The Importance of *N*-linked Glycosylation on the N-domain of Angiotensin-I Converting Enzyme

Colin Scott Anthony

Thesis presented for the degree of:

Doctor of Philosophy

In the division of Medical Biochemistry
University of Cape Town

February 2011

Supervisor: Prof. E.D. Sturrock

Acknowledgements

Firstly, I would like to thank my supervisor Prof. Ed Sturrock for his guidance throughout the last five years. The many insightful, interesting and often humorous conversations were a great benefit to my development as a scientist. Completing a project like this would not have been possible without such oversight, support and guidance.

Mercie mille fois pour Sylva Schwager. Not only does her presence keep the lab running smoothly, but her guidance with the protein and mass spectrometry sections has been invaluable, and all this in addition to the reams of moral support and friendship she has provided.

Thanks to past and present members of the ACE lab Trudi O'Neill, Jean Watermeyer, Ayesha Parker, Nailah Conrad, Tony Chang, Chris Yates, Kate Larmuth, Kerry Gordon, Riyad Domingo, and Raymond Moholisa for helpful discussions, advice and entertainment. Special mention must be made of Ross Douglas, the dream team, and RosCol Ltd. for many hours of humorous escapism, interesting conversations and helpful comments.

I would like to acknowledge our collaborators Prof. S. Danilov for supplying the ACE anti-bodies and Prof. K.R. Acharya and his group at the University of Bath who worked on the crystal structure. I would also like to thank Dr. Sharon Prince and Prof. Pete Meissner for their support and their assistance with various funding applications over the years.

Many thank to my family and friends for their support and encouragement and to God for adding the mystery element.

Finally I wish to thank the South African National Research Foundation, DAAD, Ernst and Ethel Erikson Trust, Stella and Paul Loewenstein Trust, University of Cape Town for funding, without which I would almost certainly not have lasted this long.

Abbreviations

A-beta42 - amyloid-β protein$_{1-42}$
ACE - angiotensin-I converting enzyme
AcSDKP - *N*-acetyl-Ser-Asp-Lys-Pro
Ang I - angiotensin I
Ang II - angiotensin II
Ang(1-7) - angiotensin(1-7)
Ang1-5 - angiotensin 1-5
Asn - asparagine
Asp - aspartic acid
BK - bradykinin
CD - circular dichroism
CHO-K1 - Chinese hamster ovary K1
CRD - carbohydrate recognition domain
CRT - calreticulin
CNX - calnexin
CVD - cardiovascular disease
DSC - differential scanning calorimetry
DMEM - Dulbecco's modified Eagle's medium
DMSO - dimethyl sulphoxide
E. coli - *Escherichia coli*
EDEM - ER degradation enhancing α-mannosidase-like protein
EDTA - ethylenediaminetetraacetic acid
ER - endoplasmic reticulum
ERAD - ER associated degradation pathway
ERGIC-53 - ER Golgi intermediate compartment
ERp57 - ER protein 57
FCS - foetal calf serum
FSH - follicle-stimulating hormone
Gln - glutamine
GlcNAc - *N*-acetyl glucosamines

GnRH - gonadotropin-releasing hormone

HEPES - N-2-hydroxyethylpiperazine-N'-2-ethanesulphonic Acid

KKS - kallikrein-kinin system

LH - luteinizing hormone

mAbs - monoclonal antibodies

MALDI-ToF/ToF - matrix assisted laser desorption-ionisation time of flight

mg-sACE - minimally glycosylated sACE

NB-DNJ - *N*-butyl deoxynojirimycin

OST - oligosaccharidal transferase complex

PBS - phosphate-buffered saline

PCR - polymerase chain reaction

PDB - Protein Data Bank

PMSF - phenylmethylsulphonyl fluoride

PNGase-F - peptide-*N*-glycosidase-F

Pro - proline

RAS - renin-angiotensin system

RE - restriction endonuclease

RMSD – root-mean square deviation

RT-PCR - reverse transcription PCR

sACE - somatic ACE

SDS-PAGE - sodium dodecyl sulphate polyacrylamide gel electrophoresis

Ser - serine

tACE - testis ACE

TCEP - triscarboxyethyl phosphine

TEDTA - trypsin-EDTA

TFA – trifluoroacetic acid

Thr - threonine

Tris - tris-(hydroxymethyl)-aminomethane

VIP36 - Vesicular integral protein 36

VIPL - VIP36-like protein

WT - wild-type

Z-FHL - Z-Phe-His-Leu

Abstract

Angiotensin-I converting enzyme (ACE) is an important drug target in the treatment of heart disease due to its role in the regulation of blood pressure. ACE contains two domains, the N- and C-domains, both of which are catalytically active and heavily glycosylated. Glycosylation is one of the most important forms of post-translational modification, having a wide range of functions including protein folding, modulation of the immune response, and providing targeting signals. Glycosylation is required for the expression of active ACE and structural studies of ACE have been fraught with severe difficulties because of surface *N*-glycosylation of the protein. This problem has been addressed to a large extent with respect to the C-domain, where the role of glycosylation has been extensively characterised and a minimally glycosylated form was able to crystallise reproducibly. As yet, little is known about the degree and importance of *N*-linked glycosylation on the N-domain. The generation of minimally glycosylated N-domain, however, requires a greater understanding of the relative importance of the individual *N*-linked glycosylation sites.

In order to determine the role of glycosylation and to create suitable forms of the N-domain for crystallization studies, the ten potential *N*-linked glycan sites on the N-domain were investigated. Glycan site occupancy was assessed using enzymatic deglycosylation, limited proteolysis and mass spectrometry. A number of glycosylation mutants were generated *via* site-directed mutagenesis and expressed in CHO cells. The expression of active N-domain protein was analysed by an enzymatic activity assay and by Western blotting. The effect of glycosylation on thermal stability was assessed by determining the level of residual activity after thermal denaturation as well as by circular dichroism spectroscopy. In this way the various active glycosylation mutants were assessed and compared to wild type N-domain. Nine out of the ten potential *N*-glycan sites were found to be glycosylated, while only site 10 was found to be unglycosylated. The presence of three C-terminal glycosylation sites was sufficient for the expression of active N-domain, although an N-terminal site with two of the C-terminal sites was also found to be effective. Decreasing levels of glycosylation correlated with a general decrease in thermal

stability, however, two N-terminal sites, sites 2 and 3, were found to be critically important for thermal stability.

Thus, C-terminal glycosylation was found to play a key role in the expression of active N-domain, while N-terminal glycosylation was required for thermal stability. Additionally, the minimally glycosylated N-domain variant, Ndom389 (containing glycosylation at sites 3, 8 and 9), was highly suitable for crystallization studies, and the structure of this variant was solved to 2.0 Å resolution in the presence of an N-domain selective phosphinic inhibitor, RXP407.

Table of Contents

Chapter 1 : Introduction

1.1 Angiotensin-I converting enzyme (ACE)

Angiotensin-I converting enzyme is a membrane-bound dipeptidyl carboxypeptidase with an extra cellular orientation [1]. ACE is an important biological catalyst with a number of diverse functions and occurs as two distinct isoforms: somatic ACE (sACE) and testis ACE (tACE). Somatic ACE contains two highly homologous domains, termed the N- and C-domains, and is expressed mainly in the vascular endothelial cells of the lungs, heart and kidney and brush border membranes of the kidneys and intestines [2-4]. Testis ACE consists of the C-domain portion of sACE and is only expressed in male germinal cells [1,5,6]. A naturally occurring N-domain isoform has been detected in ileal fluid, and is thought to arise from a proteolytic cleavage event in the region linking the N- and C-domains of sACE [7,8]. The N- and C-domains both contain a functional zinc binding active site and share a high degree of overall sequence identity, approximately 60% [1], yet despite this, subtle variations at key active-site residues mean that the two domains have notable differences with regard to substrate and inhibitor selectivity.

1.1.1 ACE gene structure

The ACE gene is located on chromosome 17q23.3 and contains a total of 26 exons and two promoters (Figure 1.1) [9,10]. Somatic ACE contains 25 of the 26 exons, with exon 13, coding for a heavily *O*-glycosylated region, being spliced out during mRNA processing [10,11]. Testis ACE extends from exons 13 to 26 and is expressed off an alternate promoter located in intron 12, which is only activated in male germinal cells [1,10,12]. Both isoforms contain exon 26 which codes for a membrane anchoring region [10,13].

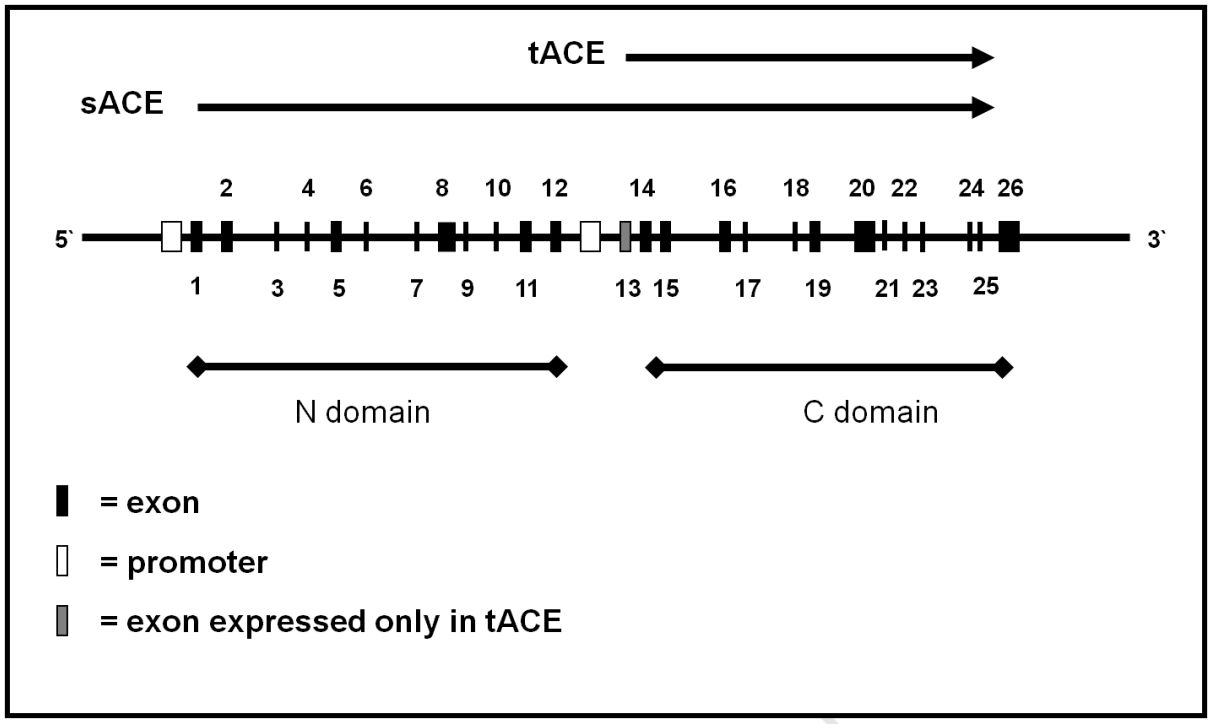


Figure 1.1: Structure of the ACE gene. Relative positions of exons, promoters and domains are indicated (Adapted from Hubert et al. 1991)

1.1.2 Biological roles of ACE

1.1.2.1 ACE and hypertension

Hypertension is one of the main causes of cardiovascular disease (CVD) [14], which is itself one of the main causes of morbidity and mortality worldwide, with up to 30% of all deaths attributable to CVD [15,16]. Hypertension is said to be the most important treatable factor with regard to the prevention of cardiovascular diseases such as myocardial infarction, stroke, heart failure and related renal diseases [15,17,18]. CVD is an important health concern, especially for developing countries where the burden of these diseases is greatest due to the large population sizes and poor access to public health facilities and treatments [15]. Investing resources into treatment and prevention strategies should be a matter of priority for these countries which often overlook the importance of CVD due to the high profile of infectious disease such as HIV-AIDS and TB.

The most celebrated role of ACE is its action in the renin angiotensin (RAS) and kallikrien-kinin (KKS) systems, where it plays an important role in the regulation of blood pressure. ACE cleaves angiotensin I (AngI), generating the vasoconstrictive

peptide angiotensin II (AngII), while bradykinin (BK), a vasodilatory peptide, is inactivated by ACE (Figure 1.2).

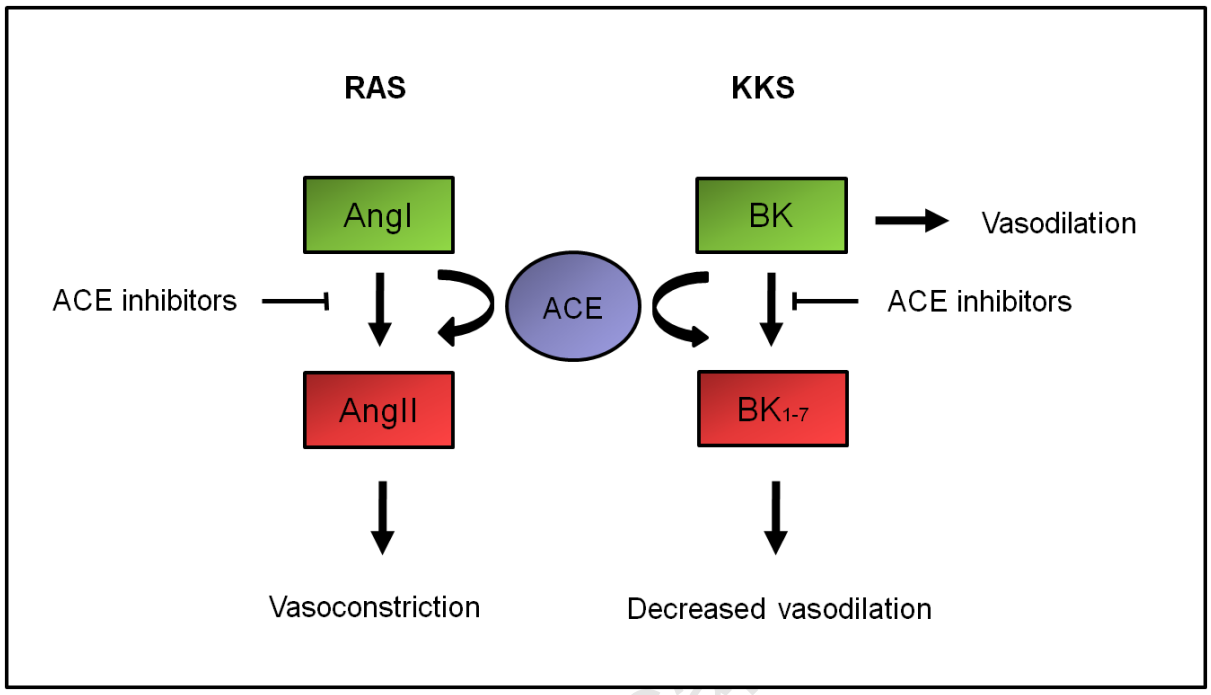


Figure 1.2: Diagram showing the role of ACE in the RAS and KKS. ACE cleaves AngI to the vasoactive AngII eliciting vasoconstriction. ACE cleaves the vasoactive peptide BK to an inactive form (BK_{1-7}), causing decreased vasodilation. The point of action of ACE inhibitors are indicated.

Recent findings indicate that the major site of AngI cleavage is at the C-domain active site, while both domains cleave BK with equal efficiency [19-23]. This finding has given rise to the pursuit of C-domain selective inhibitors, which are likely to provide control of blood pressure, while leaving the N-domain active to prevent BK accumulation, the accumulation of which is associated with some of the side effects of chronic ACE inhibition [1,24-27].

1.1.2.2 The biological role of the N-domain of sACE

Somatic ACE is able to cleave a wide range of substrates, including angiotensin-I, bradykinin, substance-P, gonadotropin-releasing hormone (GnRH), *N*-acetyl-Ser-Asp-Lys-Pro (AcSDKP), angiotensin 1-7 (Ang1-7) and amyloid -β $protein_{1-42}$ (A-beta42). Of these substrates, the N-domain has been show to be the main site of cleavage for Ang1-7, A-beta42, GnRH and AcSDKP [28 29,30].

The AngII-derived peptide, Ang1-7, elicits vasodilatory effects upon binding the Ang1-7 receptor. The N-domain is able to cleave Ang1-7 converting it to the inactive form (Ang1-5). This Ang1-7 degrading activity adds another means of ACE mediated blood pressure regulation through the RAS pathway [29].

The N-domain is known to cleave abeta42, producing the shorter A-beta40 peptide and consequently, ACE inhibition has been shown to increase abeta42 levels [30,31]. A recent review of a number of randomised controlled clinical trials has shown that treatment with ACE inhibitors, as well as diuretics, resulted in improved outcomes for patients suffering from various forms of dementia, including Alzheimer's, although the exact mechanism for this is not clear [32].

GnRH, is a short peptide hormone which instigates the release of follicle-stimulating hormone (FSH) and luteinizing hormone (LH) from the pituitary gland and is important for sexual maturation and reproductive function. Results have shown that LHRH is inactivated by ACE, and that cleavage is preferentially carried out by the N-domain [33-35].

Perhaps the most important N-domain specific substrate, Ac-SDKP, was first identified as a regulatory peptide involved in hematopoietic stem cell proliferation [36], but has also been show to have a number of beneficial effects on the cardiac system. These include preventing or reversing cardiac, vascular and renal inflammation and fibrosis [37]. Ac-SDKP mediates these effects, in part, by inhibiting macrophage differentiation, activation, migration and cytokine release. Ac-SDKP has also been shown to prevent end-organ damage resulting from hypertension [38], as well as protecting against diabetic cardiomyopathy [39]. Additionally, increased levels of AcSDKP, following ACE inhibition, are thought to contribute to the beneficial effects of ACE inhibitors by a novel mechanism whereby AcSDKP inhibits collagen deposition in the left ventricle of the heart following vascular injury [40]. N-domain knock-out mice, and mice undergoing chronic ACE inhibitor treatment, have AcSDKP levels six to seven times that of wild-type (WT) mice [41,42] [22,41], illustrating the importance of the N-domain in the regulation of AcSDKP levels.

Thus, these findings highlight the need for the development of inhibitors selective for the N-domain of ACE which may prove useful in the treatment of certain diseases such as renal and pulmonary fibrosis, where inhibition of the N-domain would be beneficial, while the C-domain would remain active to allow for normal regulation of blood pressure.

1.1.3 ACE inhibitors

ACE inhibitors have long been used in the treatment of hypertension. A number of ACE inhibitors are currently on the market, with the peptide analogue, lisinopril being widely used. ACE inhibitors currently available to clinicians do not significantly discriminate between the two domains and were largely produced before the different roles of the two domains of sACE had been identified. Conventional ACE inhibitors are commonly associated with side effects such as loss of taste, skin rash, persistent coughing and angioedema [1,43-46]. The first two side-effects are mainly attributed to the sulfhydryl inhibitors such as captopril, accounting for the decline in the use of this class of ACE inhibitor [1,25,47]. Persistent cough is the most common side effect, affecting up to 20% of patients, while the occurrence of the potentially fatal angioedema is rare in the general population, affecting up to 0.5% of patients [1,25,47]. These adverse effects have largely been attributed to the accumulation of BK [47]. Thus an exciting shift in the development of ACE inhibitors is the generation of C-domain selective inhibitors, as it has been postulated that these will result in sufficient inhibition of sACE to reduce hypertension, while allowing the N-domain to remain active, thus prevent the excessive accumulation of BK and the associated side effects [1,24-27]. At present there are no commercially available domain selective inhibitors, although certain inhibitors which show high levels of domain selectivity have been identified. These include the phosphinic inhibitors, RXPA380, an inhibitor which is 3000-fold more C-domain selective, and RXP407 which is 1000-fold more selective for the N-domain (Figure 1.3) [1,48]. RXP407, however, is not a good drug candidate due to its large size and poor bioavailability [20].

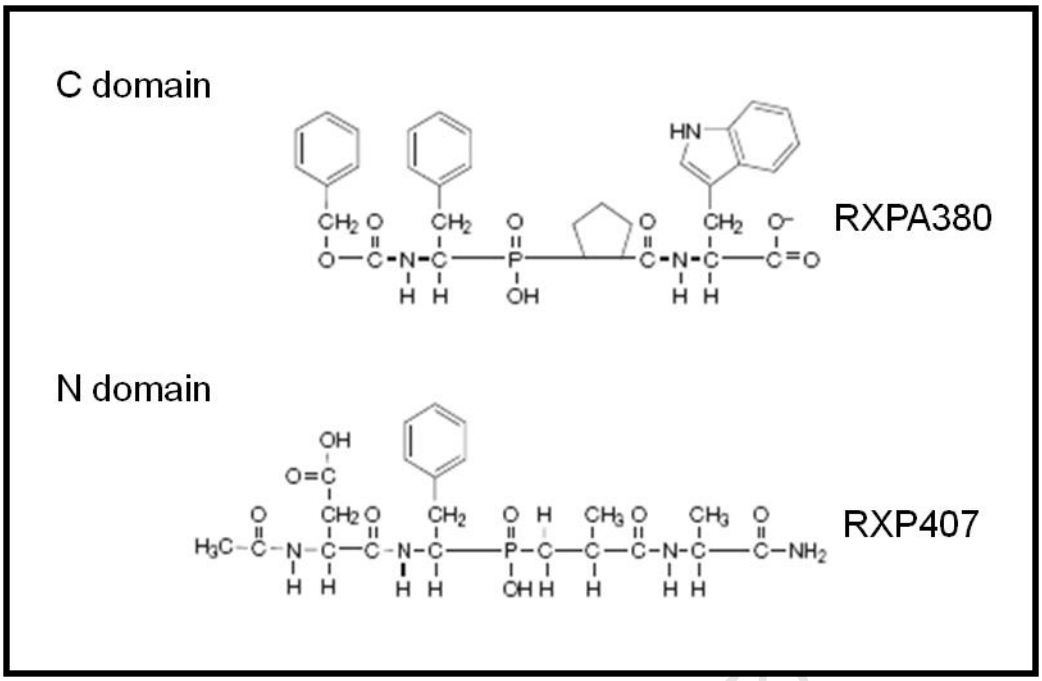


Figure 1.3: Chemical structure of the N- and C-domain selective ACE inhibitors RXP407 and RXPA380.

1.2 Protein glycosylation

Glycosylation is a highly important form of post-translational modification in eukaryotic systems and has been shown to play a role in numerous, crucial cellular functions [49,50]. The importance of protein glycosylation is further highlighted by the fact that 50% of eukaryotic proteins are predicted to be glycosylated, with 90% of these containing *N*-linked glycans [51].

1.2.1 Types of glycosylation

Enzymatic glycosylation of polypeptides occurs in two forms, namely: *O*-linked and *N*-linked glycosylation. *O*-linked glycosylation varies greatly, but generally, the glycan is attached via the linking of an N-acetylgalactosamine to the oxygen atom in the side chain of a serine (Ser) or threonine (Thr) residue [52,53]. In *N*-linked glycosylation, attachment is via an *N*-acetylglucosamine residue linked to the nitrogen on the side chain of an asparagine (Asn) residue of the Asn-Xaa-Ser/Thr sequon (where Xaa is any amino acid except proline (Pro)) [54]. Attachment of *N*-linked glycans is carried out co-translationally by the oligosaccharidal transferase complex (OST). *N*-linked glycans are generally much more bulky than their *O*-linked counterparts (Figure 1.4),

occur much more frequently and are also thought to play a greater role in protein folding and function [52].

1.2.1.1 Types of *N*-linked glycosylation

N-linked glycans fall into three general categories: high mannose, complex and hybrid glycans (Figure 1.4).

High-mannose glycans are so called because they are largely comprised of mannose residues. These glycans are generally made up of a pentasaccharide core (two *N*-acetyl glucosamines (GlcNAc), followed by three mannose resides in a branched formation), plus a varying number of branched mannose residues [53].

Complex glycans generally contain two GlcNAc residues, two galactose and up to two sialic acid resides, with one fucose sometimes attached to the first GlcNAc, in addition to the pentasaccharide core. However, the composition of complex glycans may vary considerably and are largely determined by the presence of specific glycosyltransferases in the *medial* and *trans* faces of the Golgi [53].

Hybrid glycans are comprised mainly of mannose residues, but also contain a number of other sugars. These glycans contain the pentasaccharide core, with the addition of a few mannose residues, but are also prone to the addition of fucose, galactose, GlcNAc and sialic acid residues, although addition is generally restricted to a single branch of the core glycan [53].

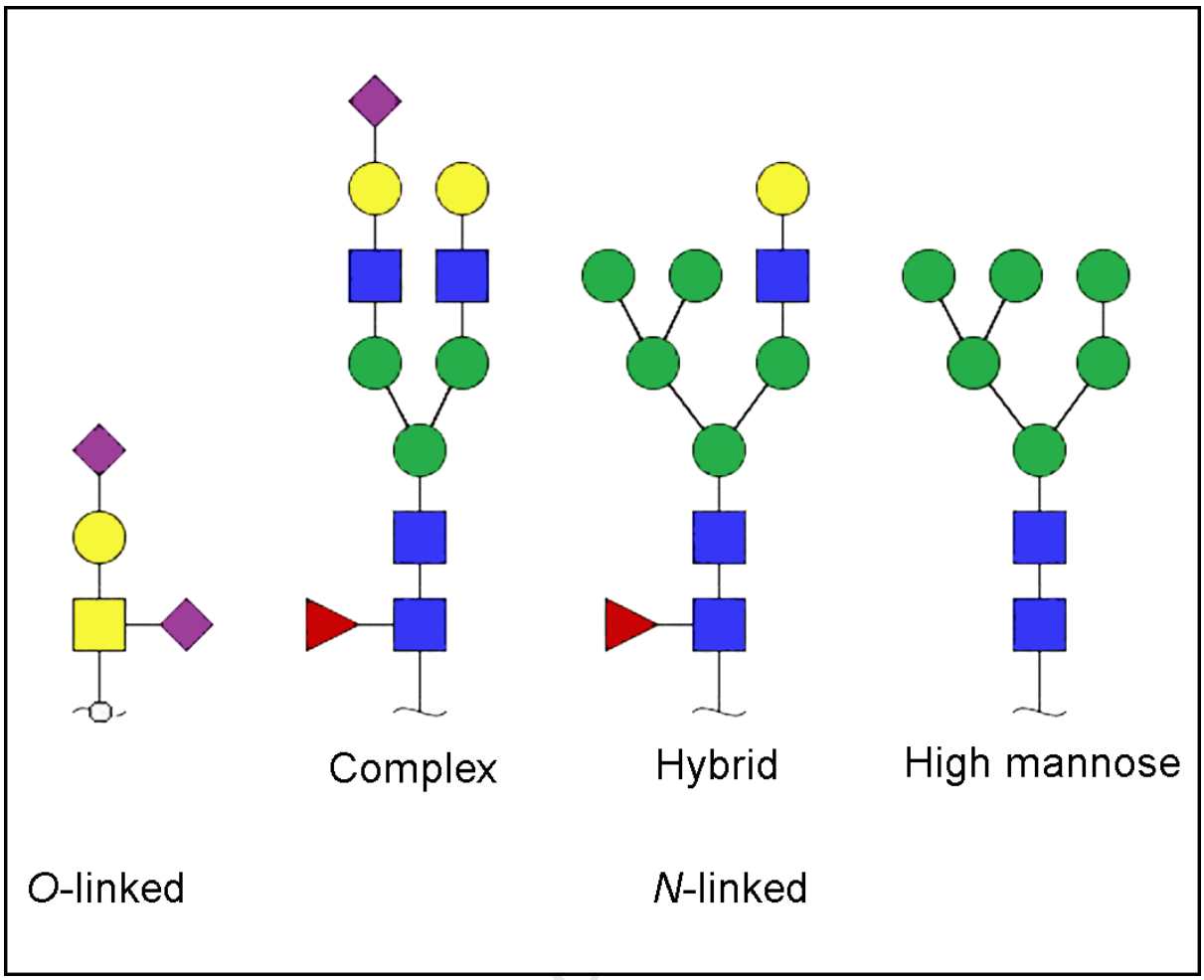


Figure 1.4: Structures of *O*- and *N*-linked glycans. Blue squares show GlcNac; Green circles depict mannose; Yellow circles indicate galactose; Red triangles indicate fucose, Purple diamonds show sialic acid residues; Yellow squares indicate N-acetyl galactosamine.

1.2.2 Requirements for protein glycosylation

While most proteins which contain *N*-linked sequons have some level of glycosylation, many proteins have more than one potential glycosylation site and not all of these are necessarily glycosylated. Analysis of glycoproteins in the curated SWISS-PROT database revealed that on average, *N*-linked sequon occupancy is close to 66% [51]. There are a number of reasons why *N*-linked sequons are not always glycosylated. These include the rate of protein folding, the subunit composition of the OST, the location of the sequon on the polypeptide chain and most notably, the sequence of the residues surrounding the glycan sequon [50]. Binding of the OST to the growing polypeptide requires the fulfilment of a set of highly specific steric constraints. These constraints are accommodated by the classic *N*-glycan recognition motif Asn-Xaa-Ser/Thr [55]. Apart from this signature motif, no

additional structural or sequence requirements appear to be critical for glycosylation, however, the presence of a Pro residue at the Xaa site (Asn-Pro-Ser/Thr), or immediately following the glycan sequon (Asn-Xaa-Ser/Thr-Pro), is known to prevent glycosylation [56]. In addition, the residues at positions close to the sequon are able to positively or negatively modulate the likelihood of glycosylation at a particular site [50,55,57]. A number of studies have examined this phenomenon and the results show that certain residues at positions Xaa^{-2}, Xaa^{-1}, Xaa and Xaa^{+1} in the sequence Xaa^{-2}-Xaa^{-1}-Asn-Xaa-Ser/Thr-Xaa^{+1} are able to influence the likelihood of whether a sequon will be glycosylated or not [49,57-62]. A recent analysis of over 25000 glycoproteins containing over 2500 experimentally annotated glycosylation sites has revealed that in addition to the sequence of residues surrounding the glycosylation sequon, the distance of the sequon from the C-terminus also has a significant effect on sequon usage [63]. Additionally, since glycosylation occurs co-translationally, glycosylation of a particular site is likely to be limited to those sequons that are able to achieve favourable conformations, enabling the glycan to be transferred to the growing polypeptide through short-range interactions with the OST [55].

1.2.3 Glycosylation and protein folding

As mentioned, *N*-glycosylation occurs co-translationally, before a significant degree of folding can take place, meaning that the timing of protein glycosylation allows for it to have a significant effect on protein folding [64-66]. Without doubt, assisting the folding of proteins in the endoplasmic reticulum (ER) is one of the most important functions of *N*-linked glycosylation in eukaryotes [67-69]. It is well established that glycosylation is required for a number of proteins to fold correctly. Interestingly there are a number of ways in which glycans are able to carry out this function, affecting both local and global structure stability, as well as playing an important role in the targeting of misfolded proteins for degradation.

1.2.3.1 Direct effects of glycosylation on protein folding

N-linked glycans are known to have a direct effect on protein folding [54,68,70]. An elegant study using polypeptide fragments with and without glycosylation sites has shown that the attachment of a glycan to a polypeptide fragment can promote a conformational shift [71]. This shift was able to cause the peptide to adopt a more compact structure, comparable to that which it occupies in the full length folded

protein. This effect was not seen with the unglycosylated peptides. It is proposed that these conformational shifts could act as “nucleation” events, limiting the transition to certain (possibly unfavourable) conformations, while making the adoption of others possible, thereby forcing the protein through a particular folding pathway. This is important since protein folding occurs rapidly, thus requiring certain mechanisms to streamline the folding process and thereby prevent the aggregation of unfolded or partially folded proteins [71,72]. Furthermore, interactions between the first few glycan residues of the attached glycan and the polypeptide backbone are able to promote the formation of β–turns, a key structural motif, further highlighting the direct role glycans play in the folding process [50,73]. The hydrophilic polysaccharides on the protein backbone are also able to promote folding by orientating the sequence proximal to the attached glycan towards the surface of the protein structure. This effect is a result of the thermodynamic force of entropy: the hydrophobic regions of a protein tend to orient themselves to the internal region of the protein where they can interact with other hydrophobic residues, thus being isolated from the polar solution surrounding the protein. In contrast, the hydrophilic regions are positioned such that they are located on the outer surface of the protein, enabling interactions with the polar environment of the ER. Thus, the attachment of hydrophilic glycans to the protein can alter the hydrophobicity of a specific region of the sequence, affecting its orientation as the protein folds [68]. In many instances the location of the glycosylation site(s) is important, although there often exists a level of redundancy, such that the importance of a particular site may only be revealed when other sites are removed [68].

1.2.3.2 Indirect effects of glycosylation on folding

1.2.3.2.1 Glycosylation and the calnexin and calreticulin cycle

The calnexin-calreticulin cycle is probably the most significant aspect of glycosylation mediated folding of glycoproteins [50]. The calnexin-calreticulin story begins with the attachment of the glycan chain to the growing polypeptide in the ER. This glycan chain consists of the classical pentasaccharide core plus a further six mannose residues, forming a branched structure. This heavily mannosidated structure is capped by three glucose residues attached to one branch of the glycan chain [50]. These glucose residues provide the key to the calnexin-calreticulin cycle. The first two glucose residues are cleaved off by glucosidase-I and II in a stepwise fashion (Figure 1.5), leaving a single glucose residue [74-76]. The third glucose residue is

subsequently removed by glucosidase-II and reattached by the glucosyltransferase in a cyclic fashion [67,77,78]. The cyclic removal and reattachment of this glucose residue is critical for the cells protein-folding quality control system as glycoprotein binding to the lectin chaperones calnexin and calreticulin is dependent on the presence of this glucose residue [65].

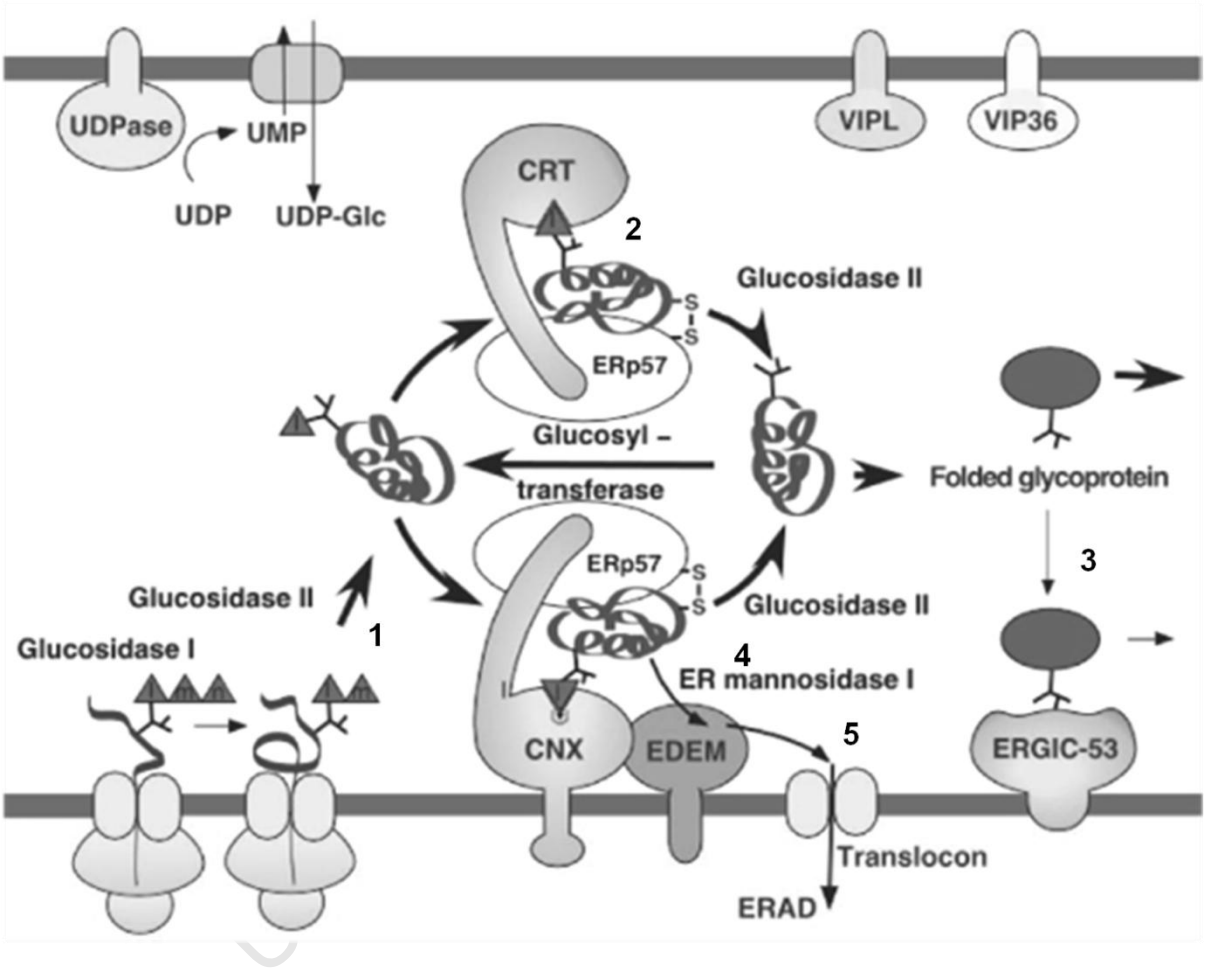


Figure 1.5: Diagram of glycoprotein folding via the calnexin-calreticulin cycle. 1) Monoglucosylated proteins bind to calnexin (CNX) and calreticulin (CRT), 2) which assist glycoprotein complex formation with ERp57, enabling correct disulfide bond formation. 3) Correctly folded proteins are translocated to the golgi by ERGIC-53, VIP36 or VIPL. 4) Missfolded proteins are demannosylated by ER manosidase-1 and translocated to the cytoplasm by EDEM, 5) for degradation by the ERAD pathway (Adapted from Helenius 2004).

Typically, the monoglucosylated glycoprotein binds to calnexin or calreticulin [78], which facilitates the formation of a complex with ER protein 57 (ERp57), a thiol-disulfide oxidoreductase. ERp57 assists disulfide bond formation, enabling the glycoprotein to form disulfide bridges in the correct positions (Figure 1.5) [50,79].

Proteins which have achieved their correct conformation are not recognised by the glucosyltransferase, and thus remain unglucosylated, exiting the calnexin-calreticulin cycle [80]. These proteins are then free to be translocated to the golgi with the assistance of ER golgi intermediate compartment (ERGIC-53), vesicular integral protein 36 (VIP36) or VIP36-like protein (VIPL) [50,79].

Protein which have failed to fold correctly are targeted for degradation [50]. This occurs when the misfolded glycoprotein is acted on by the ER manosidase-1. The removal of a single terminal mannose residue from the glycan chain allows the misfolded protein to bind the ER degradation enhancing α-mannosidase-like protein (EDEM). EDEM facilitates translocation of the misfolded protein to the cytosol for degradation as part of the ER associated degradation pathway (ERAD), thereby preventing the accumulation of misfolded proteins in the ER [50,81]. Thus, the ERAD pathway, and the controls regulating it, are elegantly designed to provide a critical means of removing misfolded proteins from the ER, without exposing resident proteins to unwanted proteolysis [50].

1.2.3.2.2 The influence of glycosylation on protein solubility and aggregation

The importance of glycosylation with respect to improving the solubility of a folding protein has more to do with the presence of the glycans and less to do with their location. Folding intermediates generally have a higher hydrophobicity profile as compared to the fully folded protein [67]. This is because the majority of the hydrophobic residues are orientated towards the interior of the protein in the folded state while, in contrast, they are exposed to the solvent during the folding process. The exposed hydrophobic regions of a polypeptide intrinsically prefer to interact with other hydrophobic residues in the polar environment of the cell, due to the lower state of entropy this induces. Thus, the presence of large polar glycans on the polypeptide chain are a convenient way for the cell to reduce the hydrophobicity of the folding intermediates, thereby increasing their solubility and reducing the likelihood of irreversible protein aggregation [54,67]. These factors allow the folding intermediates a greater chance of achieving the correct conformation than would otherwise have been possible [67].

In summary, glycosylation assists protein folding by direct and indirect means and although often critical for the correct folding of glycoproteins, the glycans are generally not needed for maintaining the structure once the protein has folded, as evidenced by the fact that removal of the attached glycans does not affect protein structure or function in a number of cases [54,67,68,82]. However, some proteins do require the presence of glycans to retain their folded state and biological function, as is the case with human CD2 protein [67,83]. In short, some proteins are dependent on attached glycans for folding, while some are able to fold regardless of whether glycans are attached or not. Some proteins are dependent on glycosylation only in certain cell types [84], while others become temperature dependant for folding if the glycans are removed, only being able to fold correctly at lower temperatures [50,67,68,82].

These findings highlight the fact that while glycosylation is critical for the folding of many glycopeptides, this is not always the case, and that even where it is, some glycosylation sites may be more important than others. Generally the more surface glycans present on a protein, the more dependent that protein is likely to be on glycosylation for folding [50].

1.2.4 Glycosylation and proteolysis

Protein glycosylation is known to retard proteolysis of a number of proteins, including deoxyribonuclease-1, butlycholinesterase, lactoferrin, fibronectin and saposin-B [85-89]. There are two likely mechanisms for a glycan mediated increase in resistance to proteolysis. The attachment of glycans have, in some cases, been associated with the formation of β–turns in the protein structure proximal to the glycan site and may provide conformational rigidity at these β–turns [52,90], making the protein a poorer substrate for proteases [88]. Alternatively, the presence of bulky *N*-linked glycans is thought to shield potential protease recognition sites on the surface of the protein [82,85-87], thus increasing the half life of the protein by reducing protease access.

1.2.5 Glycosylation and protein-protein interactions

Glycosylation is known to play a role in protein-protein interactions. This is most notably evidenced by the fact that the ability to bind sugars categorises an entire protein family, termed lectins, of which calnexin and calreticulin are classic examples. The biological role of lectins vary greatly, with proteins from this group involved in

biological processes ranging from protein folding, immune response and fertilisation to cell-cell and cell-matrix adhesion, amongst others [91,92].

An important aspect of glycan mediated protein-protein interactions is the phenomenon of dimerisation and oligomerisation. The formation of dimers and oligomers is critical to the functioning of certain proteins [93-95]. Glycans can promote the formation of dimers in a number of ways. Firstly, glycans have been shown to facilitate protein dimerisation by mediating direct interactions with the protein backbone of the other subunit(s), thus stabilising the dimer (or oligomer) and holding the various subunits together [52,96]. Additionally, evidence suggests that for certain proteins, folding is rapidly followed by oligomerisation [67]. Indeed, numerous proteins have been shown to oligomerise while attached to calnexin [67,78,97-100]. In these cases it would seem that glycosylation plays an indirect role in the formation of the oligomers by facilitating polypeptide binding to calnexin, which allows multiple subunits to fold in close proximity to one another, thereby promoting subunit assembly into dimer or oligomer complexes [67].

