O-GLNCAC MODIFICATION STUDY BY IN VITRO GLYCOSYLATION

O-GLNCAC MODIFICATION STUDY BY IN VITRO GLYCOSYLATION:

A MASS SPECTROMETRY APPROACH

By

Xi Simon Wang, B. Sc.

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfillment of the Requirements

for the Degree

Master of Science

McMaster University

MASTER OF SCIENCE (2007) McMaster University

(Biochemistry and Biomedical Sciences) Hamilton, Ontario

TITLE: O-GlcNAc Modification Study by In Vitro Glycosylation: A Mass Spectrometry Approach

AUTHOR: Xi Simon Wang, B. Sc. (Jilin University)

SUPERVISOR: Dr. Graham A. McGibbon

NUMBER OF PAGES: xi, 113

Abstract

O-GlcNAc modification is a single N-acetylglucosamine (GlcNAc) modification on Ser or Thr residue on protein. The addition and removal of the O-GlcNAc molecule are controlled by two enzymes (OGT and NCOAT). In this study, I expressed and purified the two enzymes involved in the O-GlcNAc modification. A method was developed for the synthesis and purification of the peptide substrate YSDSPSTST for in vitro glycosylation and characterized the OGT enzyme activity by the in vitro glycosylation and H^3 labeling. A method was developed based on detection of glycosylation peptide by mass spectrometry after separation by capillary liquid chromatography (CapLC). The optimization of mass spectrometry parameters was done using synthesized standard glycopeptide YSDSPSgTST ("Sg" represents O-GlcNAc modified Serine). The in vitro modification site was determined by CID after alkaline β-elimination. Furthers experiment could include detection of O-GlcNAc modification of protein substrate both in vitro and in vivo. This will give a better understanding of the dynamics of O-GlcNAc modification.

Acknowledgements

I would like to thank my supervisor, Dr. Graham McGibbon, for providing me the opportunity to work with him and for his help and guidance throughout these two years. I would also like to thank my committee members, Dr. David Andrews and Dr. Paul Berti, thanks for their advice during the project. Especially Dr. Berti for kindly allowing me used the equipment in his lab. I am especially thankful to Dr. Nodwell for carefully reading my manuscripts and for his many helpful advices about the thesis. I would like to thank Dr. Kirk Green and Dr. Jiaxi Wang at McMaster Regional Centre for Mass Spectrometry for all their assistance with the mass spectrometer.

I also want to thank Hongchao Zheng, another graduate student in our group who synthesized the standard glycopeptide. I really enjoyed the discussion with him about the project and the encouragement during these years.

Finally, I need to thank my family for their constant encouragement. For mum and dad, thanks for giving me support and encouragement by the phone every week. And for my wife Shelley, thanks for your love and support along the way. This is the best memory in my life. I would achieve nothing without any of you.

Table of Contents

List of Abbreviations

ARM-	Armadillo;
ESI-	Electrospray ionization;
GPGTF-	Glycogen phosphorylase / glycosyl transferase;
GNAT-	Gcn5-related acetyltransferase
HAT-	Histone acetyltransferase;
IPTG -	isopropyl-beta-D-thiogalactopyranoside;
PTM-	Post-translational modification;
MGEA5-	Meningioma expressed antigen 5;
MALDI-	Matrix-assisted laser desorption ionization;
mOGT-	Mitochondrial OGT;
NCOAT-	Nuclear cytoplasmic O-GlcNAcase and acetyltransferase;
ncOGT-	Nucleus and cytosol OGT;
NLS-	Nuclear localization sequence;
O-GlcNAc-	O-linked β-N-acetylglucosamine;
OGT-	O-linked GlcNAc transferase;
PEST-	Proline, glutamic acid, serine, threonine;
Q-Tof-	Quadrupole time of flight;
RCA-	Ricinus Communis Agglutinin;
sOGT-	Short OGT;
TOF-	Time-of-flight;
TPR-	Tetratricopeptide repeat;
WGA-	Wheat germ agglutinin;
MS -	Mass Spectrometer;

List of Figures

For Shelley

1 Introduction

1.1. Protein post-translational modifications

After being synthesised, unmodified polypeptides are sent to different cellular compartments such as the endoplasmic reticulum, Golgi apparatus, cytosol, and nucleus to undergo single or multiple modifications to form a fully functional protein, a process is called post-translational modification (PTM). Post-translational modifications (PTM) are important feature of proteins [1]. It is a crucial step to give a specific biological activity to a protein. More than 100 different types of PTM have been found and shown to play important roles in protein function; more than 80% of the mammalian proteins undergone post-translational modified [2]. Proteolytic cleavage, sulfation, methylation, phosphorylation, and glycosylation are typical post-translational modifications [3]. Unlike irreversible modifications such as proteolytic cleavage, phosphorylation and glycosylation are dynamic modifications in that they are reversible.

