  P  V  S  Q  A  T  P  T  N  T  D  L  P  A  R  Q  T  G

aatgttgaatggaaagagaaaagaggggaagcaggaaaatga
 N  V  E  W  K  E  K  R  G  E  A  G  K  -
```

Appendix II

pET32a_HTBH Vector Information



(a) Structure of pET-32a(+) vector

pET32a sequence [T7 promoter, lac operator, rbs, NdeI, trx-Tag (109 amino acids)..]

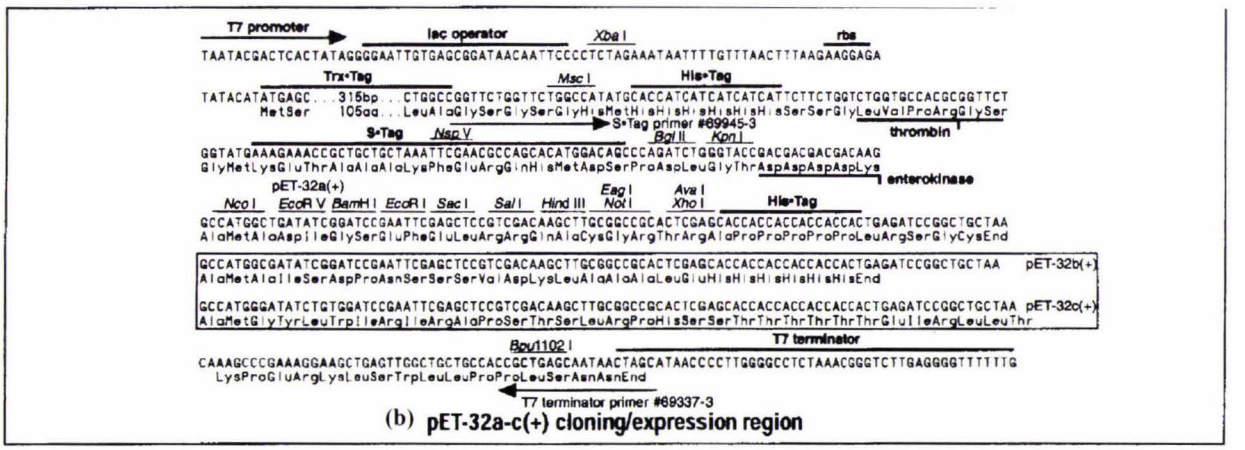
MscI *NdeI* His Tag
CTG GCC GGT TCT GGT TCT GGC CAT ATG CAT CAC CAT CAC CAT CAC GAT TAC
L A G S G S G H M H H H H H H D Y

Spacer region rTev cleavage site *BamHI* *EcoRI*
GAT ATC CCA ACG ACC GAA AAC CTG TAT TTT CAG GGC GGA TCC GAA TTC
D I P T T E N L Y F Q G G S E F

EagI *AvaI*
SacI *SalI* *HindIII* *NotI* *XhoI* His Tag
GAG CTC CGT CGA CAA GCT TGC GGC CGC ACT CGA GCA CCA CCA CCA CCA
E L R R Q A C G R T R A P P P P

CCA CTG AGA TCC GGC TGC TAA.....
P L R S G C End

(c) pET32a_HTBH cloning/expression region



(b) pET-32a-c(+) cloning/expression region

Figure 1 a) Whole plasmid map of the pET-32a-c(+) vector from Novagen™ b) Cloning/expression region of the pET-32a-c(+) vector c) Modification of the pET-32a-c(+) vector to produce the pET32_HTBH vector The pET32a_HTBH vector (created by Dr. Rose Brown) used in this study is identical to the pET32a-c(+) vector except that a *Msc*I-*Bam*HI fragment (His-tag, thrombin cleavage site, S-tag, enterokinase cleavage site **(a)**) was excised and replaced with a PCR amplified section from pProex-HTb (His-tag, spacer region, rTev protease cleavage site **(c)**. Primers included sites for *Msc*I (5') and *Bam*HI (3'). The restriction sites flanking the introduced region are shown in bold. In this study, the *PNG1Sp* gene was cloned into *Bam*HI-*Hind*III restriction sites of pET32a_HTBH vector.

Appendix III

Protocol for Preparation of Substrate Ovalbumin Glycopeptide (Norris *et al.*, 1994a)

Cyanogen Bromide Digestion

1) 12 g of hen egg ovalbumin (Sigma grade 2) is dissolved in 120 mL 50% formic acid solution. 2.7 g cyanogen bromide (CNBr) dissolved in acetonitrile is added to the ovalbumin solution and left overnight at room temperature with stirring. The air in the reaction container was removed by argon in order to avoid oxidation of CNBr (CNBr is toxic. All manipulation should be carried out in sealed vessels or in a fume hood).

2) Add 300-400 mL of water and reduce to the original volume (~120 mL) by rotary evaporation. Repeat once to remove remaining CNBr.

Removal of Insoluble Peptides

3) Add TCA aqueous solution into the CNBr free digest solution to a final concentration of 5% to precipitate unreacted proteins and large insoluble peptides. Centrifuge at 10,000*g* for 15 minutes.

4) Extract the supernatant with the same volume of diethyl ether for three times to remove any excess TCA, then, remove ether and concentrate the sample to a small volume (~10 mL) by rotary evaporation.

5) Add acetic acid to a final concentration of 0.5% to precipitate peptides, remove them by centrifugation at 14,000*g* for 10 minutes.

Separation of Glycopeptide by Size Exclusion Chromatography (SEC)

6) Apply 2.5 mL of the supernatant to a 100×2.5cm column packed with gel filtration resin (Bio-Rad®, P-4, 50-100 mesh) in 0.1 M acetic acid. Wash with 0.1 M acetic acid at a flow rate of 2 mL/min, and detect at 280 nm. Fraction of 8 mL is collected.

7) Assay fractions for reducing sugar by phenol/sulphuric acid test (add 12.5 μL 80% phenol and 1.25 mL of concentrated H_2SO_4 into 0.5 mL of sample. An orange to brown colour should develop within 15 minutes). The fractions containing glycopeptides were then pooled, lyophilised (In this study, the recovery rate of 11-mer glycopeptide is ~ 1-2% after SEC purification). After dissolving in water, the glycopeptides were further purified by RP-HPLC (250×10mm, Φ 5 μm; C18, Jupiter Series; Phenomenex, UK). A 15-minute

gradient elution from 20% acetonitrile / 0.1% trifluoroacetic acid (TFA) to 40% acetonitrile / 0.08% TFA was applied; flow rate of 4 mL/min; detection at 214 nm. Eluted fractions were collected, lyophilised and analysed using PNGase activity assay (section 2.6). The fraction containing glycopeptides was stored at -20℃ until use (the purification product includes both the homoserine 'open' form and its isomer, lactone form, of 11-mer glycopeptide). Mass spectrometry analyses of the HPLC-purified product showed it has an average mass of ~2.59 kDa as shown in Figure 2.