1.2.6 Glycosylation and thermal stability

The thermal stability of proteins is a topic of much interest, particularly in the context of the biotechnology industry, where higher temperatures generally relate to increased reaction rates [101]. There are a number of methods for analysing the thermal stability of proteins. The three most commonly used methods include Fourier transform infrared spectroscopy (FTIR), differential scanning calorimetry (DSC) and circular dichroism (CD) spectroscopy [102].

Thermal stability analysis of total protein from mammalian cells show that proteins vary greatly in their degree of thermal stability, with melting temperatures (T_m) ranging from about 40 °C to 90 °C [102,103]. This high degree of variability in the level of thermal stability for different proteins can be due to a number of factors. These factors include α-helical content, hydrophobicity, the level of internal packing and aromatic clustering, the number of ion pair interactions, proline content, as well as the presence of disulfide and salt bridges [104-106]. In addition, *N*-linked glycosylation is also able to provide proteins with increased resistance to thermal stress. It is thought that glycosylation is able to influence thermal stability by providing stabilising

interactions between the glycans and the residues on the surface of the protein [52,67,82,85]. Various studies have shown that deglycosylated, or unglycosylated, forms of certain proteins have reduced thermal stabilities compared to their fully glycosylated counterparts, as is the case for proteins such as deoxyribonuclease-1, immunoglobulin G_1 and cathepsin E [85,107,108].

1.2.7 Glycosylation and protein crystal structure determination

Glycosylation presents a major obstacle to protein crystallography. This is because the glycans attached to the protein surface are bulky, flexible and often of heterogeneous in composition, thus interfering with crystal packing by disrupting the protein-crystal forming surfaces [109-111]. However, glycosylation is a vital form of post-translational modification and is often a requirement for correct folding and processing [67], as is the case with ACE [112].

There are a number of methods to circumvent this dilemma: Firstly, non-mammalian expression systems such as expression in yeast and insect cells may be used. These systems have the advantage of possessing simpler glycosylation pathways, resulting in a reduced level of sample heterogeneity, something which is beneficial for protein crystallisation [109-111]. However, insect cell culture is costly and is often associated with technical difficulties, while a number of glycoproteins fail to fold correctly in insect and yeast cells or require complex glycosylation for proper function [113,114]. The failure of glycoproteins to fold correctly in these systems can be corrected by genetic manipulation, providing these systems with a humanised glycosylation pathway, although this does not aid crystallisation [113,114]. In addition, folding problems related to expression in yeast cells can be overcome by exploiting the unfolded protein response pathway, which activates ERAD of glycoproteins [115]. Recombinant protein overexpression is often used to generate large quantities of a protein of interest. However, this approach can saturate the cells expression machinery, leading to sharp increase in the amount of unfolded proteins and consequently activation of the ERAD pathway, thus leading to lower total yields. By adjusting the expression rate of the construct, by using different promoters, one is better able to match the expression of the protein of interest with the expression capacity of the cell, leading to a greater efficiency in the expression of correctly folded protein [115]. Additionally, saturation of the folding machinery can be alleviated

by co-overexpression of folding chaperones, leading to a greater yield of folded protein [116]. Another alternative is to express the protein in a glycosylation deficient mammalian cell line, such as Chinese hamster ovary (CHO) *Lec* 3.2.8.1 cells, which produce only high mannose glycans [117,118]. However, poor protein yields from *Lec* 3.2.8.1 cells is a drawback to this approach. Furthermore, while these systems are useful for the expression of glycoproteins, the level of glycan heterogeneity can still be problematic for crystallisation studies.

Problems associated with glycan complexity can be overcome by enzymatic deglycosylation of the purified protein [119], although this requires re-purification of the target protein and deglycosylation is not always 100% efficient. Alternatively, expressing the protein of interest in the presence of glycosylation inhibitors such as tunicamycin, swainosin, kifunensin and *N*-butyl deoxynojirimycin (NB-DNJ) [117], have been used to reduce the level of protein glycosylation [109]. The use of tunicamycin however is not suitable for the expression of all glycoproteins, as this inhibitor prevents dolichol oligosaccharide precursor synthesis, thus eliminating all *N*-linked glycosylation [67,120]. NB-DNJ inhibits α-glucosidase I and II, preventing cleavage of the three terminal glucose residues from the core oligosaccharide structure, thus, blocking further glycosylation from taking place and leaving the attached glycans in a high mannose state [109]. Swainosin and kifunensin are α-mannosidase inhibitors and generally allow for better protein yields than NB-DNJ since they act downstream of the folding pathway [117]. The use of glycosylation inhibitors has been successful, particularly when combined with an enzymatic deglycosylation step, typically involving endogylcosidase-H [117], although complete inhibition of the glycosylation machinary is not always achieved, meaning that a degree of glycan complexity and heterogeneity may still exist after treatment [109]. Attempts to address this problem have focussed on combining the use of the glycosylation deficient *Lec* 3.2.8.1 cells, with expression in the presence of a glycosidase inhibitor. This approach has proved successful in reducing the glycan complexity and heterogeneity and allowing the production of proteins suitable for crystallisation [109], although expression levels are often very low. Given the high protein concentrations required for protein crystallisation and the high cost of the glycosylation inhibitors, this amounts to a significant drawback.

Alternatively, to prevent the formation of large, complex glycans on the surface of the protein, or their removal by laborious means, a minimally glycosylated protein can be engineered. This involves determining which of the potential glycosylation sites present on the surface of the protein are actually required for correct folding. This is achieved by sequentially disrupting the *N*-glycosylation sequons using site-directed mutagenesis, where the Asn of the Asn-Xaa-Ser/Thr glycosylation sequon is mutated to a glutamine (Gln). This conservative mutation eliminates OST recognition of the sequon, while preserving the physical properties of the sequence. This method has been used successfully to generate proteins amenable to crystallisation in a number of cases [121-123]. The benefit of this method is that once set up, the expression-to-crystallisation pipeline contains relatively few steps, allowing for a high throughput approach to be employed. This is beneficial particularly in cases where the rational design of domain selective inhibitors is being sought, allowing numerous potential inhibitors to be crystallised and the interactions optimised based on the data from the X-ray crystal structure.

1.3 Glycosylation of ACE

Human ACE is a heavily glycosylated protein, approximately 30% by weight, and contains a total of 17 *N*-linked glycosylation sites [1,7,124], with ten of these being located in the N-domain. Studies on the C-domain (tACE), recombinantly expressed in CHO cells, have shown that of the seven potential sites on this domain, the three N-terminal sites (sites 1, 2 and 3) are always glycosylated, while the following three sites (sites 4, 5 and 6) are glycosylated with partial efficiency, being found in both the glycosylated and unglycosylated forms. Site 7 was found to be unglycosylated in all instances [124]. Analysis of sACE, purified from human kidney cells, showed that glycosylation was present at sites 1, 2, 4, 5 and 9 of the N-domain, while only site 5 (N913) of the C-domain was glycosylated [124] (Figure 1.6). While this study did not detect glycosylation at a number of potential *N*-linked sites on both the N- and C-domains of sACE, this does not automatically imply that they are unglycosylated. Nevertheless, analysis based on calculations of the mass contribution of the glycans to sACE indicate that approximately seven to eight sites are glycosylated [84], indicating that most of the glycosylated sites were detected.

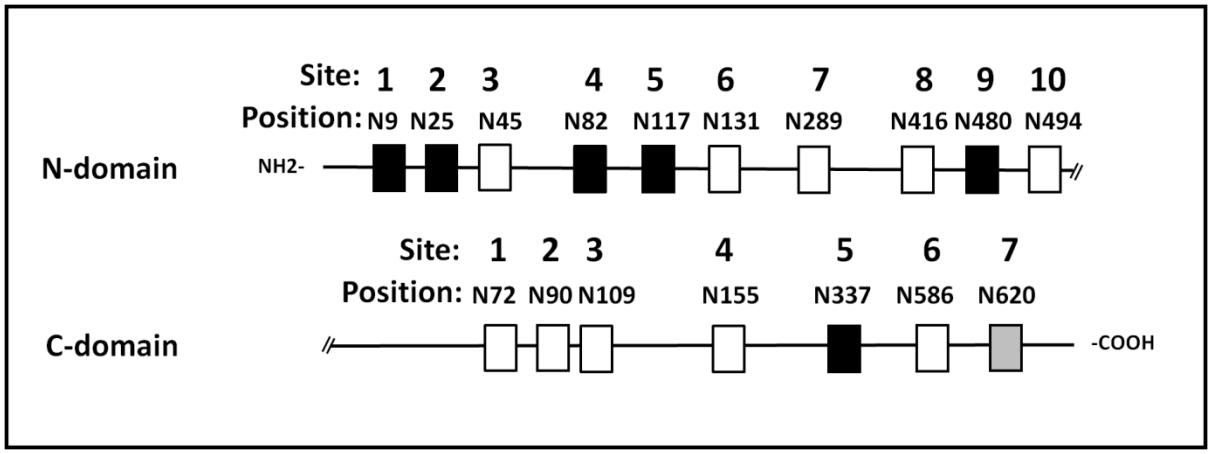


Figure 1.6: Glycosylation of the N- and C-domains of sACE. *N*-linked glycosylation sites on the N- and C-domains shown as boxes. Black boxes indicate sites which are glycosylated, white boxes depict sites with unidentified glycosylation status and grey boxes indicate sequons found to be unglycosylated. Glycosylation sites are numbered according to the sequence of the individual domains.

Additionally, Yu *et al*. (1997) analysed the sequence of the attached glycans on tACE. This was achieved using mass spectrometry to analyse the masses of digested peptides, treated with and without protein *N*-glycosidase-F (PNGase-F), which removes *N*-linked glycans. Observed masses corresponding to the sum of a glycopeptide mass, plus that of a theoretical glycan, were identified in a number of cases, but not after PNGase-F treatment, indicating that those masses were the result of glycan attachment to potential glycopeptides. From these data they were able to predict the composition of the glycans. Their results show that tACE expressed in CHO cells contain mostly biantennary, fucosylated complex glycans, with either one or two terminal sialic acid residues (as per the structure of complex glycans indicated in Figure 1.4). It is expected that the glycans of the N-domain and sACE would fall into a similar category.

1.3.1 Glycosylation requirements for ACE expression

Glycosylation is known to be important for the correct folding of ACE proteins. When tACE was expressed in bacterial cells lacking glycosylation machinery, only inactive protein was produced [112]. Additionally, when tACE was expressed in HeLa cells in the presence of tunicamycin (which prevents the attachment of *N*-linked glycans), or when all of the *N*-linked glycosylation sites were mutated, expressed tACE proteins were inactive and underwent rapid intracellular degradation [112,122]. Similarly, with the

expression of recombinant N-domain, inactive protein was produced when a number of the *N*-linked glycosylation sites were mutated [125].

Extensive site-directed mutagenesis of the *N*-glycosylation sites on human tACE has revealed that glycosylation of only site 1 or 3 is needed for the production of active protein [122]. These findings correlate with those of Yu *et al.* (1997), which showed that the three N-terminal sites were always glycosylated, while the remaining sites were either found to be unglycosylated, or were glycosylated with partial efficiency, highlighting the presence of a selective pressure on the N-terminal sites. Studies on rabbit tACE, revealed that only sites 1 or 2 were required for expression of functional protein, while site 3 alone was not sufficient. However, when expressed in yeast cells, glycosylation at site 3 alone was sufficient to allow for correct folding, indicating that the exact glycosylation requirements vary between different cell types and culture systems [112]. These findings again highlight the importance of *N*-glycosylation for the correct folding and processing of ACE proteins.

While glycosylation is essential for the folding and expression of active ACE proteins, studies have shown that if the attached glycans are enzymatically removed under native conditions, the protein retains its activity, displaying catalytic constants comparable to that of the fully glycosylated protein [124]. Similarly, when varying numbers of the *N*-glycosylation sites were mutated by site-directed mutagenesis, the catalytic constants for substrate hydrolysis remained within the same range [122]. Thus, glycosylation does not seem to play a role in the enzymatic activity of ACE.

1.3.2 ACE glycosylation and dimerisation

Somatic ACE is known to occur as a monomer, but has also been shown to occur as a homodimer in a reverse micelle membrane model, as well as in cell culture experiments with CHO and endothelial cells [126,127]. Somatic ACE also occurs as a soluble form, which is known to arise from a cleavage event at the “stalk” region, between the C-domain and the transmembrane domain, carried out by an as yet unidentified secretase [13,128,129]. Dimerisation of sACE has been found to negatively influence shedding of the enzyme from the surface of CHO cells [130]. Interestingly glycosylation has also been shown to affect the level of sACE shedding, with partially deglycosylated sACE being released into the medium with a much greater efficiency,

compared to fully glycosylated sACE [131]. The importance of this finding is evidenced by the fact that elevated levels of soluble sACE, resulting from increased cleavage of membrane bound sACE, have been associated with granulomatous disorders such as sarcoidosis [132].

Currently two theories have been proposed for the mechanism of ACE dimerisation. The first suggests that dimerisation is mediated by glycans on the N-domain binding a putative carbohydrate recognition domain (CRD) also located on the N-domain [126]. This is based on findings which showed that glycosylated sACE was able to form dimers, while this ability was lost after sACE was enzymatically deglycosylated with PNGase-F. In addition, Kost *et al.* (2000) showed that dimerisation could be inhibited by the addition of excess free glycans (specifically galactose and N-acetylneuraminic acid), as well as by the whole pool of sACE oligosaccharide chains. These data suggest that dimerisation is mediated by an interaction between certain sACE glycans and a CRD on the protein. The use of several monoclonal antibodies (mAbs) with differing epitopes allowed the localisation of this CRD to be mapped to the N-domain, as two mAbs directed to different, but overlapping regions of the N-domain were able to prevent carbohydrate mediated dimerisation [130].

The second theory is based on data that showed an increased level of dimerisation after inhibitor treatment of endothelial cells, an effect which was not seen with a C-domain knock-out form of sACE [127]. This suggests that ACE inhibitor binding to the C-domain of sACE promotes dimerisation as part of a signalling cascade leading to activation of the c-Jun kinase pathway [127]. Thus it would seem that sACE is able to form dimers via two different mechanisms, one mediated by the N-domain and dependent on glycosylation and the other initiated by inhibitor binding to the C-domain of sACE. However, a role for glycosylation in inhibitor induced sACE dimerisation has not been ruled out [133].

1.3.3 ACE glycosylation and thermal stability

As mentioned previously, sACE contains two domains, the N- and C-domains. The N-domain is known to be the more stable of the two domains. This idea arose from observations that an N-domain fragment was present in ileal fluid, presumably due to cleavage of the linker region joining the two domains, while a C-domain fragment

was not detected [7]. Furthermore, *in vitro* digestion with the endoproteinase Asp-N was found to cleave the linker region between the two domains, while limited proteolysis with trypsin, plasmin or kallikrein was found to produce a truncated form of sACE corresponding to the N-domain [7,8]. The greater stability of the N-domain was further demonstrated when it was shown that rat tACE was less thermally stable than the sACE form, indicating that the N-domain portion was providing increased stability to the sACE form [104,134]. Denaturation studies have revealed that this was also the case for human ACE, with tACE being less stable than the sACE form [104,135].

Differential scanning calorimetry (DSC) has revealed two distinct T_m values for sACE at 55 °C and 72 °C [104]. Analysis of sACE, as well as the individual N-domain fragment, revealed that the higher T_m corresponds to that of the N-domain. It is interesting to note that the authors were able to selectively denature the C-domain of ACE by incubating sACE at 55 °C for 1hr, implying that the two domains are able to fold, or unfold, relatively independently of one another [104].

Given the high degree of sequence and structural similarity between the two domains, it is surprising that they display such markedly different T_m values. A few reasons have been postulated to explain this difference in thermal stability, namely: a higher Pro content between amino acids 29-133 in the N-domain, compared to the equivalent region of the C-domain, potentially conferring a greater degree of structural rigidity; a slightly greater degree of α-helical formation and an increased level of N-linked glycosylation [104]. Recently it has been shown that glycosylation plays a critical role in maintaining the thermal stability of the C-domain. Using circular dichroism to monitor the temperature at which the unfolding transition occurs, the thermal stability of various C-domain glycoforms were determined. It was found that decreasing levels of glycosylation correlated with a decrease in T_m, with the absence of four of the six glycosylated sites in tACE resulting in a lowering of the T_m by about 8 °C [136]. The effect of glycosylation on thermal stability was also noted in an ACE homologue, AnCE. Expression of an unglycosylated form of AnCE in yeast cells showed that glycosylation was important for maintaining thermal stability, although it did not affect the folding of the enzyme or its catalytic functioning [137].

1.3.4 Glycosylation and the solution of ACE crystal structures

The use of protein crystal structures solved in complex with specific inhibitors is a valuable tool for rational drug design. Given the importance of ACE inhibitors in the treatment of cardiovascular disease, the solution of the sACE crystal structure is an important pursuit. Attempts to crystallise sACE have as yet been unsuccessful, most likely due to the high level of glycosylation and the flexibility of the region joining the two domains of this enzyme. However, when faced with a complex problem, a deconstructionist approach often proves successful, and indeed success has been achieved in crystallising the individual domains of sACE.

As is often the case with membrane bound glycoproteins, crystallisation of the individual domains of ACE has not been trivial. After initial attempts to crystallise tACE (equivalent to the C-domain of ACE) were unsuccessful, a truncated form was developed with the transmembrane region at the C-terminus, and a heavily *O*-glycosylated region at the N-terminus (not present in the C-domain of sACE), removed. This shortened, soluble form of the enzyme, expressed in the presence of NB-DNJ and treated with endoglycosidase-H (which leaves just a single GlcNAc residue attached to the Asn of the glycosylation sequon) was able to form diffractable crystals and led to the solution of the first X-ray crystal structure for human tACE [122,138]. The solution of the N-domain structure was also hampered by the high degree of surface glycosylation and once again, expression in the presence of the glucosidase inhibitor NB-DNJ was required to produce a crystallisable form of the N-domain [139]. Although the use of NB-DNJ allowed for the solution of the crystal structure for the N- and C-domains, crystallisation in both cases, was not reproducible, and combined with the high cost of the NB-DNJ, this approach was deemed unfeasible for routine use.

To circumvent the problems associated with reproducible crystallisation and the use of expensive glycosylation inhibitors, attempts were made to produce forms of the N- and C-domains which contained the fewest possible number of intact glycosylation sites. This approach would reduce glycan complexity and avoid the need for expensive glycosylation inhibitors. Based on the glycosylation site occupancy of tACE, revealed by Yu *et al.* (1997), extensive site-directed mutagenesis was carried out. As mentioned previously, data from this study revealed

that glycosylation of only one or two of the N-terminal sites was required for expression of active protein [122]. Using one of the hypoglycosylated forms of tACE (tACEg13), reproducible crystal formation was achieved [140]. An alignment of the two C-domain crystal structures (using PyMOL v0.92), shows that they are essentially identical, with a root-mean square deviation (RMSD) of 0.19 Å for all atoms, indicating that mutation of the glycan sequons did not affect the structure of the enzyme (Figure 1.7).

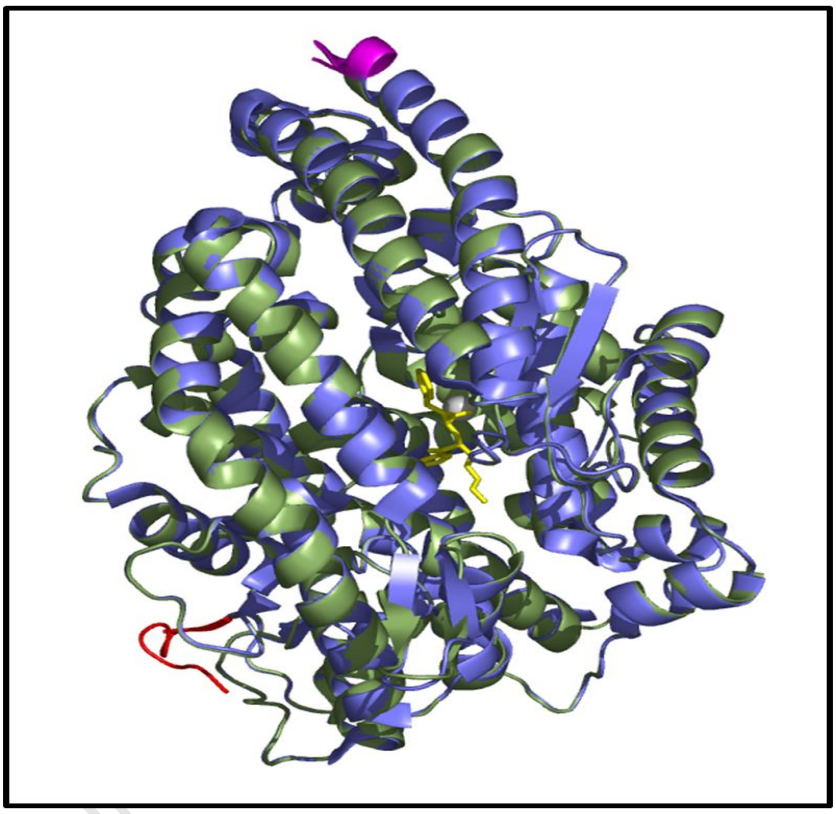


Figure 1.7: Comparison of the crystal structures of the C-domain for the glycosylation WT (tACE) (light blue) and minimally glycosylated forms (tACEg13) (Green). N- and C-termini labelled purple and red respectively. Lisinopril (yellow), bound to the active site zinc (grey) is shown. Figure was drawn with PyMOL v0.92 using the PDB structures 1O86 and 2IUL [138,140] (DeLano Scientific, San Carlos, CA, USA).

Given the high degree of structural similarity between the N- and C-domains (RMSD of 0.72 Å for all atoms) it is expected that the generation of a similarly hypoglycosylated form of the N-domain would be suitable for high throughput inhibitor-enzyme crystallisation studies. The generation of a minimally glycosylated N-domain however, requires a greater understanding of the relative importance of the individual N-linked glycosylation sites.

1.4 Aims and objectives

The effect of glycosylation on the C-domain has been well characterised, yet very little is known about the importance of glycosylation for the N-domain. This work aims to investigate the functional significance of the N-linked glycans and generate a form of the N-domain suitable for reproducible protein crystallisation. The generation of a readily crystallisable form of the N-domain will aid a high throughput approach to the rational design of N-domain selective inhibitors.

To achieve these aims, the following objectives were identified:

1. To determine the glycan site occupancy of the ten *N*-glycosylation sequons on the N-domain.
2. To determine which glycosylation sites are required for expression of active N-domain protein.
3. To examine the effect of glycosylation on the thermal stability of the N-domain.
4. To generate minimally glycosylated N-domain variants for crystallisation trials and carry out kinetic analyses on these mutants.
5. To combine minimally glycosylated N- and C-domains to generate a minimally glycosylated form of sACE
6. To assess the expression, thermal stability and kinetic parameters of minimally glycosylated sACE.

Chapter 2 : Investigating the glycosylation site occupancy of the N-domain of ACE

2.1 Introduction

The glycosylation profile of the C-domain has been extensively investigated, with the glycosylation status of each site having been determined [124]. Less is known about the glycosylation profile of the N-domain, although Yu et al. (1997) were able to identify the presence of glycosylation at sites 1, 2, 4, 5 and 9 on the N-domain of sACE purified from human kidney cells. However, this does not mean that the other sites were not glycosylated. While these sites were identified on the N-domain of full length sACE [124], the status of the ten potential glycosylation sites on recombinant N-domain, which has been used for structure determination, have not been characterised. This is important as the glycosylation site occupancy for the N-domain of full length sACE may not be identical to that of the N-domain expressed on its own. Indeed this was the case with the C-domain of sACE where only one site was found to be glycosylated, while up to six sites were glycosylated in the C-domain expressed on its own [124]. Furthermore, it is well known that protein glycosylation can vary between different cell types [68,75]. Since recombinant N-domain is commonly expressed in CHO cells, it may have a different glycosylation pattern to sACE purified from human kidney cells.

Identifying which of the ten potential glycosylation sites on the N-domain are glycosylated is an important step toward understanding the effect(s) these glycans have on enzyme processing, expression, stability and kinetics. Profiling the glycosylation site occupancy of the N-domain is also the first step toward the generation of a minimally glycosylated N-domain variant suitable for high throughput inhibitor-enzyme crystallisation trails.

The online prediction tool NetOGlyc 3.1 did not predict any *O*-linked glycosylation of the N-domain [141]. There are however ten potential *N*-linked glycosylation sites on the N-domain and an *N*-glycosylation prediction tool, NetNGlyc 1.0 [142], indicates that

nine of these sites are likely to be glycosylated (Table 2.1). The NetNGlyc software makes use of "artificial neural networks" which are trained to recognise the amino acid sequence surrounding the sequons of sites whose glycosylation status in known. These patterns are then compared to imputed sequences. In this way the likelihood of sequon usage is calculated. This method reports an accuracy of approximately 76% for *N*-linked sequon predictions [142].

Site	Position	Sequence	Potential	N-Glyc result
1	N9	NFSA	0.7272	Yes
2	N25	NSSA	0.6187	Yes
3	N45	NITA	0.6890	Yes
4	N82	NFTD	0.7827	Yes
5	N117	NMSR	0.7126	Yes
6	N131	NKTA	0.6503	Yes
7	N289	NATH	0.6084	Yes
8	N416	NDTE	0.5042	Yes
9	N480	NETH	0.5917	Yes
10	N494	NVTP	0.1559	No

Table 2.1: The predicted likelihood of *N*-linked sequon usage for the N-domain of ACE. (Generated from the N-domain amino acid sequence (see Appendix Figure A1) using the NetNGlyc web tool [142]). Glycosylation potential must exceed 0.5 for glycosylation to be predicted.

We attempted to map the *N*-glycosylation sites of purified recombinant N-domain using mass spectrometry and enzymatic deglycosylation with peptide-*N*-glycosidase-F (PNGase-F). Since PNGase-F is an amidase, the Asn (114 Da) residue of the glycosylation sequon is deamidated to an Asp (115 Da) during the removal of an attached glycan. This increase in mass of 1 Da can be used to identify the presence of *N*-glycosylation at potential glycosylation sites using mass spectrometry. To generate peptides amenable to analysis by mass spectrometry, the deglycosylated protein was digested with either trypsin or endoproteinase Glu-C (Glu-C). This allowed the formation of peptides that contained only a single

glycosylation sequon, simplifying interpretation of the data. Thus, the resulting peptides were analyzed by MALDI-ToF/ToF, while deamidation of the Asn to Asp was further confirmed by sequencing the identified peptide using MS/MS fragmentation.

Objective:

1. To determine which of the potential *N*-glycosylation sites on the N-domain are glycosylated, by identifying potential glycopeptides formed after tryptic or Glu-C digestion of PNGase-F treated N-domain, using mass spectrometry analysis.

2.2 Experimental procedures

2.2.1 Materials

PNGase-F (proteomics grade), trypsin (modified, sequencing grade) and endoproteinase Glu-C (proteomics grade) were purchased from Sigma (MO, USA).

2.2.2 Transfection and expression of N-domain in CHO cells

Soluble recombinant human N-domain (Ndom629D), comprising the first 629 residues of sACE, in the mammalian expression vector pcDNA3.1+ (Invitrogen, CA, USA) [143], was transfected into Chinese hamster ovary-K1 (CHO) cells using the $CaPO_4$ method (ProFection® Mammalian Transfection Systems kit - Promega, WI, USA) as per the manufacturer's instructions. Prior to transfection, CHO cells were grown in CHO growth medium *(10% FCS, 43% DMEM, 43% F-12 HAMS, 20 mM HEPES pH 7.5)* in 10 cm^3 culture dishes (37 °C, 5% CO_2) until approximately 50% confluent. Cells were incubated with fresh CHO growth medium for three hours prior to transfection. DNA-$CaCl_2$ crystals were prepared by bubbling 10 µg DNA and 248 mM $CaCl_2$ through an equal volume of HEPES buffered saline *(50 mM HEPES, 280 mM NaCl, 1.5 mM Na_2HPO_4 pH 7.1)*, incubated at room temperature for 30 min and added drop-wise to the cells. After 4 hrs incubation at 37 °C cells were treated with 1 ml glycerol shock solution *(15% (v/v) glycerol in 1x PBS)* for 1 min. Glycerol was removed by washing with 1x PBS *(137 mM NaCl, 2.7 mM KCl, 20 mM Na_2HPO_4, 2 mM KH_2PO_4, pH 7.4)*. Cells were then incubated with CHO growth medium (37 °C, 5% CO_2) overnight, after which medium was replaced with CHO growth medium containing 0.8 mg/ml Geneticin (G418, Sigma). Antibiotic resistant colonies were picked into a 24 well culture plate using sterile swabs and grown until confluent. Twenty four hour medium was collected and assayed for ACE activity (described in Chapter 4.2.2). Active clones were lifted into T25 culture flasks using trypsin-EDTA *(0.5% (w/v) Trypsin, 0.5 mM EDTA, 13 mM NaCl, 10 mM $Na_2HPO_4.2H_2O$, 2.7 mM KCl, 1.5 mM $KH_2PO_4.2H_2O$ pH 7.4)*. Once confluent, cells were washed with 1x PBS, lifted with TEDTA and resuspended in 10% (v/v) DMSO in FCS for storage in liquid nitrogen. For bulk expression, cells were grown to confluence in T175 culture flasks and placed under CHO medium containing 2% FCS. Medium was harvested into sterile 50 ml tubes every two to three days. Samples were stored at -20 ˚C until used.

2.2.3 Purification of recombinant N-domain protein

Soluble recombinant N-domain proteins were purified from conditioned medium by lisinopril-Sepharose affinity chromatography, as described previously [144], with the exception that the harvested medium and wash buffer were adjusted to 800 mM NaCl, to facilitate N-domain binding to the column [7]. Briefly, purification columns were washed with 20 mM HEPES, 800 mM NaCl (pH 7.5) to remove unbound components. N-domain protein was eluted with 50 mM borate (pH 9.5). After purification, samples were dialysed against 2x 2 L 50 mM HEPES buffer (pH 7.5) (containing 0.1 mM PMSF protease inhibitor) at 4 °C and concentrated using 30 000 kDa cutoff Amicon® Ultra spin centrifugal filters (Millipore, Cork, Ireland). Protein concentrations were determined by measuring absorbance at 280 nm on a Nanodrop® ND1000 using a molecular mass of 72 kDa and an extinction coefficient of 161877 M^{-1} cm^{-1}. N-domain molecular mass and extinction coefficient were calculated from the N-domain amino acid sequence (see Appendix Figure A1) using the tool ProtParam on the ExPASy web site (http://au.expasy.org/cgi-bin/protparam). Sample purity and integrity was confirmed by fractionation on 10% polyacrylamide gels using standard sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) (see Appendix A4). Protein was stained with Coomassie *(0.2% (w/v) Coomassie Blue; 7% (v/v) glacial acetic acid; 50% (v/v) ethanol)* and visualised by de-staining

(7% (v/v) glacial acetic acid; 25% (v/v) ethanol) for 90 min. De-staining solution was changed at least once. Samples were stored in 1 ml aliquots at -20 °C until used.

2.2.4 Deglycosylation of purified N-domain protein

Purified N-domain (1 mg/ml) in 50 mM potassium phosphate (0.2% SDS, 1% β-mercaptoethanol) was denatured by boiling for 5 min and then cooled on ice. Deglycosylation was then carried out by the addition of PNGase-F (100 units/ml) and incubated at 37 °C overnight.

2.2.5 In-gel protease digestion for mass spectrometry analysis

Deglycosylated ACE N-domain was fractionated using 10% polyacrylamide gels as per standard SDS-PAGE protocols (see Appendix A4). Protein was visualised by

Coomassie staining as described above. The N-domain band was excised, cut into 1 mm^3 gel slices and further de-stained in 200 mM NH_4HCO_3, 50% acetonitrile. Gel slices were dehydrated using a Speed vac®, before reduction with 5 mM triscarboxyethyl phosphine (TCEP) in 100 mM NH_4HCO_3 for 30 min at 56 °C. Excess TCEP was removed and the gel slices dehydrated. Cysteines were protected with the addition of 100 mM iodoacetamide in 100 mM NH_4HCO_3 and incubated at room temperature in the dark for 30 min. Gel slices were dehydrated and washed with 50 mM NH_4HCO_3 and dehydrated again. Gel slices were rehydrated by the addition of 20 ng/µl protease (trypsin or endoproteinase Glu-C) and digested at 37 °C overnight. Peptides were solubilised with 50 µl 0.1% trifluoroacetic acid (TFA). The peptides were lyophilised, rehydrated in 50 µl water, and concentrated to less than 20 µl in a Speed vac®, to remove residual NH_4HCO_3.

2.2.6 Mass Spectrometry

Mass spectra were recorded on an ABI 4800 matrix assisted laser desorption-ionisation time of flight (MALDI-ToF/ToF) mass spectrometer (Applied Biosystems Foster City, CA, USA), using positive reflector mode. Spectra were generated with 400 laser shots/spectrum, at an intensity of 3800 (arbitrary units), with a grid voltage of 16 kV. All peptide containing spots were internally calibrated using trypsin autolytic fragments. The matrix used was α-cyano-4-hydroxycinnamic acid, at a final concentration of 5mg/ml matrix in 40% acetonitrile, 0.1% TFA, 10 mM $NH_4H_2PO_4$ (Fluka, Buchs, Switzerland).

2.3 Results

To identify which of the *N*-linked sites on the N-domain were glycosylated, WT N-domain was deglycosylated with PNGase-F, which removes the entire glycan from the Asn attachment site. Complete deglycosylation of purified N-domain was confirmed by SDS-PAGE, where an increased electrophoretic mobility relative to glycosylated N-domain protein was noted (Figure 2.1).

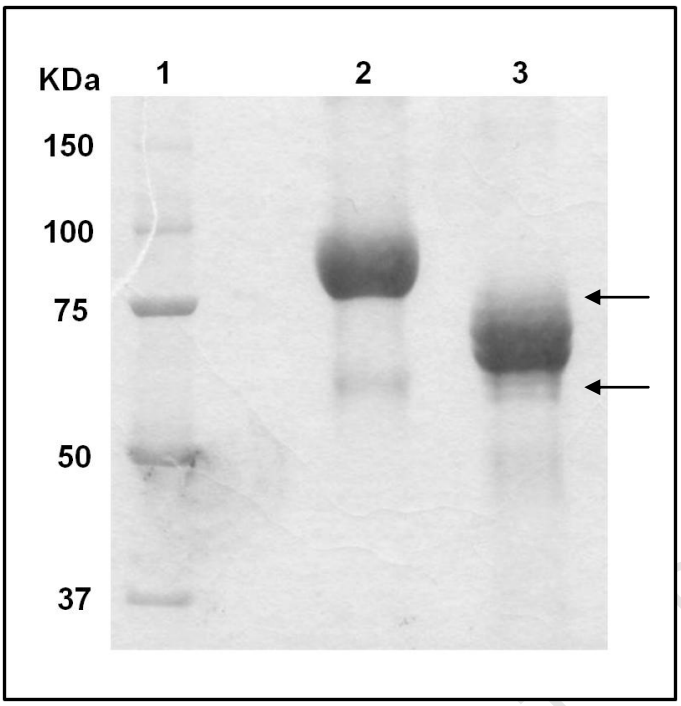


Figure 2.1: PAGE gel showing purity of N-domain protein and increased electrophoretic mobility after enzymatic deglycosylation. Lane 1 - Biorad Precision plus protein™ standard, Lane 2 - untreated purified N-domain. Lane 3 - PNGase-F treated N-domain. Additional bands indicated by arrows.

A diffuse band was noted for the PNGase treated sample, as well as a few additional bands. This may have been the result of proteolysis (lower bands) or partial deglycosylation (upper bands). However, this did not affect the analysis, as the protein band of interest was specifically excised. Disulfide bridges were reduced and alkylated in the excised gel fragments to prevent disulfide bond formation, which would confound the mass spectrometry data by linking cleaved peptides together. Following cysteine protection, the excised band was subjected to in-gel digestion with either trypsin or Glu-C and the resulting peptides analysed my MALDI-ToF/ToF mass spectrometry.

Mass spectrometry analysis of PNGase-F treated N-domain peptides resulting from tryptic and Glu-C digestion successfully revealed a number of N-domain peptides. Mass spectrometry of the tryptic peptides resulted in a sequence coverage of about 75%, while analysis of the spectra produced from Glu-C peptides revealed another 14%, meaning that approximately 90% of the 629 amino acid residues of the

N-domain were covered by this approach (Table 2.2). Importantly, all of the potential glycosylation sites were detected.

N-domain peptide	Calculated mass [MH+]	Observed mass (m/z) [MH+]	Digest Type	N-domain peptide	Calculated mass [MH+]	Observed mass (m/z) [MH+]	Digest Type
1-14	**1459.65**	**1460.56**	**Glu-C**	327-341	1881.86	1881.91	Trypsin
15-29	**1543.68**	**1544.60**	**Glu-C**	341-346	806.48	806.43	Trypsin
30-49	**2188.04**	**2188.92**	**Glu-C**	351-373	2821.34	2821.34	Trypsin
50-56	902.44	902.40	Glu-C	374-381	955.56	955.55	Trypsin
53-71	2191.07	2191.09	Trypsin	381-407	2851.47	2852.47	Trypsin
64-74	1264.62	1264.57	Glu-C	390-403	1371.73	1371.65	Glu-C
72-89	**2248.13**	**2249.12**	**Trypsin**	408-413	686.41	686.43	Trypsin
74-89	**2049.00**	**2050.00**	**Trypsin**	**414-427**	**1624.80**	**1625.79**	**Trypsin**
74-90	**2205.10**	**2206.11**	**Trypsin**	433-446	1724.91	1724.94	Trypsin
90-108	2006.23	2006.00	Trypsin	447-458	1346.69	1346.71	Trypsin
108-120	**1580.79**	**1581.79**	**Trypsin**	454-467	1900.91	1900.87	Trypsin
109-120	**1424.69**	**1425.69**	**Trypsin**	470-479	1190.59	1190.61	Trypsin
121-126	681.37	681.41	Trypsin	**480-489**	**1089.49**	**1090.50**	**Trypsin**
121-132	**1393.74**	**1394.62**	**Trypsin**	**490-500**	**1342.72**	**1342.74**	**Trypsin**
133-151	2121.02	2121.03	Trypsin	518-532	1807.80	1807.83	Trypsin
188-199	1416.61	1416.63	Trypsin	533-541	931.56	931.51	Trypsin
232-236	652.40	652.40	Trypsin	536-542	743.48	743.43	Trypsin
236-240	666.32	666.34	Trypsin	546-557	1357.71	1357.74	Trypsin
246-295	**2447.14**	**2448.20**	**Trypsin**	558-572	1597.84	1598.82	Trypsin
296-326	3526.65	3526.65	Trypsin	597-609	1655.74	1655.66	Glu-C
300-305	743.35	743.34	Glu-C	597-619	2698.22	2698.06	Glu-C
306-320	1749.78	1749.58	Glu-C	623-629	914.41	914.43	Trypsin
327-340	1753.77	1753.80	Trypsin				

Table 2.2: Tryptic and Glu-C peptides of PNGase-F treated N-domain showing calculated and observed MH^+ ions. Singly charged ions are indicated. Peptides containing potential glycosylation sites are highlighted in bold. PNGase-F deamidates the Asn (114 Da) residue of the glycosylation sequon to an Asp (115 Da). This increase of 1 Da can be used to identify the presence of *N*-glycosylation at potential glycosylation sites.

2.3.1 Glycosylation sites identified from deglycosylated tryptic peptides of the N-domain

As mentioned above, various peptides containing potential glycosylation sites were detected. The majority of these were peptides resulting from tryptic digestion of the N-domain (Figure 2.2).

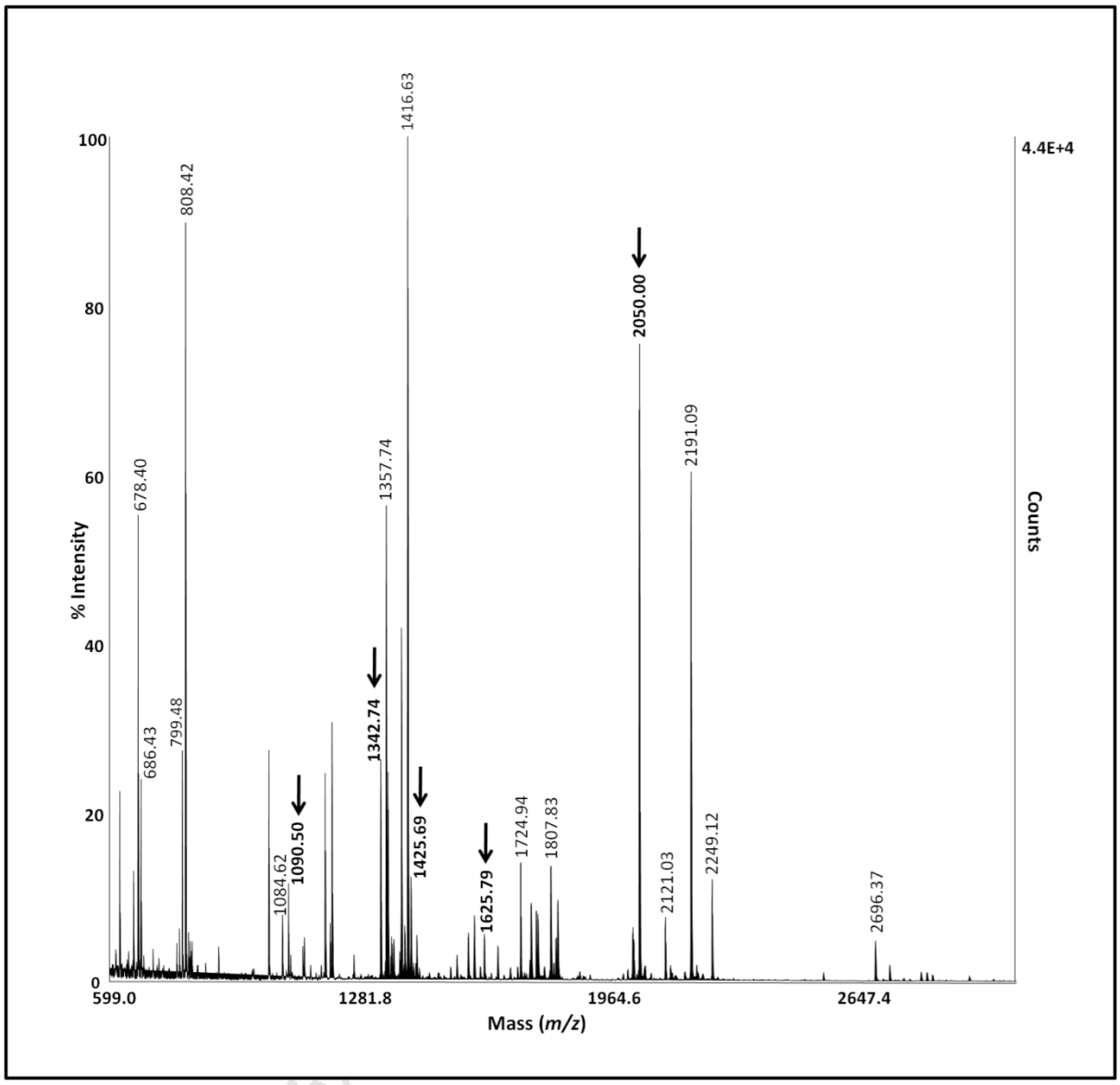


Figure 2.2: Representative MALDI-ToF/ToF spectrum of tryptic peptides from PNGase-F digested ACE N-domain. Masses indicate singly charged ions. Peptides resulting from tryptic digestion of the N-domain are labeled. Glycopeptides containing sites 4 (*m/z* 2050.00), 5 (*m/z* 1425.69), 8 (*m/z* 1625.79), 9 (*m/z* 1090.50) and 10 (*m/z m/z* 1342.74) are indicated by arrows.