Glycosylation is found in both prokaryotes and eukaryotes [4]. There are different kinds of glycosylations. One is called N-linked glycosylation, where the sugar is attached to an Asn residue amide group. The modification sites have a conserved sequence "Asn-X-Ser/Thr", where X cannot be proline. Another glycosylation is called O-linked glycosylation. The sugar is attached to the oxygen atom from hydroxyl group of a Ser or Thr [5]. The glycoproteins were

thought to only exist on the lumenal compartment or cell surface [6]. Nevertheless, over the past twenty years, accumulating evidence of nuclear and cytoplasmic O-glycosylation has broken that dogma [7].

1.2. O-GlcNAc modification

In 1984, Hart and Torres found a novel protein glycosylation [8]. The protein was modified at Ser or Thr residue by a single N-acetylglucosamine (GlcNAc) residue, as shown in Figure 1.1. Later on studies showed that this kind of modification happened on nuclear and cytoplasmic proteins rather than O-linked carbohydrates attached to cell surface protein [9]. Since then, more and more proteins have been found containing this modification [10].

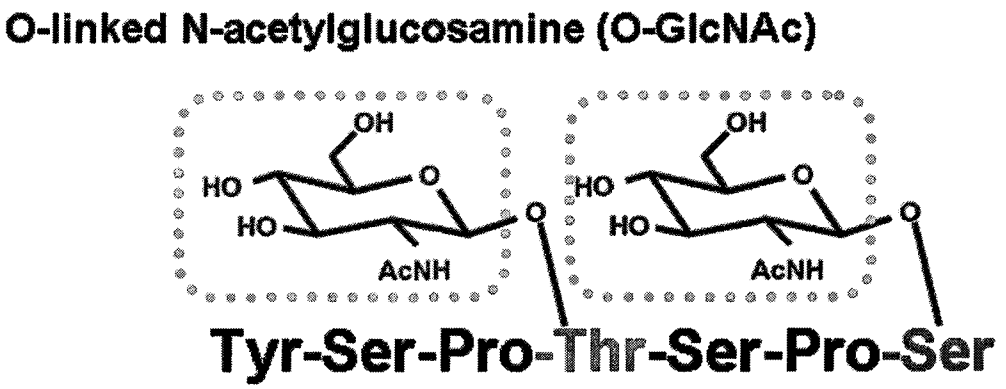


Figure 1.1 O-linked N-acetylglycosamine model.

1.3. Dynamic O-GlcNAc modification

Accumulating evidence shows that the O-GlcNAc modification is a dynamic process. The addition and removal of the O-GlcNAc molecule are controlled by two enzymes [11]. Evidence has shown that the O-GlcNAc modification of nuclear pore protein p62 can happen both cotranslationally and posttranslationally [9,12]. Rapid changes of the O-GlcNAc level in T-lymphocytes under mitogen stimulation demonstrated that this modification is responsive to extracellular signals [13]. Further evidence indicates that the turnover of O-GlcNAc can be regulated during development and between cell cycle. For instance, cell cycle studies revealed that the level of O-GlcNAc modification on certain cytosolic and nuclear pore proteins vary during the cell cycles [14]. Additionally, mitotic arrest increased the O-GlcNAc modification level of the keratins 8 and keratins 18. The O-GlcNAc turnover rate studies showed that both cytokeratins and the small heat shock protein alpha-crystallin have a much faster turnover rate than that of the proteins themselves [11]. In summary all available evidences suggest that of O-GlcNAc modification is a dynamic regulatory modification.