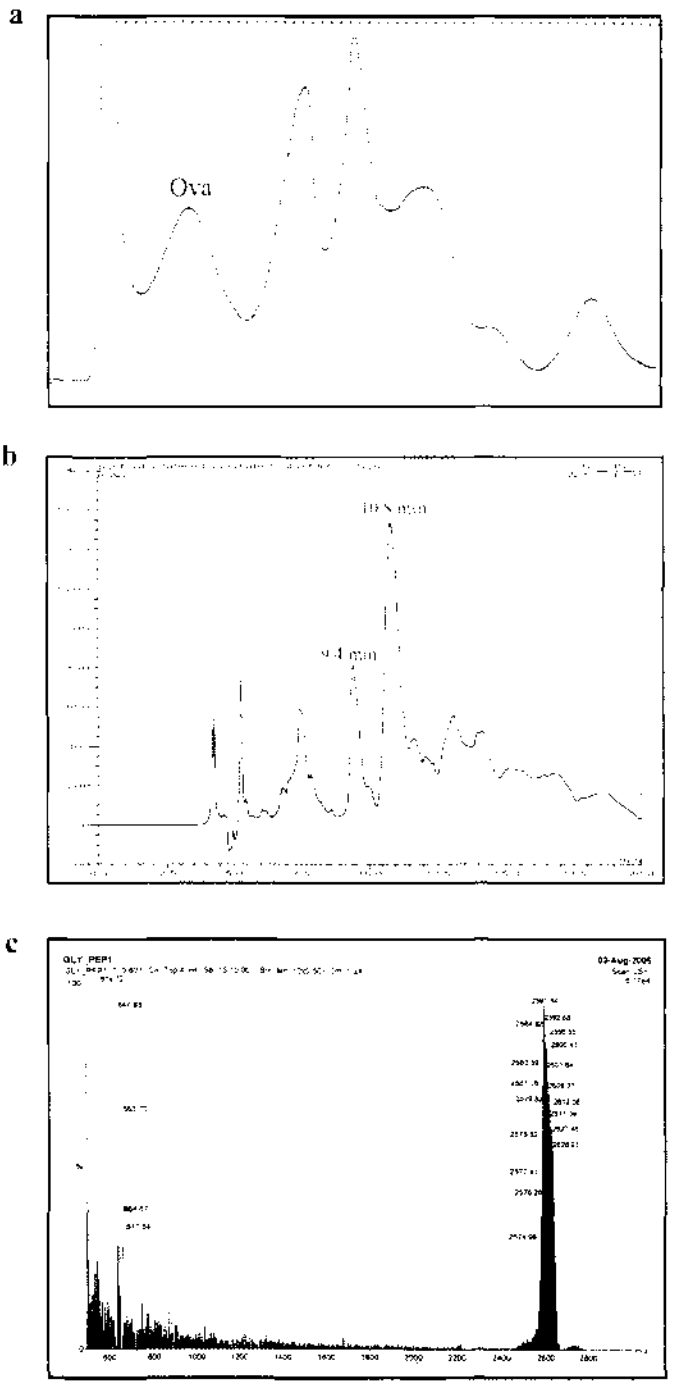


Figure 2. SEC purification (a), HPLC purification (b) and ESI mass spectrum (c) of the 11-mer glycopeptide substrate, Ova. In SEC purification, the retention volume of Ova was ~ 250 mL. In HPLC purification, the retention time of the lactone form and homoserine 'open' form of Ova were 10.8 and 9.4 minutes respectively. ESI mass spectrometry analyses indicated that the Ova substrate has an average molecular weight of ~ 2.59 kDa.

Appendix IV

In-gel Tryptic Digest for Protein ID by Mass Spectrometry

This protocol is based on Shevchenko, A., Wilm, M., Vorm, O., and Mann, M., Mass Spectrometric sequencing of proteins from silver-stained polyacrylamide gels. Anal. Chem. 1996, 68:850-8, and modified by David Miyamoto.

1) Excise band from Coomassie or silver stained gel. Cut gel band into 1 mm cubes using clean razor bland on a clean glass surface. Transfer to an Eppendorf tube.
2) Remove excess water with pipet. Add 25-35 μL acetonitrile to tube to cover gel pieces. Incubate at room temperature for 10 minutes to dehydrate and shrink gel pieces.
3) Remove acetonitrile with pipet. Speed-vacuum to dryness for 10 munutes.
4) Swell gel particles in 150 μL 10 mM DTT in 100 mM NH_4HCO_3. incubate for 1 hour at 56℃.
5) Cool to room temperature. Replace DTT solution with 150 μL 55 mM iodoacetamide in 100 mM NH_4HCO_3. Incubate for 45 minutes at room temperature in the dark with occasional vortexing.
6) Remove solution and wash gel pieces with 150 μL 100 mM NH_4HCO_3. Incubate for 10 minutes at room temperature.
7) Remove NH_4HCO_3 solution with pipet. Add 150 μL acetonitrile to dehydrate gel pieces. Incubate for 10 minutes at room temperature.
8) Repeat wash steps 6) through 7). Remove acetonitrile and speed-vacuum to dryness for 10 minutes.
9) Place tubes on ice and swell gel particles in 25-35 μL digestion buffer (12.5 ng/μL trypsin in 50 mM NH_4HCO_3). Incubate 45 minutes on ice. To make the digestion buffer, dissolve 20 μg trypsin (Promega sequence-grade modified porcine trypsin, Cat. #V511A) in 80 μL Promega trypsin buffer solution (50 mM acetic acid), and dilute with 50 mM NH_4HCO_3 to 12.5 ng/μL. Remove trypsin-containing buffer. Add 5-10 μL 50 mM NH_4HCO_3 without trypsin to keep pieces wet during cleavage. Incubate overnight at 37℃.
10) Centriguge for 1 minute at 14,000 rpm to spin down gel pieces. Save supernatant in a separate Eppendord tube.
11) Add 20 μL 20 mM NH_4HCO_3 to cover gel pieces. Incubate 10 minutes at room temperature. Transfer supernatant to the Eppendorf tube from step 10.
12) Add 25μL 5% formic acid and 50% acetonitrile to the gel pieces. Incubate for 20 minutes at room temperature.
13) Centrifuge for 1 minute at 14,000 rpm. Remove formic acid / acetonitrile solution and save in the same Eppendorf tube from step 10.
14) Repeat formic acid extraction (steps 12 through 13) twice more.
15) Dry the Eppendorf tube in speed-vacuum to complete dryness. Store at –20℃ until analyses.