Observed molecular ions from the tryptic digest at *m/z* 2050.00, 1425.69, 1625.79 and 1090.50 (Figure 2.2), were 1 Da greater than the calculated masses of $[MH]^+$ 2049.00 (Glu^{74}-Arg^{89}), 1424.69 (Gln^{109}-Arg^{120}), 1624.80 (Val^{414}-Lys^{427}) and 1089.49 (Asp^{480}- Lys^{489}), consistent with PNGase-F deamidation of Asn at sites 4 (N82), 5 (N117), 8 (N416) and 9 (N480), respectively. The observed ion at *m/z* 1342.76 agrees well with the expected mass of $[MH]^+$ 1342.72 for Phe^{490}-Arg^{500} indicating that site 10 (N494) is not glycosylated. To confirm these results, the relevant

peptides were fragmented by MS/MS (Figure 2.3), where the Asn to Asp deamidation was confirmed in each case. The exception, of course, was site 10 (N494), where an Asn was detected, consistent with the finding that this site is not glycosylated.

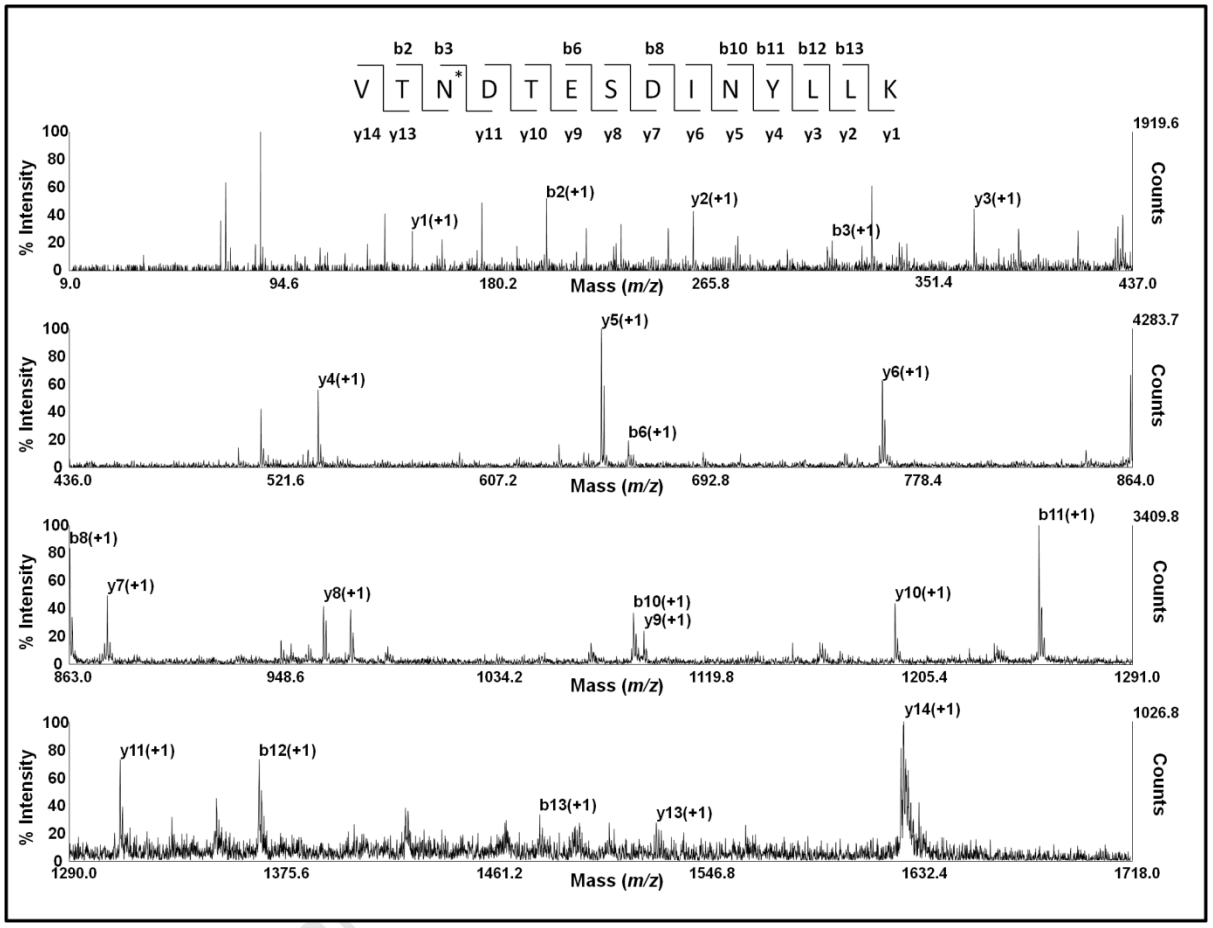


Figure 2.3: Representative MALDI-ToF/ToF-MS/MS spectra for the sequencing of potential glycopeptides. The sequence of the fragmented glycopeptide *m/z* 1625.79, containing glycosylation site 8 (N416) is indicated. The "y" and "b"-ion series are shown. The asterisk indicates a deamidated Asn (N416), causing a mass increase of 1Da.

An observed mass at *m/z* 1394.61 corresponds well with the expected mass of 1393.51 for (Ile^{127}- Lys ^{132}) given the presence of glycosylation at site 6 (N131), while the remaining site, site 7 (N289), was detected at *m/z* 2448.20 which corresponds well with the calculated mass of 2448.14 (Pro^{246}-Arg^{295}), assuming glycosylation at site seven (N289) and tryptic cleavage after Lys^{274}, which is followed by a Pro. While it is typically thought that trypsin does not cleave after a Lys or Arg when the residue is followed by a Pro, recent work has shown that a number of such sites are in fact

cleaved [145]. The likelihood of such cleavage is increased when the protein is exposed to extended overnight digestion, as was the case here. MS/MS sequencing of these two peptides was hampered by the fact that they could not be fragmented.

2.3.2 Glycosylation sites identified from deglycosylated Glu-C peptides of the N-domain

While analysis of the mass spectrometry data resulting from tryptic digestion of the N-domain enabled the identification of the glycosylation site occupancy of seven of the potential glycosylation sites. The glycosylation status of sites 1 (N9), 2 (N25) and 3 (N45) could not be determined, as these sites all occur on a single large tryptic peptide. Further analysis of the amino acid sequence of this region revealed four Glu-C cleavage sites. Fortuitously, these cleavage sites were such that glycosylation sequons 1, 2 and 3 all occur on separate Glu-C peptides, the sizes of which were between 1000 and 3000 Da, making them ideal candidates for identification by mass spectrometry. Mass spectra of peptides resulting from Glu-C digestion of PNGase-F treated N-domain showed a peak at *m/z* 1460.56 which corresponds closely with the calculated masses of $[MH]^+$ 1459.65 for the peptide Leu^1-Glu^{14}, assuming site 1 (N9) was glycosylated and the Asn converted to Asp by PNGase-F. Similarly, observed peaks at *m/z* 1544.60 and 2188.92 agree with calculated masses of $[MH]^+$ 1543.68 and 2188.04 for peptides Ala^{15}-Glu^{29} and Gln^{30}-Glu^{49}, assuming glycosylation at sites 2 (N25) and 3 (N45), respectively. The sequence of these peptides were further confirmed by MS/MS fragmentation of the relevant parent ions and showed deamidation of the Asn to Asp by PNGase-F in each case (Figure 2.4), confirming that these sites were in fact glycosylated.

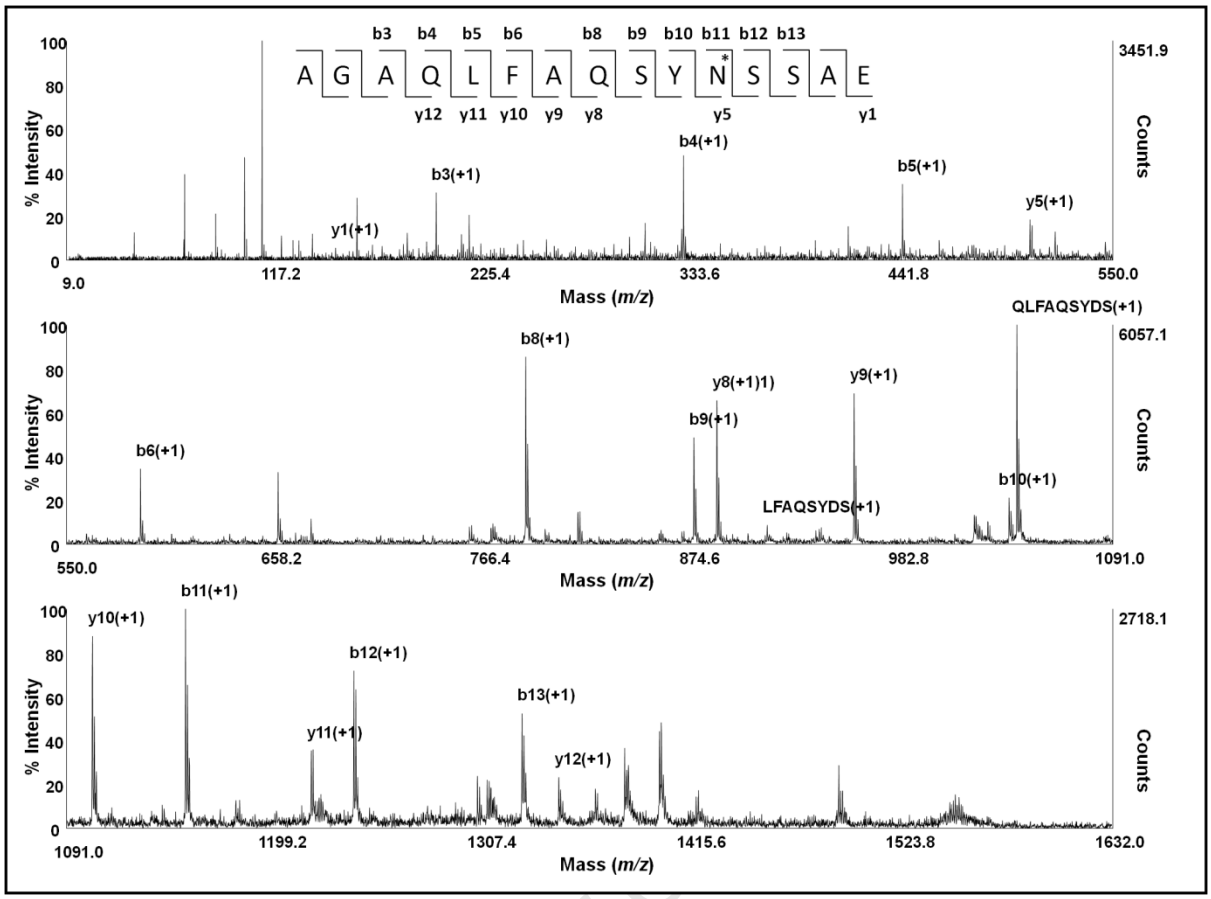


Figure 2.4: Representative MALDI-ToF/ToF-MS/MS spectra for the sequencing of potential glycopeptides. The sequence of the fragmented glycopeptide *m/z* 1544.60, containing site 2 (N25), is indicated. The "y" and "b"-ion series are shown. The asterisk indicates a deamidated Asn (N25), causing a mass increase of 1Da.

2.4 Discussion

Protein glycosylation plays a major role in regulating protein folding and function and is one of the most important post-translational modifications to occur in the endoplasmic reticulum and golgi systems of eukaryotic organisms [68]. Determining the glycosylation site occupancy is an important step toward understanding the relative importance of the various glycosylation sites on a protein, as glycosylation of specific sites is often crucial for the folding and function of glycoproteins [146]. In this regard, mass spectrometry combined with enzymatic deglycosylation and endoproteinase digestion has proven to be a useful tool in identifying the presence of attached glycans [146,147]. Indeed this approach was used to identify glycan containing sequons for another carboxypeptidase, (glutamate carboxypeptidase II) which also contains ten potential glycosylation sites [148]. In that study mass spectrometry,

enzymatic deglycosylation and endoproteinase digestion revealed that all of the ten sequons contained carbohydrates.

While sACE has 17 potential *N*-linked glycosylation sequons, it is estimated that only between seven and eight of these actually contain glycans. These findings are based on the mass difference between fully glycosylated and enzymatically deglycosylated sACE and the average size of a typical complex glycan chain [84]. Furthermore, work carried out by Yu et al. (1997) on somatic ACE purified from human kidney cells, showed that glycosylation was present at sites 1 (N9), 2 (N25), 4 (N82), 5 (N117) and 9 (N480) of the N-domain. While it is likely that Yu *et al.* were able to identify most of the glycans present on sACE, it is possible that not all of the glycosylated sequons were identified as the authors were not able to identify all of the glycosylation sequon containing peptides. In addition, the glycosylation site occupancy of the N-domain of sACE is not necessarily the same as that of an individually expressed recombinant N-domain molecule, as was found to be the case for the C-domain [124]. Interestingly, the crystal structure of the recombinant N-domain (expressed in the presence of NB-DNJ) shows ordered density for GlcNac residues at sites 2 (N25), 5 (N117) and 9 (N480) [139].

Our approach of using mass spectrometry and enzymatic deglycosylation combined with endoproteinase digestion was able to achieve up to 90% sequence coverage for the N-domain, and has allowed for the profiling of all ten of the potential *N*-glycosylation sites. Additionally, a number of non-glycopeptides were identified from the mass spectra, with a similar level of accuracy (approximately 60 ppm).

Glycan site no.	Glycan site residue	Peptide residues	Calculated mass $[MH]^+$ *	Observed mass (m/z) $[MH]^+$	Peptide identification (determined by MS/MS)	Digest type
1	N9	01-14	1459.65	1460.56	LDPGLQPG**D**FSADE	Glu-C
2	N25	15- 29	1543.68	1544.60	AGAQLFAQSY**D**SSAE	Glu-C
3	N45	30- 49	2188.04	2188.92	QVLFQSVAASWAHDT**D**ITAE	Glu-C
4	N82	74- 89	2049.00	2050.00	ELYEPIWQ**D**FTDPQLR	Trypsin
5	N117	109-120	1424.69	1425.69	QQYNALLS**D**MSR	Trypsin
6	N131	121-132	1393.74	1394.62	N/D	Trypsin
7	N289	246-295	2447.14	2448.20	N/D	Trypsin
8	N416	414-427	1624.80	1625.79	**D**DTESDINYLLK	Trypsin
9	N480	480-489	1089.49	1090.50	**D**ETHFDAGAK	Trypsin
10	N494	490-500	1342.72	1342.74	FHVP**N**VTPYIR	Trypsin

Table 2.3: Observed molecular ions for potential glycopeptides from tryptic and endoproteinase Glu-C digests of PNGase-F treated ACE N-domain. Singly charged ions are indicated. PNGase-F deamidates the Asn (114 Da) residue of the glycosylation sequon to an Asp (115 Da). This increase of 1 Da can be used to identify the presence of *N*-glycosylation at potential glycosylation sites. * Refers to calculated mass if the peptide was not glycosylated and therefore not deamidated by PNGase-F.

Here we report that nine of the ten potential glycosylation sites on the N-domain contain glycans when expressed from CHO cells, with only site 10 (N494) being found to be unglycosylated (Table 2.3). Site 10 was not expected to be glycosylated since this site is located in the interior of the protein close to the active site and additionally, the glycosylation sequon is followed by a Pro residue which is known to interfere with oligosaccharidal transferase recognition of the glycosylation sequon [56]. For these reasons, this site was also predicted by NetNGlyc to be unglycosylated (Table 2.1).

Thus, recombinantly expressed N-domain has a high degree of glycosylation site occupancy (90%); much greater than that of the C-domain which has only three out of seven potential glycan sites in the glycosylated form, although an additional three sites have been found in both glycosylated and unglycosylated forms [124]. The identification that the three N-terminal sites on the C-domain were always glycosylated, highlighted these sites as being more important than the three

C-terminal sites that were found in both glycosylated and unglycosylated forms. This fact was born out in mutagenic studies involving the removal of a number of glycosylation sites from the C-domain, where it was found that the three N-terminal sites were crucial for expression of active protein, while removal the other three had little effect on the expression of the C-domain [122]. Our findings, however, did not highlight particular glycosylation sites which may be important for the processing, folding or thermal stability of the N-domain, since virtually all sites were found to be glycosylated. However, the fact that site 10 is not glycosylated allowed us to exclude this site from further experiments aimed at identifying which glycosylation sites are important for enzyme processing and thermal stability.

Chapter 3 : Construction of N-domain glycosylation mutants

3.1 Introduction

To determine the role of the various *N*-glycosylation sites with respect to the expression of functional protein and thermal stability, a number of N-domain mutants (with varying number of intact glycosylation sequons) needed to be generated.

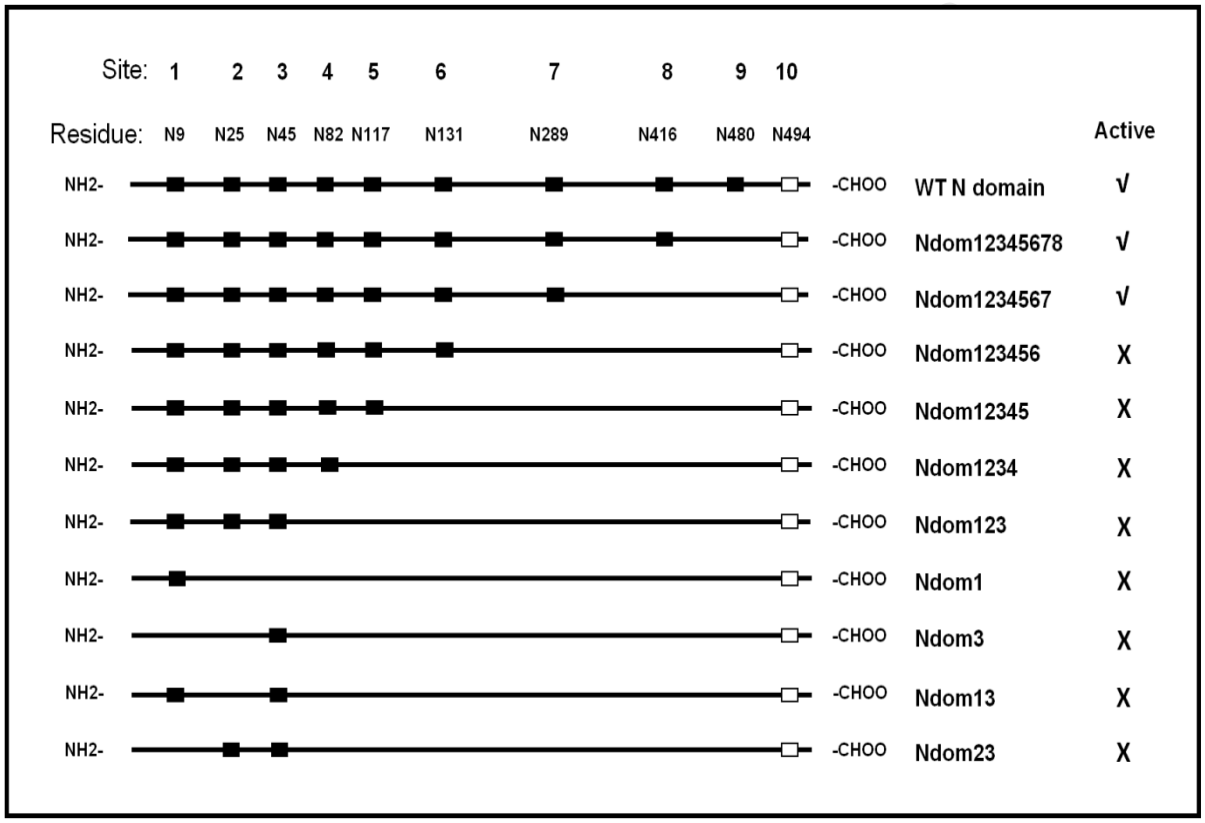


Figure 3.1: Schematic diagram of glycosylation mutants generated by P. Redelinghuys (2006), with the activity status of each mutant indicated. Black boxes indicate intact *N*-glycosylation sequons and white boxes show sites that are not glycosylated. Ndom stands for N-domain, while the numbers indicate the intact glycosylation sites.

Previous work focused on the removal of C-terminal glycosylation sites from the N-domain, based on the fact that the N- and C-domains share a high degree of sequence and structural homology and the finding that N-terminal sites were important for the expression of the C-domain [122,125]. However, results from that study

showed that C-terminal rather than N-terminal sites were important for the expression of N-domain proteins. While the presence of only one or two N-terminal sites was sufficient for expression of C-domain proteins, the removal of three C-terminal sites from the N-domain resulted in the expression of inactive protein (Figure 3.1).

Thus, a series of mutants were generated to assess the importance of the C-terminal glycosylation sites of the N-domain. A specific requirement for C-terminal sites, however, does not preclude a role for N-terminal sites in the processes of thermal stability and expression of functional protein. Therefore, a series of mutants were generated to investigate the importance of these sites in the expression of functional N-domain proteins. While C-terminal sites are important, the most C-terminal site, site ten (Asn 480) was shown to be unglycosylated (see Chapter 2.4), thus, it was not deemed necessary to investigate the importance of this site and consequently, it was not altered during mutagenesis.

Objectives:

1. To reintroduce various C-terminal sites into an existing construct to determine whether C-terminal sites rather than N-terminal sites are required.
2. To selectively remove various N-terminal sites to determine whether they affect enzyme processing and thermal stability.
3. To generate glycosylation mutants with three or fewer intact glycosylation sites to determine the minimum glycosylation requirements for the N-domain and thereby aid reproducible crystallization.

3.2 Experimental procedures

3.2.1 Materials

O'generuler DNA Ladder mix, restriction endonucleases *EcoRI, Hind III, Eco47III, Eco91I, BglII, BsmI, PvuI, NotI and XbaI* as well as T4 DNA ligase were purchased from Fermentas (Ontario, Canada). *E.coli* DH5α cells, *Dpn I* and the Wizard® SV gel and clean-up kit were obtained from Promega. *MluI, BalI and RcaI* were obtained from Roche.

3.2.2 N-domain plasmid templates

Site-directed mutagenesis of the glycosylation sequons was carried out on three previously generated human N-domain glycosylation mutants, namely Ndom123456 Ndom123 and Ndom0, where "Ndom" stands for N-domain and numerals indicate intact glycosylation sites [125]. Mutagenic products were used as templates for further mutagenesis in an iterative cycle.

3.2.3 Mutagenic oligonucleotides

Primers were designed using Watcut (Michael Palmer, University of Waterloo, Canada. http://watcut.uwaterloo.ca/watcut) and were synthesised by Inqaba biotech South Africa. Primers were engineered to introduce an Asn to Gln codon change in the Asn-Xaa-Ser/Thr recognition sequon (or to reverse such a change from a previously mutated sequon). Additionally, restriction sites were either introduced or removed via the introduction of silent mutations, to enable screening of positive clones, as indicated in Table 3.1.

Mutant	Tm (°C)	Sequence	RE Site
N9Q	83	F 5'CCCCGGGCTGCAGCCTGGC**cag**TTTTCTGCTGACGAG 3' R 3'GGGGCCCGACGTCGGACCG**gtc**AAAAGACGACTGCTC 5'	*BalI*
N25Q	77	F 5'GCAGAGCTAC**caa**TCCAGCGCTGAACAGGTGCTG 3' R 3'CGTCTCGATG**gtt**AGGTCGCGACTTGTCCACGAC 5'	*Eco47III*
N45Q	84	F 5'GGCCGCCAGCTGGGCTCATGACACC**cag**ATCACCGCGGAGAATG 3' R 3'CCGGCGGTCGACCCGAGTACTGTGG**gtc**TAGTGGCGCCTCTTAC 5'	*RcaI*
Q289N	82	F 5'GCTGCAGCAGGGCTGG**aat**GCCACGCATATGTTCCG 3' R 3'CGACGTCGTCCCGACC**tta**CGGTGCGTATACAAGGC 5'	*BsmI*
Q416N	78	F 5'CAAAATCGGCCTGCTGGACCGTGTCACC**aat**GACACGGAAAGTG 3' R 3'GTTTTAGCCGGACGACCTGGCACAGTGG**tta**CTGTGCCTTTCAC 5'	*PvuI**
Q480N	83	F 5'GGATCTGTCCTCCTGTTACCCGA**aac**GAAACCCACTTTGATGCTG 3' R 3'CCTAGACAGGAGGACAATGGGCT**ttg**CTTTGGGTGAAACTACGAC 5'	*MluI**

Table 3.1: Mutagenic oligonucleotides for the disruption or restoration of glycosylation sequons. Introduced, or removed (*) restriction sites are underlined and codon changes are indicated in bold lowercase. Forward (F) and reverse (R) primers shown.

3.2.4 Site-directed mutagenesis and cloning of glycosylation mutants

3.2.4.1 Site-directed mutagenesis

A modified version of the Quickchange™ (Stratagene, CA, USA) *DpnI* method of site-directed mutagenesis [149] was used to disrupt, or repair, glycosylation sequons in the N-domain. Reactions were carried out in a Hybaid PCRSprint thermocycler (Mandel Scientific Company Inc, Ontario, Canada) using the Kapa Hifi PCR kit (Kapa Biosystems, Cape Town, South Africa) according to the manufacturer's instructions. Briefly, 10 ng of the relevant template DNA was used in each 50 µl PCR reaction. The cycling parameters consisted of one cycle of 95 °C for 2 min, then 20 cycles of 95 °C for 0.3 min, X °C for 0.5 min (where X = T_m of the primer-target complex), 68 °C for 3 min, followed by a final extension cycle of 5 min at 68 °C. Primer-target T_m values were calculated using a the following equation: T_m = 81.5 + 0.41(% GC) - 675/N -% mismatch, where N = total number of bases. Thermal cycling was carried out with the relevant mutagenic primers. Successful mutagenesis was confirmed by analysing 10 µl of the PCR reaction on 1% (w/v) agarose gels in TBE *(0.09 M Tris,*

0.09 M Boric acid, 2.0 mM EDTA, pH 8.0, 0.3 µg/ml ethidium bromide). Following confirmation of successful mutagenesis, methylated and hemi-methylated template DNA was removed by *DpnI* digestion (Figure 3.2 A) and transformed into competent *E.coli* DH5α cells (see Appendix A1 and A2).

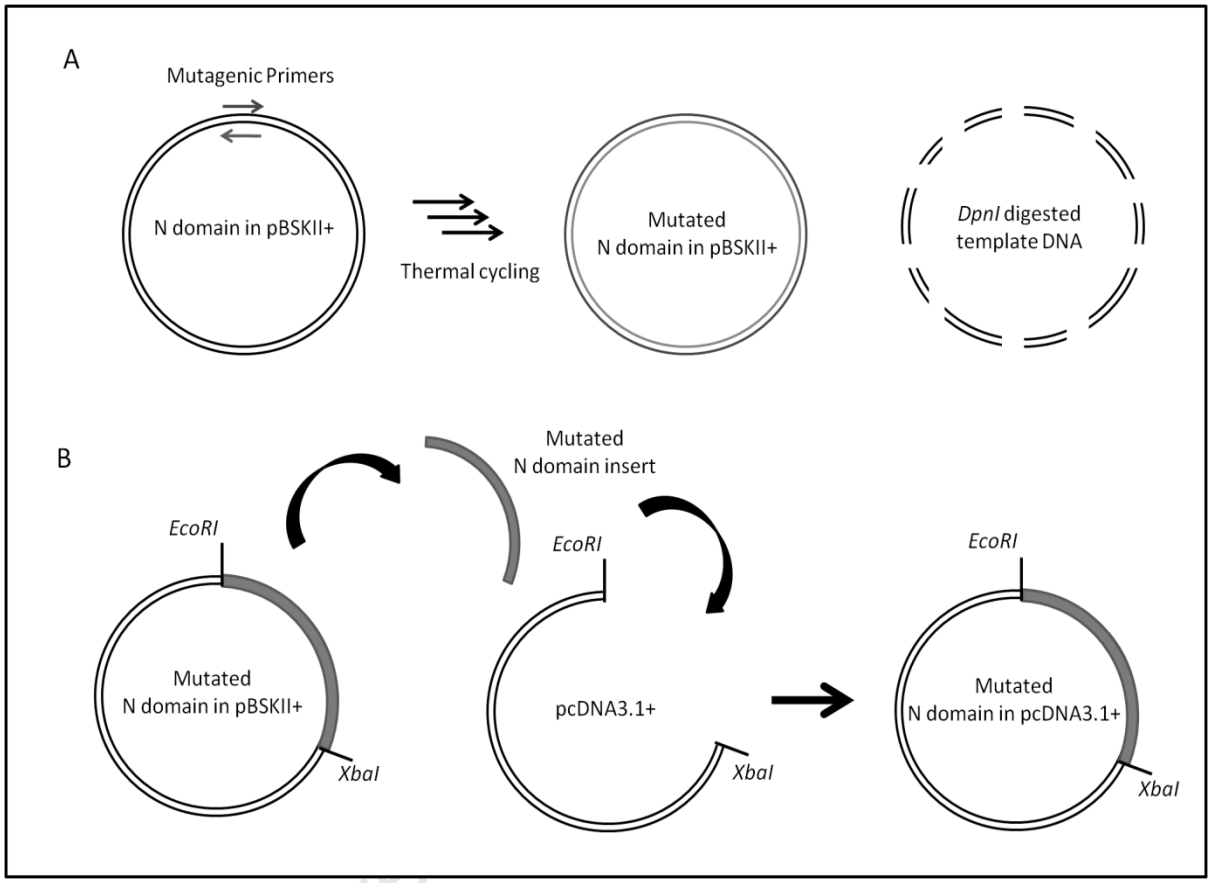


Figure 3.2: Schematic of the *Dpn I* method of site-directed mutagenesis (A) and the subcloning strategy (B). The *EcoRI* and *XbaI* restriction sites used for subcloning are indicated.

3.2.4.2 Crude isolation of plasmid DNA for screening of positive clones

Colonies were picked into 3 ml Luria broth (LB) containing 50 µl/ml ampicillin and incubated overnight at 37 °C with shaking. Plasmid DNA was isolated using the STET boiling-lysis method [150]. Overnight culture (1 ml) was transferred to an eppendorf tube and centrifuged for 2 min at 12000 rpm. The cell pellet was re-suspended in 250 µl of STET buffer *(8% (w/v) sucrose; 5% (w/v) triton X-100; 50 mM EDTA; 50 mM TRIS, pH 8.0)* containing 1 mg/ml lysozyme. Suspensions were boiled for 1 min and then centrifuged for 8 min. The pellet was removed with a sterile

toothpick and discarded. Isopropanol (0.25 volumes) was added to the remaining supernatant and mixed by inversion. Following addition of isopropanol, samples were centrifuged for 8 min to pellet the DNA. The supernatant was removed and the plasmid DNA pellet was air-dried. Finally, the DNA pellet was re-suspended in 1/10x tris-EDTA *(1.0 mM Tris-HCl, 0.1 mM EDTA, pH 7.4)* and heated to 68 °C for 10 min to ensure solubilisation of the DNA. Each DNA preparation was subjected to restriction enzyme (RE) digestion (see Appendix A3), to determine the presence of the introduced restriction site. Digest products were visualised on 1% (w/v) agarose gels.

3.2.4.3 Small scale preparation of plasmid DNA for sequencing and sub-cloning

Positive clones, identified by restriction enzyme digest screens, were picked into 2 ml LB containing 50 µg/ml ampicillin, and incubated overnight at 37 ˚C with shaking. Plasmid DNA isolation was carried out using the Zippy plasmid miniprep kit (Zymo research, USA) as per the manufacturer's instructions. The DNA pellet was re-suspended in 20 µl nuclease free H_2O (NF H_2O), and the concentration calculated from the absorbance at 260nm, using a Nanodrop® ND1000 (*NanoDrop* Technologies, Wilmington, USA). Successful mutagenesis was further confirmed by automated DNA sequencing carried out by Macrogen Inc. (Seoul, Korea). Sequencing primers shown in Appendix Figure A2.

3.2.4.3 Sub-cloning of N-domain constructs for mammalian transfection

Glycosylation mutant N-domain inserts were subcloned from the pBSKII cloning vector into the mammalian expression vector pCDNA3.1+ (Invitrogen, CA, USA), using the *EcoRI* and *XbaI* restriction sites (Figure 3.2 B). Glycosylation mutants in pBSKII, as well as the expression vector pcDNA3.1+, were digested with *EcoRI* and *XbaI* and DNA fragments separated by agarose gel electrophoresis. Relevant DNA fragments were excised using the Wizard® SV gel and clean-up kit (Promega) according to the manufacturer's instructions, resuspended in 50 µl NF H_2O and DNA concentrations were quantified using a Nanodrop® ND1000 as mentioned above. Ligations were carried out overnight at room temperature (RT), or over three days at 4 ˚C, using an insert to vector ratio of 3:1, under the following conditions: 1x T4

buffer (Fermentas, Burlington, Canada), X µl insert DNA, 3X µl pCDNA 3.1+, 1.5 U T4 ligase (Fermentas Burlington, Canada), made up to 20 µl with N.F. H_2O Competent DH5α cells were transformed with 10 µl ligation mixture. DNA from resistant clones was prepared using the STET method above. Prepared DNA was screened for the presence of the relevant restriction endonuclease sites as mentioned above.

3.2.4.4 Scaled up preparation of plasmid DNA for transfection

A 100 µl aliquot of bacterial culture, identified as positive for the desired mutation by the restriction digest screen, was inoculated into 50 ml LB containing 50 µg/ml ampicillin, and incubated overnight at 37 ˚C with shaking. Bulk plasmid DNA isolation was carried out using the Qiagen Midi-preparation kit (GmBH, Hilden, Germany), as per the manufacturer's instructions. The DNA pellet was re-suspended in 100 µl NF H_2O, and the concentration calculated from the absorbance at 260nm, using a Nanodrop® ND1000. Isolated DNA was screened for the presence of the relevant restriction endonuclease site introduced during mutagenesis, as mentioned above.

3.3 Results

Previous work has highlighted the role of C-terminal glycosylation sites in the expression of active N-domain proteins [125]. That study produced a number of N-domain glycosylation mutants (Figure 3.1), most of which were inactive due to the fact that glycosylation sites were sequentially removed from the C-terminus (glycosylation of the C-terminus was subsequently found to be important for the expression of active N-domain). Rather than investigate the importance of the C-terminal glycan sites by sequentially removing the *N*-linked sequons from the N-terminus, it seemed more efficient to make use of the previously generated glycosylation mutants, using these as templates for reintroducing the C-terminal glycan sites via site-directed mutagenesis. Primarily making use of the previously generated N-domain glycosylation variants Ndom123456, Ndom123 and Ndom0 (which has no intact *N*-linked glycosylation sites), a range of additional N-domain glycoforms were proposed (Figure 3.3).

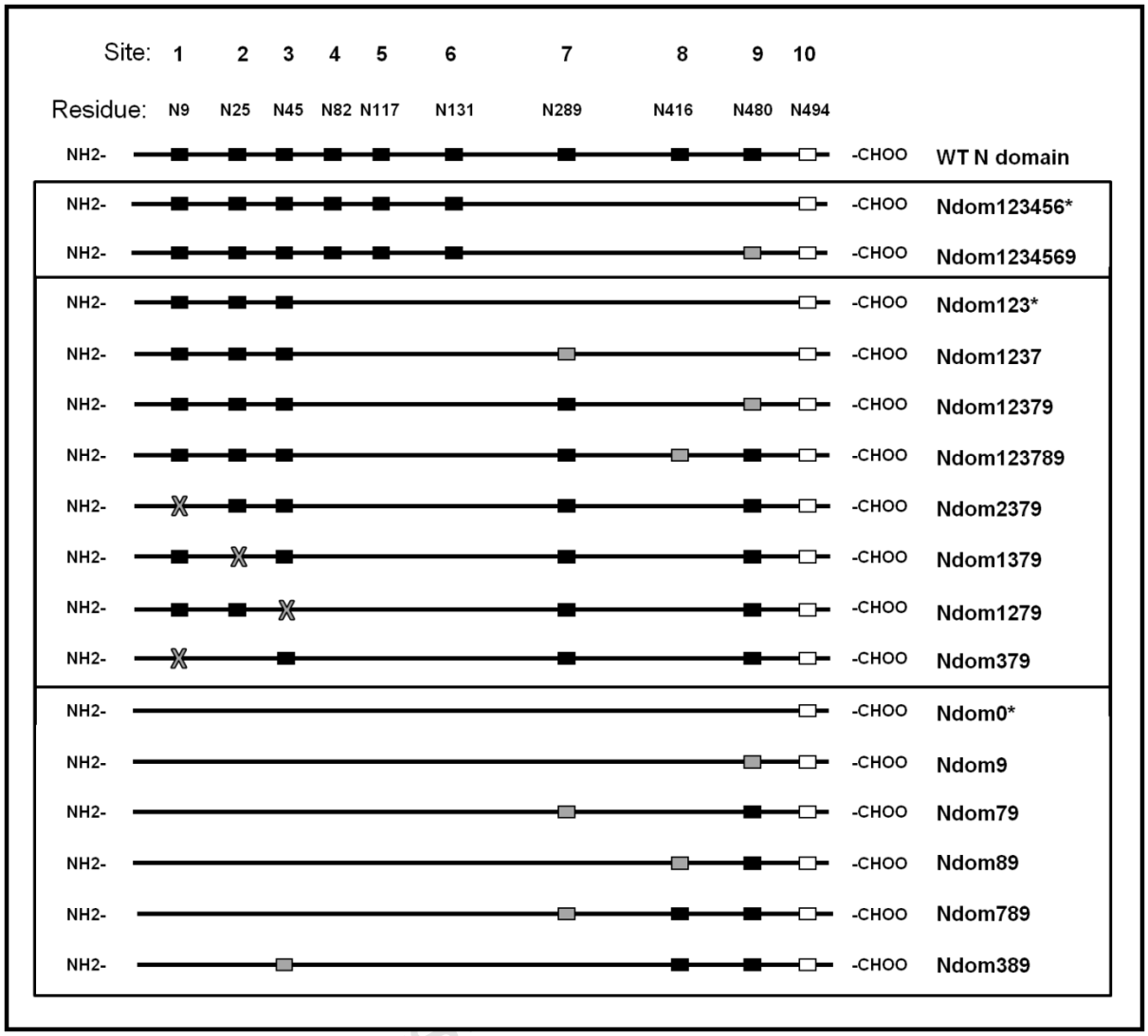


Figure 3.3: Schematic representation of constructed glycosylation mutants. Black boxes indicate intact glycosylation sequons, Grey boxes show reintroduced sequons, Grey crosses indicate removed sequons. Site 10 (white box) is not glycosylated. "*" indicates previously generated mutants [125] used as templates. Mutants are grouped according to the template they were constructed from.

Previous mutagenic studies have shown that Ndom1234567 is expressed as an active protein, while Ndom123456 is not. These data suggesting that site 7 is important for the expression of active N-domain protein. In order to assess whether other C-terminal sites were important, we decided to reintroduce site 9 into the N-domain variant Ndom123456, to see whether the presence of site 9 could rescue the inactive mutant. Thus, site 9 was introduced into Ndom123456 to create Ndom1234569 (as in Figures 3.1 and 3.3, the numbers indicate the intact *N*-glycosylation sites), in the process, the *MluI* restriction site at position 2491 was removed. Positive clones were identified by digestion with *MluI* and *EcoRI*, where the

loss of 5.1 and 1.5 kb DNA fragments and the formation of a 6.7 kb fragment is evidence of the removal of the *MluI* site and thus incorporation of the desired Glu to Asn mutation (Figure 3.4).

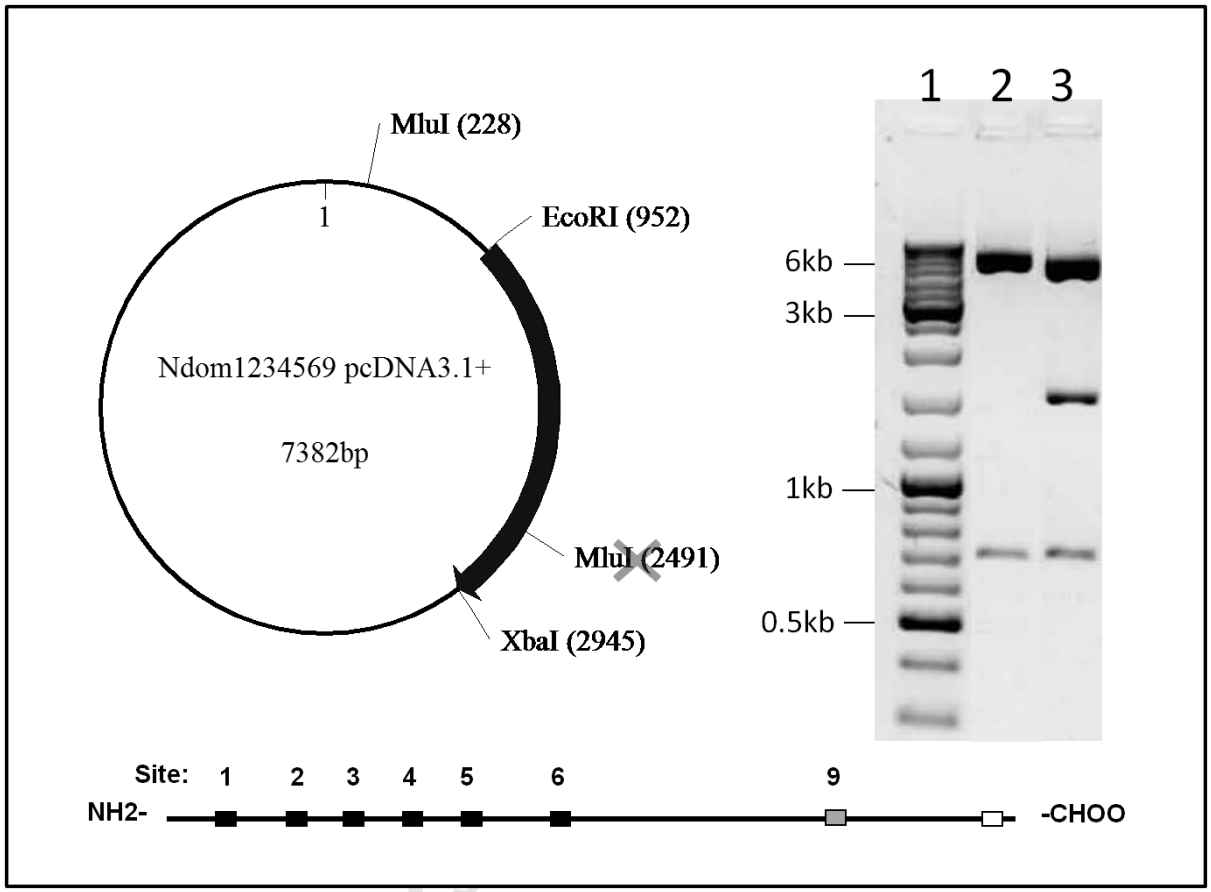


Figure 3.4: Vector map and restriction endonuclease digest of Ndom1234569, generated from Ndom123456. Digested with *MluI* and *EcoRI*. The RE site *MluI* (2491) is removed during mutagenesis. Lane 1 - O'generuler DNA ladder mix, Lane 2 - Ndom1234569, Lane 3 - Ndom123456 (negative control).