The identification of the enzymes responsible for O-GlcNAc modification supports the hypothesis that the modification is a dynamic regulatory modification. There are two enzymes involved in this regulation. One is O-linked GlcNAc transferase (OGT), which transfers a GlcNAc residue to the protein. The other is

nuclear cytoplasmic O-GlcNAcase and acetyltransferase (NCOAT), which removes the GlcNAc residue from the modified protein. These enzymes can be considered the functional analogs of the protein phosphotases and kinases. One possible model for the relations between O-GlcNAc modification and phosphorylation is called the replacement model. In this model, one protein can be one of the three possible forms: (1) unmodified, (2) phosphorylated, (3) O-GlcNAc modified. Different forms may have distinct functions or biological properties. The different properties of the isoforms could alter protein structures, changing enzyme activity or regulate the protein localization and affect the functions.

1.4. Addition of O-GlcNAc: OGT

A uridine diphospho-N-acetylglucosamine peptide β-N-acetylglucosaminyl transferase (O-GlcNAc transferase, OGT), was identified using a synthetic peptide acceptor from rat liver and rabbit reticulocyte membranes [15]. Further studies confirmed that OGT enzyme is present in both the cytosol and nucleus [16]. The OGT enzyme was purified from rat liver [17] and shown to be a dimer of two subunits, one of 78 kDa, the other 110 kDa. Antibodies generated from the 78 kDa subunit also recognized the 110 kDa subunit, indicating that the two subunits have high similarity at the sequence level. The similarity may be the result of alternative splicing or proteolysis of 110 kDa subunit to generate the 78 kDa subunit [18].

The cDNA sequence encoding the OGT enzyme was cloned from rat pancreas by Hart's group in 1997 [18], and from human and C. elegans by Hanover's group in the same year [19]. The bioinformatics study indicated that the O-GlcNAc transferase was highly conserved through evolution. Sequence alignment revealed homology between the O-GlcNAc transferase and the GPGTF (glycogen phosphorylase / glycosyl transferase) motif [20].

The OGT enzyme has three parts, an N-terminal tetratricopeptide repeat (TPR) domain, a bipartite nuclear localization sequence (NLS), and a C-terminal catalytic domain. Human OGT has 12 tetratricopeptide repeats (TPR) at its N-terminal [21]. The TPR domain is responsible for specific protein-protein interactions [22]. Studies of human cDNA clones indicated that mammalian *ogt* gene encodes three alternative splicing variants [19, 21,23]. One major variant, mOGT (m for mitochondrial), contains three parts. These are an N-terminal mitochondrial targeting sequence, a nuclear localization motif, and a 9 tetratricopeptide repeats. The second major variant, ncOGT (nc for nucleus and cytosol), does not have the mitochondrial targeting motif but contains 12 tetratricopeptide repeats. An even smaller variant, sOGT (s for short), contain only 2 tetratricopeptide repeats and is also found in human. The ncOGT and mOGT have been better studied while less is known about sOGT [24,25]. Synthetic genes using the *E.coli* preferred codons for all the three variants have been constructed into an *E.coli* expression system. As the result, a much higher yield of the recombinant proteins was obtained [25]. The TPR also affects the

selectivity of the substrate. Partial deletion of tetratricopeptide repeats illustrates an important role for the interaction between proteins and substrates [21,26].

The X-ray crystal structure of the human OGT TPR domain at a resolution of 2.85 Å has been published. The tetratricopeptide repeats forms a superhelical architecture. The inner concave surface is constrained by completely conserved asparagines which are quite similar to the armadillo (ARM)-repeats proteins such as importin α and β-catenin [27]. The OGT itself is both O-GlcNAc and tyrosine-phosphate modified [18, 26]. It has been proposed that different post-translational modification patterns may regulate the enzyme function although the details are not fully understood yet.

The OGT enzyme transfers the GlcNAc molecule from UDP-GlcNAc to Ser/Thr. The enzyme activity is inhibited by UDP, the by product of this transferase reaction. The enzyme is sensitive to changes in of UDP-GlcNAc/ UDP ratios. Therefore, the fluctuations of UDP-GlcNAc or UDP levels inside certain organs or throughout the organism may vary the OGT activity [17,28]. The enzyme is also effectively inhibited by NaCl (IC_{50} = 50 mM), UDP (IC_{50} = 5 mM) [29], and uracil analog alloxan, IC_{50} = 0.1 mM [30]. Three more effective inhibitors than alloxan were discovered by high throughput screening, but the in vivo effect of these compounds is not clear [25].