Appendix V

FITC-labeling of Ovalbumin Glycopeptide

The protocol for FITC-labeling of ovalbumin glycopeptide was adapted from [112]

Procedure:

About 50 mg of 11-mer glycopeptide (Ova) purified from HPLC step (Appendix III) was dissolved in 10 mL 0.1 M Na_2HPO_3/NaH_2PO_3 buffer (pH 7.0) in a small roundbottom flask wrapped with foil (FITC is light-sensitive). To this solution, 2 mL 0.5% FITC in acetone (w/v) was added dropwise with stirring. The reaction was left overnight at room temperature and purified on a C18 HPLC column (250×10mm, Φ 5 μm; Jupiter Series; Phenomenex) with the following gradient: (solvent A: 0.1% TFA in water; solvent B: 0.08% TFA in acetonitrile) 1) isocratic flow at 80% A, 20% B for 2 min.; 2) gradient to 60% A, 40% B over 5 min.; 3) gradient to 30% A, 70% B over 10 min.; 4) gradient to 100% B over 5 min.; 5) isocratic flow at 100% B for 5 min.; 6) gradient to 80% A, 20% B over 10 min. Flow rate of 4 mL/min; detected using UV (214 nm) or fluorescence (excitation: 495 nm, emission: 520 nm) detector. Fractions containing 11-mer glycopeptide were identified using PNGase activity assay (section 2.6).

Like Ova, the FITC-labeled substrate (FITC-Ova) is equilibrates between two isomers (homoserine 'open' form and homoserine lactone form). Thus, the FITC-labeled substrate was lyophilised and then totally converted (hydrolysed) to the 'open' form by boiling in 0.1 M ammonia bicarbonate (pH 8.5, adjusted using NH_4OH) followed by lyophilisation and a final purification step using reverse phase HPLC described above. The final product was lyophilised to give a yellow powder that could be stored at -20℃.

a

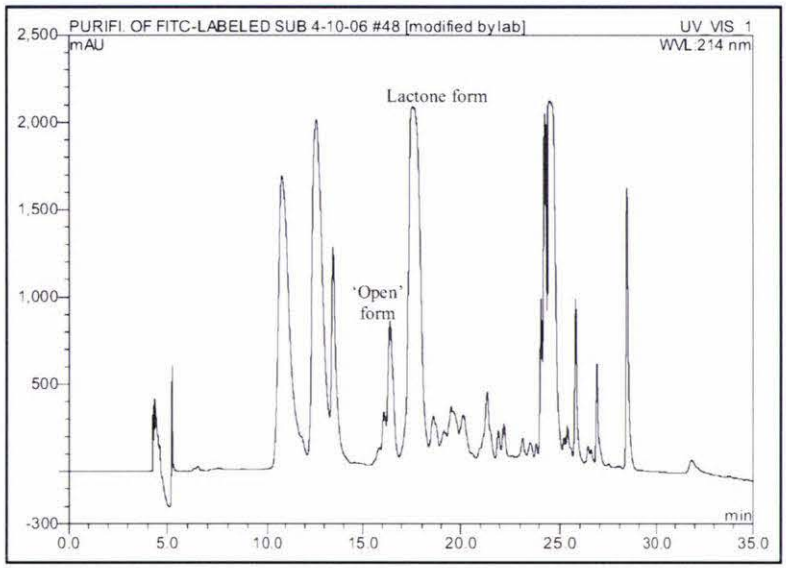


Figure 3. HPLC-chromatogram for the purification of the labelling product (FITC-Ova) using preparative HPLC The retention times of the homorserine 'open' form and lactone form of FITC-Ova were 16.4 and 17.6 minutes respectively.

b

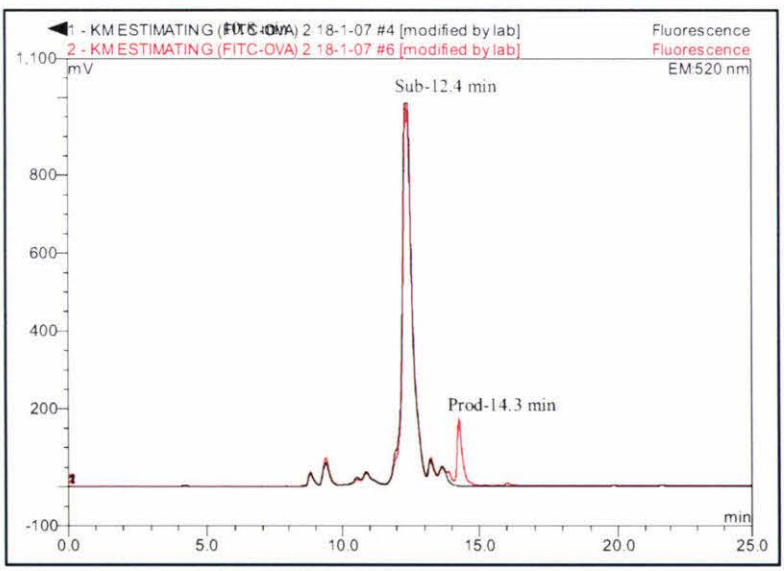


Figure 4. A typical HPLC-chromatogram for the assay using FITC-Ova The retention times of the substrate (hormoserine 'open' form of FITC-Ova) and the product were 12.4 and 14.3 minutes respectively.