Given the suggested importance of the C-terminal site 7, it was decided to introduce this site into the hypoglycosylated N-domain variant Ndom123 to generate Ndom1237, and assess whether the presence of site 7 could rescue activity for this variant. An additional *BsmI* site at position 1929 was incorporated during mutagenesis. Given that there were two existing *BsmI* sites present in the template, digestion with *BsmI* alone was sufficient to identify positive clones. The loss of a DNA fragment at 3.9 kb and the formation fragments at 3.2 and 0.7 kb, after *BsmI* digestion, represents successful incorporation of the desired mutation (Figure 3.5).

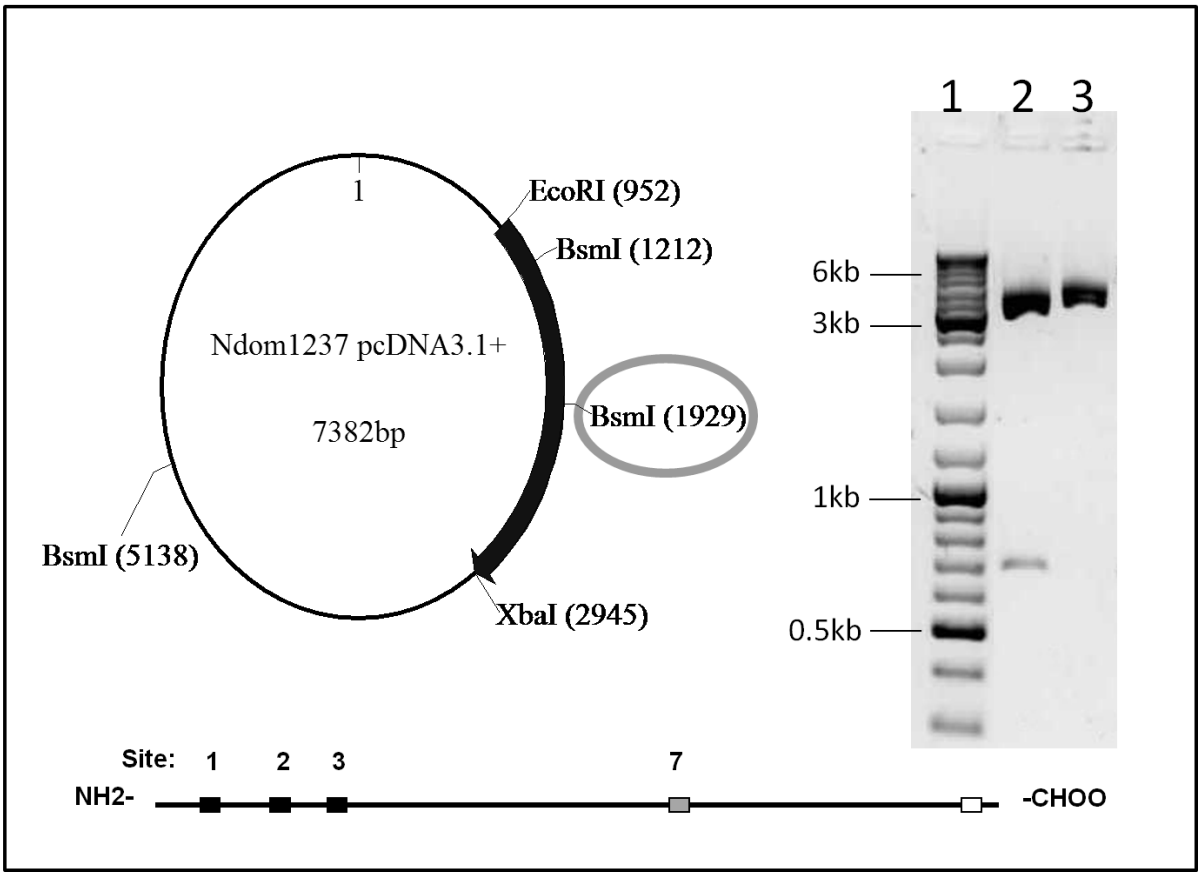


Figure 3.5: Vector map and restriction endonuclease digest of Ndom1237, generated from Ndom123. Digested with *BsmI*. The RE site *BsmI* (1929) is introduced during mutagenesis. Lane 1 -O'generuler DNA ladder mix, Lane 2 -Ndom1237, Lane 3 - Ndom123.

In the event that site 7 alone is not sufficient to rescue activity for the Ndom1237 mutant, it was decided to introduce site 9 into Ndom1237. In this way Ndom12379 was generated. An *MluI* restriction enzyme site, at position 2491, was removed during mutagenesis, thus the loss 5.1 and 1.5 kb DNA fragments and the formation of a 6.7 kb fragment, after digestion with *MluI* and *EcoRI*, indicates successful mutagenesis (Figure 3.6).

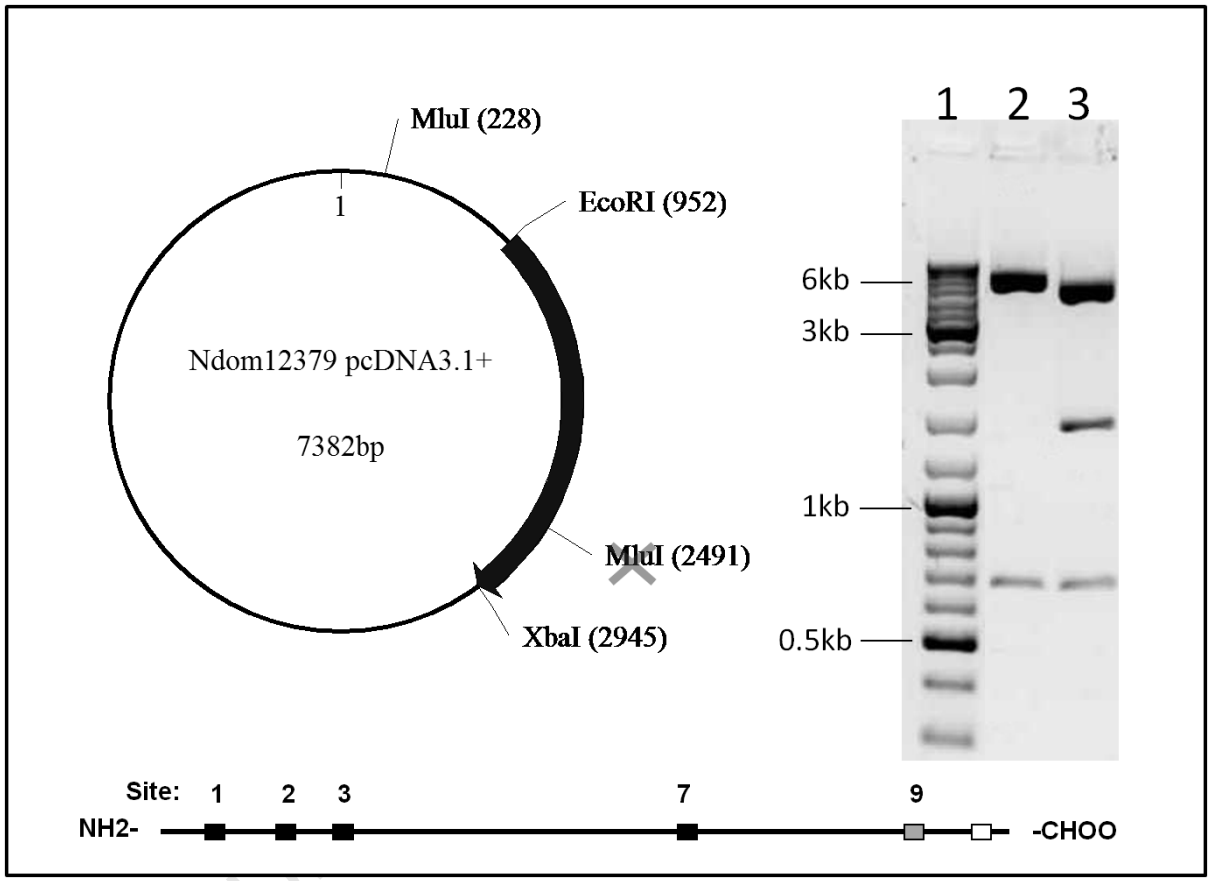


Figure 3.6: Vector map and restriction endonuclease digest of Ndom12379, generated from Ndom1237. Digested with *MluI*. The RE site *MluI* (2491) is removed during mutagenesis. Lane 1 - O'generuler DNA ladder mix, Lane 2 - Ndom12379, Lane 3 - Ndom1237.

To ensure that all of the C-terminal sites were investigated, site 8 was introduced into Ndom12379 to generate Ndom123789. A *PvuI* site at position 2295 was removed during mutagenesis. An additional *PvuI* site is present in the template DNA, thus removal of the site at position 2295 results in the loss of a fragments at 4.5 and 2.8 kb and the formation of a new fragment of 7.3 kb, consistent with linearisation of the vector (Figure 3.7).

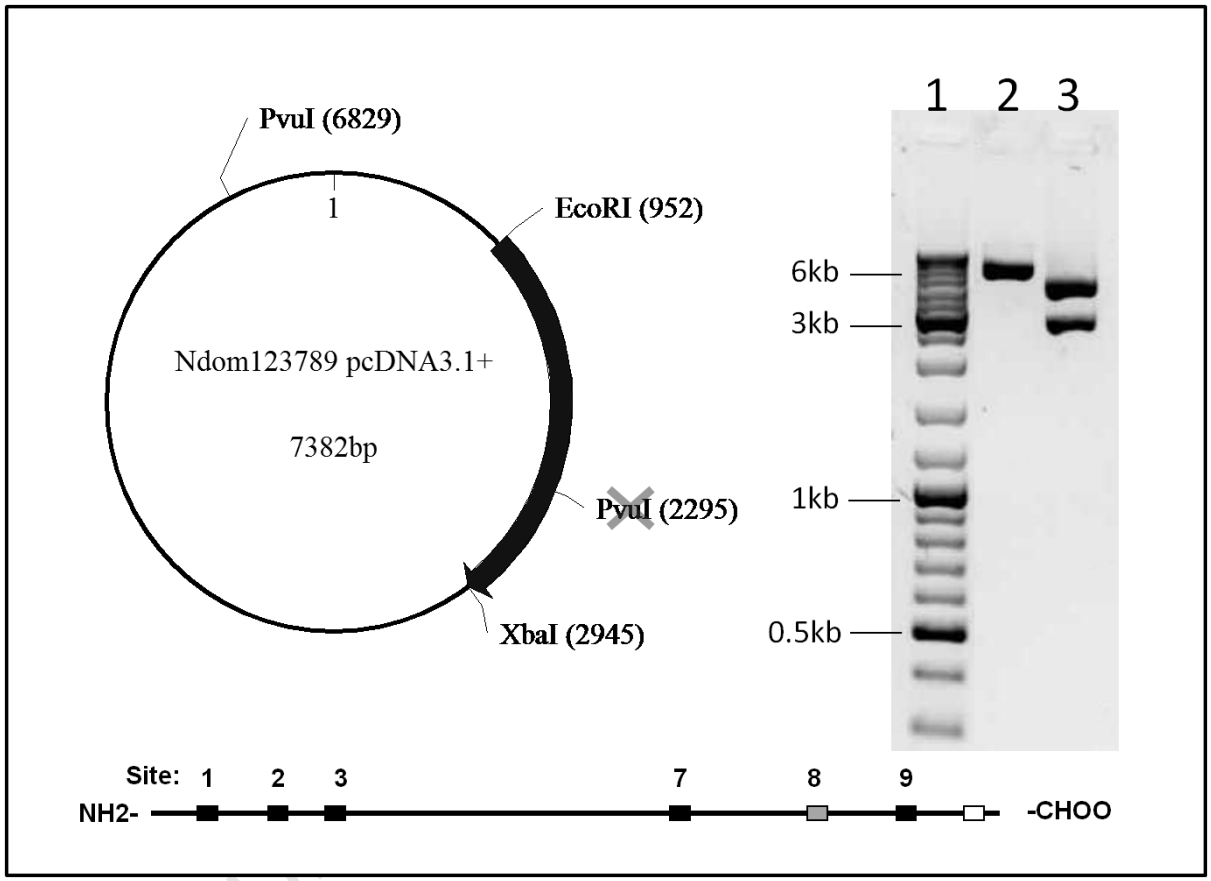


Figure 3.7: Vector map and restriction endonuclease digest of Ndom123789, generated from Ndom12379. Digested with *PvuI*. The RE site *PvuI* (2295) is removed during mutagenesis. Lane 1 - O'generuler DNA ladder mix, Lane 2 - Ndom123789, Lane 3 - Ndom12379.

While C-terminal sites were previously shown to be important for the expression of N-domain protein, the importance of the N-terminal sites (or the lack thereof) cannot be taken for granted. Thus, the three N-terminal sites were removed from Ndom12379 individually, to assess their relative importance for the expression of active N-domain protein.

Site 1 was removed from Ndom12379 to create Ndom2379, where a *BalI* site at position 1082 was introduced during mutagenesis. Thus, digestion with *XbaI* and *BalI* results in the loss of a fragment at 6 kb and the formation of fragments at 4.2 and 1.8 kb in positive clones (Figure 3.8).

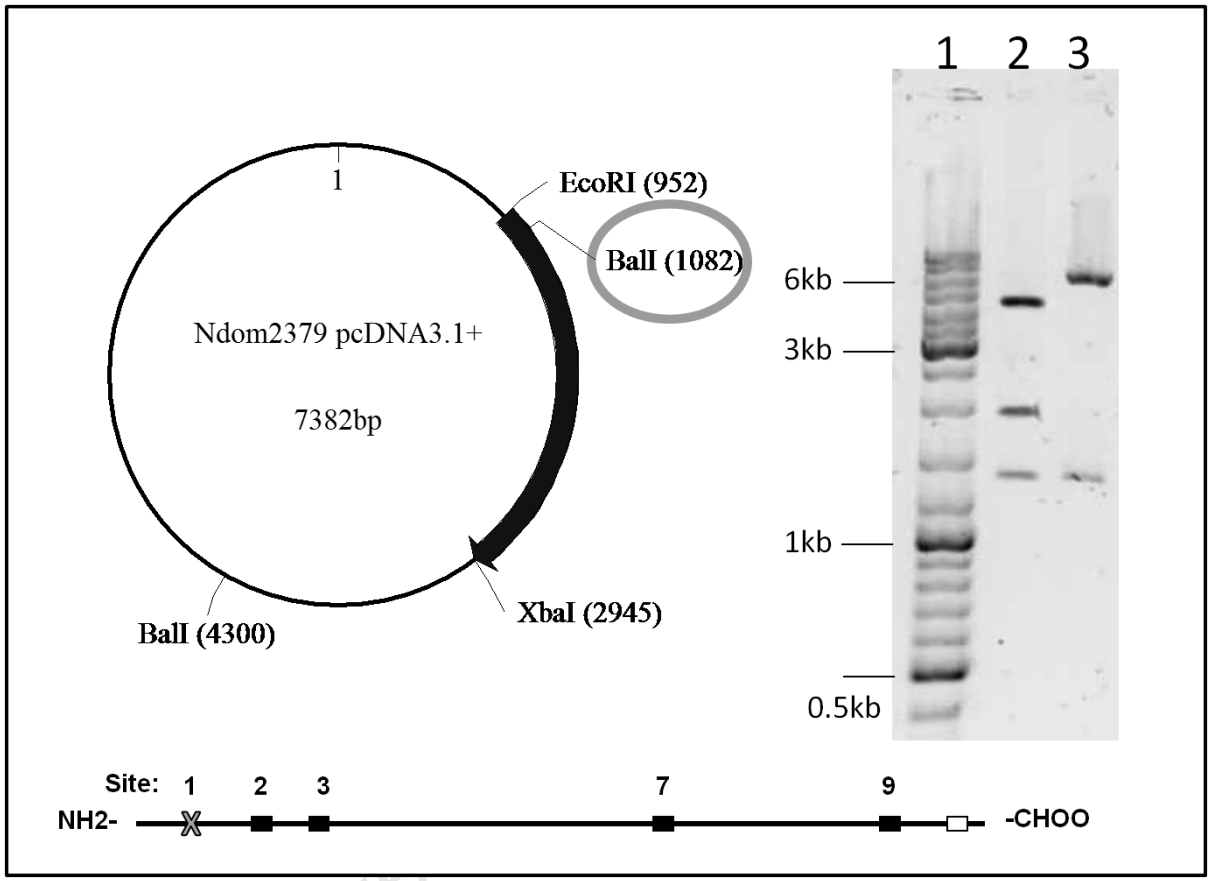


Figure 3.8: Vector map and restriction endonuclease digest of Ndom2379, generated from Ndom12379. Digested with *XbaI* and *BalI*. The highlighted RE site *BalI* (1082) is introduced during mutagenesis. Lane 1 - O'generuler DNA ladder mix, Lane 2 - Ndom2379, Lane 3 - Ndom12379.

Site 2 was removed from Ndom12379 to create Ndom1379. An *Eco47III* site at position 1140 was introduced for the screening of positive clones. Digestion with *XbaI* and *Eco47III* results in the loss of a fragment at 7.4 kb and the formation of fragments at 5.6 and 1.8 kb (Figure 3.9) indicating successful mutagenesis.

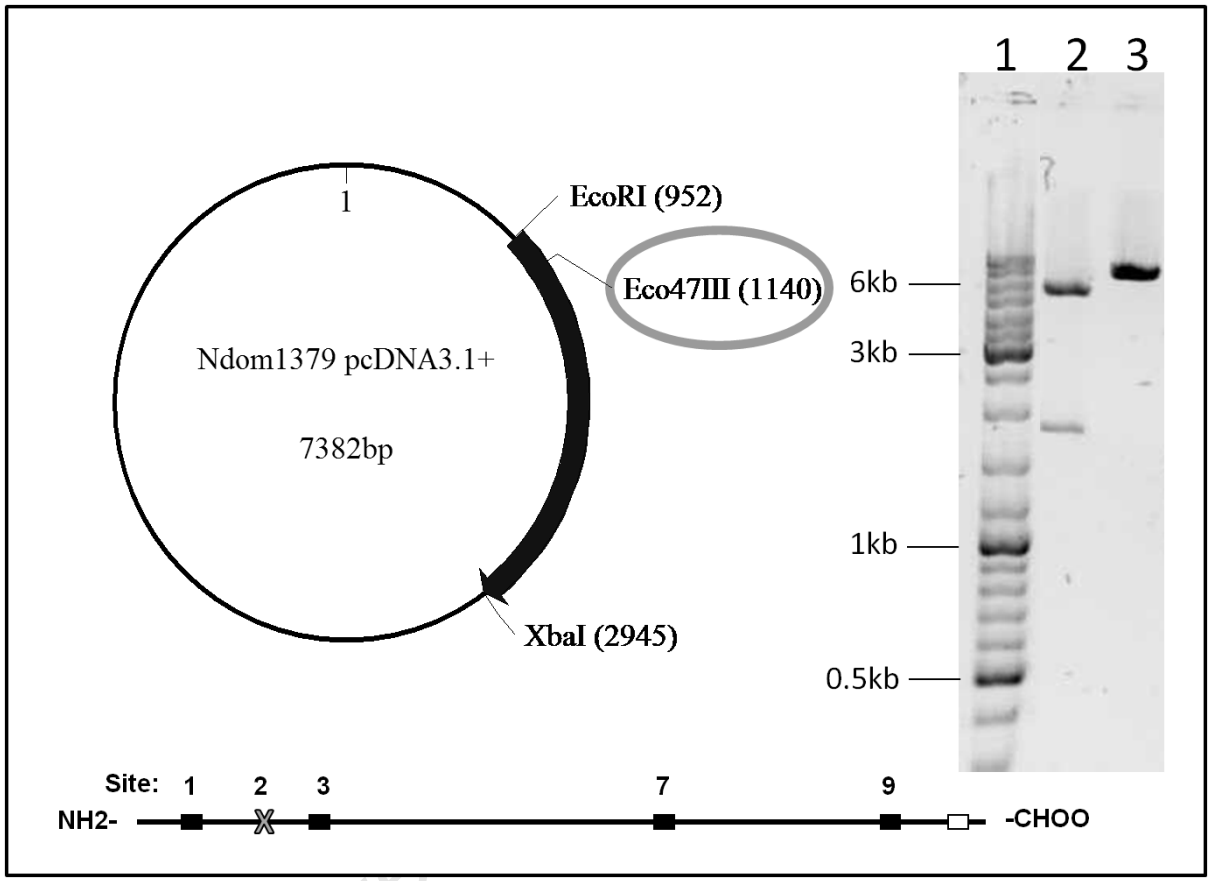


Figure 3.9: Vector map and restriction endonuclease digest of Ndom1379, generated from Ndom12379. Digested with *XbaI* and *Eco47III*. The highlighted RE site *Eco47III* (1140) is introduced during mutagenesis. Lane 1 - O'generuler DNA ladder mix, Lane 2 - Ndom1379, Lane 3 - Ndom12379.

Site 3 was removed from Ndom12379 (N45Q) to generate Ndom1279. An additional *RcaI* site at position 1182 was introduced during mutagenesis. Given that there were two existing *RcaI* sites present in the template, digestion with *RcaI* alone was sufficient to identify positive clones. The loss of DNA fragments at 6.4 kb and the formation fragments at 5.1 and 1.3 kb, after *RcaI* digestion, represents successful incorporation of the desired mutation (Figure 3.10).

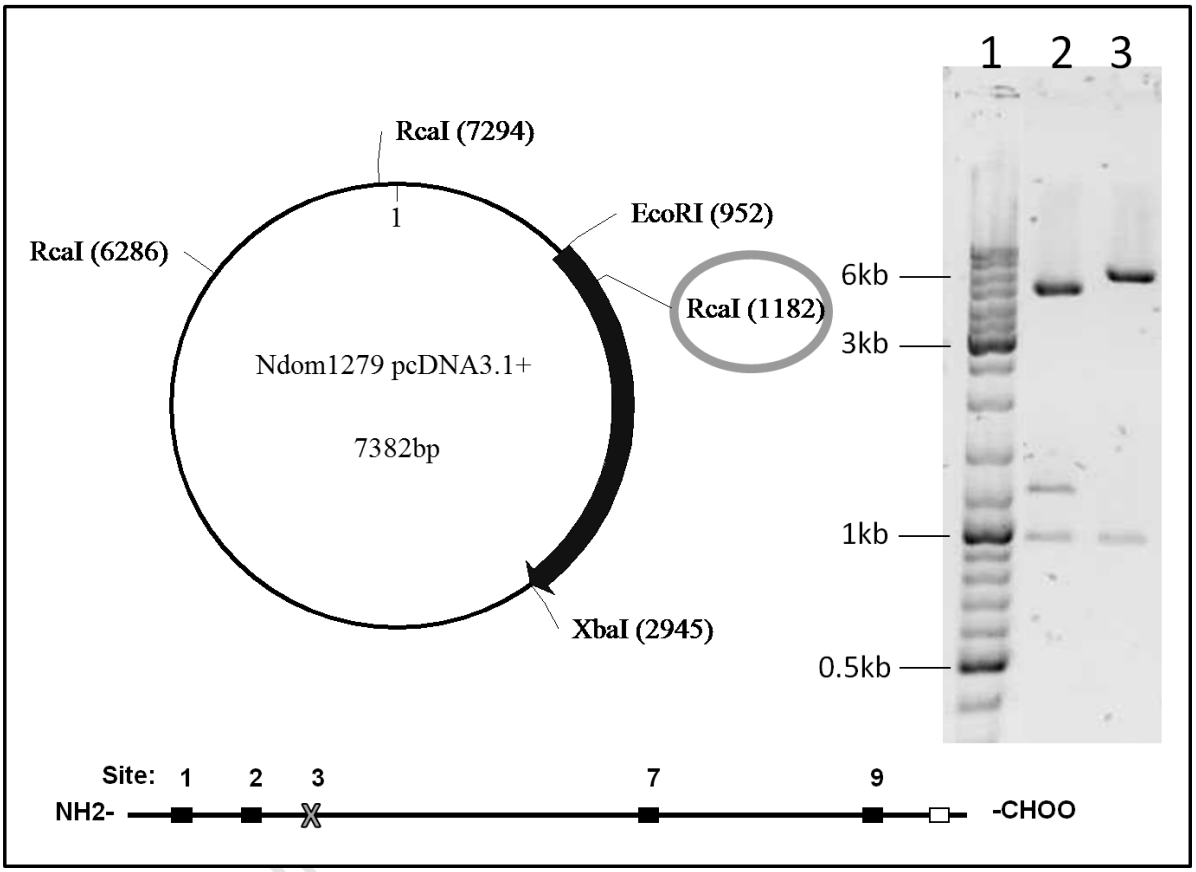


Figure 3.10: Vector map and restriction endonuclease digest of Ndom1279, generated from Ndom12379. Digested with *RcaI*. The highlighted RE site *RcaI* (1182) is introduced during mutagenesis. Lane 1 - O'generuler DNA ladder mix, Lane 2 - Ndom1279, Lane 3 - Ndom12379.

To generate a minimally glycosylated N-domain variant, we made used of an N-domain construct with all glycosylation sites removed (Ndom0). Ndom0 was thus used as a template for the reintroduction of various glycosylation sites, including combinations of the three C-terminal sites 7, 8 and 9.

Thus, site 9 was introduced into Ndom0 to create Ndom9. As in the case of Ndom1234569 and Ndom12379 above, the loss of 5.1 and 1.5 kb DNA fragments and the formation of a 6.7 kb fragment, after digestion with *MluI* and *EcoRI*, indicates successful mutagenesis (Figure 3.11).

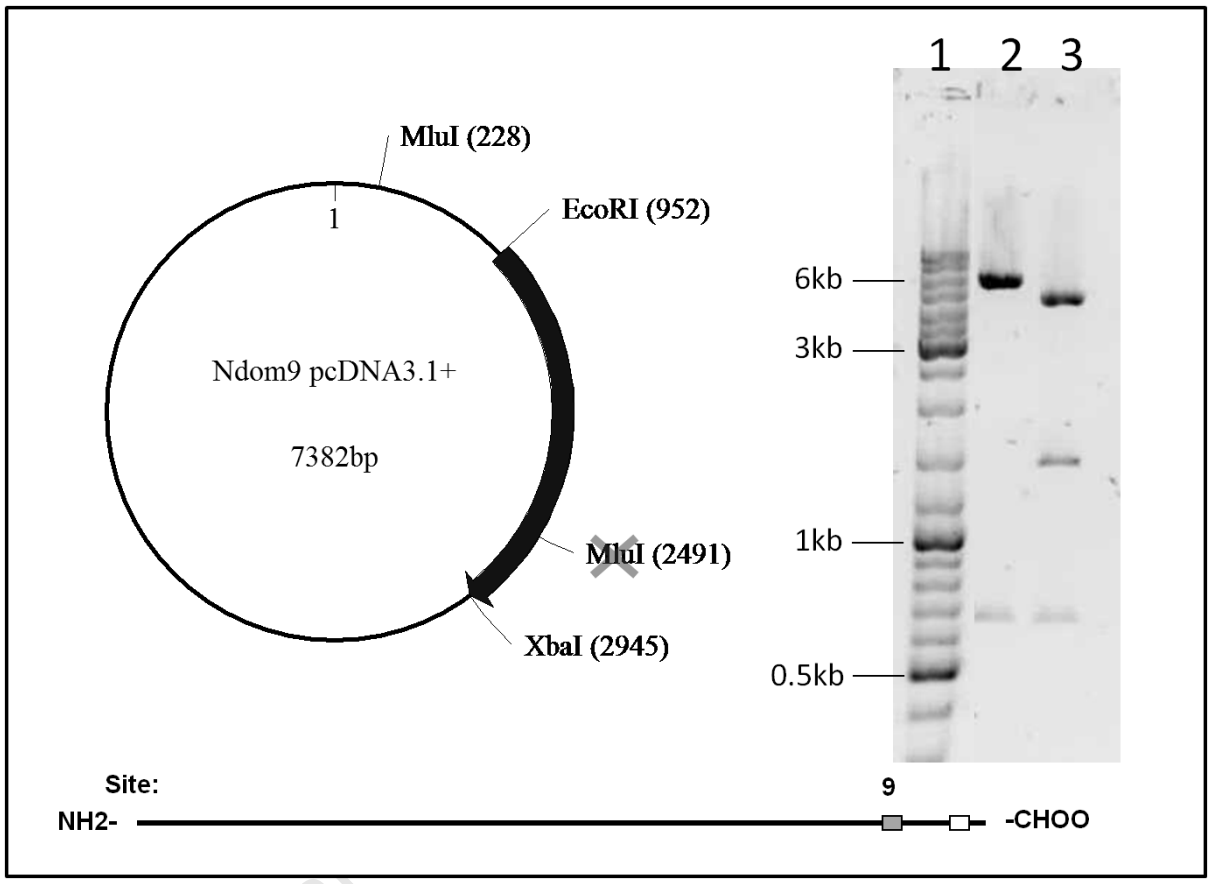


Figure 3.11: Vector map and restriction endonuclease digest of Ndom9, generated from Ndom0. Digested with *EcoRI* and *MluI*. The highlighted RE site *MluI* (2491) is removed during mutagenesis. Lane 1 - O'generuler DNA ladder mix, Lane 2 - Ndom9, Lane 3 - Ndom0.

Site 7 was introduced into Ndom9 to generate Ndom79. As for Ndom1237 the loss of a DNA fragment at 3.9 kb and the formation fragments at 3.2 and 0.7 kb, after digestion with *BsmI*, represents successful incorporation of the desired mutation (Figure 3.12).

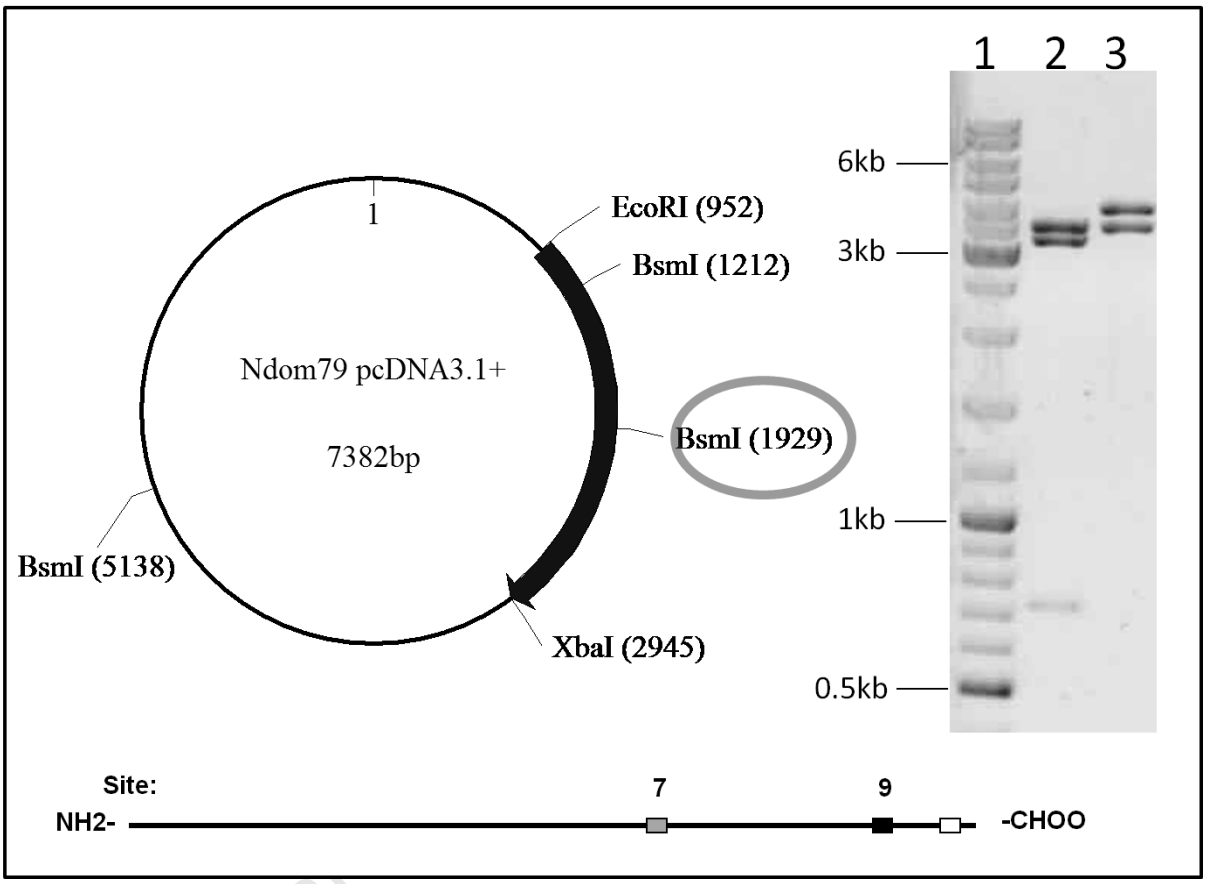


Figure 3.12: Vector map and restriction endonuclease digest of Ndom79, generated from Ndom9. Digested with *BsmI*. The highlighted RE site *BsmI* (1929) is introduced during mutagenesis. Lane 1 - O'generuler DNA ladder mix, Lane 2 - Ndom79, Lane 3 - Ndom9.

Site 8 was introduced into Nom79 (Q416N) to generate Ndom789. As for Ndom123789, the removal of the *PvuI* site results in the loss of a fragments at 4.5 and 2.8 kb and the formation of a new fragment of 7.4 kb (Figure 3.13).

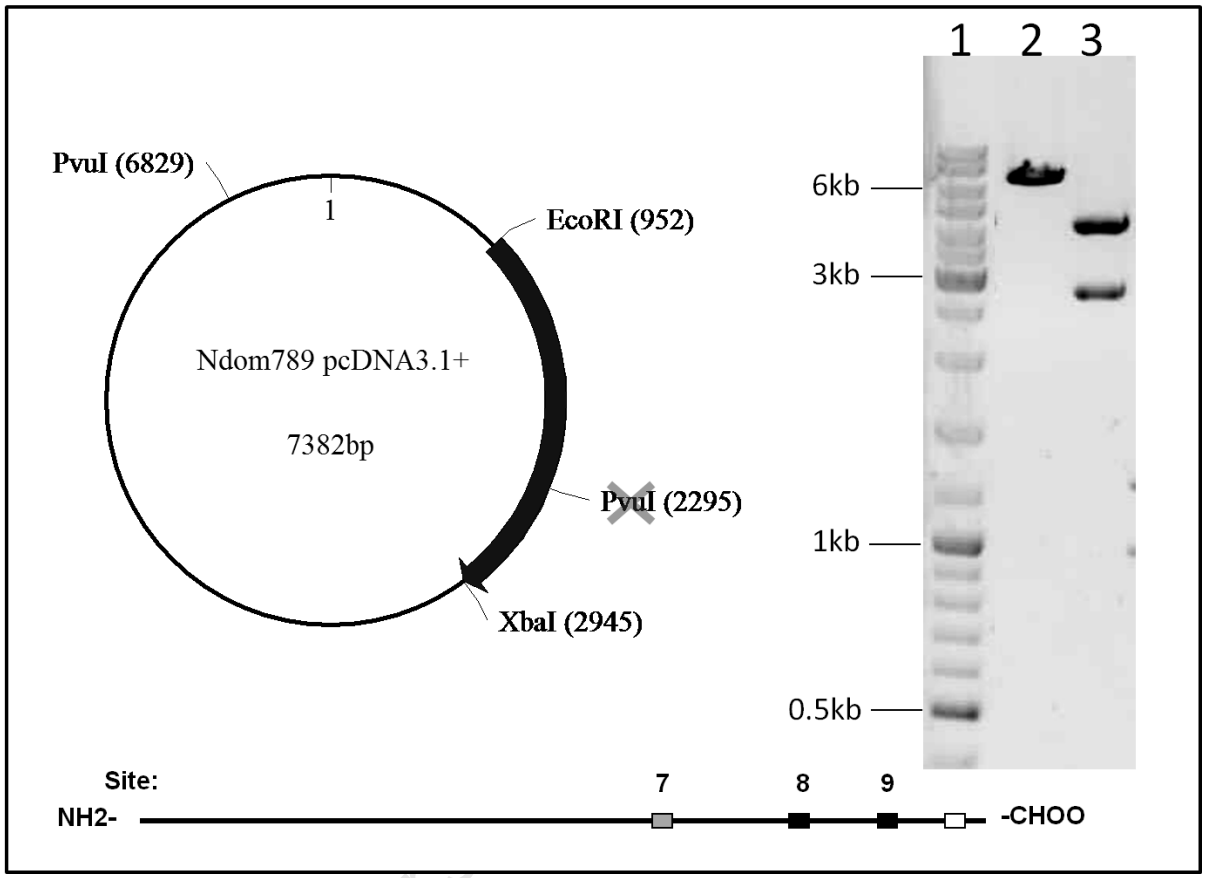


Figure 3.13: Vector map and restriction endonuclease digest of Ndom789, generated from Ndom79. Digested with *PvuI*. The RE site *PvuI* (2295) is removed during mutagenesis. Lane 1 - O'generuler DNA ladder mix, Lane 2 - Ndom789, Lane 3 - Ndom79.

Similarly, site 8 was introduced into Nom9 (Q416N) to generate Ndom89. As in the case of Ndom123789 and NDom789 above, the removal of the *PvuI* site results in the loss of a fragments at 4.5 and 2.8 kb and the formation of a new fragment of 7.4 kb (Figure 3.14).

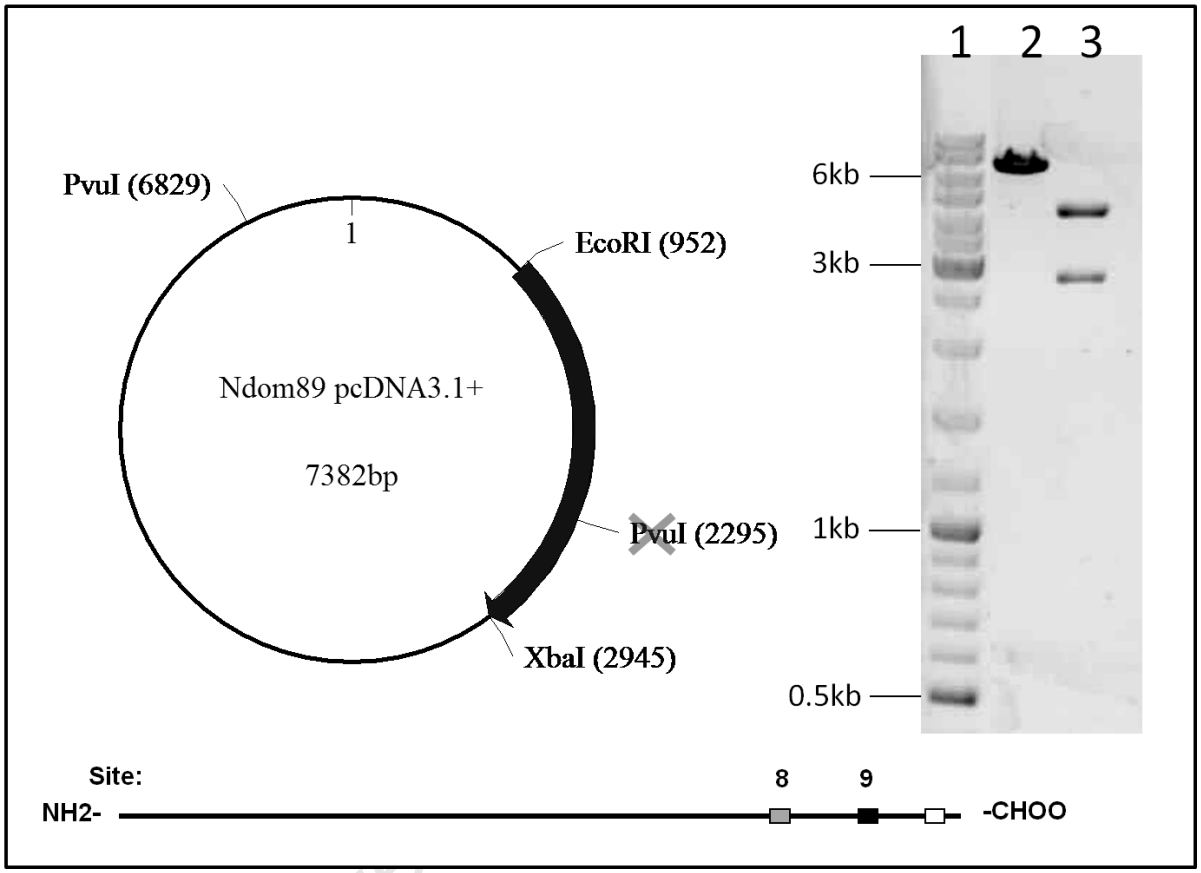


Figure 3.14: Vector map and restriction endonuclease digest of Ndom89, generated from Ndom9. Digested with *PvuI*. The RE site *PvuI* (2295) is removed during mutagenesis. 1-O'generuler DNA ladder mix, 2-Ndom89, 3-Ndom9.