1.5. Removal of O-GlcNAc: NCOAT

A nuclear cytoplasmic O-GlcNAcase and acetyltransferase (NCOAT, EC 3.2.1.52) was first purified from rat spleen [36] and the cDNA was cloned from bovine brain [31]. The purified enzyme was able to release the O-GlcNAc molecule from modified proteins. The enzyme was different from the lysosomal hexosaminidases in the following ways: (1) NCOAT had a neutral optimum pH around 6 which is distinct from lysozomal hexosaminidases (optimum pH around 4); (2) NCOAT only recognized GlcNAc as substrate and did not show activity toward GalNAc or its analogs; (3) Overexpressed NCOAT protein is localized in nucleus and cytoplasm [31,32]. Little else was known about NCOAT for over a decade, but recently, the protein has become the subject of intense scrutiny by several research groups leading to characterization of its catalytic residues, along with their structural and mechanistic details.

The NCOAT protein is encoded by the meningioma expressed antigen 5 (MGEA5) located on chromosome 10q24.1-q24.3, the same locus as AD6 (Alzheimer disease 6) [33]. The NCOAT protein consists of 916 amino acids with a calculated molecular weight of 103 kDa and a pI value about 4.63 [31]. The N-terminal domain (aa.1-342) was homologous to bacterial hyaluronidases and exhibited the O-GlcNAcase activity [34]. Further sequence analysis indicated that the NCOAT enzyme also has a weak homology to histone acetyltransferases (HAT) of the Gcn5-related acetyltransferase (GNAT) family [35] and experiments

confirmed the histone acetyltransferase activity in the C-terminal domain (aa. 719-916) [32].

About 90% of O-GlcNAcase activity was localized in cytosol and about 10% in the nucleus. The NCOAT enzyme is expressed in most human tissues examined and is especially highly expressed in brain, pancreas, and placenta [31,36]. NCOAT is a substrate for the executioner apoptotic caspase-3 and is cleaved during apoptosis [32]. Interestingly, cleavage of the full-length NCOAT into two fragments does not affect its O-GlcNAcase activity in vitro [32]. This suggests that the C-terminal 75 kDa component may serve as a regulatory subunit, which may be involved in subcellular localization and protein-protein associations, and may also play a role for NCOAT during apoptosis.

A recent study showed that NCOAT could tightly associate with its counter partner OGT enzyme. The complexes could regulate gene transcription by binding to repressed promoters [37]. This result is interesting because the association of two functional opposite enzymes is not commonly observed in biological system.

The bifunctional NCOAT protein could activate transcription in different ways. For instance, NCOAT could regulate the O-GlcNAc modified level of some transcriptional factor such as Sp1 [10]. It could also assist gene transcription by acetylating histones through its histone acetyltransferase domain.

The enzyme kinetic studies revealed that the NCOAT enzyme mechanism involving substrate assisted catalysis is similar to other family 84 glycosylhydrolases [38]. The crystal structure of NagJ (a member of the family 84 glycosylhydrolases) revealed the active sites of the NCOAT enzyme [39,40]. Site-directed mutants of Asp174 and Asp 175 confirmed these as the key catalytic residues [41]. At the start of this project little was known about NCOAT, the DNA encoding human NCOAT was available and the recombinant protein could reportedly be expressed in bacteria [31], so it was an attractive target for characterization and for use in studying O-glycosylation in vitro. However the above developments made OGT a priority for further study instead.

1.6. The role of O-GlcNAc modification

1.6.1. Nuclear O-GlcNAc modification

Increasing numbers of proteins are identified to be O-GlcNAc modified. Subcellular localization studies found higher levels O-GlcNAc modification on the nuclear membrane and nucleus than in cytosol components [42,43]. This suggests that the modification may play a key role in transcriptional regulation.