Appendix VI

Experimental Data Sheets

Data Sheet #1 — Experimental Data of Optimum pH Determination

pH	v_1 (*)	v_2	v_3	Mean	Std. Deviation
5.0	11.19	12.71	9.92	11.27	1.397
5.5	94.98	95.07	99.73	96.59	2.717
6.0	103.32	98.71	103.3	101.8	2.656
6.5	64.1	64.16	61.2	63.15	1.692
7.0	64.69	64.63	62.58	63.97	1.201
7.5	—	80.55	75.71	78.13	3.422
8.0	43.09	51.57	55.23	49.96	6.227
8.5	26.56	27.74	31.46	28.59	2.557
9.0	14.47	11.28	13.24	13	1.609

Data Sheet #2 — Data used for the determination of Km for FITC-Ova (pH 6.0)

[S] (μM)	v_1 (nmol/min)	v_2	v_3	Mean	Std. Deviation
3.32	55.7	76.8	62.9	65.2	10.7
6.65	146	152	148	149	2.62
13.3	303	300	301	301	1.87
26.6	665	621	597	628	34.7
39.8	886	846	840	857	24.9
53.1	1053	1062	1089	1068	18.7
106	1413	1450	1468	1444	27.5
159	1449	1422	1387	1420	31.6
212	1233	1252	1270	1252	18.0
265	1134	1116	1116	1122	10.4
398	689	666	632	662	28.6

Data Sheet #3 — Data used for the determination of Km for Ova (pH 7.0)

[S] (μM)	v_1 (nmol/min)	v_2	v_3	Mean	Std. Deviation
26.5	401	406	—	403	4.07
53	794	756	747	766	25.2
106	1302	1360	1380	1347	40.8
212	2057	2045	2169	2090	68.3
319	2837	2715	2878	2810	85.0
638	3755	3715	3674	3715	40.8
850	3980	4016	—	3998	25.8
1700	4368	4449	4470	4429	54.0

Data Sheet #4 — Data used for the determination of Km for Ova (pH 6.0)

[S] (μM)	v_1 (nmol/min)	v_2	v_3	Mean	Std. Deviation
26.5	270	276	—	273	4.07
53	531	493	484	503	25.2
106	893	931	927	917	21.2
212	1336	1344	1386	1355	26.6
319	2184	2037	2225	2148	98.8
425	2118	2107	2099	2108	9.19
638	3123	3102	2980	3068	77.3
850	3380	3096	—	3238	201
1700	3041	3061	3000	3034	31.2

Data Sheet #5 — Experimental Data of 'Effects of Metal Ions' analyses

	v_1 (*)	v_2	v_3	Mean	Std. Deviation
Control	16.5	16.4	16.7	16.5	0.146
5mM Cu^{2+}	0.00	0.00	0.00	0.00	0.00
5mM Fe^{3+}	16.1	16.8	15.5	16.2	0.642
5mM Mn^{2+}	18.7	18.1	17.2	18.0	0.718
5mM Ni^{2+}	10.5	10.5	11.9	11.0	0.803
5mM Zn^{2+}	2.27	2.31	2.23	2.27	0.040
5mM Cd^{2+}	1.72	1.67	1.58	1.66	0.0709

Data Sheet #6 — Experimental Data of 'Effects of EDTA and Zn^{2+}' analyses

	v_1 (*)	v_2	v_3	Mean	Std. Deviation
Control (100%)	12.1	10.1	10.9	11.0	1.01
Apo-E (**)	3.14	3.00	3.23	3.12	0.116
Apo-E + 1.2 μM Zn^{2+}	2.15	2.27	2.22	2.21	0.0603
Apo-E + 120 μM Zn^{2+}	1.98	1.97	1.87	1.94	0.0608
Apo-E + 1.2 mM Zn^{2+}	0.734	0.648	0.639	0.674	0.0524
Apo-E + 5 mM Zn^{2+}	0.182	0.197	0.183	0.187	0.00839

Data Sheet #7 — Experimental Data of 'Effects of Mn^{2+}' analyses

	v_1 (*)	v_2	v_3	Mean	Std. Deviation
Ctrl (100%)	13.0	13.0	13.6	13.2	0.374
Apo-E	8.53	8.02	7.83	8.13	0.363
Apo-E + 1.2 μM Mn^{2+}	7.43	7.73	7.37	7.51	0.193
Apo-E + 120 μM Mn^{2+}	6.97	6.88	7.36	7.07	0.253
Apo-E + 5 mM Mn^{2+}	7.05	7.18	6.88	7.04	0.151

Data Sheet #8 — Experimental Data of 'Effects of Mn^{2+}' analyses

	v_1 (*)	v_2	Mean	Std. Deviation
Ctrl (100%)	10.7	9.96	10.4	0.558
Apo-E	7.35	7.67	7.51	0.226
Apo-E + 1.2μM Zn^{2+}	5.86	5.93	5.89	0.0508
Apo-E + 1.2μM Zn^{2+} / 1.2μM Mn^{2+}	5.27	5.43	5.35	0.116
Apo-E + 1.2μM Zn^{2+} / 120μM Mn^{2+}	4.91	5.16	5.04	0.173
Apo-E + 1.2μM Zn^{2+} / 1.2mM Mn^{2+}	5.59	4.62	5.10	0.685
Apo-E + 1.2μM Zn^{2+} / 5mM Mn^{2+}	4.13	3.86	4.00	0.191

* v : Initial velocity (V_0) of the deglycosylation reaction was presented in the peak area of the reaction products calculated from the HPLC chromatogram (Chromeleon™ Client software).

** **Apo-E:** the rPNGase Sp sample treated with 20 mM EDTA (See section 2.9.5)

References

1. Helenius, A., Aebi, M. *Intracellular functions of N-linked glycans.* Science, 2001. **291**: p. 2364-2369.

2. Varki, A., Cummings, R., Esko, J., Freeze, H., Hart, G. and Marth, J. *Essentials of Glycobiology.* Cold Spring Harbor Laboratory Press, 1999. NY.

3. Tarentino, A.L., Plummer, T.H., Jr. *Deglycosylation of asparagine-linked glycans by peptide:N-glycosidase F.* Trends Glycosci. Glycotechnol., 1993. **5**: p. 163-170.

4. Suzuki, T., Park, H., Kitajima, K., Lennarz, W.J. *Peptides glycosylated in the endoplasmic reticulum of yeast are subsequently deglycosylated by a soluble peptide: N-glycanase activity.* J. Biol. Chem., 1998. **273**: p. 21526-30.

5. Plummer, T.H., Jr., Elder, J.H., Alexander, S., Phelan, A.W., Tarentino, A.L. *Demonstration of peptide:N-glycosidase F activity in endo-beta-N-acetylglucosaminidase F preparations.* J. Biol. Chem., 1984. **259**: p. 10700-10704.

6. Norris, G.E., Flaus, A.J., Moore, C.H., Baker, E.N. *Purification and crystallization of the endoglycosidase PNGase F, a peptide:N-glycosidase from Flavobacterrium meningosepticum.* J. Mol. Bio., 1994a. **241**: p. 624-6.

7. Norris, G.E., Stillman, T.J., Anderson, B.F. Baker, E.N. *The three dimensional structure of PNGase F, a glycosyl asparaginase from Flavobacterium meningosepticum.* Structure, 1994b. **2**: p. 1049-1059.

8. Kuhn, P., Guan, C., Cui, T. Tarentino, A.L., Plummer, T.H., Jr., Roey, P.V. *Active site and oligosaccharide recognition residues of peptide-N4- (N-acetyl-ß-D-glucosaminyl) asparagine amidase F.* J. Biol. Chem., 1995. **270**: p. 29493-29497.

9. Takahashi, N., *Demonstration of a new amidase acting on glycopeptides.* Biochem. Biophys. Res., 1977. **76**: p. 1194-1201.

10. Ftouhi-Paquin, N., Hauer, C.R., Stack, R.F., Tarentino, A.L., Plummer T.H., Jr. *Molecular Cloning, Primary Structure, and Properties of a New Glycoamidase from the Fungus Aspergillus tubigensis.* Biological Chemistry J., 1997. **272**: p. 22960-22965.

11. Altmann, F., Paschinger, K., Dalik, T., Vorauer, K. *Characterisation of ß-D-glucosaminyl) asparagine amidase A and its N-glycans.* Eur J Biochem., 1998. **252**: p. 118-23.

12. Baker, H.M., Day, C.L., Norris, G.E., Baker, E.N. *Enzymatic deglycosylation as a tool for crystallization of mammalian binding proteins.* Acta Crystallogr D Biol Crystallogr, 1994. **50**: p. 380-4.