To assess whether an N-terminal site may be important for the expression of a minimally glycosylated N-domain form, Ndom389 was constructed. The construction of Ndom389 involved subcloning the N-terminal portion of a mutant containing only site 3 at the N-terminus. The previously generated mutant Ndom3 could unfortunately not be used for this purpose as it was found to contain a spurious coding mutation, W40L. Thus, Ndom379 was prepared by removing site 1 from Ndom1379 (as per the generation of Ndom2379 above). The N-terminal fragment of Ndom379 was then subcloned into Ndom89 using the RE sites *BglII* and *EcoRI* (Figure 3.15).

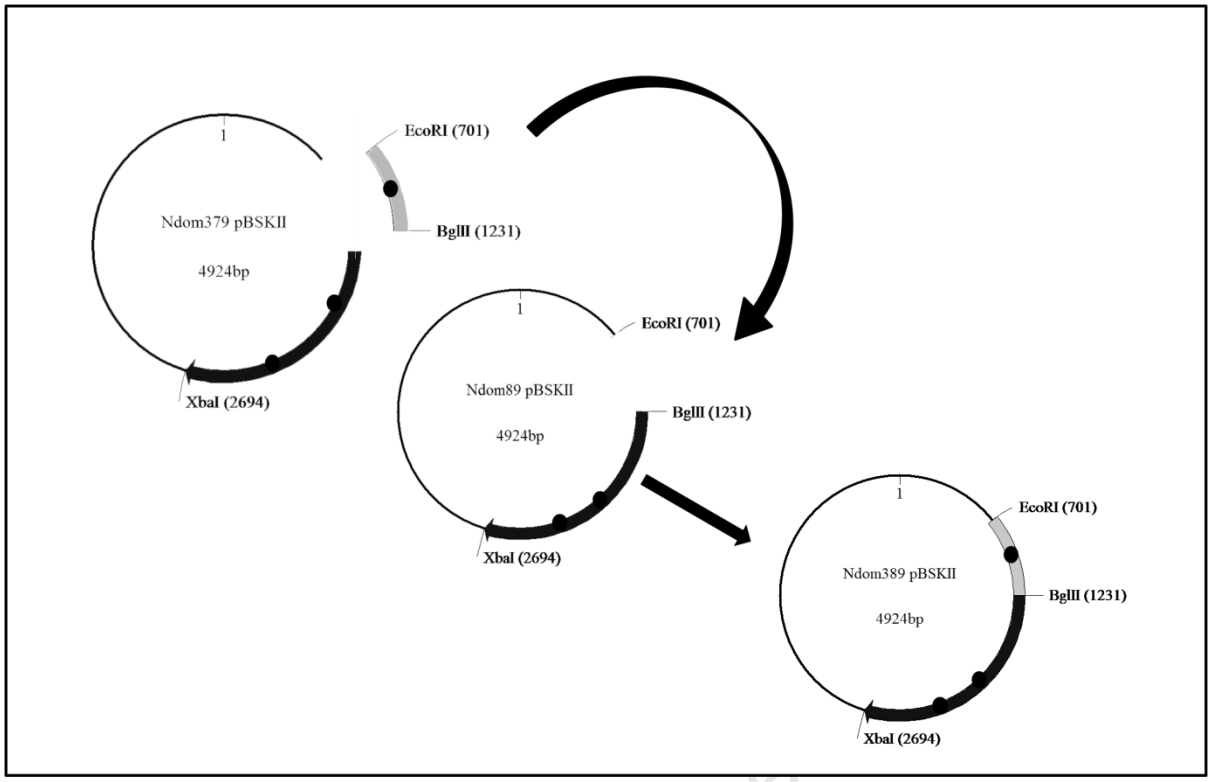


Figure 3.15: Cloning strategy for the generation of Ndom389, from Ndom379 and Ndom89. Relevant RE sites are indicated, black dots represent intact glycosylation sites.

Since Ndom89 had site 3 removed during earlier mutagenesis, the *RcaI* restriction site at position 1182 is present. Ndom379 however, does not have this site. Thus, successful subcloning of site 3 into Ndom89 could be confirmed by the absence of this *RcaI* site. Positive clones were therefore identified by digestion with *RcaI*, where the loss of 5.1 and 1.3 kb DNA fragments and the formation of a 6.4 kb fragment is evidence of the absence of the relevant *RcaI* site and thus the presence of an intact glycosylation sequon at site three (Figure 3.16).

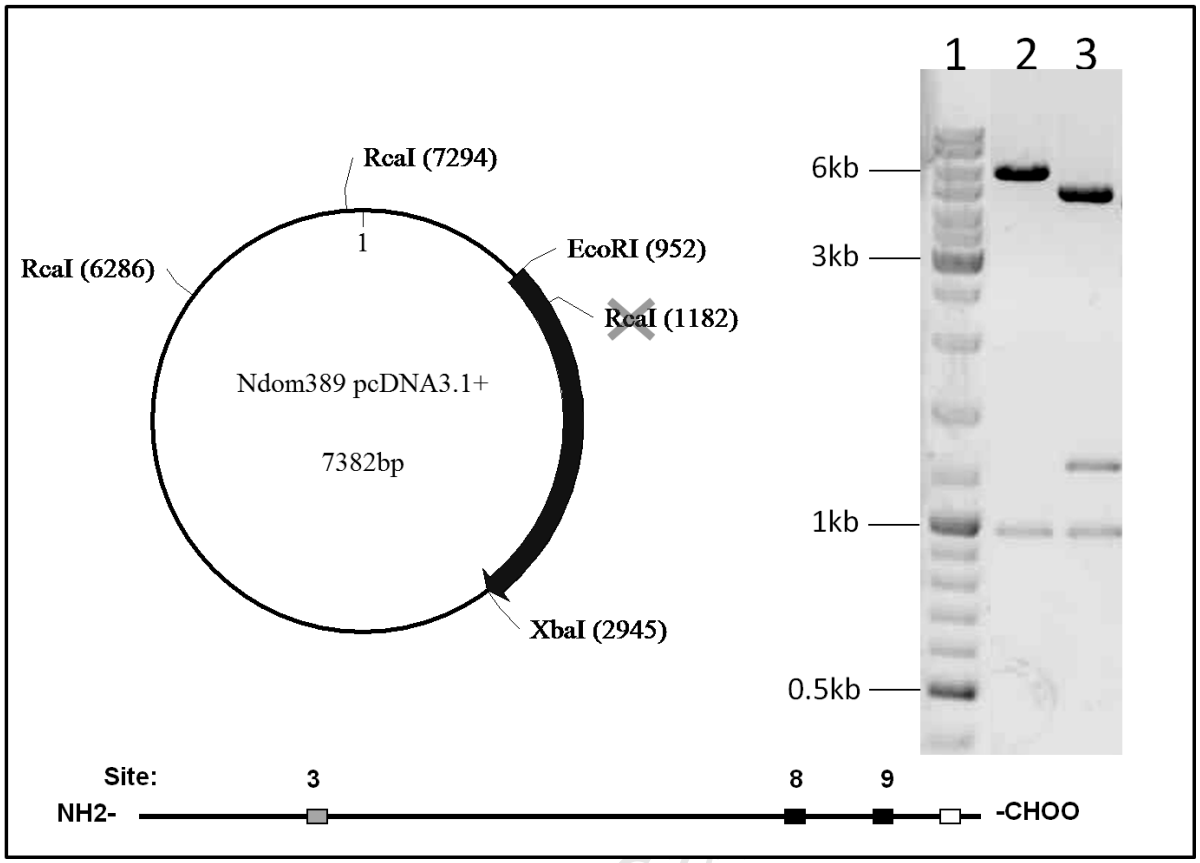


Figure 3.16: Vector map and restriction endonuclease digest of Ndom389, generated from Ndom89 and Ndom1379. Digested with *RcaI*. The highlighted RE site *RcaI* (1182) is eliminated from Ndom89 during subcloning. Lane 1 - O'generuler DNA ladder mix, Lane 2 - Ndom389, Lane 3 - Ndom89.

3.4 Discussion

The aim of this work was to produce a series of N-domain forms with differing levels of intact glycosylation sites. The elimination of a glycosylation site can be achieved by the introduction of a conservative Asn to Gln substitution at the Asn of the Asn-Xaa-Ser/Thr glycosylation consensus sequence. While there are many approaches to site-directed mutagenesis, three polymerase chain reaction (PCR) based approaches are most commonly used, namely overlap extension, megaprimer and modified forms of the Quickchange™ method [151]. Both the megaprimer and overlap extension methods require two rounds of PCR [152,153]. The Quickchange™ method, and its modified forms, require only a single PCR reaction [149,154]. In our approach, the newly generated mutant was used as a template for subsequent mutagenesis of the glycosylation sites in an iterative cycle. Thus, the single step method of the Quickchange™ protocol offers a more streamlined approach to the generation of

multiple mutant constructs. Therefore it was decided that this method was best suited for our study.

In the case of the C-domain, the rationale for which sites were to be mutated was guided by previous work which showed that the three N-terminal sites of the C-domain were always glycosylated, while the following three sites were glycosylated with only partial efficiency [124]. These findings suggested a preference for N-terminal glycosylation, which was later confirmed by mutagenesis of the glycosylation sites [122]. Profiling of the glycosylation site occupancy of the N-domain did not reveal a similar pattern, with all of the sites containing glycosylation, excepting site 10 (see Chapter 2.3). Thus, no inferences could be made as to the relative importance of the different *N*-linked glycosylation sites on the N-domain, except that site 10 is not of importance.

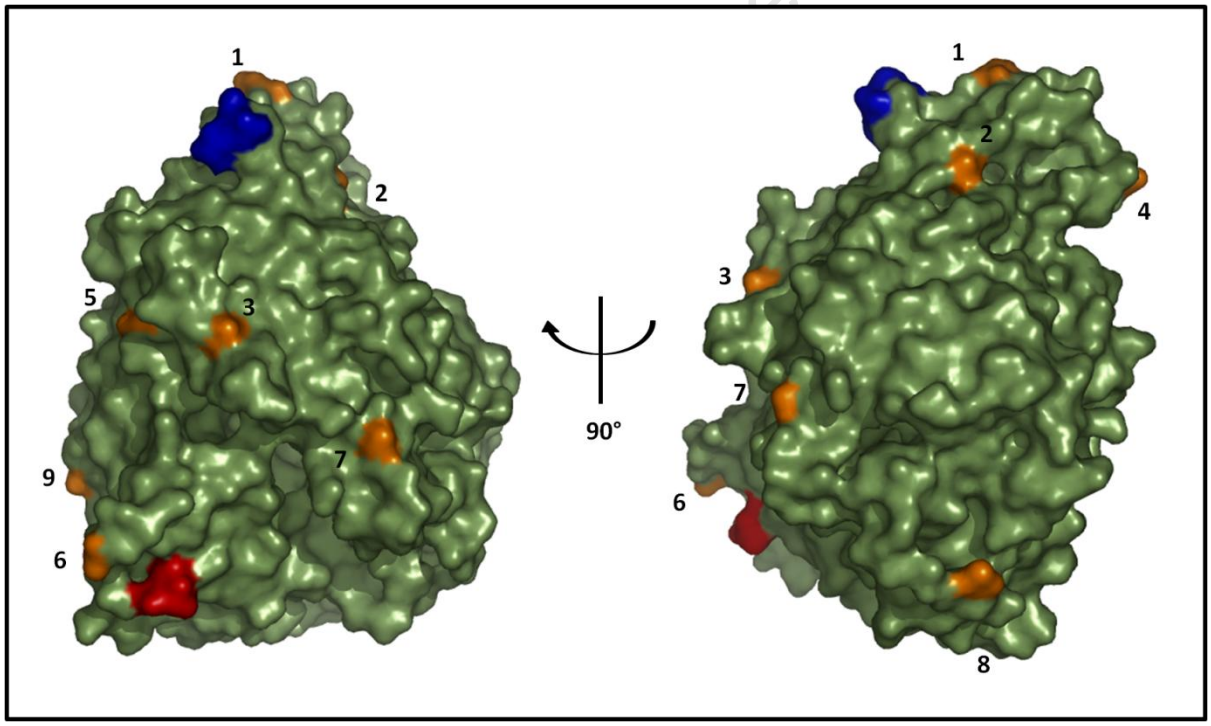


Figure 3.17: Surface representation of the N-domain of ACE, showing the location of the different N-linked glycosylation sites (orange). N- and C-termini are labelled in blue and red respectively. Glycosylation sites are numbered. Figure was drawn with PyMOL v0.92 using the PDB structure 2C6N [139] (DeLano Scientific, San Carlos, CA, USA).

Analysis of the N-domain crystal structure showed that the different *N*-linked sites were widely distributed across the surface of the protein (Figure 3.17). However, the

N-terminal sites 1 to 4 were found to be located on the lid helices, which are thought to play an important role in the proposed hinge movement (discussed further in Chapter 4.4.2). Therefore, initial attempts to generate glycosylation mutants of the N-domain were carried out as per the C-domain approach, with the sequential removal of glycan sites from the C-terminus [125]. Results from that initial study revealed that, contrary to what was seen in the C-domain, the C-terminal sites were important for the expression of active N-domain protein. Given these findings, the primary focus of our study was to generate a range of hypoglycosylated mutants containing various degrees of glycosylation at the C-terminus. In addition, a number of mutants with reduced N-terminal glycosylation were generated to assess the relative contribution of these sites on the expression and thermal stability of the N-domain.

In total 13 mutants were successfully generated (Figure 3.3), of which 11 were subsequently used for further investigation, while two constructs, Ndom9 and Ndom379, were used as intermediates in the generation of the different glycosylation mutants, and were not expressed in CHO cells. The various hypoglycosylated mutants can be grouped into three categories, namely: mutants aimed at establishing the role of C-terminal sites (Ndom1234569, Ndom1237, Ndom12379 and Ndom123789); mutants designed to evaluate the role of N-terminal sites (Ndom2379, Ndom1379 and Ndom1279) and mutants designed to produce minimally glycosylated N-domain variants (Ndom79, Ndom89, Ndom789 and Ndom389).

Chapter 4: Characterisation of N-domain glycosylation variants

4.1 Introduction

Protein *N*-glycosylation is known to affect a number of properties, including protein folding, processing and thermal stability [50,67,155]. The expression of functional ACE protein requires the presence of *N*-glycosylation. This is evidenced by the fact that ACE isoforms expressed in bacterial systems which lack glycosylation machinery, and ACE expressed in the presence of tunicamycin, which blocks all *N*-linked glycosylation, are inactive and rapidly degraded [112]. It has been determined that only one or two of the N-terminal glycosylation sites are required for the expression of functional C-domain [122]. The glycosylation requirements for the expression of the N-domain have not been determined, although it is known that C-terminal glycosylation is important [125].

Glycosylation was found to play a significant role in conferring thermal stability to the C-domain of ACE, where a series of glycosylation mutants showed that decreasing levels of glycosylation correlated with a decrease in thermal stability [136]. While glycosylation of the N-domain is known to affect its stability [136], little is known about the glycan sites involved and the extent to which glycosylation affects thermal stability. However, previous work has shown that the N-domain is significantly more thermostable than the C-domain, with a T_m of 72 °C, as compared with 55 °C for the C-domain [104,136].

There are various methods for assessing the level of protein thermal stability, including quantitative approaches such as differential scanning calorimetry (DSC), circular dichroism (CD), and the more qualitative assessment of residual activity following various lengths of exposure to thermal denaturation. CD spectroscopy is particularly useful in that it reveals signature profiles for protein secondary structures such as α-helix, β-sheet and random coil (unfolded) (Figure 4.1). Thus, performing CD spectroscopy over a temperature gradient allows one to monitor the unfolding

transition from high α-helical content, for example, to an unfolded state (random coil). Plotting this change in ellipticity as a function of temperature allows one to calculate the T_m from the point of inflection on the resulting curve.

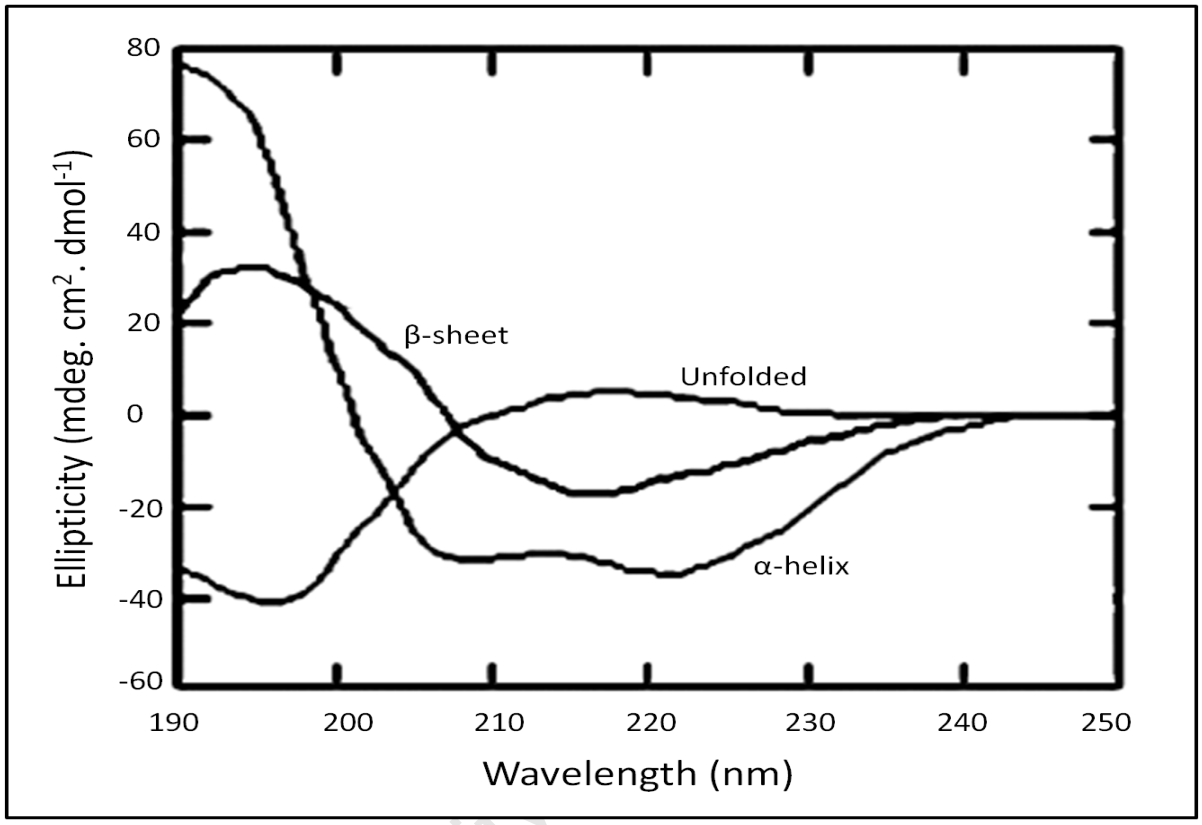


Figure 4.1: Theoretical CD spectra for all α-helix, all β-sheet or all random coil (unfolded) proteins (adapted from Correa *et al*. 2009)**.**

Objectives:

1. To determine the role of different N-linked glycosylation sites in the expression of N-domain proteins.
2. To determine the minimum glycosylation requirements of the expression of functional N-domain proteins.
3. To assess the effect of reduced glycosylation on the enzyme kinetics of the N-domain.
4. To determine the effect of glycosylation on the thermal stability of the N-domain and to assess which glycosylation sites are important for conferring thermal stability.

4.2 Experimental procedures

4.2.1 Transfection, expression and purification of N-domain glycosylation variants

Transfection, expression and purification of N-domain glycosylation variants was carried out as described in Chapter 2.2.2 and 2.2.3.

4.2.2 Enzymatic activity assay and Western blotting

ACE activity was assayed using the substrate Z-Phe-His-Leu (Z-FHL) (Bachem, Budendorf, Switzerland) (see Appendix A6). Protein concentrations were determined as described in Chapter 2.2.3. Three hundred nano grams of protein were fractionated by SDS-PAGE (see Appendix A4) and immunoblotting of purified recombinant ACE N-domain proteins was carried out as described in Appendix A5. Data were plotted using Graphpad Prism 4.01.

4.2.3 Reverse-transcription PCR (RT PCR)

Total RNA was isolated from CHO cells using Trizol Reagent (Invitrogen, CA, USA) according to the manufacturer's instructions. RNA concentrations were determined by spectrophotometry at 260 nm. Reverse transcription of total mRNA was carried out using the Advantage RT-for-PCR kit (Clonetech, CA, USA), in a Hybaid PCR Sprint thermocycler (Mandel Scientific Company Inc, Ontario, Canada), according to the manufacturer's instructions. Specific amplification of ACE N-domain cDNA was achieved using the forward and reverse primers 5'-TTTGCCTGGGAGGGCTGGCA-3' and 5'-TGGTCCACCAAGTAGCCAAA-3' which correspond to nucleotides 581-600 and 1421-1440, of the ACE N-domain sequence (Genbank accession number NM_000789.2). Ten micro litres of total cDNA product was used as a template in each 50 µl PCR reaction. The cycling parameters consisted of one cycle of 95 °C for 2 min, then 20 cycles of 95 °C for 0.3 min, 58 °C for 0.5 min, 68 °C for 1.5 min followed by a final extension cycle of 3 min at 68 °C. PCR products were electrophoresed on 1% (w/v) agarose gels in TBE, using O'GeneRuler DNA ladder mix (Fermentas, Ontario, Canada) as a molecular size standard.

4.2.4 Thermal denaturation assay

Purified recombinant ACE proteins were incubated at 55 °C for up to 60 min in 50 mM HEPES, pH 7.4. Residual activity was determined using the above mentioned ACE assay.

4.2.5 Determination of melting temperatures (T_m)

Far-UV CD spectra were recorded on a Jasco J-810 spectrometer (Jasco Corporation, Easton, USA). Ellipticity was monitored from 210 to 260 nm using a protein concentration of 0.04 mg/ml (in 100 mM phosphate buffer, pH 7.5). Ellipticity at lower wavelengths (190 - 210 nm) could not be measured due to instrument limitations, as a result of the high voltage required to scan these wavelengths under our conditions (see Chapter 4.4.3). Proteins were exposed to thermal denaturation in 5 °C increments from 30 to 90 °C at a rate of 0.43 °C/min. Ellipticity was plotted as a function of temperature using Graphpad Prism 5.01 (Graphpad software, CA, USA). Melting points were obtained from the first derivative of the unfolding transition curve and are reported as mean values of five different sigmoidal fits. Previous work has shown that thermal inactivation of ACE is irreversible [104], therefore, this aspect was not investigated further.

4.2.6 Determination of kinetic constants for the hydrolysis of Z-FHL

Kinetic parameters for the hydrolysis of Z-FHL (Bachem, Budendorf, Switzerland) were determined using the standard ACE assay (see Appendix A6), except that initial reaction rates were calculated using Z-FHL concentrations ranging from 0.00 to 2.54 mM. Assays were repeated at least three times in triplicate. Data fitting and Michaelis-Menten constants, K_m, V_{max}, k_{cat} and k_{cat}/K_m, were determined using GraphPad Prism 5.01. Specific activities of samples at the time of assaying were compared to values obtained immediately after purification and the percent active protein calculated. Kinetic values were then normalised according to the percentage of active protein. Error bars indicate SE.

4.3 Results

4.3.1 The effect of different glycosylation profiles on the expression of active N-domain

As shown in Chapter 3, a range of glycosylation mutants were generated (Figure 4.2).

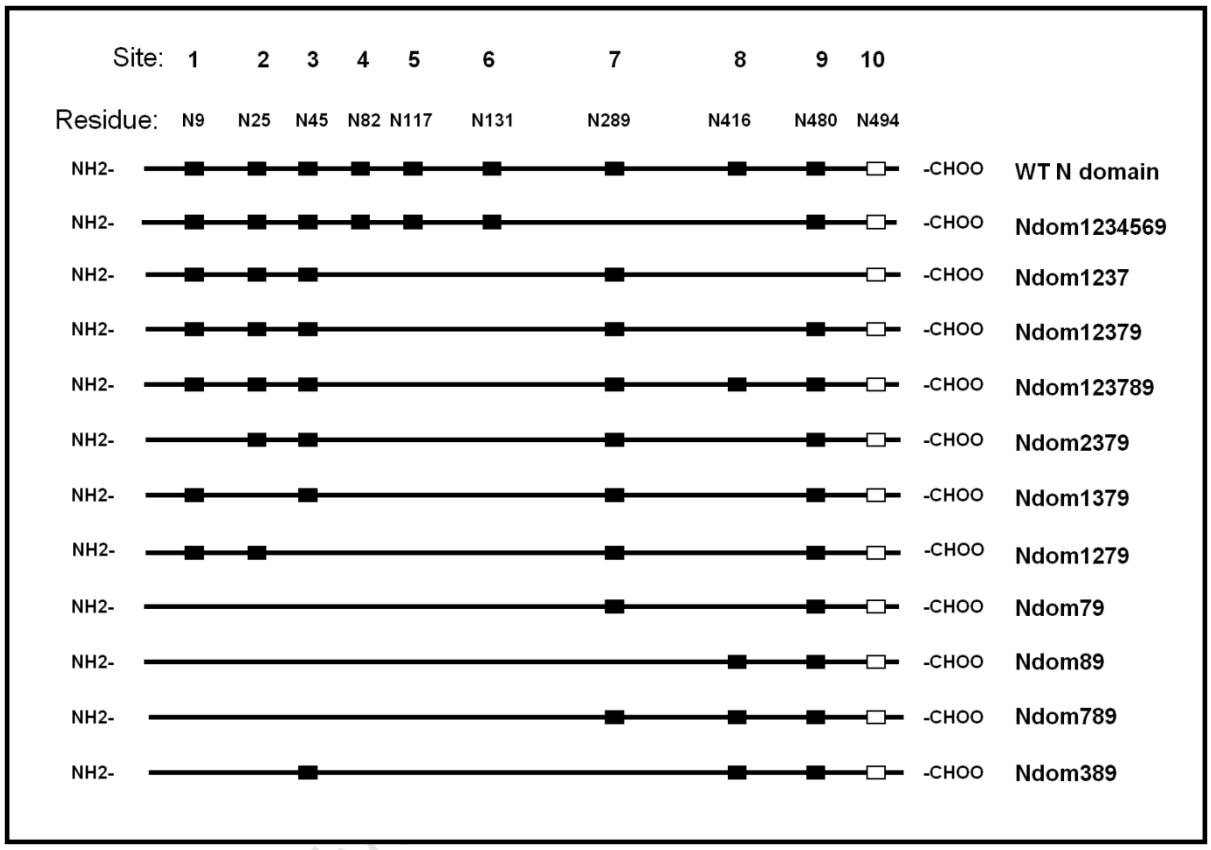


Figure 4.2: Schematic diagram of the N-domain glycosylation mutants. Black boxes indicate intact glycosylation sites, while the white boxes indicate an unglycosylated site.

WT N-domain and the various glycosylation mutants were expressed in CHO cells. As per Chapter 2.2.3, ACE N-domain proteins were purified from culture media and subjected to SDS-PAGE analysis, where the presence of a single prominent band and negligible degradation confirmed sample purity and integrity (Figure 4.3, Appendix Figure A3). Following purification, the effect of the differential glycosylation of these N-domain variants on protein expression, specific activity and thermal stability were investigated.

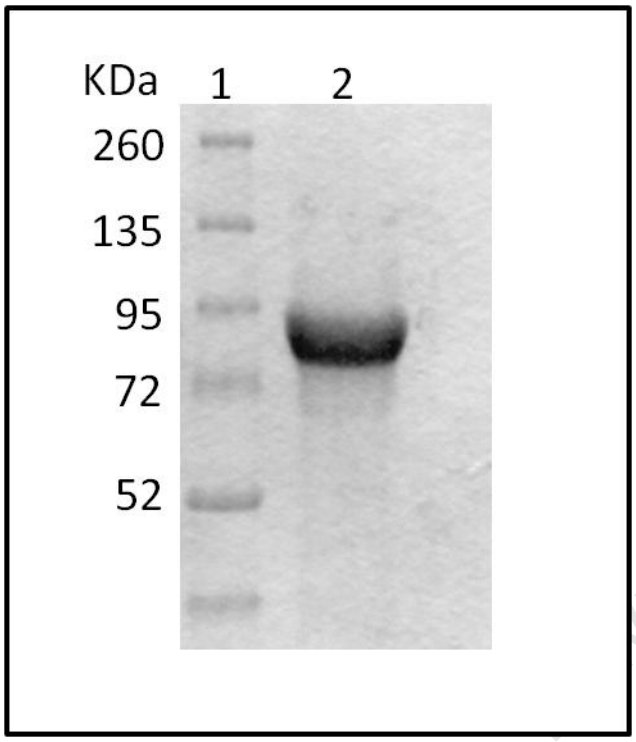


Figure 4.3: Representative SDS-PAGE gels stained with Coomassie, showing successful purification of ACE N-domain protein. Lane 1 - Spectra™ broad range molecular marker. Lane 2 - Ndom1279.

Previously Ndom123456 had been found to be inactive as a direct result of the removal of the three C-terminal glycosylation sites [125]. The reintroduction of site 9 into this mutant (Ndom1234569), however, was able to rescue activity (Figure 4.4A). Similarly, the hypoglycosylated mutant Ndom123 is also known to be inactive [125]. However, it was found that activity could be rescued by the reintroduction of C-terminal sites 7, 8 and 9 (Ndom123789), 7 and 9 (Ndom12379) and, even site 7 alone (Ndom1237) (Figure 4.4A). Removal of individual N-terminal glycosylation sites (site 1, 2 and 3) from the hypoglycosylated mutant Ndom12379 (creating Ndom2379, Ndom1379 and Ndom1279) had no major effect on the expression of active protein (Figure 4.4A). The hypoglycosylated variants Ndom79 and Ndom89 were both found to be inactive. However, the reintroduction of either site 3 or site 7 (Ndom389 and Ndom789) was able to rescue activity and produced functional protein in both cases (Figure 4.4A).

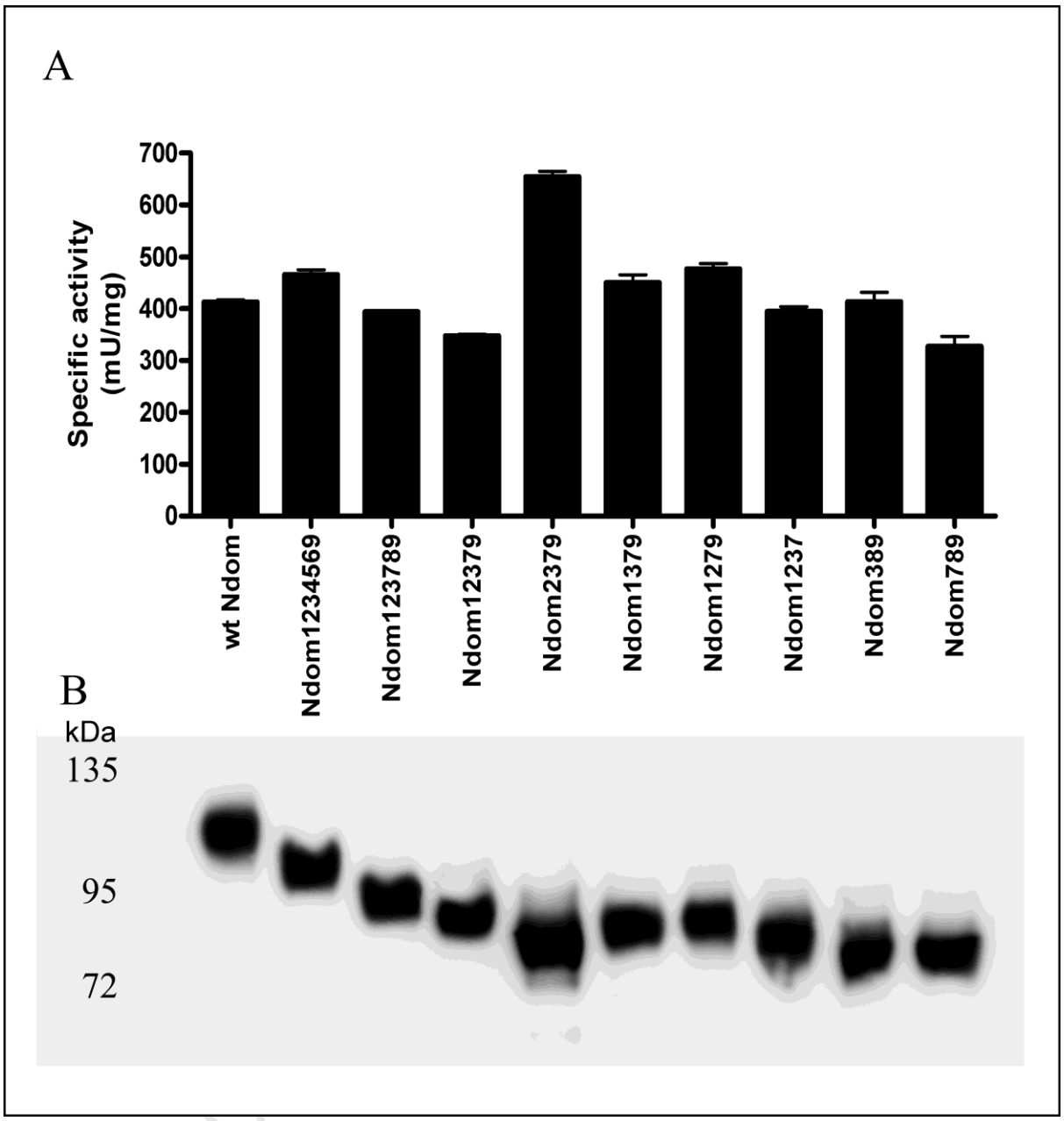


Figure 4.4: Expression of N-domain glycosylation variants. A) Specific activity of ACE N-domain and various glycosylation mutants are shown. B) Western blot of purified recombinant N-domain proteins, probed with an anti-N-domain antibody (4G6), Spectra™ broad range molecular marker indicates molecular sizes.

In addition, Western blot analysis of the glycosylation mutants showed that an increase in electrophoretic mobility correlated with a decrease in the number of intact glycosylation sequons and thus *N*-glycan content of the proteins (Figure 4.4B). While an increase in electrophoretic mobility was noted, the increase was not directly proportional for all mutants, with Ndom1379 and Ndom1279 migrating slightly higher than expected for proteins carrying only four glycan sites. This phenomenon has

been noted before and it is thought that the difference in molecular size/migration may be due to the presence of bulkier glycan groups at a particular site (or sites) [112], in this case, potentially sites 1 and 9.

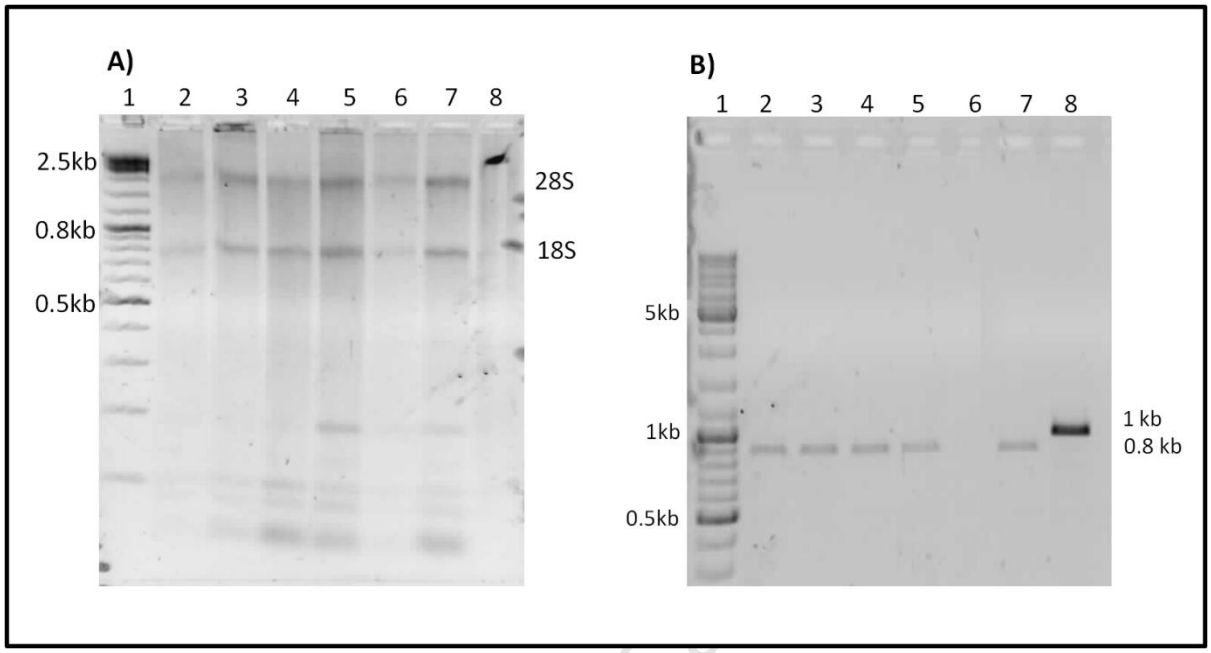


Figure 4.5: Agarose gel of RT-PCR products showing RNA isolation and mRNA transcription for inactive mutants. Lane 1 - O'generuler DNA ladder; Lane 2 - Ndom123456; Lane 3 - Ndom123; Lane 4 - Ndom79; Lane 5 - Ndom89; Lane 6 – untransfected CHO cells. Lane 7 - WT N-domain; Lane 8 - total RNA control. A) Agarose gel of total RNA isolated from transfected CHO cells, with 18S and 28S rRNA fragments indicated. B) Agarose gel showing specific amplification of N-domain mRNA.

Inactive proteins Ndom123, Ndom123456, Ndom79 and Ndom89 could not be detected by Western blot. However, mRNA expression was detected by RT-PCR. Total RNA was isolated from CHO cells stably transfected with each of the inactive glycosylation mutants, as well as from WT N-domain and untransfected CHO cells as controls. RNA integrity was confirmed in each case by the presence of distinct fragments, corresponding to the S18 and S28 RNA sizes (Figure 4.5A). These fragments migrated to approximately 0.8 and 2.5 kb according to the double stranded DNA molecular ladder, which corresponds to the expected sizes of 1.8 and 5 Kb, as RNA is single stranded. Messenger RNA was detected in each case by the specific amplification of a 0.8 kb fragment for the N-domain mutants, or a 1 kb fragment in the case of the control RNA sample (Figure 4.5B). These data indicate that protein detection by Western blot may have been hampered by rapid

intracellular degradation, a factor which is known to affect terminally misfolded membrane proteins in the ER, including underglycosylated and unglycosylated ACE [67,112].

4.3.2 The effect of glycosylation on the thermal stability of the N-domain

4.3.2.1 Thermal inactivation of N-domain glycosylation variants

The thermal denaturation of the N-domain and its hypoglycosylated mutants was assessed following incubation at 55 °C, with the subsequent determination of residual activity using the ACE substrate Z-FHL. Wild type N-domain showed a remarkable level of thermal stability, retaining almost 80% of its activity after 30 min thermal inactivation at 55 °C, even after 60 min at 55 °C there was little change in enzyme activity (Figure 4.6). The removal of sites 4, 5 and 6 did not affect the thermal stability, with Ndom123789 retaining about 80% of its activity after 30 min. The removal of sites 7 and 8 (Ndom1234659), sites 4, 5, 6 and 8 (Ndom12379) and sites 4, 5, 6, 8 and 9 (Ndom1237) reduced the stability to between 70 and 55% after 30 min. The removal of site 1 (Ndom2379), did not greatly alter the thermal stability from that of Ndom12379, with the variant retaining approximately 60% of its activity after 30 min at 55 °C. However, removal of either sites 2 or 3 from Ndom12379 (that is, Ndom1379 and Ndom1279) caused a remarkable loss of thermal stability, with just under 20% residual activity remaining after 10 min. These mutants were essentially inactive after 30 min of thermal denaturation. The hypoglycosylated mutant Ndom389 was even more unstable, with virtually no activity remaining after 10 min thermal inactivation. Ndom789 was the most unstable of all the glycosylation variants and had virtually no residual activity after 5 min of thermal inactivation.

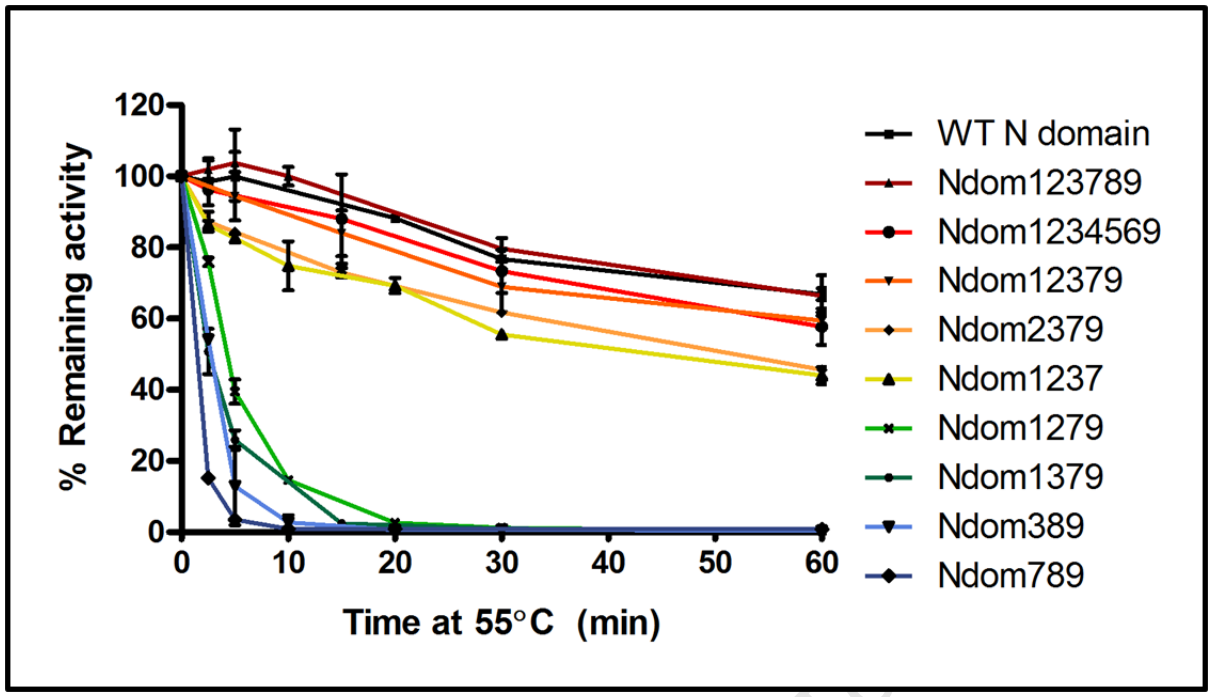


Figure 4.6: Thermal denaturation curve for ACE N-domain and glycosylation variants. Percentage residual activity was calculated using the standard ACE assay with Z-FHL as a substrate (see Chapter 4.2.2). Error bars indicate SE.