1.6.2. Regulating protein phosphorylation

The O-GlcNAc modification, similar to phosphorylation, regulates critical aspects of protein biology, including protein stability, subcellular localization and protein-protein interactions. The modification sites of O-GlcNAc are the same as

for phosphoserine or phosphothreonine and the reciprocal phosphorylation /glycosylation site of some proteins has been demonstrated [44,45]. One hypothesis is that O-GlcNAc could alter the phosphorylation patterns and protein function by blocking the available Ser/Thr residues as shown in Figure 1.2. This hypothesis was supported by in vitro studies of RNA polymerase II. The addition of an O-GlcNAc residue blocked phosphorylation at the same position and phosphorylation also blocked subsequent O-GlcNAc modification [46]. Further studies of casein kinase II has shown that the reciprocal modification can also happen on adjacent sites. These data suggest that glycosylation can inhibit phosphorylation at nearby sites, possibly though steric hindrance.

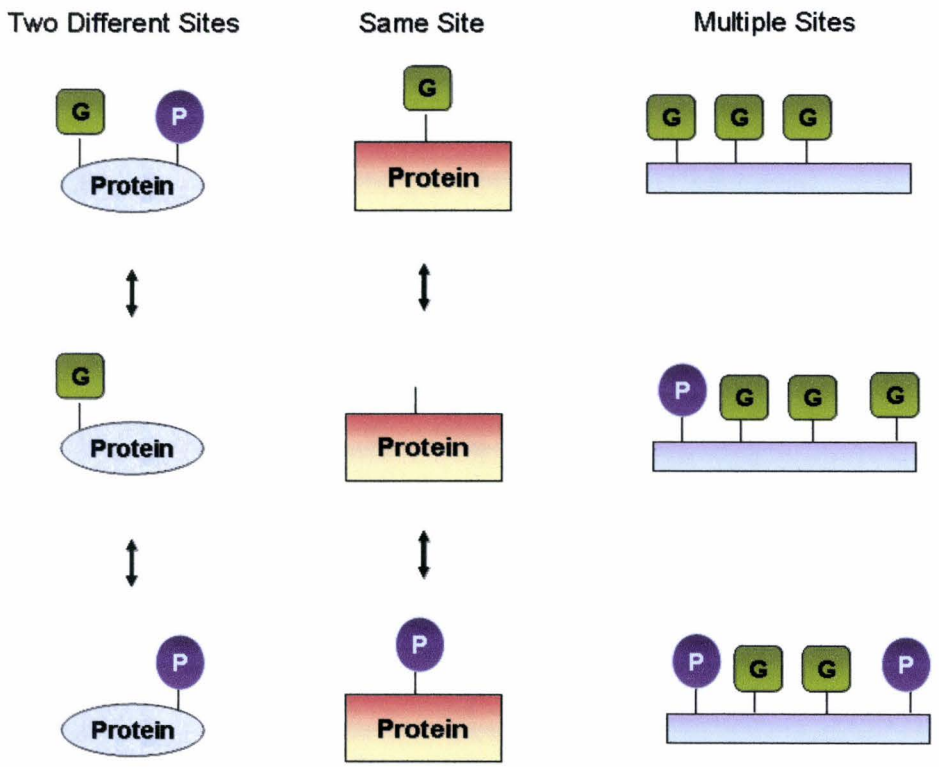


Figure 1.2 Dynamic modification of glycosylation and phosphoralation.

1.6.3. Regulating protein degradation

The in-vivo half-lives of proteins vary from minutes to days. Proteolysis plays a key role in removing damaged or abnormal proteins, activating protein precursors, and maintaining the free amino acid pool. There are two major kinds of proteolytic pathways: ATP-independent pathway and ubiquitin mediated ATP-dependent pathway [47]. Researchers have found that protein sequences enriched in proline (P), glutamic acid (E), serine (S) and threonine (T) undergo fast degradation [48]. Further studies demonstrated that phosphorylation of the PEST domains played an important role in targeting protein degradation.

Phosphorylation of Estrogen receptor (ER)-β PEST region, for example, leads to ubiquitination and is followed by proteasome-mediated degradation [49]. However, glycosylation of the PEST domain of estrogen receptor (ER)-β extended its half-life. This evidence suggests that the glycosylation may regulate protein degradation by preventing further phosphorylation of certain domains [50]. Interestingly, in vitro peptidase assay has shown that O-GlcNAc modification can also regulate protein degradation by altering the proteasome activity [51].