13. Suzuki, T., Park, H., Hollingsworth, N.M., Sternglanze, R. Lennarz, W.J. *PNG1, a yeast gene encoding a highly conserved peptide:N-glycanase.* Cell Biology J., 2000. **149**: p. 1039-1051.

14. Makarova, K.S., Aravind, L., Koonin, E.V., *A superfamily of archaeal, bacterial, and eukaryotic proteins homologous to animal transglutaminases.* Protein Science, 1999. **8**: p. 1714-1719.

15. Katiyar, S., Suzuki, T., Balgobin, B.J., Lennarz, W.J. *Site-directed Mutagenesis Study of Yeast Peptide: N-Glycanase.* J. Biol. Chem., 2002. **277**: p. 12953-12959.

16. Suzuki, T., Park, H. Lennarz, W.J. *Cytoplasmic peptide: N-glycanase (PNGase) in eukaryotic cells: occurrence, primary structure, and potential functions.* FASEB J., 2002. **16**: p. 635-641.

17. Hirsch, C., Blom, D., Ploegh, H.L. *A role for N-glycanase in the cytosolic turnover of glycoproteins.* The EMBO J., 2003. **22**: p. 1036-1046.

18. Kopito, R.R., *ER quality control: the cytoplasmic connection.* Cell, 1997. **88**: p. 427-30.

19. Wiertz, E.J., Jones, T.R., Sun, L., Bogyo, M., Geuze, H.J., Ploegh, H. L. *The human cytomegalovirus US11 gene product dislocates MHC class I heavy chains from the endoplasmic reticulum to the cytosol.* Cell, 1996. **84**: p. 769-79.

20. Blom, D., Hirsch, C., Stern, P., Tortorella, D., Ploegh, H.L. *A glycosylated type I membrane protein becomes cytosolic when peptide: N-glycanase is compromised.* EMBO J., 2004. **23**: p. 650-8.

21. Suzuki, T., Kitajima, K., Emori, Y., Inoue, Y., Inoue, S. *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.* Proc Natl Acad Sci U S A., 1997. **94**: p. 6244-9.

22. Seko, A., Kitajima, K., Inoue, Y., Inoue, S. *Peptide:N-glycosidase activity found in the early embryos of Oryzias latipes (Medaka fish).* Biological Chemistry J., 1991. **266**: p. 22110-22114.

23. Suzuki, T., Seko, A., Kitajima, K., Inoue, Y., Inoue, S. *Identification of peptide:N-glycanase activity in mammalian-derived cultured cells.* Biochem Biophys Res Commun., 1993. **194**: p. 1124-1130.

24. Suzuki, T., Park, H., Till, E.A., Lennarz, W.J. *The PUB domain: a putative protein-protein interaction domain implicated in the ubiquitin-proteasome pathway.* Biochem Biophys Res Commun., 2001a. **287**: p. 1083-7.

25. Lorand, L., Conrad, S. M. *Transglutaminases.* Mol Cell Biochem., 1984. **58**: p. 9-35.

26. Griffin, M., Casadio, R., Bergamini, C.M. *Transglutaminases: nature's biological glues.* Biochem. J., 2002. **368(Pt 2)**: p. 377-96.

27. Lee, J.H., Choi, J.M., Lee, C., Yi, K.J., and Cho, Y. *Structure of a peptide:N-glycanase-Rad23 complex: Insight into the deglycosylation for denatured glycoproteins.* PNAS. USA, 2005. **102**: p. 9144-9149.

28. Kitajima, K., Suzuki, T., Kouchi, Z., Inoue, S., Inoue, Y. *Identification and distribution of peptide:N-glycanase (PNGase) in mouse organs.* Arch Biochem Biophys., 1995. **319**: p. 393-401.

29. Zhao, G., Zhou, X., Wang, L., Li, G., Kisker, C., Lennarz, W.J., and Schindelin, H. *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., 2006. **281**: p. 13751-13761.

30. Misaghi, S., Pacold, M.E., Blom, D., Ploegh, H.L., and Korbel, G.A. *Using a small molecule inhibitor of peptide:N-glycanase to probe its role in glycoprotein turnover.* J. Chem. Biol., 2004. **11**: p. 1677-1687.

31. Misaghi, S., Korbel, G.A., Kessler, B., Spooner, E., Ploegh, H.L. *z-VAD-fmk inhibits peptide:N-glycanase and may result in ER stress.* Cell Death and Differentiation, 2006. **13**: p. 163-165.

32. Suzuki, T., Hara, I., Nakano, M., Zhao, G., Lennarz, W.J., Schindelin, H., Taniguchi, N., Totani, K., Matsuo, I., Ito, Y. *Site-specific labeling of cytoplasmic peptide:N-glycanase by N,N'-diacetylchitobiose-related compounds.* J. Biol. Chem., 2006. **281**: p. 22152-22160.

33. Weng, S., Spiro, R.G. *Demonstration of a peptide:N-glycosidase in the endoplasmic reticulum of rat liver.* Biochem. J., 1997. **322**: p. 655-61.

34. Enenkel, C., Lehmann, A., Kloetzel, P.M. *Subcellular distribution of proteasomes implicates a major location of protein degradation in the nuclear envelope-ER network in yeast.* EMBO J., 1998. **17**: p. 6144-54.

35. Katiyar, S., Li, G., Lennarz, W.J. *A complex between peptide:N-glycanase and two proteasome-linked proteins suggests a mechanism for the degradation of misfolded glycoproteins.* Proc. Natl. Acad. Sci. USA, 2004. **101**: p. 13774-9.

36. Suzuki T, Park, H., Kwofie, M.A. Lennarz, W.J., *Rad23 Provides a Link between the Png1 Deglycosylating Enzyme and the 26 S Proteasome in Yeast.* J. Biol. Chem., 2001b. **276**: p. 21601-7.

37. Biswas, S., Katiyar, S., Li, G., Zhou, X., Lennarz, W.J., Schindelin, H. *The N-terminus of yeast peptide:N-glycanase interacts with the DNA repair protein Rad23.* Biochem Biophys Res Commun., 2004. **323**: p. 149-155.

38. Kim, I., Ahn, J., Liu, C., Tanabe, K., Apodaca, J., Suzuki, T., Rao, H. *The Png1-Rad23 complex regulates glycoprotein turnover.* J. Cell Biol., 2006. **172**: p. 211–219.

39. Bertolaet, B.L., Clarke, D.J., Wolff, M., Watson, M.H., Henze, M., Divita, G., Reed, S.I. *UBA domains of DNA damage-inducible proteins interact with ubiquitin.* Nature Structural Biology, 2001. **8**: p. 417 - 422.

40. Caroline, R.M., Wilkinson, Seeger, M., Hartmann-Petersen, R., Stone, M., Wallace, M., Semple, C., Gordon, C. *Proteins containing the UBA domain are able to bind to multi-ubiquitin chains.* Nature Cell Biology, 2001. **3**: p. 939 - 943.