4.3.2.1 T_m determination of N-domain glycosylation variants by CD spectroscopy

Four N-domain glycosylation variants were selected for melting point (T_m) determination by CD spectroscopy, namely WT N-domain, Ndom2379, Ndom1279 and Ndom389. These variants were selected such that glycosylation mutants with high, intermediate and low thermal stabilities were represented (based on the thermal denaturation studies above).

The crystal structure of the N-domain of ACE reveals that the enzyme is comprised largely of α-helices [139]. CD spectroscopy of α-helical proteins typically involves scanning the protein from 190 to 260 nm. This is because much of the change in ellipticity occurs over the 190 to 230 nm range (Figure 4.1) [156]. However, we were only able to scan from 210 nm. This was due to the relationship between the voltage and the level of sample/buffer absorbance in the low ultra violet (UV) range. The voltage gives a direct measure of the amount of light being passed through the sample. Our sample and buffer preparation were moderately absorbent of circular light at low wavelengths, which necessitated an increased voltage to produce the

required amount of transmitted light. When the voltage exceeds 700 V, the CD output becomes unreliable due to a decrease in the signal to noise ratio [157]. Since sample/buffer absorbance increases with decreasing wavelength, we were thus unable to scan ellipticity below 210 nm (Figure 4.7). Typically, a buffer dilution is able to resolve this issue [157]. However, scans of buffer only revealed that the buffer was not responsible for the increased absorbance, indicating that the sample itself was the cause. Attempts to scan dilutions of the sample were not successful due to the low signal obtained for these scans.

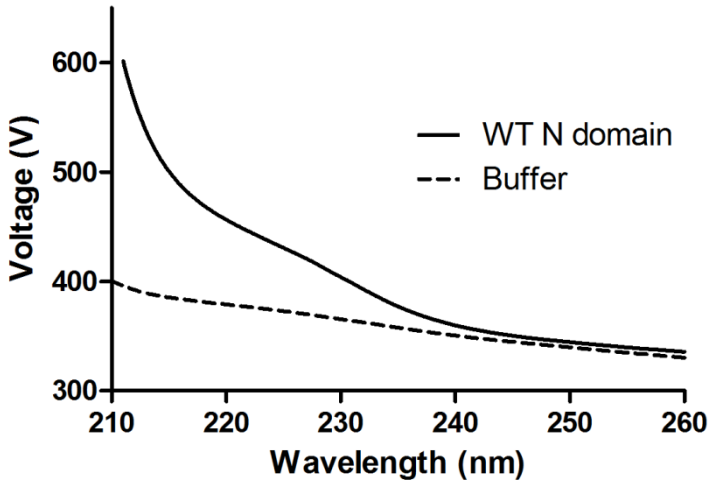


Figure 4.7: Representative graph of voltage as a function of wavelength for a CD scan of WT N-domain. Voltage required for CD scans at 35 °C for WT N-domain and buffer**.**

Despite the reduced range of wavelengths scanned, sufficient change in ellipticity was observed between 210 and 260 nm to allow calculation of T_m values. Initial circular dichroism scans of the different glycosylation variants at low temperature reported spectra typical of proteins with high α-helical content, as expected for the N-domain. However, as the temperature was increased, a decrease in the level of ellipticity was noted, indicating a loss of α-helical content, typical of an unfolding transition

(Figure 4.8). The WT N-domain protein began thermal transition at approximately 55 °C and reached plateau at about 80 °C, while the same transition occurred between 50 and 75 °C for Ndom2379. The thermally unstable variants Ndom1279

and Ndom389, however, underwent transition at a much lower temperature (40 °C) and had reached a plateau at about 50 °C.

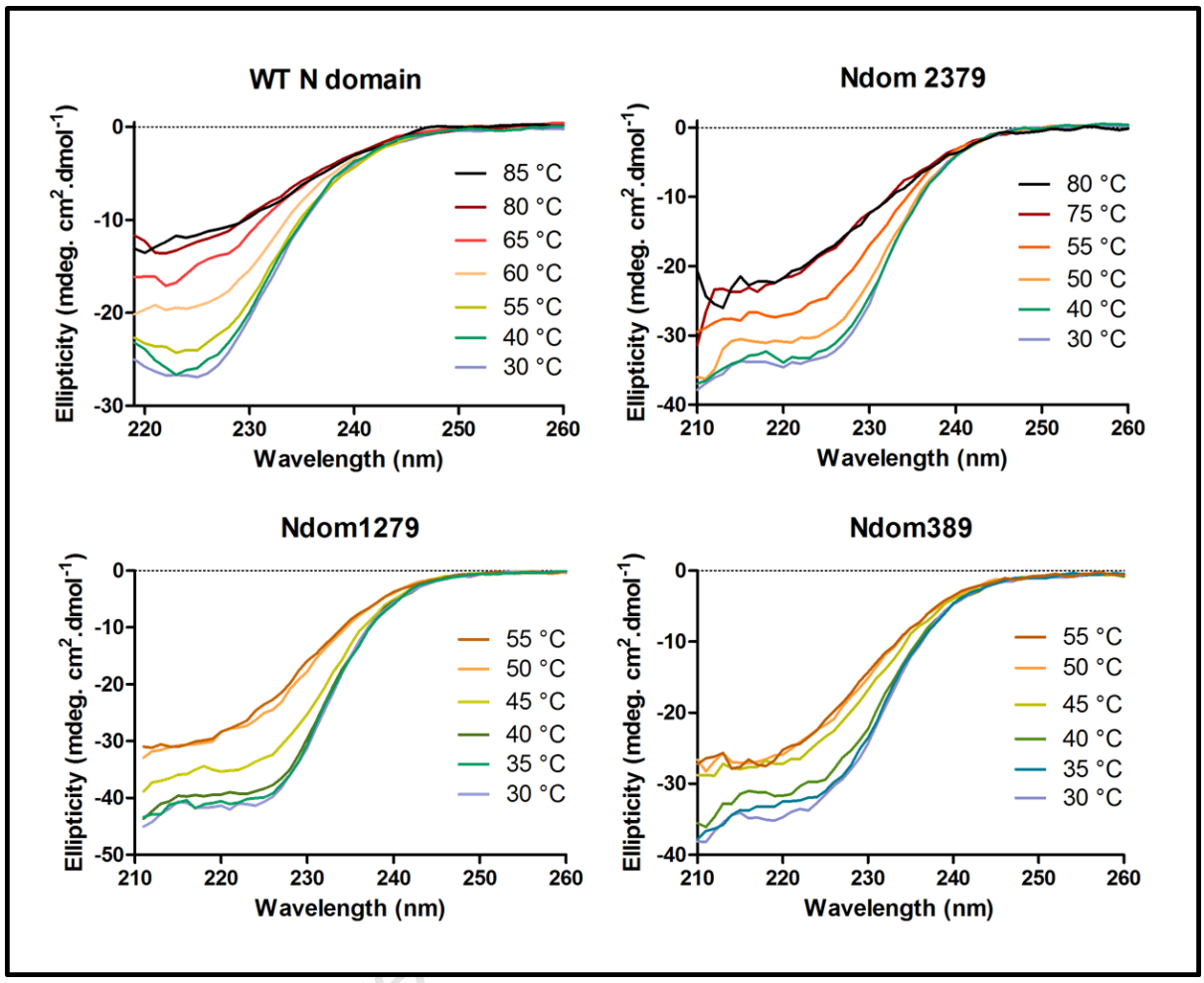


Figure 4.8: Circular dichroism spectra of N-domain variants showing temperature induced denaturation. Unfolding transitions for WT N-domain, Ndom2379, Ndom1279 and Ndom389 are shown over 210 to 260 nm. Scans for selected temperatures shown.

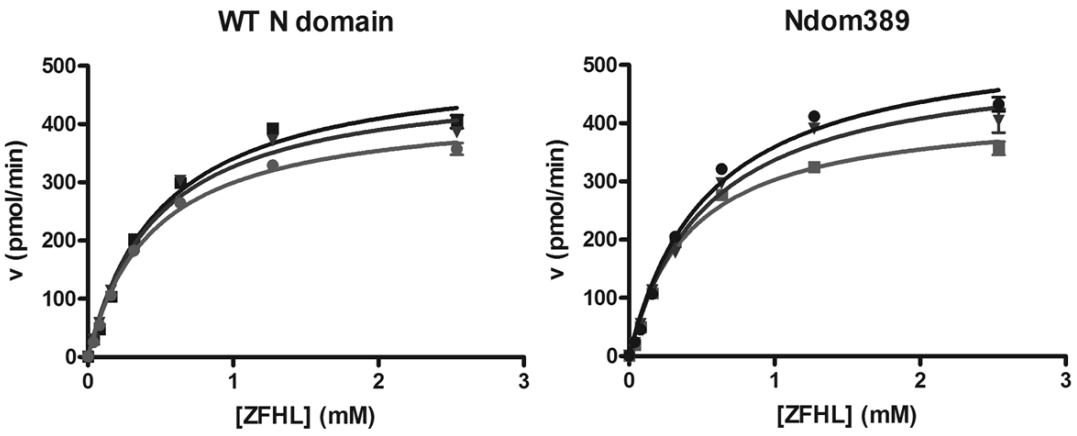
The calculated T_m for WT N-domain was 60.3 °C (Figure 4.9), which is lower that the literature value for the N-domain (±70 °C), yet still higher than that of the C-domain (±52 °C) [104,136]. Ndom2379 reported a T_m somewhat lower than that of WT N-domain (53.9 °C ±0.3), while Ndom1279 and Ndom389 had markedly lower values of 45.9 °C ±0.1 and 42.7 °C ±0.2 respectively. As a control, the T_m for BSA fraction-V (Sigma) was determined (see Appendix Figure A4) and found to be ±67 °C, which closely matches the literature value of ±69 °C [158].

N domain Variant	Tm (°C)
WT N domain	60.3 ±0.1
Ndom2379	53.9 ±0.3
Ndom1279	45.9 ±0.1
Ndom389	42.7 ±0.2
BSA	66.9 ±0.7

Figure 4.9: Thermal denaturation of N-domain glycosylation variants. Tm values for WT N-domain, Ndom2379, Ndom1279, Ndom389, and BSA were calculated from ellipticity plots monitored from 210 to 260 nm using a protein concentration of 0.04 mg/ml. T_m values are expressed ± standard deviation (SD).

4.3.3 Kinetic characterisation of minimally glycosylated N-domain

Since the minimally glycosylated mutant Ndom389 was likely to be a suitable candidate for crystallisation studies, it was important to ensure that the mutagenesis of the glycosylation sites did not compromise the function or structure of the enzyme. Analysis of the specific activities of the various glycosylation mutants gave some indication that the catalytic functioning of the enzyme was preserved despite reduced glycosylation. However, given the importance of Ndom389 for crystal structure determination, a more detailed kinetic analysis was performed to ensure that functional integrity of Ndom389 was retained, relative to glycosylation WT N-domain. Analysis of the kinetic parameters revealed only minor differences between Ndom389 and WT N-domain for the hydrolysis of Z-FHL. The K_m values for WT N-domain were highly comparable to those of Ndom389, being 0.48 and 0.5 mM respectively (Figure 4.10). The k_{cat} values for WT N-domain and Ndom389 (15.09 and 24.32 s^{-1} respectively) showed a slightly higher turnover rate for the minimally glycosylated mutant. A comparison of the k_{cat}/K_m ratio revealed comparable catalytic efficiencies between the two forms, with values of 31.58 and 49.64 $mM^{-1}s^{-1}$ for WT and Ndom389 respectively. Previous work had put the K_m of WT N-domain for the hydrolysis of Z-FHL at between 0.9 and 1.2 mM [159,160]. While the kinetic constants shown here were slightly different from previous findings. This was not deemed to be problematic given that slightly different assay conditions were used in this study. In addition the kinetic parameters for WT N-domain and Ndom389 were comparable.



	kcat/Km	Km (mM)	kcat (s-1)
WT N domain	31.58 (±0.46)	0.48 (±0.0174)	15.09 (±0.90)
Ndom389	49.64 (±1.0028)	0.50 (±0.041)	24.32 (±2.20)

Figure 4.10: Michaelis-Menten plot of WT N-domain and Ndom389, with kinetic constants K_m, k_{cat} and k_{cat}/K_m indicated. The Michaelis-Menten kinetic constants K_m, k_{cat} and k_{cat}/K_m were determined for WT N domain and Ndom389 using the standard ACE assay (see Appendix A6), except that final Z-FHL concentrations ranging from 0.00 to 2.54 mM were used. Assays were repeated three times in triplicate. Error bars indicate SE.

4.4 Discussion

Previous work has shown that glycosylation is important for N-domain expression [125] as well as thermal stability [136]. Here we show that the glycosylation sites required for expression of active protein differ from those sites which are critical to thermal stability.

4.4.1 Identification of glycosylation sites important for expression of enzymatically active N-domain

Reintroducing one or more C-terminal sites into an inactive hypoglycosylated variant was sufficient to restore the expression of active N-domain, as seen with Ndom1234569 and Ndom123789, Ndom12379 and Ndom1237. This indicates that C-terminal glycosylation is crucial for expression of N-domain proteins in CHO cells. These data also seem to indicate that there may be an element of redundancy with regard to the specific sites required, since both site 7 and site 9 were able to rescue

inactive mutants when reintroduced. The removal of individual N-terminal sites 1, 2 and 3 had no major effect on the expression of Ndom2379, Ndom1379 and Ndom1279, indicating that, at least individually, these sites were not critical for the production of active protein. These findings agree with the hypothesis that C-terminal glycosylation is the major contributor to effective folding and processing of the enzyme. However, Ndom79 and Ndom89 were not active, while Ndom789 and Ndom389 were. This would suggest that there is a requirement for glycosylation of at least three sites, and that while some of these sites must be toward the C-terminal region of the sequence, glycosylation of an N-terminal site is able to compensate for the absence of glycosylation at one of the three C-terminal sites. The location of the three C-terminal sites 7, 8 and 9, as well as the N-terminal site 3, does not immediately suggest a mechanism underlying their importance, as they are not clustered in a single region, but rather they are located at positions distant from one another (Figure 4.11). This is not surprising, however, given the various roles that N-linked glycosylation plays in the folding process (see Chapter 1.2.3). It is interesting to note the preference for C-terminal glycosylation of the N-domain, particularly in light of the fact that the C-domain (tACE) requires N-terminal glycosylation for the expression of active protein [122]. A number of other proteins, such as butyrylcholinesterase and chondromodulin-I, also show an N-terminal requirement for *N*-linked glycans [89,161], a finding corroborated by a study on glycosylation sequon usage which found that distance from the C-terminus affects the likelihood of glycosylation of a particular sequon [63]. While the N-domain has a preference for C-terminal glycosylation, both N- and C-terminal glycosylation sites have been shown to be glycosylated (see Chapter 2.4). Thus, while N-terminal glycosylation of proteins is generally preferred, this does not necessarily preclude the fact that C-terminal glycan sites may also be important in some cases. Another factor to consider is that the N-domain is not expressed as a single unit *in vivo*, but rather occurs as part of the larger two-domain sACE. Indeed, results have shown that glycosylation of sACE was preferentially located on the N-domain, with 5 of the 6 glycans detected, being located on the N-domain [124].

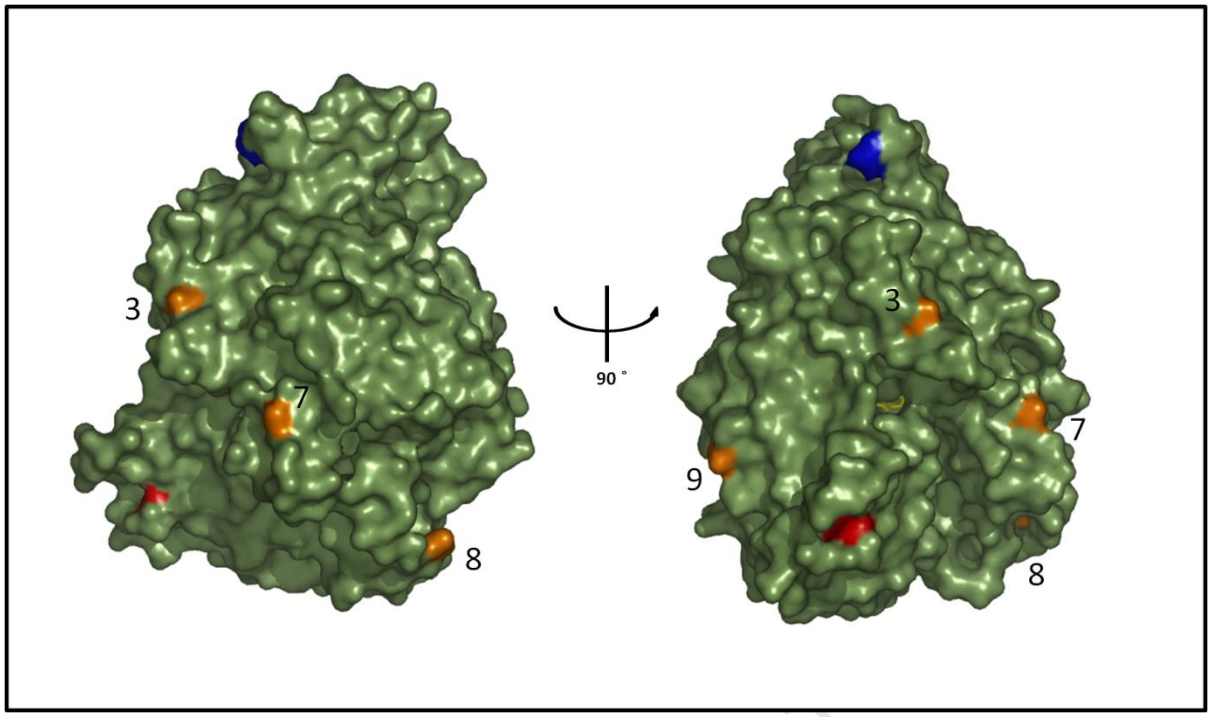


Figure 4.11: Surface representation of the N-domain of ACE indicating the location of glycosylation sites 3, 7, 8 and 9. N- and C-termini are highlighted in blue and red respectively. Glycosylation sites are numbered. Figures were drawn with PyMOL v0.92 using the PDB structure 2C6N [139] (DeLano Scientific, San Carlos, CA, USA).

4.4.2 The effect of glycosylation on the thermal stability of the N-domain

The different N-domain glycosylation variants can be broadly placed into two groups with respect to thermal stability. The first group consists of mutants which retained a high level of thermal stability, namely Ndom123469, Ndom123789, Ndom12379, Ndom1237 and Ndom2379. The second group comprises those mutants which displayed a marked decrease in thermal stability, namely Ndom1379, Ndom1279, Ndom389 and Ndom789. While the most interesting findings lie with those mutants comprising the second group, some deductions can be made about the contribution of certain glycan sites from the first group. Firstly, the removal of sites 4, 5 and 6 had virtually no effect, with the thermal stability of Ndom123789 being comparable to that of WT N-domain. This indicates that these glycosylation sites are not involved in conferring thermal stability to the N-domain. Secondly, the absence of site 8 (N416) in Ndom12379, Ndom1237 and Ndom2379 or sites 7 and 8 (Ndom1234569) resulted in a slight decrease in stability (Figure 4.6), indicating that these two sites play a limited role in conferring thermal stability.

The second group contains those variants where the removal of glycan sites caused a marked loss of thermal stability (namely Ndom1379, Ndom1279, Ndom389 and Ndom789). The major characteristic distinguishing these variants from their more stable counterparts is the removal of either site 2 (N25) or site 3 (N45), or both sites 2 and 3. These data clearly highlight the critical importance of these two N-terminal sites in maintaining thermal stability of the N-domain. Glycosylation sites 2 and 3 are located on α-helix two, which forms part of the lid region of the protein. It has been noted that glycosylation sites are often located at positions where the secondary structure changes [50]. The active site entrance channel in the crystal structure of both domains is not large enough to allow substrate entry, thus necessitating some degree of movement within this region [162]. Modelling of the C-domain structure to open and closed conformations of an ACE homologue, ACE2, which has been shown to adopt a catalytically important hinge movement upon substrate binding, revealed a number of potential hinge movements in the C-domain [162]. Due to the structural similarity between the two domains, it has been proposed that both domains of sACE also undergo a hinge movement to allow substrate access to the active site [162]. The location of the main hinge movements were mapped to the regions 98 to 125, 400 to 409 and 569 to 578 of the C-domain [162]. These regions correspond to the N-domain residues 70 to102, 378 to 386 and 546 to 554 respectively (Figure 4.12). Interestingly, the glycosylation sites 1 to 4 on the N-domain are located on the lid region, which is thought to undergo a conformational change during the proposed hinge movement [140]. Furthermore, site 2 is located adjacent to the two main hinge regions. Glycosylation sites 1 and 4 did not have an effect thermal stability, however, it is possible that sites 2 and 3 are involved in stabilizing interactions involving the lid region and those areas undergoing hinge movements, thus, protecting the catalytically important lid region from thermal denaturation.

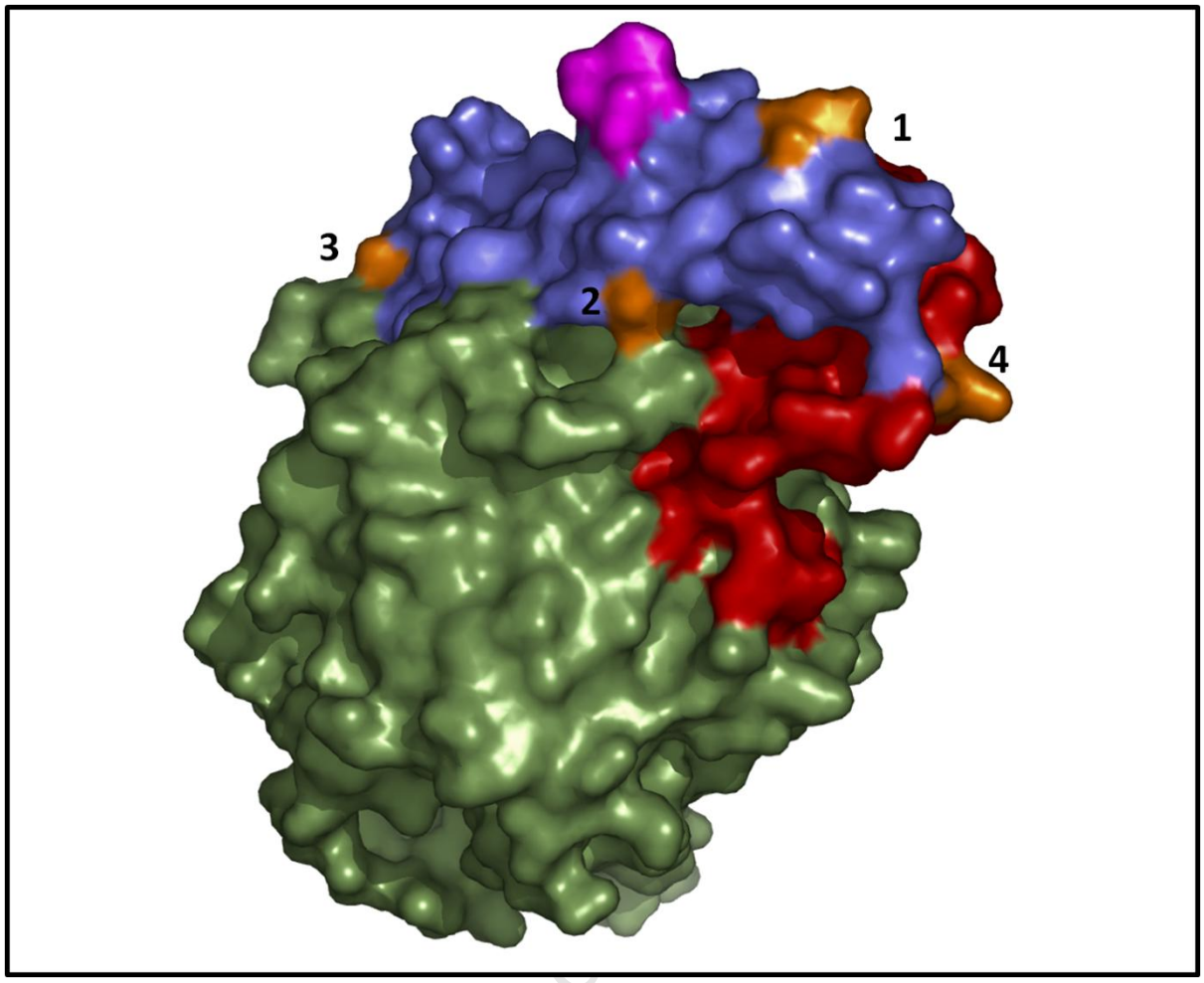


Figure 4.12: N-domain structure showing the location of the proposed hinge movement and the glycosylation sites on the lid region. Glycosylation sites are labelled and indicated in orange. Lid helices shown in blue. Regions involved in the proposed hinge movements of are marked in red. The N-terminal region of the N-domain is shown in purple. Figures were drawn with PyMOL v0.92 using the PDB structure 2C6N and 1O86 [138,139] (DeLano Scientific, San Carlos, CA, USA).

The N-domain has a reported T_m of about 70 °C [104,136], yet we determined the T_m to be slightly lower. Since the BSA control reported a T_m consistent with literature values, it is unlikely that this difference is due to instrumental or handling error. However, the trend of reduced glycosylation corresponding to a decrease in thermal stability, as well as the importance of sites 2 and 3, is strongly evident in these data, further confirming the results from the thermal denaturation assays. Interestingly, minimally glycosylated C-domain (tACEg13) has a T_m of 43 °C [136], virtually identical to that of minimally glycosylated N-domain (Ndom389) 42.7 °C. Given the high degree of sequence and structural identity between the two domains, it is possible that this value represents the base level of thermal stability for the two domains

resulting from factors other than glycosylation. These findings further highlight the significant role of glycosylation with respect to the thermal stability of ACE.

It has previously been suggested that the marked difference in thermal stability between the N- and C-domains (15 °C) is attributable to the fact that the N-domain has a greater number of α-helices and a greater degree of glycosylation and an increased proline content over the region D29 to T133, as compared to the corresponding region of the C-domain [104]. Recent work has shown that the *N*-linked glycans of the C-domain contribute significantly to its thermal stability, while in contrast, the presence of O-linked glycans had no effect [136]. While the increased number of α-helices and the greater degree of proline content of the N-domain, over the region D29 to T133, may play a role in conferring thermal stability, we have shown that surface glycosylation is the major determining factor with regard to the greater level of thermal stability of the N-domain. This may be due to a greater degree of glycan site occupancy for the N-domain, about 90%, compared with 50 to 85% for the C-domain. However, Voronov et al. (2002), also mention that the greater degree of glycosylation on the N-domain over the region D29 to T133 (six sites, as opposed to three for the C-domain) is also likely to be an important factor. Our findings show that while glycosylation of this region is indeed the primary factor causing the increased level of thermal stability for the N-domain, it is in fact the presence of glycosylation at sites 2 (N25) and 3 (N45) that is critical for the high level of thermal stability of the N-domain and not that of the entire D29 to T133 region. This finding is further highlighted when one considers that the loss of either of these two sites effectively reduces the thermal stability of the N-domain by 8 °C (Figure 4.9).

4.4.3 Kinetics characterization of minimally glycosylated N-domain

As mentioned previously deglycosylation often has little or no effect on enzyme catalysis (see Chapter 1.2.3.3.). Previous work has shown that this was indeed the case with enzymatically deglycosylated C-domain (tACE) [124], as well as for a range of C-domain glycosylation mutants generated by site-directed mutagenesis, which all exhibited similar catalytic efficiencies for the ACE substrate hippuryl-L-histidyl-L-leucine (HHL), when compared with WT C-domain [122]. Kinetic studies on Ndom389 and WT N-domain have shown that the two glycoforms exhibit identical binding

affinities and also have similar catalytic efficiencies for the ACE substrate Z-FHL. Thus, glycosylation does not appear to have an effect on the kinetic properties of either the N- or C-domain.

4.4.4 Crystallisation of minimally glycosylated N-domain

The use of a hypoglycosylated form of a protein to enable reproducible crystallization has been shown to be an effective approach in a number of cases [89,122,163,164]. However, reduced glycosylation can translate to reduced stability and Ndom389 has been shown to be relatively unstable as compared to the WT form. However, its thermal stability was sufficient to justify its use in crystallization trials, typically carried out at 16 °C. Recent crystallization trials with this variant have shown that it is able to crystallize reproducibly, with crystals forming in multiple wells of the crystallization screening plate, the best of which diffracted to 2 Å [123]. It is often a concern that introducing mutations may affect the structure of the protein, which would be problematic for this approach to crystal structure determination. However, the kinetics of the minimally glycosylated mutant are similar to the WT form, which would imply that a structural determination of Ndom389 is likely to be highly informative of the structure of WT N-domain. Furthermore, aligning the structures of WT N-domain and Ndom389 in PyMOL v0.92 (Figure 4.13), revealed an RMSD of 0.56 Å for all atoms. Given that the structures were solved to a resolution of approximately 2 Å, this RMSD value indicates that the two structures are essentially identical.

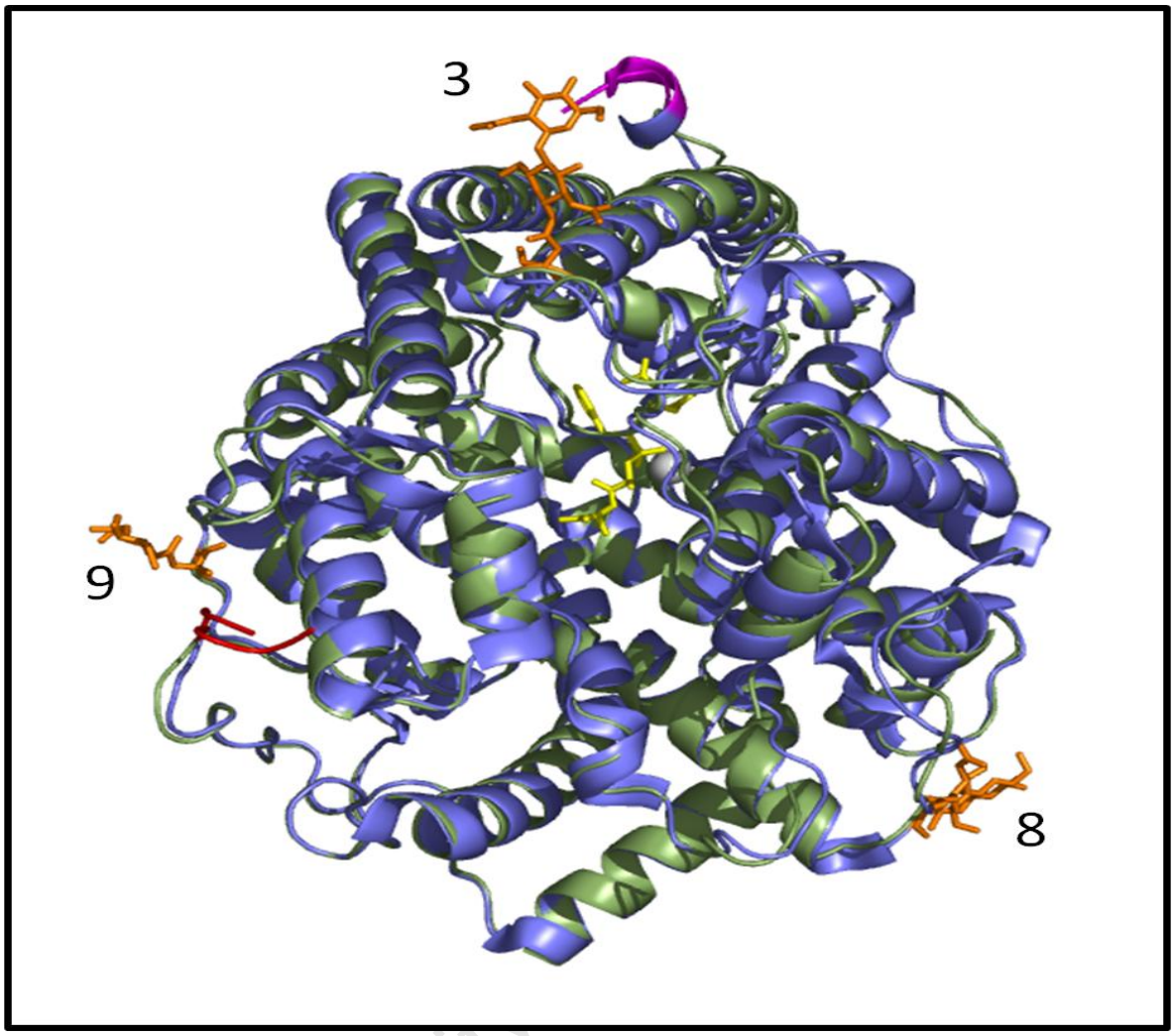


Figure 4.13: Comparison of the NB-DNJ and minimally glycosylated crystal structures of the N-domain. Density corresponding to GlcNac residues at sites N45, N416 and N480 are indicated in orange on the Ndom389 structure, the N-domain selective inhibitor RXP407 (yellow) bound to the active site zinc (grey) of Ndom389 is shown. N- and C-termini highlighted in purple and red respectively. Figures were drawn with PyMOL v0.92 using the PDB structures 2C6N [139] and 3NXQ [123] (DeLano Scientific, San Carlos, CA, USA). *The crystallisation and solution of the Ndom389 structure was carried out by R. Acharya's research group (University of Bath, UK).

4.4.5 Conclusions

This study has revealed that C-terminal glycosylation sites are required for the expression of active N-domain proteins, while N-terminal sites are critical for maintaining thermal stability. It is interesting that different glycosylation sites are involved in these key roles. This is likely due to different selective pressures acting on the different glycosylation sites. Glycosylation sites required for expression may be selected for based on the degree to which they facilitate lectin-chaperone binding, such as calnexin and calreticulin, and thus the degree to which they aid protein

folding in the ER [78-80]. Those sites involved in thermal stability (sites 2 and 3) are most likely to be selected for based on their ability to stabilize catalytically and structurally important regions of the protein, such as the proposed flexible hinge region of ACE. We have shown that reduced glycosylation does not greatly affect the kinetics of the enzyme and that producing a minimally glycosylated variant of the N-domain is an effective means of obtaining a form of the protein suitable for high throughput crystallization studies, as evidenced by the recently solved crystal structure of Ndom389 [123].

Chapter 5: Cloning, expression and characterisation of minimally glycosylated sACE

5.1 Introduction

There is a drive towards the rational design of domain selective ACE inhibitors. The crystal structures of the individual N- and C-domains have greatly aided this effort. However, these structures do not give an indication as to the orientation of the two domains relative to one another, a factor which is important given the negative cooperativity between the two domains [160,165,166]. In addition, the solution of the crystal structure of sACE would greatly benefit our understanding of the mechanism of the reported negative cooperativity, as well as providing insight into inhibitor-enzyme interactions within the full-length protein. Thus far it has not been possible to crystallise sACE. This is likely due to the high degree of *N*-glycosylation present on the surface of the enzyme, as sACE is expected to contain glycosylation at between seven and eight sites, based on calculations of the mass contribution of the glycans to sACE [84]. In addition, N-terminal sequencing of sACE has identified the presence of glycans at six sites, most of which were located on the N-domain [124]. This would suggest that the glycosylation state of the N-domain is critical for the expression of active sACE.

However, there is also a degree of inter-domain flexibility –the capacity for two domains to move relative to one another, due to the flexible linker region joining them [139]. Both the interdomain flexibility as well the presence of extensive surface glycosylation are factors that could inhibit the formation of tightly packed molecules in a crystal lattice [109-111].

While the X-ray crystal structure for sACE has yet to be determined, a low resolution structure using electron microscopy has been published [167]. This work modelled the crystal structures of the N- and C-domains to electron micrograph reconstructions of

negatively stained sACE particles. These results indicate that the domains adopt an extended orientation relative to one another. While these findings give some indication of the relative positions of the two domains, there are a few limitations to the study. Firstly, assignment of the N- and C-domains to the electron micrograph reconstructions of sACE involves a certain amount of subjectivity as the differences between the two domains are not clear at the achieved resolution of 2.3 nm. Secondly, the mechanism underlying domain negative cooperativity was not addressed. Thus, the solution of a sACE crystal structure is still needed to clarify the exact orientation of the domains and identify the mechanism of domain negative cooperativity.

It is hoped that creating a form of sACE with minimal glycosylation will allow the production of a crystallisable form of the enzyme. As mentioned previously, sACE is expected to contain up to eight glycans, most of which are likely to be located on the N-domain, with only one being identified on the C-domain [84,124]. This seems to suggest that glycosylation of the N-domain is especially important for the expression of sACE. Thus, we attempted to investigate whether the minimum glycosylation requirements for the N-domain, could be applied to sACE. To test this hypothesis, we decided to fuse minimally glycosylated N-domain (Ndom389) together with minimally glycosylated C-domain (since very little glycosylation was found on the C-domain of sACE) and determine whether this construct could produce active sACE.

Objectives:

1. To introduce a restriction endonuclease site into minimally glycosylated N- and C-domains to enable subcloning of the two domains into a single construct.
2. To express, purify and characterise the minimally glycosylated sACE forms.
3. Carry out crystallisation trials on the minimally glycosylated sACE forms.

5.2 Experimental procedures

5.2.1 Materials

Restriction endonucleases *Eco91I*, *EcoRI*, *KpnI*, *NotI* and T4 ligase were purchased from Fermentas. ACE N-domain monoclonal anti-body 4G6 used to detect sACE was a gift from S. Danilov.

5.2.2 Plasmid templates

Previously generated minimally glycosylated mutants, Ndom389 (N-domain), and tACEg13 (C-domain) [122], were used as templates. tACEg13 is a soluble construct, lacking a transmembrane region. Soluble sACE in pcDNA3.1+ was used for comparison with hypoglycosylated sACE.

5.2.3 Mutagenic oligonucleotides

Primers were engineered to introduce a restriction site via the introduction of silent mutations (Table 5.1), as described in Chapter 3.2.3.

Mutant	Sequence	RE Site
N-C domain linker	F5'CCTGGTGACCGATGAGGCTGAGGCCAGCAAGTTTGTGG 3' R3'GGACCACTGGCTACTCCGACTCCGGTCGTTCAAACACC 5'	*Eco91I*

Table 5.1: Mutagenic oligonucleotides for the introduction of a restriction endonuclease cloning site. The introduced *Eco91I* restriction site is underlined.

5.2.4 Site-directed mutagenesis and cloning of hypoglycosylated sACE

Carried out as described previously (Chapter 3.2.4)

5.2.5 Expression and purification of sACE glycosylation variants

Carried out as described previously (Chapter 4.2.1)

5.2.6 Enzymatic activity assay and Western blot

Carried out as described previously (Chapter 4.2.2)

5.2.7 Thermal denaturation assay

Carried out as described previously (Chapter 4.2.4)

5.2.8 Determination of kinetic constants for the hydrolysis of Z-FHL

Carried out as described previously (Chapter 4.2.6)

5.3 Results

5.3.1 Construction of minimally glycosylated sACE

The recombinant N-domain constructs (including Ndom389) contain a sequence of eight amino acids at the C-terminus which overlaps with the N-terminal region of the recombinant C-domain constructs (including tACEg13). This region is ideal for joining Ndom389 and tACEg13 into a single construct, essentially creating minimally glycosylated sACE. However, due to a lack of unique endonuclease restriction sites in this region, it was necessary to introduce one via site-directed mutagenesis. Using the Watcut webtool (see chapter 3.2.3) it was determined that an *Eco91I* restriction site could be introduced into this region in both Ndom389 and tACEg13, without affecting the coding sequence. The presence of this new site would then facilitate subcloning of the two domains into a single minimally glycosylated construct.

Successful introduction of the cloning site was observed by digestion of Ndom389pBSKII with *KpnI* and *Eco91I* and tACEg13pBSK with *NotI* and *Eco91I*. The resulting digests produced DNA fragments at approximately 2 and 3 Kb in both cases, as expected (Figure 5.1A and B).

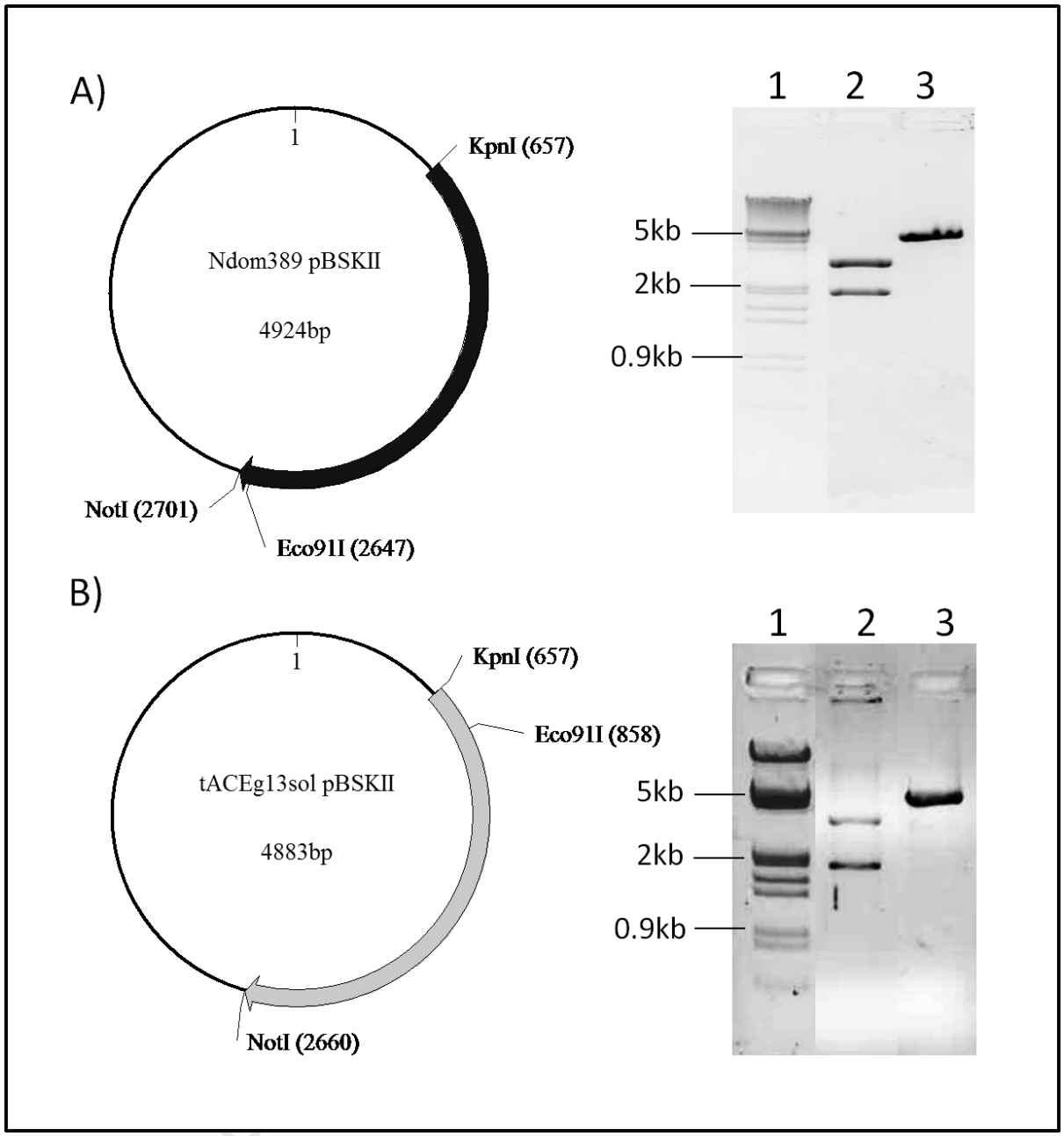


Figure 5.1: Vector map and restriction endonuclease digest of Ndom389 and tACEg13. A) Ndom389 digested with *KpnI* and *Eco91I*. Lane 1 - λDNA-*EcoRI*/*HindIII* ladder, Lane 2 - Ndom389+*Eco91I* showing fragments at 2 and 3 kb, Lane 3 - Ndom389 (negative control) showing DNA fragment at about 5 kb. B) tACEg13 digested with *NotI* and *Eco91I*. Lane 1 - λDNA-*EcoRI*/*HindIII* ladder, Lane 2 - tACEg13+*Eco91I* showing fragments at 1.8 and 3 kb, Lane 3 - tACEg13 (negative control) showing DNA fragment at about 5 kb.

Following successful mutagenesis, tACEg13 was subcloned into Ndom389pBSKII using *Not I* and the newly generated *Eco91I* restriction site (Figure 5.2 A). This full-length minimally glycosylated sACE construct was then cloned into the mammalian expression vector pcDNA3.1+ using the *KpnI* and *NotI* restriction sites in the multiple

cloning site region. The successful cloning of Ndom389-tACEg13 into pCDNA3.1+ was confirmed by digestion with *EcoRI* and *Eco91,* which revealed the presence of signature DNA fragments at 6, 2 and 0.9 kb (Figure 5.2 B).

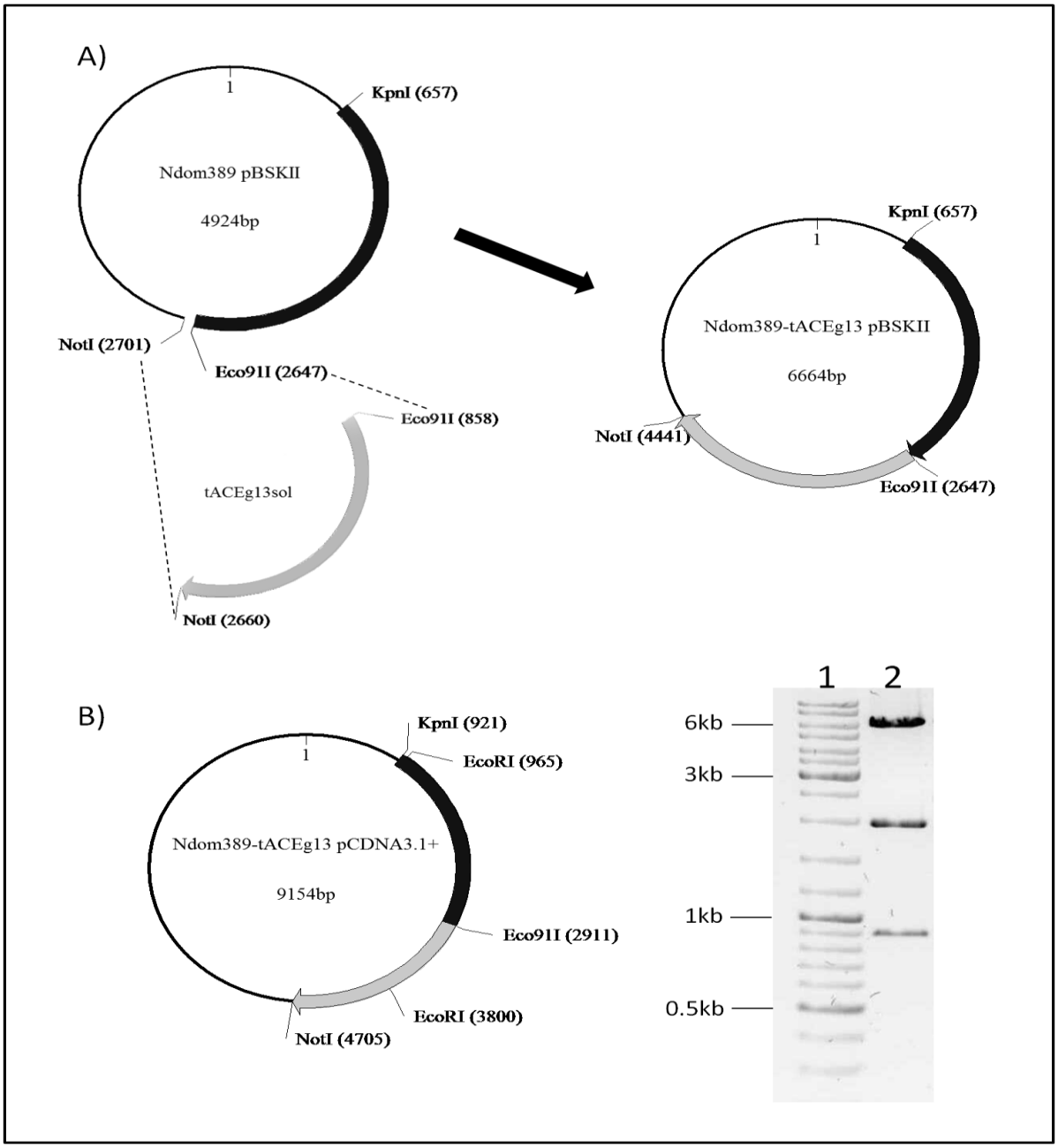


Figure 5.2: Cloning strategy, vector map and restriction endonuclease digest of Ndom389-tACEg13 in pcDNA3.1+. A) Schematic diagram of the cloning strategy used to generate Ndom389-tACEg13. Relevant restriction sites and their positions indicated. B) Vector map and digest screen for Ndom389-tACEg13 pCDNA3.1+ digested with *EcoRI* and *Eco91I*. Lane 1 - O'generuler DNA ladder mix, Lane 2 - Ndom389+*Eco91I*.

5.3.2 Expression and purification of Ndom389-tACEg13 in CHO cells

Ndom389-tACEg13, referred to as minimally glycosylated sACE (mg-sACE) from here on, was expressed in CHO cells and found to be enzymatically active. Indeed purified mg-sACE reported a specific activity comparable to that of WT sACE, with values of 1058 and 1106 mU/mg respectively (Figure 5.4A). These findings are consistent with work on the individual domains which showed that kinetic parameters for Ndom389 (see Chapter 4.3.1) and tACEg13 [122] were similar to those for the WT form of the two domains.

The purity of mg-sACE was confirmed by SDS-PAGE, where a single band was observed. However, additional bands between 72 and 95 kDa and just below 72 kDa were also seen (Figure 5.3). Interestingly, the minimally glycosylated N-Domain migrates to a similar position between 72 and 95 kDa (see Chapter 4, Figure4.4), while minimally glycosylated C-domain migrates to just under 72 KDa [122]. This indicates that a small portion of mg-sACE may have been cleaved at the linker region, producing the individual domains.

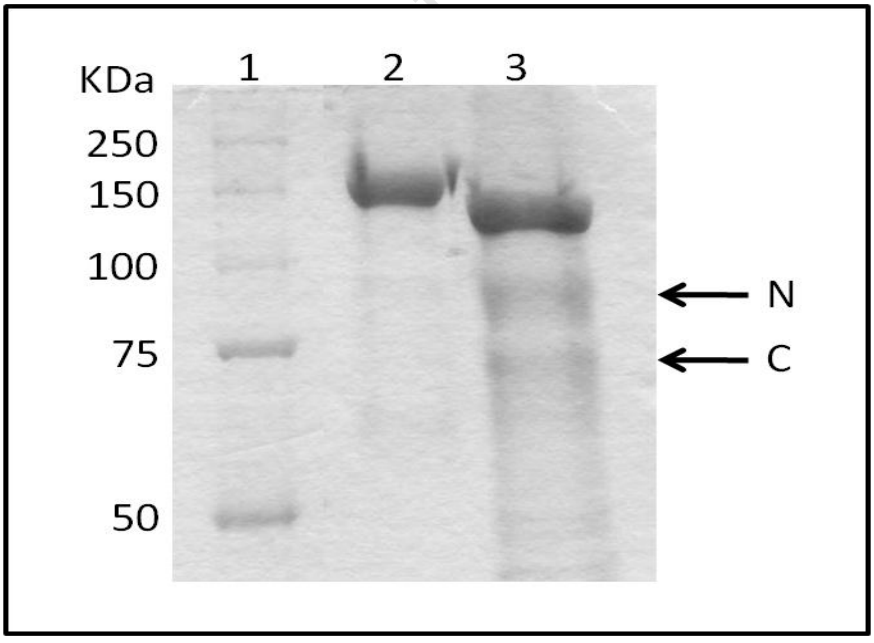


Figure 5.3: SDS-PAGE of sACE variants stained with Coomassie brilliant blue. Lower molecular weight bands corresponding to the N- and C-domains are indicated. Lane 1 - Biorad Prescision plus protein™ standard, Lane 2 - sol sACE, Lane 3 - mg-sACE. Cleaved N- and C-domain fragments are indicated.

Western blot detection of purified mg-sACE was carried out using an N-domain specific monoclonal antibody (see Appendix A5). Immunoblotting showed that mg-sACE migrated to a lower position as compared with WT sACE, consistent with the reduced level of glycosylation of this enzyme (Figure 5.4B). Interestingly, an additional band was observed between 72 and 95 kDa for mg-sACE, but not for WT sACE. This indicates that the corresponding fragment seen in the SDS-PAGE (Figure 5.3) was that of the individual N-domain, and implies that the lower band must indeed have been the C-domain.

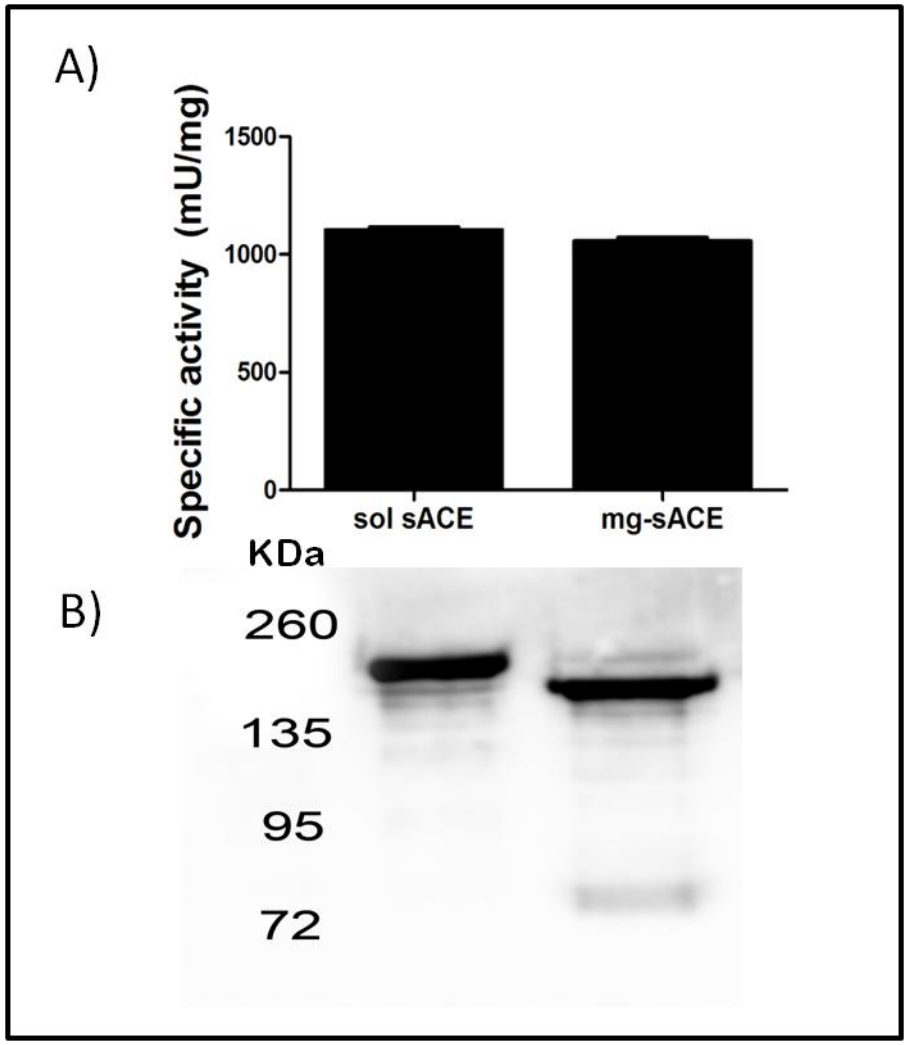
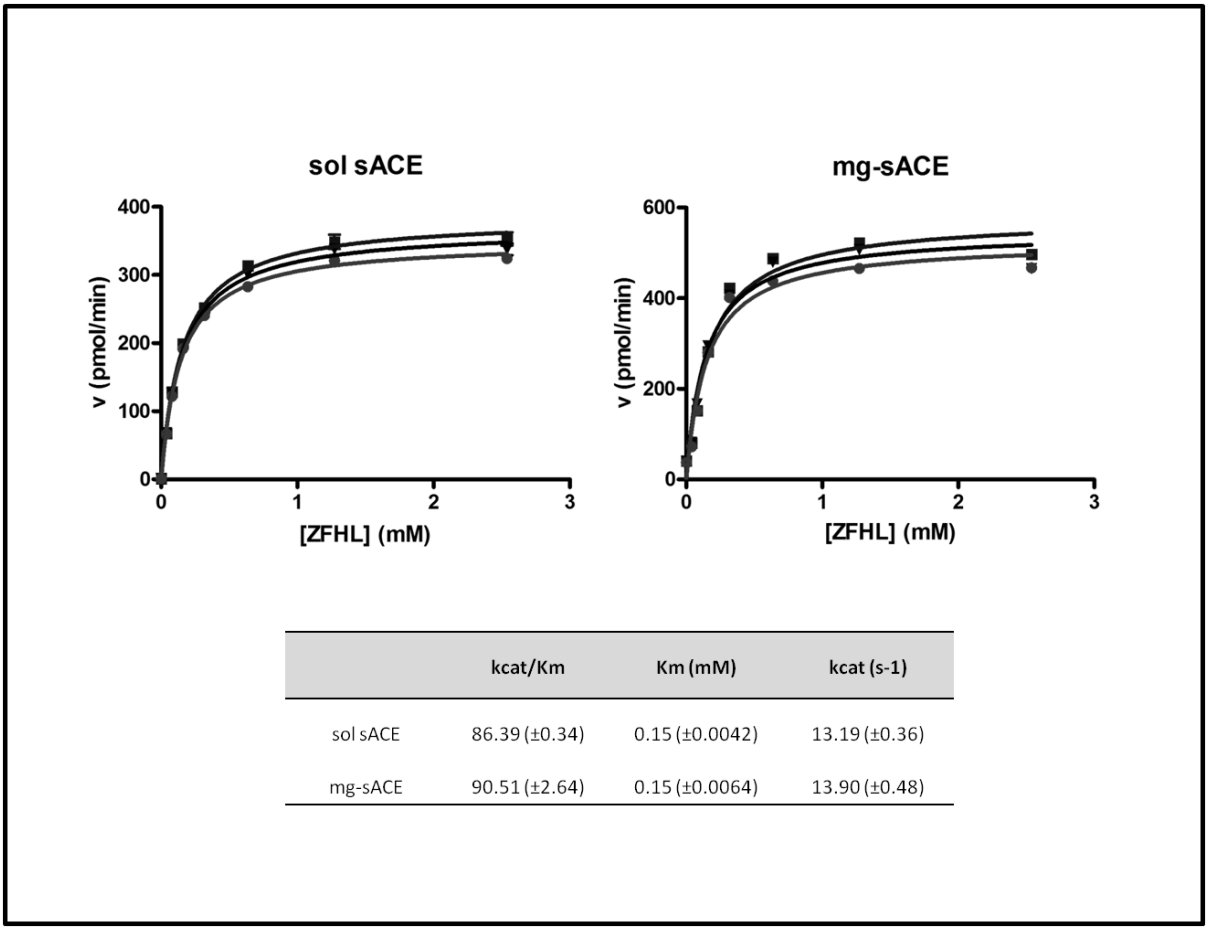


Figure 5.4 Activity and expression of mg-sACE and WT sol sACE. A) Specific activity of sACE glycosylation variants. Error bars indicate SE. B) Western blot of purified recombinant sACE proteins, probed with an anti-N-domain monoclonal antibody (4G6), Spectra™ broad range molecular marker.

5.3.3 Kinetic characterisation of mg-sACE

The kinetic parameters governing the hydrolysis of the substrate Z-FHL by sACE were determined for soluble sACE as well as for mg-sACE. Determination of the binding constants revealed identical K_m's, with both forms reporting values of 0.15 mM (Figure 5.5). These findings are consistent with those previously determined for sACE where a K_m of 0.16 mM was reported [159]. Furthermore, it was found that the turnover number and catalytic efficiency of WT sACE and mg-sACE were also highly comparable, with k_{cat} values of 13.19 and 13.90 s^{-1} and k_{cat}/K_m values of 86.39 and 90.51 $mM^{-1}s^{-1}$, respectively (Figure 5.5).



	kcat/Km	Km (mM)	kcat (s-1)
sol sACE	86.39 (±0.34)	0.15 (±0.0042)	13.19 (±0.36)
mg-sACE	90.51 (±2.64)	0.15 (±0.0064)	13.90 (±0.48)

Figure 5.5: Michaelis-Menten plot of WT sACE and mgsACE, with kinetic constants K_m, k_{cat} and k_{cat}/K_m indicated. The Michaelis-Menten kinetic constants K_m, k_{cat} and k_{cat}/K_m were determined for soluble sACE and mg-sACE using the standard ACE assay (see Appendix A6), except that final Z-FHL concentrations ranging from 0.00 to 2.54 mM were used. Assays were repeated three times in triplicate. Error bars indicate SE.

5.3.4 The effect of glycosylation on the thermal stability of sACE

The effect of reduced glycosylation on the thermal stability of sACE was assessed using the thermal inactivation assay mentioned previously (Chapter 4.2.4). Thermal inactivation of WT sACE showed that the enzyme was moderately stable, as compared with the highly stable WT N-domain, retaining approximately 20% activity after 30 min thermal inactivation (Figure 5.6), consistent with previous findings [136]. Minimally glycosylated sACE however, showed much reduced thermal stability, retaining 20% activity after only 5 min and reaching complete inactivation after approximately 15 min thermal denaturation. Not surprisingly, the thermal stability profile of mg-sACE was virtually identical to that of the minimally glycosylated N-domain (Ndom389) and C-domain (tACEg13) [136].

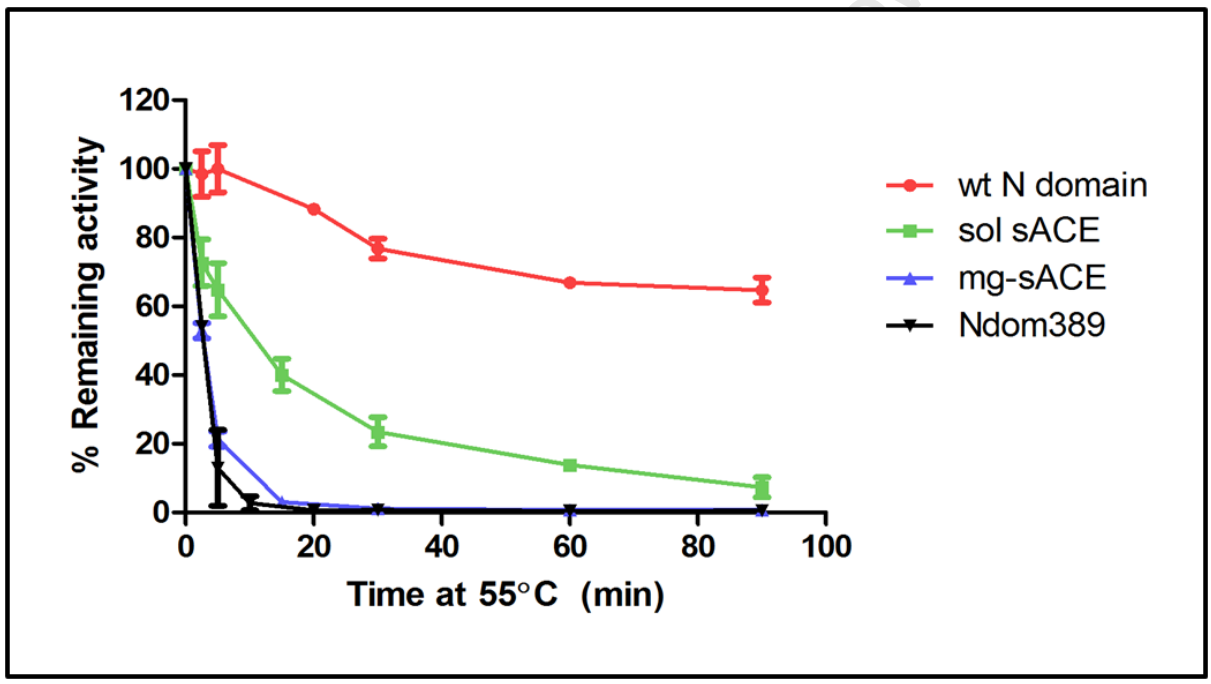


Figure 5.6: The effect of glycosylation on the thermal stability of sACE and mg-sACE as compared to the N-domain and Ndom389. Percentage residual activity as a function of time at 55 °C was determined by enzymatic assay using the ACE substrate Z-FHL (see Appendix A6). Error bars indicate SE.

5.4 Discussion

5.4.1 Cloning, expression and kinetic characterisation of minimally glycosylated sACE

The generation of a soluble sACE variant with minimal glycosylation was successfully carried out by subcloning the minimally glycosylated N-domain mutant,

Ndom389, in frame with the minimally glycosylated C-domain mutant, tACEg13. In this way, a sACE form containing only five functional glycosylation sequons, out of a potential 17, was produced. The expression of this mg-sACE variant in CHO cells resulted in the production of active enzyme. This indicates that the folding requirements (in terms of glycosylation) of full-length sACE are at least equal to the sum of those for the individual domains. Given that Yu *et al*. (1997) report that most of the surface glycosylation on sACE was on the N-domain, it is likely that the minimum glycosylation requirements for the expression of active N-domain are a major determining factor in producing an active form of minimally glycosylated sACE. Furthermore, mg-sACE and WT sACE displayed identical kinetic constants (K_m, k_{cat} and k_{cat}/K_m) indicating that the presence, or absence, of the majority of glycans on surface of sACE does not affect its primary catalytic functioning. Thus, by inference, the structure of the two forms, particularly with regard to the active site residues, is likely to be highly similar. This would imply that a structural determination of mg-sACE is likely to be representative of WT sACE.

5.4.2 The effect of glycosylation on the thermal stability of sACE

While glycosylation did not affect the activity or kinetic constants of the enzyme, it did have a major affect on its thermal stability. The thermal stability of sACE was found to be consistent with previous findings [136], retaining roughly 20% activity after 30 min thermal inactivation. This was a comparable level of thermal stability to that of WT tACE, yet somewhat lower than that of WT N-domain which retains 80% activity over the same period. However, the thermal stability of tACE was determined by inactivation at 50 °C, while thermal denaturation of N-domain and sACE in this study was carried out at 55 °C. This would indicate that sACE is in fact more stable than the individual C-domain (tACE), while still being notably less stable than the N-domain. Thus, the stability of sACE is likely between that of the N- and C-domains. The stability of mg-sACE was much lower than that of WT sACE, retaining only 20% activity after just 5 min of thermal inactivation, making the level of thermal stability of mg-sACE identical to that of the individual minimally glycosylated N- and C-domains (Ndom389 and tACEg13). This is hardly surprising, however, since mg-sACE is in fact made up of the minimally glycosylated domains Ndom389 and tACE13, which themselves have identical thermal stabilities (Chapter 4.4.2)

5.4.3 Crystallisation trails on mg-sACE

Minimally glycosylated sACE was submitted for crystallisation trials, which were carried out by our collaborator, R. Acharya's research group (Bath, UK). However, mg-sACE has thus far not produced diffractable crystals under a variety of conditions tested. This could imply that reducing glycan complexity alone is not sufficient in this case, and that inter-domain flexibility may also be a significant factor regarding the inhibition of tight crystal packing. To this end, the introduction of a cysteine residue into the linker region may facilitate the formation of a novel disulfide bridge, which in turn may be effective in stabilising the protein by restricting the inter-domain flexibility. However, it is, unlikely that this approach will be successful on its own, but introducing this disulfide bridge into a minimally glycosylated sACE variant may well be a more successful approach.

Alternatively, the failure of mg-sACE to form crystals could also be due to the presence of heterogeneous sACE populations. Non homogenous protein populations are not ideal for crystallisation experiments as these rely on multiple units of the same molecule arranging themselves in a highly ordered fashion [109-111]. SDS-PAGE and Western blot analysis revealed the presence of protein fragments at sizes comparable to those of the minimally glycosylated N- and C-domains. This is likely to have a negative effect on mg-sACE crystallisation. The cleavage of sACE at the linker region joining the two domains has been reported previously [7,8]. Previous work comparing glycosylated and unglycosylated two-domain cellulases reported that the two variants were similar in most characteristics, except that the glycosylated form was found to be resistant to interdomain cleavage, where the unglycosylated form was not [168]. Thus, it is possible that the reduced level of surface glycosylation in mg-sACE allows for an increased level of cleavage at the linker region, effectively resulting in a greater proportion of the sACE population occurring as individual domains. Examination of the N- and C-domain crystal structures revealed that two *N*-linked glycosylation sites on each domain were located close to the linker region. These include sites 6 and 9 for the N-domain and sites 2 and 3 for the C-domain [138,139]. Since only two of the four sites are present in our construct (site 9 on the N-domain and site 3 on the C-domain), the missing glycans may be responsible for protecting the linker region from proteolytic cleavage.

However, a short term solution to this problem, in terms of solving the sACE crystal structure, would be to selectively elute only the full length sACE during the protein purification step. Chen *et al*. (2010) report that they were able to separate individual N- and C-domains from a sample containing full length sACE by altering the buffer and pH during elution from the sepharose-lisinopril affinity column. According to their strategy, it was possible to elute the N- and C-domains with 20mM Mes-NaOH (pH 6) and 20 mM HEPES (pH 6) respectively, before finally eluting the remaining full length sACE proteins with 50 mM borate (pH9.5). The selective elution of full length mg-sACE may result in an improved outcome for crystallisation trials on this enzyme.

It is interesting to note that the kinetic constants of mg-sACE were comparable to those of sACE, despite the increased level of cleavage in the linker region producing a sub population of individual N- and C-domains. One reason for this is that although a portion of mg-sACE appears to be cleaved into its individual domains, it is likely that these domains remain active [7,8,167]. Thus, the individual domains are still able to contribute to substrate cleavage. In addition, the proportion of mg-sACE molecules cleaved to the individual domains, while large enough to potentially inhibit crystallisation, may not have been large enough to greatly affect the kinetics of mg-sACE.

Conclusions and future work

Mass spectrometry based profiling of the glycosylation site occupancy of the N-domain glycosylation sequons revealed that all of the potential glycosylation sites on the N-domain do in fact carry glycosylation, with the exception of site 10 (N494). Site 10 was not expected to be glycosylated given that it is not located on the surface, but buried in the protein, as well as the fact that its Asn-Xaa-Ser/Thr sequon is followed by a Pro, a factor known to prevent glycosylation of a particular sequon [56].

Furthermore, we have determined that glycosylation of the C-terminal region of the N-domain is critical for expression of active protein, and that the presence of three C-terminal glycans is sufficient for expression of active protein. While we found that the presence of at least three C-terminal glycosylation sites were necessary for the expression of active protein, we also discovered that there was some redundancy with regard to which sites were necessary. This is based on our finding that an N-terminal site (site 3) was able to compensate for the loss of glycosylation at a C-terminal site, such that both the Ndom789 and Ndom389 glycosylation mutants were enzymatically active.

Recent reports have shown that glycosylation contributes significantly to the thermal stability of the C-domain of ACE [136]. Our findings indicate that the thermal stability of the N-domain is also significantly affected by the presence of surface glycosylation, with decreased levels of glycosylation corresponding to decreased thermal stability. Moreover, the presence of glycosylation at sites 2 and 3 are critical for the maintenance of high levels of thermal stability, with a loss of glycosylation at either site proving severely detrimental to the thermal stability of the N-domain.

In contrast to its critical role in thermal stability, glycosylation appeared to have little or no effect on the enzymatic activity of the N-domain, with specific activities being similar between the various glycosylation mutants and the WT protein. Furthermore, Michaelis-Menten constants for Ndom389 and WT N-domain were also comparable.

The identification of the minimum glycosylation requirements for the production of active N-domain has led to the generation of an N-domain variant that is able to crystallise in a reproducible manner, which was previously not achievable. The value of having a readily crystallisable form of the N-domain cannot be overstated, as it allows high throughput inhibitor-enzyme crystallisation studies. Indeed, the crystal structure of the minimally glycosylated N-domain variant (Ndm389) has been solved in the presence of an N-domain selective inhibitor (RXP407) [123]. This structure provides further insight into the binding interactions that contribute to the domain selectivity of this inhibitor, as well as demonstrating the effectiveness of this approach for the crystallisation of glycoproteins.

The fact that mg-sACE was expressed in an active form revealed that the glycosylation requirements for full length sACE were at least equal to the sum of the requirements for the two individual domains. Previous findings indicate that the majority of the glycans on sACE are located on the N-domain [112,124]. Thus, it may well be that the glycosylation requirements for the expression of sACE are determined by the glycosylation requirements of the N-domain.

While glycosylation is important for the expression of active sACE, its effect on the catalytic functioning of the enzyme was found to be negligible since both soluble sACE and mg-sACE had similar binding affinities, turnover rates and catalytic efficiencies for the hydrolysis of Z-FH. Although the minimally glycosylated form of sACE retained its catalytic properties, the reduced level of protein glycosylation resulted in a marked decrease in thermal stability and was also most likely responsible for the increased level of cleavage in the linker region of mg-sACE.

The mg-sACE variant has thus far not been able to produce crystals under the conditions tested, indicating the need to revisit the design of the minimally glycosylated sACE variant. Given the high degree of thermal stability of Ndom2379 (Chapter 4.3.2), it may be worthwhile generating a minimally glycosylated sACE variant using this N-domain glycoform, in an attempt to increase its thermal stability. Additionally, generating N-domain variants Ndom237, Ndom238 and Ndom239 for insertion into mg-sACE may prove useful by allowing the expression of sACE

proteins with increased stability, yet having only a few intact glycosylation sites. As mentioned in Chapter 5.4.2, the introduction of a novel disulfide bridge into the linker region of such a glycosylation variant of sACE, may result in a thermally stable, minimally glycosylated variant which also has reduced inter-domain flexibility. This approach targets three important factors currently hampering the crystallisation of sACE, and is an ideal candidate to enable the solution of the presently elusive sACE crystal structure. However, while reduced glycosylation, increased thermal stability and reduced interdomain flexibility are important constraints, the presence of an increased level of proteolytic cleavage at the linker region joining the two domains of mg-sACE is a matter that must be addressed. Given that sites 6 and 9 on the N-domain, and sites 2 and 3 on the C-domain, are located in close proximity to the linker region, including these sites in a minimally glycosylated sACE mutant may restore protection against the proteolytic cleavage of the linker region.

Finally, this work has resulted in the creation of a number of N-domain constructs with varying degrees of intact glycosylation sequons. A similar group of C-domain mutants has previously been generated [122]. Combining the N- and C-domain glycosylation variants, through the cloning strategy used to make mg-sACE, will result in the creation of a library of sACE glycosylation mutants. These sACE glycosylation mutants would be a useful resource to investigate the effects of glycosylation on protein-protein interactions, such as ACE dimerisation and the proposed link between glycan mediated dimerisation and the biological process of ACE shedding from the cell surface [126,130].

Appendix

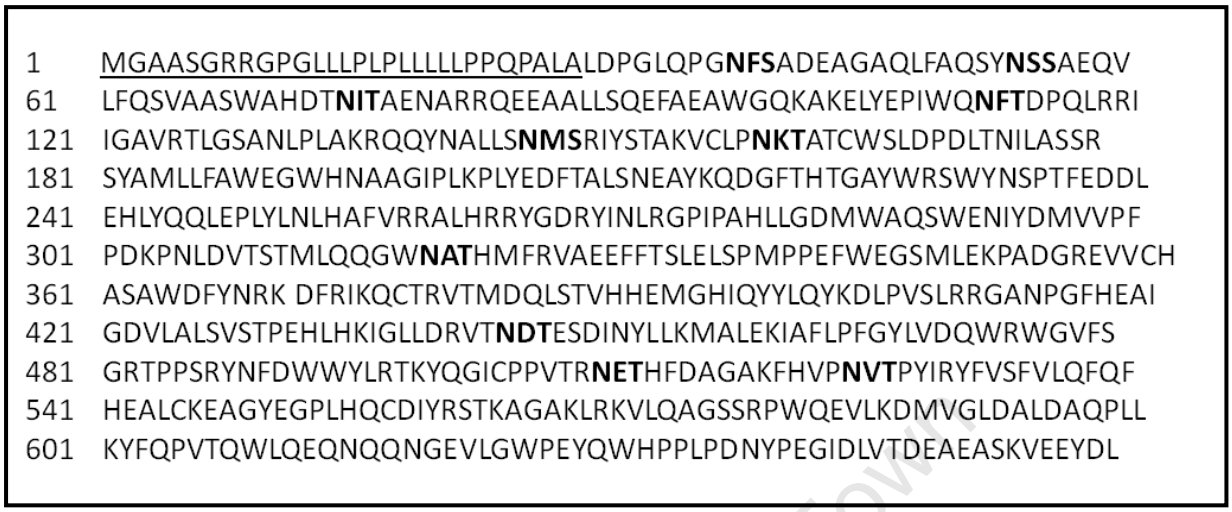


Figure A1: Amino acid sequence of recombinant N-domain. Signal peptide targeting the enzyme for secretion is underlined (not present in the mature form of the enzyme). *N*-linked glycosylation sequons highlighted in bold.

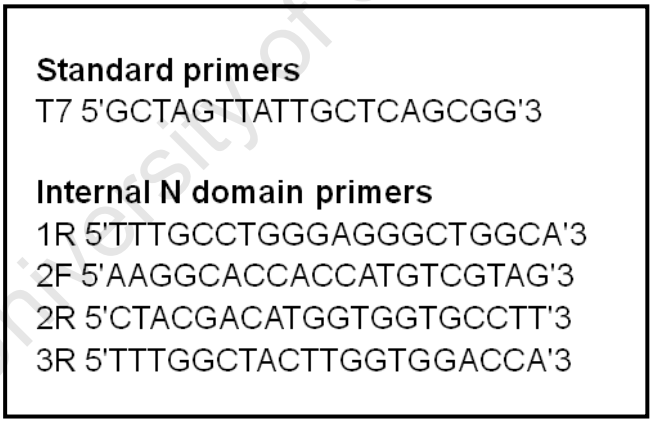


Figure A2: Primers used for the nucleotide sequencing of the N-domain.

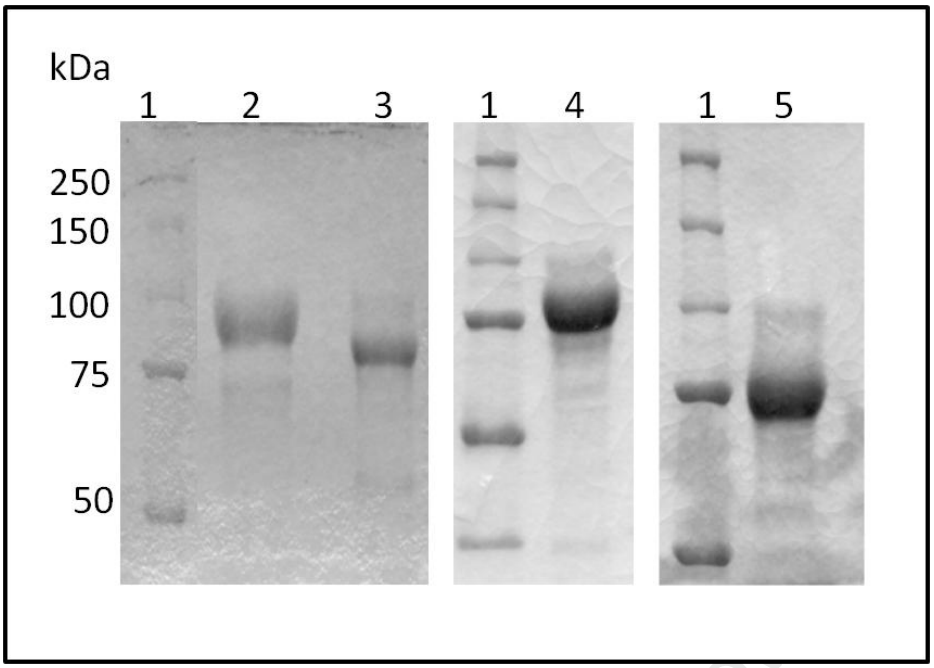


Figure A3: Representative SDS-PAGE gels stained with Coomassie, showing successful purification of ACE N-domain proteins. Lane 1 - Biorad Prescision plus protein™ standard. Lane 2 - WT N-domain, Lane 3 – Ndom1234569, Lane 4 – Ndom12379, Lane 5 – Ndom1279, Lane 6 – Ndom1237, Lane 7- Ndom389. N-domain variants migrated to different sizes due to the presence of differing levels of intact glycosylation sites.

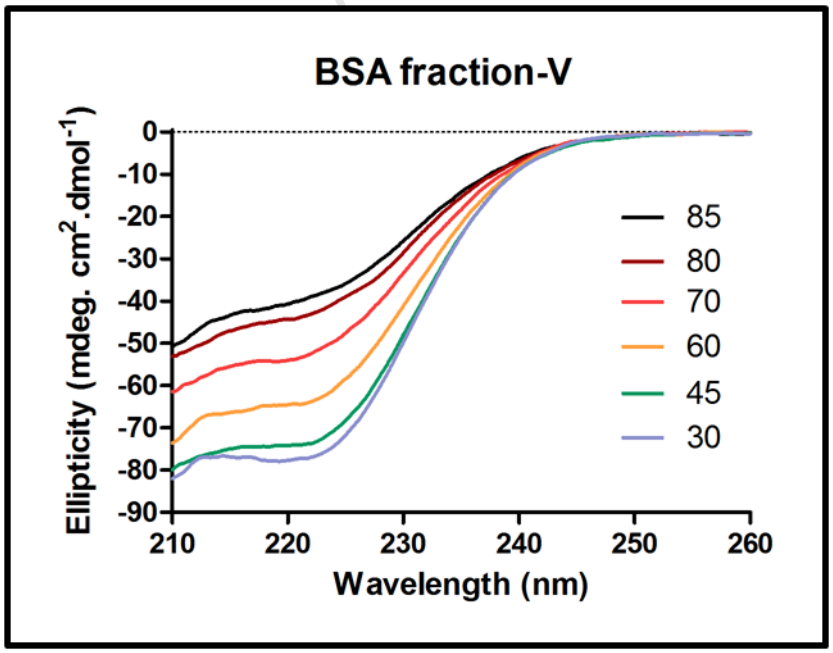


Figure A4: Circular dichroism spectra of BSA variants showing temperature induced denaturation. Unfolding transitions for WT N-domain, Ndom2379, Ndom1279 and Ndom389 are shown over 210 to 260 nm. Scans for selected temperatures shown.

A1. Preparation of competent *E.coli* DH5α cells [169]

An over-night culture of *E.coli* DH5α cells was inoculated into 50 ml of Luria broth (LB) *(1% (w/v) tryptone; 5% (w/v) yeast extract; 0.17 M NaCl)* and incubated at 37 °C until an OD_{595} of between 0.4 and 0.6 was reached. The culture was chilled on ice for 15 min. Cells were then collected by centrifugation at 6000 RPM for 5 min. The cell pellet was resuspended in 1/10 original volume of pre chilled TFBI *(30 mM KOAc; 100 mM RbCl; 10 mM CaCl2; 50 mM MnCl; 15% (v/v) glycerol; adjusted to pH 5.8 with glacial acetic acid).* The resuspended cells were incubated on ice for 15 min before centrifugation at 6000 RPM for 5 min. The cell pellet was resuspended in 1/25 original volume of pre-chilled TFBII *(10 mM MOPS; 10 mM $RbCl_2$; 75 mM CaCl; 15% (v/v) glycerol; adjusted to pH 6.5 with glacial acetic acid)*. The final resuspension was divided into 100 µl fractions and stored at -80 °C until use.

A2. Transformation of competent cells [169]

50 to 100 ng plasmid DNA was added to 100 µl of prepared competent cells were thawed on ice. The DNA-competent cells mixture was incubated on ice for at least 25 min before heat shock of 42 °C for 0.5 min. Heat shocked, cells were immediately placed on ice for 1 min, following which 900 µl LB was added to the cells. Transformed cells were grown at 37 °C for 60 min with shaking. 100 µl of the transformation was plated on Luria agar plates *(1.2% (w/v) agar, 1% (w/v) tryptone; 5% (w/v) yeast extract; 0.17 M NaCl)* containing 50 µg/ml ampicillin.

A3. Restriction enzyme digestion

Restriction enzyme digests were carried out in 20 µl reactions on 3 µl crude DNA preparation or 500 ng purified DNA, for 90 min at 37 °C. Relevant reaction buffers at 1x or 2x concentration were used as per the manufacturers instruction. Restriction enzymes were used at between 1 and 10 units per reaction, according to the manufacturers guidlines.