41. Rao, H., Sastry, A. *Recognition of Specific Ubiquitin Conjugates Is Important for the Proteolytic Functions of the Ubiquitin-associated Domain Proteins Dsk2 and Rad23.* J. Biol. Chem., 2002. **277**(14): p. 11691-11695.

42. Park, H., Suzuki, T., Lennarz, W.J. *Identification of proteins that interact with mammalian peptide:N-glycanase and implicate this hydrolase in the proteasome-dependent pathway for protein degradation.* Proc. Natl. Acad. Sci. USA., 2001. **98**: p. 11163-8.

43. Li, G., Zhou, X., Zhao, G., Schindelin, H., Lennarz, W.J. *Multiple modes of interaction of the deglycosylation enzyme, mouse peptide N-glycanase, with the proteasome.* Proc Natl Acad Sci U S A., 2005. **102**(44): p. 15809-14.

44. Richly, H., Rape, M., Braun, S., Rumpf, S., Hoege, C. Jentsch, S. *A series of ubiquitin binding factors connects CDC48/p97 to substrate multiubiquitylation and proteasomal targeting.* Cell, 2005. **120**: p. 73-84.

45. McNeill, H., Knebel, A., Arthur, J.S., Cuenda, A., Cohen, P. *A novel UBA and UBX domain protein that binds polyubiquitin and VCP and is a substrate for SAPKs.* Biochem. J., 2004. **384**: p. 391-400.

46. Allen, M.D., Buchberger, A. Bycroft, M. *The PUB domain functions as a p97 binding module in human peptide N-glycanase.* J. Biol. Chem., 2006. **281**: p. 25502-25508.

47. Suzuki, T., Seko, A., Kitajima, K., Inoue, Y., Inoue, S. *Purification and enzymatic properties of peptide:N-glycanase from C3H mouse-derived L-929 fibroblast cells.* J. Biol. Chem., 1994. **269**: p. 17611-8.

48. Suzuki, T., Kitajima, K., Inoue, Y., Inoue, S. *Carbohydrate-binding property of peptide: N-glycanase from mouse fibroblast L-929 cells as evaluated by inhibition and binding experiments using various oligosaccharides.* J. Biol. Chem., 1995. **270**: p. 15181-6.

49. Hirsch, C., Misaghi, S., Blom, D., Pacold, M.E., Ploegh, H.L. *Yeast N-glycanase distinguishes between native and non-native glycoproteins.* EMBO J., 2004. **5**: p. 201-6.

50. Joshi, S., Katiyar, S., Lennarz, W.J. *Misfolding of glycoproteins is a prerequisite for peptide:N-glycanase mediated deglycosylation.* FEBS Letters, 2005. **579**: p. 823-826.

51. Huppa, F.B., Ploegh, H.L. *The chain of the T cell antigen receptor is degraded in the cytosol.* Immunity, 1997. **7**: p. 113-122.

52. Della Mea, M., Caparros-Ruiz, D., Claparols, I., Serafini-Fracassini, D., Rigau, J. *AtPng1p. The first plant transglutaminase.* Plant Physiol, 2004. **135**: p. 2046-2054.

53. Laemmli, U.K. *Cleavage of structural proteins during the assembly of the head of bacteriophage T4.* Nature, 1970. **227**: p. 680-685.

54. Bradford, M. *A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding.* Anal. Biochem., 1976. **72**: p. 248-254.

55. Sharon, N.L. *Glycoproteins. The Proteins.* Edited by H Neurath, RL Hill. New York, Academic, 1982: p. 1-144.

56. Waelsch, H., Mycek, M.J. *Transglutaminase.* Methods in Enzymology, 1962. **V**: p. 833-838.

57. Griffin, P.R. *Direct database searching with MALDI-PSD spectra of peptides.* Rapid Commun Mass Spectrom, 1995(1546-51).

58. Hillenkamp, F., Karas, M., Beavis, R.C., and Chait, B.T. *Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers.* Anal. Chem., 1991. **63**: p. 1193A-1203A.

59. Shevchenko, A., Wilm, M., Vorm, O., Mann, M. *Mass spectrometric sequencing of proteins from silver stained polyacrylamide gels.* Anal. Chem., 1996. **68**: p. 850-858.

60. Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F., Whitehouse, C.M. *Electrospray ionization for mass spectrometry of large biomolecules.* Science, 1989. **246**: p. 64-71.

61. Han, J. *Tris[2-carboxyethyl]phosphine - A reducing agent with versatile applications including cleavage of disulfide bonds and quantitation of numerous oxidants.* Previews, 1999. **2(4)**: p. 16-21.

62. Palmer, T. *Enzymes: Biochemistry, Biotechnology and Clinic Biochemistry.* Ellis Horwood Ltd. London, New York, Toronto, Sydney, Tokyo, Singapore, Madrid, Mexiko City, Munich, 2001. **fifth edition**.

63. Drenth, J. *Principles of Protein X-Ray Crystallography.* Springer-Verlag Inc. NY, 1999: p. ISBN 0-387-98587-5.

64. Martin, I.M.B. *Scientific Data Visualization and Digital Image Processing for Structural Biology.* Purdue University, West Lafayette, 1996.

65. Jancarik, J., Kim, S.H. *Sparse matrix sampling: a screening method for crystallization of proteins.* J. Appl. Cryst., 1991. **24**: p. 409-411.

66. Stewart, E.J., Aslund, F., Beckwith, J. *Disulfide bond formation in the Escherichia coli cytoplasm: an in vivo role reversal for the thioredoxins.* EMBO J., 1998. **17**(19): p. 5543-50.

67. Fountoulakis, M. *Apparent heterogeneity of recombinant interferon γ receptors produced in prokaryotic and eukaryotic expression systems.* J. Chem. Technol. Biotechnol., 1996. **65**(2): p. 123-30.

68. Baneyx, F. *Recombinant protein expression in Escherichia coli.* Current Opinion in Biotechnology, 1999. **10**(5): p. 411-421.

69. Sorensen, M.A., Kurland, C.G., Pedersen, S. *Codon usage determines translation rate in Escherichia coli.* j. mol. Bio, 1989. **207**(2): p. 365-77.

70. Zhang, S.P., Zubay, G., Goldman, E. *Low-usage codons in Escherichia coli, yeast, fruit fly and primates.* Gene, 1991. **105**(1): p. 61-72.

71. Grossman, T.H., Kawasaki, E.S., Punreddy, S.R., Osburne, M.S. *Spontaneous cAMP-dependent derepression of gene expression in stationary phase plays a role in recombinant expression instability.* Gene, 1998. **209**(1-2): p. 95-103.

72. Manuvakhova, M., Keeling, K., Bedwell, D.M. *Aminoglycoside antibiotics mediate context-dependent suppression of termination codons in a mammalian translation system.* RNA, 2000. **6**(7): p. 1044-1055.