A4. SDS-polyacrylamide gel electrophoresis (SDS-PAGE) [170]

SDS-polyacrylamide gel electrophoresis (SDS-PAGE) was carried out as per standard laboratory protocols. Briefly, 20 µl of sample was mixed with 5 µl 5x sample buffer *(50% (v/v) glycerol, 25% (v/v) β-mercaptoethanol, 15% (w/v) SDS, 0.31 M Tris,*

0.25% (w/v) bromophenol blue) and denatured at 100 ˚C for 5 min. Denatured samples were separated on 10% polyacrylamide gels made up in TBE buffer *(89 mM Tris base, 89 mM boric acid, 2 mM EDTA)* with 0.1 % (w/v) ammonium persulphate (AMPS) and 10 µl tetramethylethylenediamine (TEMED). Electrophoresis was carried out at 25 mA/gel for 1 hr.

A5. Immunoblotting of purified proteins [169]

Following protein separation by SDS-PAGE, proteins were transferred to Hybond™-ECL nitrocellulose membranes (Amersham, Buckinghamshire, UK) in Blotting buffer *(0.5 M Tris, 1.44% (w/v) glycine, 0.2% (v/v) methanol)* at 100 V for 1 hr, using the wet gel transfer method. Membranes were incubated in 1x TBST *(50 mM Tris-HCl pH 7.4, 0.2 M NaCl, 0.1% (v/v) Tween-20)* containing 5% skim milk for 1 hr with shaking, to block unoccupied binding sites and prevent non-specific binding during the antibody incubation steps. Membranes were incubated in rat anti-ACE primary antibody (4G6) (kind donation from S. Danilov) diluted 1:200 in 1x TBST containing 5% skim milk for 1.5 hrs. Unbound antibody was removed by washing 3x 5 min with 1x TBST. Membranes were incubated in goat anti-rat peroxidase-conjugated secondary antibody, diluted 1:2000 in 1x TBST containing 5% skim milk for 1 hr. Unbound antibody was removed by washing 3x 5 min with 1x TBST. ACE was detected using the ECL Plus Western Blot Detection kit (Amersham) and visualised on a G:Box iChemi™ chemiluminescence imager (Syngene, Cambridge, UK) and analysed using the GeneSnap™ software package (Syngene).

A6. ACE activity assay [171,172]

A6.1 Z-FHL working solution

A 20 mM stock solution of *Z*-FHL was prepared by dissolving 220 mg *Z*-FHL in 2 mL 0.28 M NaOH, and made up to 20 mL with dH_2O. The 1 mM Z-FHL working solution was made up by adding 15 ml dH_2O to 4 ml 5x phosphate buffer *(0.5 M potassium phosphate, pH 8.3, 1.5 M NaCl)* with 20 µl 10 mM $ZnSO_4$, lastly 1 ml of 20 mM *Z*-FHL was added.

A6.2 Hippuryl-His-Leu standard curve

Hippuryl-His-Leu (HL) standards were prepared from a 5.7 mM stock solution in 1x phosphate buffer, pH 8.3, at concentrations of 0; 0.037; 0.074; 0.148; 0.296; 0.592

nmoles/μl HL. 35 μl of each standard were added to a 96-well fluorometric plate in triplicate. To this, 120 μl 0.4 N NaOH was added. HL was derivatised by the addition of 10.0 μl 20mg/ml *O*-pthaldialdehyde and incubated for 10 min. Derivatisation was terminated by the adition of 30.0 μl 3N HCL.

A6.3 Z-FHL activity assay

Assays were carried out, using three sets of triplicate repeats for each sample. Briefly, 30 µl Z-FHL working solution was aliquoted into a 96 well plate and pre-incubated to 37 °C. To this, 5 µl sample (pre warmed to 37 °C) was added, and the reaction incubated at 37˚C for 15 min to allow enzymatic cleavage of the Z-FHL substrate (no sample was added to the blank-time-zero (BZT)). Enzymatic cleavage was stopped by the addition of 120 µl 0.4 M NaOH and 5 µl sample was added to the BZT. Cleaved HL product was derivatised by the addition of 10 µl O-phthaldialdehyde (20mg/ml in methanol). Derivatisation was terminated after 10 min by the addition of 3N HCl. Fluorescence of the 195 µl of reaction was measured at excitation-emission wavelengths of λ_{Ex}=360 nm and λ_{Em}=485 nm, on a Cary Eclipse fluorimeter (Varian, CA, USA). The measured fluorescence readings were converted to activity (mU) using the HL standard curve (Figure A5). ACE activity (mU) is defined as the number of nmols of HL product released per ml of sample, per minute (nmol/ml/min).

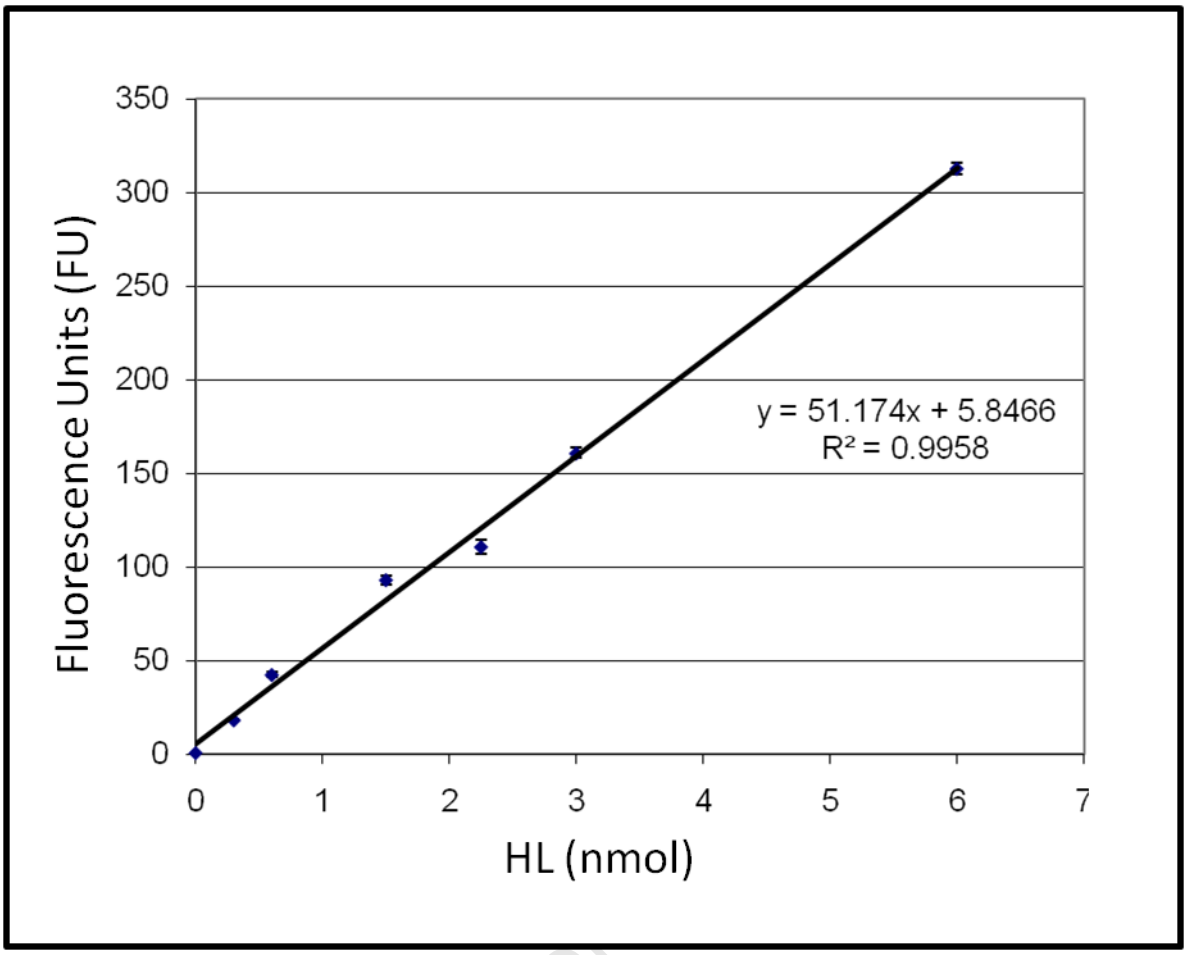


Figure A5: HL standard curve showing correlation between fluorescence intensity and different nmol of HL.

References

[1] K. R. Acharya, *et al.*, (2003) "Ace revisited: a new target for structure-based drug design", Nat. Rev. Drug Discov. **2** (11), 891 .

[2] F. Soubrier, *et al.*, (1988) "Two putative active centers in human angiotensin I-converting enzyme revealed by molecular cloning", Proc. Natl. Acad. Sci. U. S. A **85** (24), 9386 .

[3] A. J. Turner and N. M. Hooper, (2002) "The angiotensin-converting enzyme gene family: genomics and pharmacology", Trends Pharmacol. Sci. **23** (4), 177 .

[4] C. R. Esther, *et al.*, (1997) "The critical role of tissue angiotensin-converting enzyme as revealed by gene targeting in mice", J. Clin. Invest **99** (10), 2375 .

[5] L. Wei, *et al.*, (1991) "The two homologous domains of human angiotensin I-converting enzyme are both catalytically active", J Biol. Chem. **266** (14), 9002 .

[6] M. R. Ehlers, *et al.*, (1989) "Molecular cloning of human testicular angiotensin-converting enzyme: the testis isozyme is identical to the C-terminal half of endothelial angiotensin-converting enzyme", Proc. Natl. Acad. Sci. U. S. A **86** (20), 7741 .

[7] P. A. Deddish, *et al.*, (1994) "Naturally occurring active N-domain of human angiotensin I-converting enzyme", Proc. Natl. Acad. Sci. U. S. A **91** (16), 7807 .

[8] E. D. Sturrock, S. M. Danilov, and J. F. Riordan, (1997) "Limited proteolysis of human kidney angiotensin-converting enzyme and generation of catalytically active N- and C-terminal domains", Biochem. Biophys. Res. Commun. **236** (1), 16 .

[9] D. Coates, (2003) "The angiotensin converting enzyme (ACE)", Int. J. Biochem. Cell Biol. **35** (6), 769 .

[10] C. Hubert, *et al.*, (1991) "Structure of the angiotensin I-converting enzyme gene. Two alternate promoters correspond to evolutionary steps of a duplicated gene", J Biol. Chem. **266** (23), 15377 .

[11] M. R. Ehlers, Y. N. Chen, and J. F. Riordan, (1992) "The unique N-terminal sequence of testis angiotensin-converting enzyme is heavily O-glycosylated and unessential for activity or stability", Biochem. Biophys. Res. Commun. **183** (1), 199 .

[12] J. H. Krege, *et al.*, (1995) "Male-female differences in fertility and blood pressure in ACE-deficient mice", Nature **375** (6527), 146 .

[13] L. Wei, *et al.*, (1991) "Expression and characterization of recombinant human angiotensin I-converting enzyme. Evidence for a C-terminal transmembrane anchor and for a proteolytic processing of the secreted recombinant and plasma enzymes", J Biol. Chem. **266** (9), 5540 .

[14] T. Unger, (2002) "The role of the renin-angiotensin system in the development of cardiovascular disease", Am. J Cardiol. **89** (2A), 3A .

[15] P. M. Kearney, *et al.*, (2005) "Global burden of hypertension: analysis of worldwide data", Lancet **365** (9455), 217 .

[16] World Health Organization, (2003) "Global Strategy on Diet, Physical Activity and Health. Cardiovascular Disease: Prevention and Control",

[17] J. He and P. K. Whelton, (1997) "Epidemiology and prevention of hypertension", Med. Clin. North Am. **81** (5), 1077 .

[18] P. K. Whelton, (1994) "Epidemiology of hypertension", Lancet **344** (8915), 101 .

[19] S. Fuchs, *et al.*, (2008) "Angiotensin-converting enzyme C-terminal catalytic domain is the main site of angiotensin I cleavage in vivo", Hypertension **51** (2), 267 .

[20] C. Junot, *et al.*, (2001) "RXP 407, a selective inhibitor of the N-domain of angiotensin I-converting enzyme, blocks in vivo the degradation of hemoregulatory peptide acetyl-Ser-Asp-Lys-Pro with no effect on angiotensin I hydrolysis", J. Pharmacol. Exp. Ther. **297** (2), 606 .

[21] J. Cotton, *et al.*, (2002) "Selective inhibition of the C-domain of angiotensin I converting enzyme by bradykinin potentiating peptides", Biochemistry **41** (19), 6065 .

[22] S. Fuchs, *et al.*, (2004) "Role of the N-terminal catalytic domain of angiotensin-converting enzyme investigated by targeted inactivation in mice", J. Biol. Chem. **279** (16), 15946 .

[23] K. E. Bernstein, *et al.*, (2010) "Different in vivo functions of the two catalytic domains of angiotensin-converting enzyme (ACE)", Curr. Opin. Pharmacol.

[24] W. L. Kroger, *et al.*, (2009) "Investigating the domain specificity of phosphinic inhibitors RXPA380 and RXP407 in angiotensin-converting enzyme", Biochemistry **48** (35), 8405 .

[25] K. Dickstein and J. Kjekshus, (2002) "Effects of losartan and captopril on mortality and morbidity in high-risk patients after acute myocardial infarction: the OPTIMAAL randomised trial. Optimal Trial in Myocardial Infarction with Angiotensin II Antagonist Losartan", Lancet **360** (9335), 752 .

[26] D. Georgiadis, *et al.*, (2003) "Roles of the two active sites of somatic angiotensin-converting enzyme in the cleavage of angiotensin I and bradykinin: insights from selective inhibitors", Circ. Res. **93** (2), 148 .

[27] M. R. Ehlers, (2006) "Safety issues associated with the use of angiotensin-converting enzyme inhibitors", Expert. Opin. Drug Saf **5** (6), 739 .

[28] A. Rousseau, *et al.*, (1995) "The hemoregulatory peptide N-acetyl-Ser-Asp-Lys-Pro is a natural and specific substrate of the N-terminal active site of human angiotensin-converting enzyme", **270** (8), 3656 .

[29] P. A. Deddish, *et al.*, (1998) "N-domain-specific substrate and C-domain inhibitors of angiotensin-converting enzyme: angiotensin-(1-7) and keto-ACE", Hypertension **31** (4), 912 .

[30] K. Zou, *et al.*, (2009) "Abeta42-to-Abeta40- and angiotensin-converting activities in different domains of angiotensin-converting enzyme", J. Biol. Chem. **284** (46), 31914 .

[31] K. Zou, *et al.*, (2007) "Angiotensin-converting enzyme converts amyloid beta-protein 1-42 (Abeta(1-42)) to Abeta(1-40), and its inhibition enhances brain Abeta deposition", J. Neurosci. **27** (32), 8628 .

[32] K. Shah, *et al.*, (2009) "Does use of antihypertensive drugs affect the incidence or progression of dementia? A systematic review", Am. J. Geriatr. Pharmacother. **7** (5), 250 .

[33] M. R. Ehlers and J. F. Riordan, (1991) "Angiotensin-converting enzyme: zinc- and inhibitor-binding stoichiometries of the somatic and testis isozymes", Biochemistry **30** (29), 7118 .

[34] R. A. Skidgel and E. G. Erdos, (1985) "Novel activity of human angiotensin I converting enzyme: release of the NH2- and COOH-terminal tripeptides from the luteinizing hormone-releasing hormone", Proc. Natl. Acad. Sci. U. S. A **82** (4), 1025 .

[35] A. Papakyriakou, *et al.*, (2007) "Simulated interactions between angiotensin-converting enzyme and substrate gonadotropin-releasing hormone: novel insights into domain selectivity", Biochemistry **46** (30), 8753 .

[36] D. Bonnet, *et al.*, (1992) "Reversible inhibitory effects and absence of toxicity of the tetrapeptide acetyl-N-Ser-Asp-Lys-Pro (AcSDKP) in human long-term bone marrow culture", Exp. Hematol. **20** (10), 1165 .

[37] U. Sharma, *et al.*, (2008) "Novel anti-inflammatory mechanisms of N-Acetyl-Ser-Asp-Lys-Pro in hypertension-induced target organ damage", Am. J. Physiol Heart Circ. Physiol **294** (3), H1226-H1232 .

[38] Y. H. Liu, *et al.*, (2009) "N-acetyl-seryl-aspartyl-lysyl-proline prevents cardiac remodeling and dysfunction induced by galectin-3, a mammalian adhesion/growth-regulatory lectin", Am. J. Physiol Heart Circ. Physiol **296** (2), H404-H412 .

[39] G. Castoldi, *et al.*, (2010) "Prevention of myocardial fibrosis by N-acetyl-seryl-aspartyl-lysyl-proline in diabetic rats", Clin. Sci. (Lond) **118** (3), 211 .

[40] H. Peng, *et al.*, (2007) "Role of N-acetyl-seryl-aspartyl-lysyl-proline in the antifibrotic and anti-inflammatory effects of the angiotensin-converting enzyme inhibitor captopril in hypertension", Hypertension **49** (3), 695 .

[41] M. Azizi, *et al.*, (1996) "Acute angiotensin-converting enzyme inhibition increases the plasma level of the natural stem cell regulator N-acetyl-seryl-aspartyl-lysyl-proline", J. Clin. Invest **97** (3), 839 .

[42] M. Azizi, *et al.*, (1997) "High plasma level of N-acetyl-seryl-aspartyl-lysyl-proline: a new marker of chronic angiotensin-converting enzyme inhibition", Hypertension **30** (5), 1015 .

[43] D. P. Bicket, (2002) "Using ACE inhibitors appropriately", Am. Fam. Physician **66** (3), 461 .

[44] D. M. Coulter and I. R. Edwards, (1987) "Cough associated with captopril and enalapril", Br. Med. J. (Clin. Res. Ed) **294** (6586), 1521 .

[45] E. E. Slater, *et al.*, (1988) "Clinical profile of angioedema associated with angiotensin converting-enzyme inhibition", JAMA **260** (7), 967 .

[46] C. Speirs, F. Wagniart, and L. Poggi, (1998) "Perindopril postmarketing surveillance: a 12 month study in 47,351 hypertensive patients", Br. J. Clin. Pharmacol. **46** (1), 63 .

[47] A. Adam, *et al.*, (2002) "Aminopeptidase P in individuals with a history of angiooedema on ACE inhibitors", Lancet **359** (9323), 2088 .

[48] D. Georgiadis, *et al.*, (2004) "Structural determinants of RXPA380, a potent and highly selective inhibitor of the angiotensin-converting enzyme C-domain", Biochemistry **43** (25), 8048 .

[49] A. Yan and W. J. Lennarz, (2005) "Unraveling the mechanism of protein N-glycosylation", J Biol. Chem. **280** (5), 3121 .

[50] A. Helenius and M. Aebi, (2004) "Roles of N-linked glycans in the endoplasmic reticulum", Annu. Rev. Biochem. **73**, 1019 .

[51] R. Apweiler, H. Hermjakob, and N. Sharon, (1999) "On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database", Biochim. Biophys. Acta **1473** (1), 4 .

[52] N. Mitra, *et al.*, (2006) "*N*-linked oligosaccharides as outfitters for glycoprotein folding, form and function", Trends Biochem. Sci. **31** (3), 156 .

[53] H. J. An and C. B. Lebrilla, (2010) "Structure elucidation of native N- and O-linked glycans by tandem mass spectrometry (tutorial)", Mass Spectrom. Rev.

[54] B. Imperiali and S. E. O'Connor, (1999) "Effect of N-linked glycosylation on glycopeptide and glycoprotein structure", Curr. Opin. Chem. Biol. **3** (6), 643 .

[55] C. Ronin, *et al.*, (1981) "Synthetic substrates for thyroid oligosaccharide transferase. Effects of peptide chain length and modifications in the Asn-Xaa-Thr-region", Eur. J. Biochem. **118** (1), 159 .

[56] T. Roitsch and L. Lehle, (1989) "Structural requirements for protein N-glycosylation. Influence of acceptor peptides on cotranslational glycosylation of yeast invertase and site-directed mutagenesis around a sequon sequence", Eur. J. Biochem. **181** (2), 525 .

[57] S. Kasturi, *et al.*, (1994) "Role of glycosylation in the biosynthesis and activity of rabbit testicular angiotensin-converting enzyme", Biochemistry **33** (20), 6228 .

[58] J. L. Mellquist, *et al.*, (1998) "The amino acid following an asn-X-Ser/Thr sequon is an important determinant of N-linked core glycosylation efficiency", Biochemistry **37** (19), 6833 .

[59] C. Kellenberger, T. L. Hendrickson, and B. Imperiali, (1997) "Structural and functional analysis of peptidyl oligosaccharyl transferase inhibitors", Biochemistry **36** (41), 12554 .

[60] S. Peluso, *et al.*, (2002) "Neoglycopeptides as inhibitors of oligosaccharyl transferase: insight into negotiating product inhibition", Chem. Biol. **9** (12), 1323 .

[61] P. D. Eason and B. Imperiali, (1999) "A potent oligosaccharyl transferase inhibitor that crosses the intracellular endoplasmic reticulum membrane", Biochemistry **38** (17), 5430 .
[62] M. Igura and D. Kohda, (2010) "Quantitative assessment of the preferences for the amino acid residues flanking archaeal N-linked glycosylation sites", Glycobiology .
[63] M. Bano-Polo*, et al.*, (2011) "N-glycosylation efficiency is determined by the distance to the C-terminus and the amino acid preceding an Asn-Ser-Thr sequon", Protein Sci. **20** (1), 179 .
[64] L. W. Bergman and W. M. Kuehl, (1979) "Formation of an intrachain disulfide bond on nascent immunoglobulin light chains", J. Biol. Chem. **254** (18), 8869 .
[65] W. Chen*, et al.*, (1995) "Cotranslational folding and calnexin binding during glycoprotein synthesis", Proc. Natl. Acad. Sci. U. S. A **92** (14), 6229 .
[66] R. Daniels*, et al.*, (2003) "N-linked glycans direct the cotranslational folding pathway of influenza hemagglutinin", Mol. Cell **11** (1), 79 .
[67] A. Helenius, (1994) "How N-linked oligosaccharides affect glycoprotein folding in the endoplasmic reticulum", Mol. Biol. Cell **5** (3), 253 .
[68] A. Helenius and M. Aebi, (2001) "Intracellular functions of N-linked glycans", Science **291** (5512), 2364 .
[69] J. C. Paulson, (1989) "Glycoproteins: what are the sugar chains for?", Trends Biochem. Sci. **14** (7), 272 .
[70] M. R. Wormald and R. A. Dwek, (1999) "Glycoproteins: glycan presentation and protein-fold stability", Structure. **7** (7), R155-R160 .
[71] B. Imperiali and K. W. Rickert, (1995) "Conformational implications of asparagine-linked glycosylation", Proc. Natl. Acad. Sci. U. S. A **92** (1), 97 .
[72] C. R. Matthews, (1993) "Pathways of protein folding", Annu. Rev. Biochem. **62**, 653 .
[73] S. E. O'Connor and B. Imperiali, (1996) "Modulation of protein structure and function by asparagine-linked glycosylation", Chem. Biol. **3** (10), 803 .
[74] S. C. Hubbard and P. W. Robbins, (1979) "Synthesis and processing of protein-linked oligosaccharides in vivo", J. Biol. Chem. **254** (11), 4568 .
[75] R. Kornfeld and S. Kornfeld, (1985) "Assembly of asparagine-linked oligosaccharides", Annu. Rev. Biochem. **54**, 631 .
[76] H. Hettkamp, G. Legler, and E. Bause, (1984) "Purification by affinity chromatography of glucosidase I, an endoplasmic reticulum hydrolase involved in the processing of asparagine-linked oligosaccharides", Eur. J. Biochem. **142** (1), 85 .
[77] D. Brada and U. C. Dubach, (1984) "Isolation of a homogeneous glucosidase II from pig kidney microsomes", Eur. J. Biochem. **141** (1), 149 .
[78] C. Hammond, I. Braakman, and A. Helenius, (1994) "Role of N-linked oligosaccharide recognition, glucose trimming, and calnexin in glycoprotein folding and quality control", Proc. Natl. Acad. Sci. U. S. A **91** (3), 913 .
[79] E. S. Trombetta and A. Helenius, (1998) "Lectins as chaperones in glycoprotein folding", Curr. Opin. Struct. Biol. **8** (5), 587 .
[80] E. S. Trombetta and A. Helenius, (2000) "Conformational requirements for glycoprotein reglucosylation in the endoplasmic reticulum", J Cell Biol. **148** (6), 1123 .
[81] E. S. Trombetta, (2003) "The contribution of N-glycans and their processing in the endoplasmic reticulum to glycoprotein biosynthesis", Glycobiology **13** (9), 77R .
[82] K. Olden, J. B. Parent, and S. L. White, (1982) "Carbohydrate moieties of glycoproteins. A re-evaluation of their function", Biochim. Biophys. Acta **650** (4), 209 .
[83] J. M. Withka*, et al.*, (1993) "Structure of the glycosylated adhesion domain of human T lymphocyte glycoprotein CD2", Structure. **1** (1), 69 .
[84] J. E. Ripka*, et al.*, (1993) "N-glycosylation of forms of angiotensin converting enzyme from four mammalian species", Biochem. Biophys. Res Commun. **196** (2), 503 .
[85] J. Fujihara*, et al.*, (2008) "Two *N*-linked glycosylation sites (Asn18 and Asn106) are both required for full enzymatic activity, thermal stability, and resistance to proteolysis in mammalian deoxyribonuclease I", Biosci. Biotechnol. Biochem. **72** (12), 3197 .

[86] P. H. van Berkel*, et al.*, (1995) "Glycosylated and unglycosylated human lactoferrins both bind iron and show identical affinities towards human lysozyme and bacterial lipopolysaccharide, but differ in their susceptibilities towards tryptic proteolysis", Biochem. J **312 (Pt 1)**, 107 .

[87] K. A. Kretz*, et al.*, (1990) "Characterization of a mutation in a family with saposin B deficiency: a glycosylation site defect", Proc. Natl. Acad. Sci. U. S. A **87** (7), 2541 .

[88] B. A. Bernard, K. M. Yamada, and K. Olden, (1982) "Carbohydrates selectively protect a specific domain of fibronectin against proteases", J. Biol. Chem. **257** (14), 8549 .

[89] F. Nachon*, et al.*, (2002) "Engineering of a monomeric and low-glycosylated form of human butyrylcholinesterase: expression, purification, characterization and crystallization", Eur. J. Biochem. **269** (2), 630 .

[90] C. J. Bosques*, et al.*, (2004) "Effects of glycosylation on peptide conformation: a synergistic experimental and computational study", J. Am. Chem. Soc. **126** (27), 8421 .

[91] D. B. Williams, (2006) "Beyond lectins: the calnexin/calreticulin chaperone system of the endoplasmic reticulum", J. Cell Sci. **119** (Pt 4), 615 .

[92] S. S. Komath, M. Kavitha, and M. J. Swamy, (2006) "Beyond carbohydrate binding: new directions in plant lectin research", Org. Biomol. Chem. **4** (6), 973 .

[93] M. Nesper*, et al.*, (1998) "Dimers of Thermus thermophilus elongation factor Ts are required for its function as a nucleotide exchange factor of elongation factor Tu", **255** (1), 81 .

[94] M. C. Moldovan*, et al.*, (2002) "CD4 dimers constitute the functional component required for T cell activation", J. Immunol. **169** (11), 6261 .

[95] N. J. Marianayagam, M. Sunde, and J. M. Matthews, (2004) "The power of two: protein dimerization in biology", Trends Biochem. Sci. **29** (11), 618 .

[96] F. Fusetti*, et al.*, (2002) "Crystal structure of the copper-containing quercetin 2,3-dioxygenase from Aspergillus japonicus", Structure. **10** (2), 259 .

[97] E. Degen and D. B. Williams, (1991) "Participation of a novel 88-kD protein in the biogenesis of murine class I histocompatibility molecules", J. Cell Biol. **112** (6), 1099 .

[98] E. Degen, M. F. Cohen-Doyle, and D. B. Williams, (1992) "Efficient dissociation of the p88 chaperone from major histocompatibility complex class I molecules requires both beta 2-microglobulin and peptide", J. Exp. Med. **175** (6), 1653 .

[99] W. J. Ou*, et al.*, (1993) "Association of folding intermediates of glycoproteins with calnexin during protein maturation", Nature **364** (6440), 771 .

[100] K. Galvin*, et al.*, (1992) "The major histocompatibility complex class I antigen-binding protein p88 is the product of the calnexin gene", Proc. Natl. Acad. Sci. U. S. A **89** (18), 8452 .

[101] G. Vogt, S. Woell, and P. Argos, (1997) "Protein thermal stability, hydrogen bonds, and ion pairs", J. Mol. Biol. **269** (4), 631 .

[102] J. C. Bischof and X. M. He, (2005) "Thermal stability of proteins", **1066**, 12 .

[103] J. R. Lepock, (2003) "Cellular effects of hyperthermia: relevance to the minimum dose for thermal damage", Int. J. Hyperthermia **19** (3), 252 .

[104] S. Voronov*, et al.*, (2002) "Temperature-induced selective death of the C-domain within angiotensin-converting enzyme molecule", FEBS Lett. **522** (1-3), 77 .

[105] V. Renugopalakrishnan*, et al.*, "Enhancement of Protein Thermal Stability: Toward the Design of Robuse Proteins for Biotechnological Applications,"in *Bionanotechnology,*(Springer, Netherlands, 2006), pp.117-140.

[106] V. Renugopalakrishnan*, et al.*, (2005) "Rational design of thermally stable proteins: relevance to bionanotechnology", J. Nanosci. Nanotechnol. **5** (11), 1759 .

[107] H. Liu, G. G. Bulseco, and J. Sun, (2006) "Effect of posttranslational modifications on the thermal stability of a recombinant monoclonal antibody", Immunol. Lett. **106** (2), 144 .

[108] Y. Yasuda, *et al.*, (1999) "Role of N-glycosylation in cathepsin E. A comparative study of cathepsin E with distinct N-linked oligosaccharides and its nonglycosylated mutant", Eur. J. Biochem. **266** (2), 383 .

[109] T. D. Butters, *et al.*, (1999) "Effects of N-Butyldeoxynojirimycin and the Lec3.2.8.1 Mutants Phenotype on N-Glycan Processing in Chinese Hamster Ovary Cells: Application to Glycoprotein Crystallization", 8 pp.1696-1701.

[110] S. J. Davis, *et al.*, (1993) "Expression of soluble recombinant glycoproteins with predefined glycosylation: application to the crystallization of the T-cell glycoprotein CD2", Protein Eng **6** (2), 229 .

[111] A. Mcpherson, "Preparation and Analysis of Protein Crystals,"in (John Wiley and sons, New York, 1982).

[112] R. Sadhukhan and I. Sen, (1996) "Different glycosylation requirements for the synthesis of enzymatically active angiotensin-converting enzyme in mammalian cells and yeast", J Biol. Chem. **271** (11), 6429 .

[113] S. Wildt and T. U. Gerngross, (2005) "The humanization of N-glycosylation pathways in yeast", Nat. Rev. Microbiol. **3** (2), 119 .

[114] D. L. Jarvis, (2003) "Developing baculovirus-insect cell expression systems for humanized recombinant glycoprotein production", Virology **310** (1), 1 .

[115] D. A. Griffith, *et al.*, (2003) "A novel yeast expression system for the overproduction of quality-controlled membrane proteins", FEBS Lett. **553** (1-2), 45 .

[116] C. R. Midgett and D. R. Madden, (2007) "Breaking the bottleneck: eukaryotic membrane protein expression for high-resolution structural studies", J. Struct. Biol. **160** (3), 265 .

[117] V. T. Chang, *et al.*, (2007) "Glycoprotein structural genomics: solving the glycosylation problem", Structure. **15** (3), 267 .

[118] B. Chen, *et al.*, (2000) "Expression, purification, and characterization of gp160e, the soluble, trimeric ectodomain of the simian immunodeficiency virus envelope glycoprotein, gp160", J. Biol. Chem. **275** (45), 34946 .

[119] H. Liu, A. H. Shim, and X. He, (2009) "Structural characterization of the ectodomain of a disintegrin and metalloproteinase-22 (ADAM22), a neural adhesion receptor instead of metalloproteinase: insights on ADAM function", J. Biol. Chem. **284** (42), 29077 .

[120] R. Konig, G. Ashwell, and J. A. Hanover, (1988) "Glycosylation of CD4. Tunicamycin inhibits surface expression", J. Biol. Chem. **263** (19), 9502 .

[121] M. Jinek and E. Conti, (2006) "Eukaryotic expression, purification, crystallization and preliminary X-ray analysis of murine Manic Fringe", Acta Crystallogr. Sect. F. Struct. Biol. Cryst. Commun. **62** (Pt 8), 774 .

[122] K. Gordon, *et al.*, (2003) "Deglycosylation, processing and crystallization of human testis angiotensin-converting enzyme", Biochem. J **371** (Pt 2), 437 .

[123] C. S. Anthony, *et al.*, (2010) "The N domain of human angiotensin-I-converting enzyme: the role of N-glycosylation and the crystal structure in complex with an N domain-specific phosphinic inhibitor, RXP407", J. Biol. Chem. **285** (46), 35685 .

[124] X. C. Yu, *et al.*, (1997) "Identification of N-linked glycosylation sites in human testis angiotensin-converting enzyme and expression of an active deglycosylated form", J Biol. Chem. **272** (6), 3511 .

[125] P. Redelinghuys, (2006) "Structure-Function Relationship of Angiotensin-Converting Enzyme: Glycosylation and Domain-Selectivity", PhD. University of Cape Town.

[126] O. A. Kost, *et al.*, (2000) "New feature of angiotensin-converting enzyme: carbohydrate-recognizing domain", J. Mol. Recognit. **13** (6), 360 .

[127] K. Kohlstedt, *et al.*, (2006) "Angiotensin-converting enzyme (ACE) dimerization is the initial step in the ACE inhibitor-induced ACE signaling cascade in endothelial cells", Mol. Pharmacol. **69** (5), 1725 .

[128] M. R. Ehlers, Y. N. Chen, and J. F. Riordan, (1991) "Spontaneous solubilization of membrane-bound human testis angiotensin-converting enzyme expressed in Chinese hamster ovary cells", Proc. Natl. Acad. Sci. U. S. A **88** (3), 1009 .

[129] S. Y. Oppong and N. M. Hooper, (1993) "Characterization of a secretase activity which releases angiotensin-converting enzyme from the membrane", Biochem. J. **292 (Pt 2)**, 597 .

[130] O. A. Kost*, et al.*, (2003) "Epitope-dependent blocking of the angiotensin-converting enzyme dimerization by monoclonal antibodies to the N-terminal domain of ACE: possible link of ACE dimerization and shedding from the cell surface", Biochemistry **42** (23), 6965 .

[131] I. V. Balyasnikova*, et al.*, (2005) "Localization of an N-domain region of angiotensin-converting enzyme involved in the regulation of ectodomain shedding using monoclonal antibodies", J Proteome. Res. **4** (2), 258 .

[132] M. J. Maliarik*, et al.*, (1998) "Angiotensin-converting enzyme gene polymorphism and risk of sarcoidosis", Am. J Respir. Crit Care Med. **158** (5 Pt 1), 1566 .

[133] S. M. Danilov, O. Kost, and E. D. Sturrock, (2006) "The missing link: ACE dimerization and shedding", Mol. Pharmacol.

[134] P. A. Velletri, M. L. Billingsley, and W. Lovenberg, (1985) "Thermal denaturation of rat pulmonary and testicular angiotensin-converting enzyme isozymes. Effects of chelators and CoCl2", Biochim. Biophys. Acta **839** (1), 71 .

[135] B. Marcic*, et al.*, (2000) "Effects of the N-terminal sequence of ACE on the properties of its C-domain", Hypertension **36** (1), 116 .

[136] H. G. O'Neill*, et al.*, (2008) "The role of glycosylation and domain interactions in the thermal stability of human angiotensin-converting enzyme", Biol. Chem. **389** (9), 1153 .

[137] T. A. Williams*, et al.*, (1996) "Drosophila melanogaster angiotensin I-converting enzyme expressed in Pichia pastoris resembles the C domain of the mammalian homologue and does not require glycosylation for secretion and enzymic activity", Biochem. J. **318 (Pt 1)**, 125 .

[138] R. Natesh*, et al.*, (2003) "Crystal structure of the human angiotensin-converting enzyme-lisinopril complex", Nature **421** (6922), 551 .

[139] H. R. Corradi*, et al.*, (2006) "Crystal structure of the N domain of human somatic angiotensin I-converting enzyme provides a structural basis for domain-specific inhibitor design", J. Mol. Biol. **357** (3), 964 .

[140] J. M. Watermeyer*, et al.*, (2006) "Structure of testis ACE glycosylation mutants and evidence for conserved domain movement", Biochemistry **45** (42), 12654 .

[141] K. Julenius*, et al.*, (2005) "Prediction, conservation analysis, and structural characterization of mammalian mucin-type O-glycosylation sites", Glycobiology **15** (2), 153 .

[142] R. Gupta, E. Jung, and S. Brunak, (2004) "Prediction of N-Glycosylation Sites in Human Proteins", Manuscript in preparation,

[143] I. V. Balyasnikova*, et al.*, (2003) "Monoclonal antibodies to denatured human ACE (CD 143), broad species specificity, reactivity on paraffin sections, and detection of subtle conformational changes in the C-terminal domain of ACE", Tissue Antigens **61** (1), 49 .

[144] M. R. Ehlers, Y. N. Chen, and J. F. Riordan, (1991) "Purification and characterization of recombinant human testis angiotensin-converting enzyme expressed in Chinese hamster ovary cells", Protein Expr. Purif. **2** (1), 1 .

[145] J. Rodriguez*, et al.*, (2008) "Does trypsin cut before proline?", J. Proteome. Res. **7** (1), 300 .

[146] C. Albach*, et al.*, (2004) "Identification of N-glycosylation sites of the murine neural cell adhesion molecule NCAM by MALDI-TOF and MALDI-FTICR mass spectrometry", Anal. Bioanal. Chem. **378** (4), 1129 .

[147] Y. Zhang, E. P. Go, and H. Desaire, (2008) "Maximizing coverage of glycosylation heterogeneity in MALDI-MS analysis of glycoproteins with up to 27 glycosylation sites", Anal. Chem. **80** (9), 3144 .
[148] C. Barinka*, et al.*, (2004) "Identification of the N-glycosylation sites on glutamate carboxypeptidase II necessary for proteolytic activity", Protein Sci. **13** (6), 1627 .
[149] C. Papworth*, et al.*, (1996) "Site-directed mutagenesis in one day with 80% efficiency.", **9**, 3 .
[150] S. Ehrt and D. Schnappinger, (2003) "Isolation of plasmids from E. coli by boiling lysis", Methods Mol. Biol. **235**, 79 .
[151] L. U. Haiqiang, Y. U. Hongwei, and J. I. A. Yingmin, (2009) "Improvement of megaprimer method for site-directed mutagenesis and its application to phytase", Front. Agric. China **1** (3), 43 .
[152] S. H. Ke and E. L. Madison, (1997) "Rapid and efficient site-directed mutagenesis by single-tube 'megaprimer' PCR method", Nucleic Acids Res. **25** (16), 3371 .
[153] R. M. Horton and L. R. Pease, *directed mutagenesis* (Oxford University Press, New York, 1991).
[154] L. Zheng, U. Baumann, and J. L. Reymond, (2004) "An efficient one-step site-directed and site-saturation mutagenesis protocol", Nucleic Acids Res. **32** (14), e115 .
[155] R. A. Dwek, (1995) "Glycobiology - More Functions for Oligosaccharides", Science **269** (5228), 1234 .
[156] D. H. Correa and C. H. Ramos, (2009) "The use of circular dichroism spectroscopy to study protein folding, form and function", **3** (5), 164 .
[157] A. Fung, (1995) "Jasco J-810 Spectropolarimeter Basic Operating Manual",
[158] A. Michnik, (2003) "Thermal stability of bovine serum albumin DSC study", **71** (2), 509 .
[159] S. M. Danilov*, et al.*, (2008) "Simultaneous determination of ACE activity with 2 substrates provides information on the status of somatic ACE and allows detection of inhibitors in human blood", J. Cardiovasc. Pharmacol. **52** (1), 90 .
[160] Z. L. Woodman*, et al.*, (2005) "The N domain of somatic angiotensin-converting enzyme negatively regulates ectodomain shedding and catalytic activity", Biochem. J **389** (Pt 3), 739 .
[161] J. Kondo*, et al.*, (2010) "A functional role of the glycosylated N-terminal domain of chondromodulin-I", J. Bone Miner. Metab .
[162] J. M. Watermeyer*, et al.*, (2008) "Probing the basis of domain-dependent inhibition using novel ketone inhibitors of Angiotensin-converting enzyme", Biochemistry **47** (22), 5942 .
[163] F. Grueninger-Leitch*, et al.*, (1996) "Deglycosylation of proteins for crystallization using recombinant fusion protein glycosidases", Protein Sci. **5** (12), 2617 .
[164] B. C. Jones, N. J. Logsdon, and M. R. Walter, (2008) "Crystallization and preliminary X-ray diffraction analysis of human IL-22 bound to the extracellular IL-22R1 chain", Acta Crystallogr. Sect. F. Struct. Biol. Cryst. Commun. **64** (Pt 4), 266 .
[165] Z. L. Woodman*, et al.*, (2006) "Homologous substitution of ACE C-domain regions with N-domain sequences: effect on processing, shedding, and catalytic properties", Biol. Chem. **387** (8), 1043 .
[166] P. V. Binevski*, et al.*, (2003) "Evidence for the negative cooperativity of the two active sites within bovine somatic angiotensin-converting enzyme", FEBS Lett. **550** (1-3), 84 .
[167] H. L. Chen*, et al.*, (2010) "Porcine pulmonary angiotensin I-converting enzyme--biochemical characterization and spatial arrangement of the N- and C-domains by three-dimensional electron microscopic reconstruction", Micron. **41** (6), 674 .
[168] M. L. Langsford*, et al.*, (1987) "Glycosylation of bacterial cellulases prevents proteolytic cleavage between functional domains", FEBS Lett. **225** (1-2), 163 .
[169] F. M. Ausubel*, et al.*, *short protocols in molecular biology,* 4th ed. (John Wiley and Sons, Inc., New York, 1992).
[170] U. K. Laemmli, (1970) "Cleavage of structural proteins during the assembly of the head of bacteriophage T4", Nature **227** (5259), 680 .

[171] J. Friedland and E. Silverstein, (1976) "A sensitive fluorimetric assay for serum angiotensin-converting enzyme", Am. J Clin. Pathol. **66** (2), 416 .
[172] S. L. Schwager, A. K. Carmona, and E. D. Sturrock, (2006) "A high-throughput fluorimetric assay for angiotensin I-converting enzyme", Nat. Protoc. **1** (4), 1961 .