73. Namy, O., Rousset, J.P., Napthine, S., Brierley, I. *Reprogrammed Genetic Decoding in Cellular Gene Expression.* Mol Cell., 2004. **13**(2): p. 157-168.

74. McCaughan, K.K., Brown, C.M., Dalphin, M.E., Berry, M.J., Tate, W.P. *Translational Termination Efficiency in Mammals is Influenced by the Base Following the Stop Codon.* Proc. Natl. Acad. Sci. U. S. A. , 1995. **92**: p. 5431-5435.

75. Buckingham, R.H. *Codon context and protein synthesis: Enhancements of the genetic code.* Biochimie, 1994. **76**(5): p. 351-4.

76. Namy, O., Hatin, I. Rousset, J.P. *Impact of the six nucleotides downstream of the stop codon on translation termination.* EMBO Rep., 2001. **2**: p. 787-793.

77. Kang, S.H., Jung, S.T., Kang, T.J., Kim, R.G., Suh, S.H., Woo, J.H., Lee, E.Y., Choi, C.Y. *An efficient translational termination of human erythropoietin in Escherichia coli by altering the base following the stop codon.* Biotechnology Techniques, 1999. **13**(11): p. 761-764.

78. Arkov, A.L., Korolev, S.V. Kisselev, L.L. *5' Contexts of Escherichia coli and human termination codons are similar.* Nucleic Acids Res., 1995. **23**(22): p. 4712-4716.

79. Tork, S., Hatin, I., Rousset, J. P., Fabret, C. *The major 5' determinant in stop codon read-through involves two adjacent adenines.* Nucleic Acids Res., 2004. **32**(2): p. 415-21.

80. Melchior, F. *SUMO-nonclassical ubiquitin.* Annu. Rev. Cell. Dev. Biol. , 2000. **16**: p. 591-626.

81. Malakhov, M.P., Mattern, M.R., Malakhova, O.A., Drinker, M., Weeks, S. D., Butt, T. R., *SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins.* J Struct Funct Genomics, 2004. **5**(1-2): p. 75-86.

82. Jeffrey, G. *Comparison of SUMO fusion technology with traditional gene fusion systems: Enhanced expression and solubility with SUMO.* Protein Science, 2006. **15**: p. 182-189.

83. Kontos, H., Napthine, S., Brierley, I. *Ribosomal pausing at a frameshifter RNA pseudoknot is sensitive to reading phase but shows little correlation with frameshift efficiency.* Mol. Cell. Biol., 2001. **21**(24): p. 8657-8670.

84. Jenkins, N. *Role of physiology in the determination of protein heterogeneity.* Current Opinion in Biotechnology, 1996. **7**: p. 205-209

85. Yee, V.C., Pedersen, L.C., Le Trong, I., Bishop, P.D., Stenkamp, R.E. and Teller, D.C. *Three-dimensional structure of a transglutaminase : human blood coagulation factor XIII.* . Proc. Natl. Acad. Sci. U.S.A. , 1994 **91**: p. 7296-7300.

86. Pedersen, L.C., Yee, V.C., Bishop, P.D., Le Trong, I., Teller, D.C. Stenkamp, R.E. *Transglutaminase factor XIII uses proteinase-like catalytic triad to crosslink macromolecules.* Protein Sci. , 1994. **3**(1131-135).

87. Berry, M.B., Phillips, G.N., Jr. *Crystal structures of Bacillus stearothermophilus adenylate kinase with bound Ap5A, Mg2+ Ap5A, and Mn2+ Ap5A reveal an intermediate lid position and six coordinate octahedral geometry for bound Mg2+ and Mn2+.* Proteins, 1998. **32**: p. 276-288.

88. Marangoni, A.G. *Enzyme Kinetics, A modern Approch.* Wiley-Interscience, 2003.

89. Suzuki, T., Kitajima, K., Inoue, S., Inoue, Y. *Does an animal peptide:N-glycanase have the dual role as an enzyme and a carbohydrate-binding protein? .* Glycoconj J., 1994. **11**(5): p. 469-76.

90. Zhou, X., Zhao, G., Truglio, J.J., Wang, L., Li, G., Lennarz, W.J., Schindelin, H. *Structural and biochemical studies of the C-terminal domain of mouse peptide-N-glycanase identify it as a mannose-binding module.* PNAS, 2006. **103**(46): p. 17214-9.

91. Sawyer, W.H., Klonis, N. *Spectral properties of the prototropic forms of fluorescein in aqueous solution.* Journal of Fluorescence, 1996. **6**(3): p. 147-157.

92. Gregory, A.P. *Preparation of isomorphous heavy-atom derivatives.* Methods Enzymol., 1985. **144**(147-156).

93. Disbudak, A., Bektas, S., Patir, S., Genc, O., Denizli, A. *Cysteine-metal affinity chromatography: determination of heavy metal adsorption properties.* Separation and Purification Technology, 2002. **26**(2): p. 273-281(9).

94. Rulísek L, V.J. *Coordination geometries of selected transition metal ions (Co2+, Ni2+, Cu2+, Zn2+, Cd2+ and Hg2+) in metalloproteins.* J Inorg Biochem. , 1998. **71**((3-4)): p. 115-27.

95. Holland, D.R., Hausrath, A.C., Juers, D., Matthews, B.W. *Structural analysis of zinc substitutions in the active site of thermolysin.* Protein Science, 1995. **4**: p. 1955-1965.

96. Munoz, A., Laib, F., Petering, D.H., Shaw III, C.F. *Characterization of the cadmium complex of peptide 49-61: a putative nucleation center for cadmium-induced folding in rabbit liver metallothionein IIA.* JBIC, 1999. **4** p. 495-507.

97. Messerle, B.A., Schaffer, A., Vasak, M., Kagi, J.H.R., Wuthrich, K. *Comparison of the Solution Conformations of Human [Zn7]-Metallothionein-2 and [Cd7]-Metallothionein-2 Using Nuclear-Magnetic-Resonance Spectroscopy.* J. Mol. Biol., 1992. **225** p. 433-443.

98. Gomez Ortiz, M., Gomis-Ruth, F.X., Huber, R., Aviles, F.X. *Inhibition of carboxypeptidase A by excess zinc: analysis of the structural determinants by X-ray crystallography.* FEBS letters 1997. **400**(3): p. 336-340

99. Holland, D.R., Hausrath, A.C., Juers, D. and Matthews, B.W. *Structural analysis of zinc substitutions in the active site of thermolysin.* Protein Science, 1995. **4**(10): p. 1955-1965.

100. Kerr, M.A., Kenny, A.J. *The molecular weight and properties of a neutral metallo-endopeptidase from rabbit kidney brush border.* Biochem J. , 1974. **137**(3): p. 489-495.

101. Mallya, S.K., Wart, H.E.V. *Mechanism of Inhibition of Human Neutrophil Collagenase by Gold (I) Chrysotherapeutic Compounds.* J. Biol. Chem., 1989. **264**(3): p. 1594-1601.

102. Draper, A.J., Hammock, B.D. *Inhibition of soluble and microsomal epoxide hydrolase by zinc and other metals.* Toxicological Sciences, 1999 **52**: p. 26-32.

103. Maret, W., Yetman, C.A., Jiang, L. *Enzyme regulation by reversible zinc inhibition: glycerol phosphate dehydrogenase as an example* Chem Biol Interact., 2001. **130-132**(1-3): p. 891-901.

104. Maret, W., Jacob, C., Vallee, B.L., Fischer, E.H. *Inhibitory sites in enzymes: Zinc removal and reactivation by thionein.* Proc Natl Acad Sci U S A, 1999. **2**(96(5)): p. 1936-40.

105. Maret, W., Yetman, C.A., Jiang, L.J. *Enzyme regulation by reversible zinc inhibition: glycerol phosphate dehydrogenase as an example* Chem Biol Interact., 2001. **130-132**(1-3): p. 891-901.

106. Maret, W., Jacob, C., Vallee, B.L., Fischer E.H. *Inhibitory sites in enzymes: Zinc removal and reactivation by thionein.* Proc Natl Acad Sci U S A, 1999. **2**(96(5)): p. 1936-40.

107. Nyborg, J.K., Peersen, O.B. *That zincing feeling: the effects of EDTA on the behaviour of zinc-binding transcriptional regulators.* Biochem. J., 2004. **381**: p. e3-e4.

108. Sauer, K., Thauer, R.K. *Methyl-coenzyme M formation in methanogenic archaea Involvement of zinc in coenzyme M activation.* Eur. J. Biochem., 2000. **267**: p. 2498-2504.

109. Green, E.D., Adelt, G., Baenziger, J.U., Wilson, S., Halbeek, H.V. *The asparagine-linked oligosaccharides on bovine fetuin. Structural analysis of N-glycanase-released oligosaccharides by 500-megahertz 1H NMR spectroscopy.* J. Biol. Chem., 1988. **263**(34): 18253-18268.

110. Williams, R.L., Greene, S.M., McPherson, A. *The crystal structure of ribonuclease B at 2.5-A resolution.* J. Biol. Chem., 1987. **262**(33): 16020-16031.

111. Lenz, D.H. *N-linked clycopeptide mimetics as tools in kinetic, mechanistic and structural studies of peptide:N-glycanase F.* Ph.D thesis, 2003, Massey University, New Zealand.

112. Loo, T.S. *Expression, purification and characterisation of recombinant peptide:N-glycanase F.* Master thesis, 2000. Massey University, New Zealand.

113. Leffers, K.W., Schell, J., Jansen, K., Lucassen, R., Kaimann, T., Nagel-Steger, L., Tatzelt, J. Riesner, D. *The structureal transition of the prion protein into its pathogenic conformation is induced by unmasking hydrophobic sites.* J. Mol. Biol., 2004. **344**: p839-853.

114. Sreerama, N., Woody, R.W. *A self-consistent method for the analysis of protein secondary structure from circular dichroism*, Anal. Biochem., 1993. **209**(1): p 32-44.

115. Sreerama, N., Woody, R.W. *Estimation of protein secondary structure from CD spectra: Comparison of CONTIN, SELCON and CDSSTR methods with an expanded reference set.* Anal. Biochem., 2000. **287(2)**, p252-260.

ERRATA

1	Page ii	Pattchet should be Patchett
2	Page 7 onwards	yPNG1p is ScPNG1p
3	Page 19	DMPDA stands for 2,2,dimethyl-1,3,propanediamine
4	Page 27	Origami cells are *E. coli* host cells with mutations in both the thioredoxin reductase (*trxB*) and glutathione reductase (*gor*) genes, greatly enhancing disulfide bond formation in the cytoplasm compared to normal BL21s. They are conducive to protein folding in bacterial cytoplasm and are ideal for use with pET-32 vectors, since the thioredoxin fusion tag further enhances the formation of disulfide bonds in the cytoplasm
5	Page 39	a single broad range buffer was used to reduce effects due to buffer species, not buffer ions as stated.
6.	Page 53	In the sentence starting with "another possible reason" the words in brackets need to be deleted.
7.	Page 56	The numbers in the last column reflect the purification for each step relative to the one before. The correct values for each step, starting from step 1 are: 1, 3.5, 5.4, 5.5.
8.	Page 63-64	The sentence running from page 63-64 should read "–in *E. coli* with increased yields and ---"
9.	Page 71	The term GRAVY index stands for Grand Average of Hydropathy. Kyte, J. and Doolittle, R.F. (1982) A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157, 105-132. [PubMed: 7108955] The aliphatic index of a protein is defined as the relative volume occupied by aliphatic side chains (alanine, valine, isoleucine, and leucine). It may be regarded as a positive factor for the increase of thermostability of globular proteins. The aliphatic index of a protein is calculated according to the following formula [9]: Aliphatic index = X(Ala) + a * X(Val) + b * (X(Ile) + X(Leu)) where X(Ala), X(Val), X(Ile), and X(Leu) are mole percent (100 X mole fraction) of alanine, valine, isoleucine, and leucine. The coefficients a and b are the relative volume of valine side chain (a = 2.9) and of Leu/Ile side chains (b = 3.9) to the side chain of alanine.
10.	Page 74	In the legend for Figure 3.18, the reference Pedersen *et al*, 1994 is number 86 in the reference list, and Katiyar *et al* is number 15.
11.	Page 82	The substrate is an ovalbumin 11 mer.
12.	Page 87	The caption of Figure 3.25 refers to Lineweaver-Burke, instead of Burk. This error is also found elsewhere in the text of Chapter 3. The data in the plot were fitted using the substrate inhibition model shown in (b)

13. Page 87. The reference in the caption is to Alejandro G. Marangoni, "Enzyme kinetics: a modern approach". Hoboken, N.J. : Wiley-Interscience, c2003

14. Page 87, 88. Using Enzfitter, and a substrate inhibition model, the values for V_{max}, k_s, and k_{si}, are 17 ± 5.nM/minute, 7.4 ±2.3 mM and 2.0 ± 0.6 pM respectively. These are quite different from the values calculated wth Prism software. As the algorithm used is not identified, it is difficult to compare these two estimations.

15. Page 90. The last sentence of Section 3.11.1 should read "From this analysis, although inhibition of rPNGaseSp by the substrate is happening at all concentrations, it is obvious at substrate concentrations greater than 50 μM."

16. Page 91. Using EnzFitter, and a Michaelis-Menten model, K_M and V_{max} have the values) 0.342 ± 0.002 mM, 0.545 ± 0.002 nM/min repectively.

17. Page 107. The last sentence of the first paragraph should read, "All but the last target were achieved."

18. Page 108. Reference 113 in the 2nd paragraph, should be reference 112, and reference 113 should be Reference 111.