1.6.4. Mediating transcription and nuclear localization

It has been shown that the nuclear membrane contains the highest concentration of O-GlcNAc modified proteins. In fact, the nuclear pore protein (p62) was one of the first proteins described to be O-GlcNAc modified [52]. This was shown by immunoprecipitation followed by Western analysis with wheat germ agglutinin (WGA) [53]. Further studies of the glycosylated p62 revealed that O-GlcNAc may regulate nuclear transport function by altering both p62 and other nuclear pore proteins such as p54 and p58.

Other examples of O-GlcNAc modification are the transcription factors Sp1 and c-Myc. O-GlcNAc modification of Sp1 increased the transactivation ability and low levels of O-GlcNAc modification led to ubiquitination mediated Sp1 degradation [54]. C-Myc contains a two adjacent phosphorylation sites, Thr58 and Ser62. Glycosylation of Thr58 blocked the phosphorylation of both sites and prevented ubiquitin/proteasome-mediated degradation [55].

Some other proteins undergo multiple modifications. Tau protein for instance, is known as both phosphorylated and O-GlcNAc glycosylated. The hyperphosphorylation of Tau occurs in Alzheimer's disease. It has been demonstrated that there is a balance between phosphorylation and glycosylation. Dysfunction of this balance could lead to changes of Tau protein localization [56].

For another example, Andrews and co-workers have demonstrated that both β-catenin and E-cadherin are glycosylated. O-glycosylation of E-cadherin in response to ER stress prevents its transport to the cell membrane [57]. Interestingly, the crystal structure of human OGT TPR domain showed similarity to the ARM domain of importin α [27]. This indicated that OGT might interact via the TPR domain with its substrate proteins, particularly those that possess or interact with ARM domains such as plakoglobin and E-cadherin.

1.7. Analysis of O-GlcNAc

1.7.1. Methods used for detection of O-GlcNAc

Detection of O-GlcNAc modifications in vivo is quite challenging. First of all, the O-GlcNAc modification is a dynamic process. The modification pattern is not constant during the cell cycle. It may be substoichiometric and the proteins are not necessarily abundant. Useful but limited methods for identification of O-GlcNAc modified proteins have been developed: tritium-labeling, which is labor intensive and dangerous as well [58]; enrichment with lectins, which may not be specific for O-GlcNAc [59], or antibodies [60] and neither this nor chemical tagging by metabolic labelling, which requires cellular uptake [61], allow site mapping. For O-GlcNAc modified proteins from tissues or cell lysates, these methods are insufficient and even the simplest PTM patterns are beyond the reach of currently available technology.

There is a growing need for the development of a suitable system to study the functional relevance of O-GlcNAc modification. Many of O-GlcNAc modification proteins are present at low abundance, so it is difficult to get enough protein for direct analysis of the glycan attachment and glycosylation sites. Using of the reticulocyte lysate system for in vitro glycosylation studies has allowed for a great improvement of the analysis [53]. But the system complexity makes it difficult to understand and control the dynamic nature of glycosylation.

One appealing approach for analysis is to combine enrichment and/or derivatization strategies. Ricinus Communis Agglutinin (RCA) coupling used by Aebersold [62] and BEMAD by Hart, BEMAD (β-elimination followed by Michael addition with DTT) [63], approach can be used to map sites on purified proteins and protein complexes. But this is an inherently destructive technique that requires extensive controls to establish whether a peptide contains a phosphate, O-GlcNAc, or a complex O-linked carbohydrate group.

1.7.2. Mass spectrometric analysis of O-GlcNAc

Mass spectrometry has become an important analytical tool for biological molecules in the past two decades. The development of two ionization techniques; matrix-assisted laser desorption ionization (MALDI) and electrospray ionization (ESI) has enabled the transfer of charged non-volatile biomolecules to the gas phase, allowing mass spectrometry analysis of these molecules. These biomolecules included peptides, proteins, nucleotides, oligosaccharides, and lipids. The coupling of time-of-flight (TOF) analyzer with these ionization methods has significantly increased the sensitivity and has allowed higher resolution [64]. The most powerful strategies emerging for O-GlcNAc analysis are relying on the capabilities of sophisticated mass spectrometers to detect peptides derived from O-GlcNAc modified proteins.

1.7.2.1. Electrospray ionization (ESI)

In 1985, Fenn and co-workers developed the first electrospray ion sources [65,66]. The ESI source transfers solutes into the gas phase by applying a large potential over a liquid-air interface. Figure 1.3 shows a basic model of electrospray ionization process.

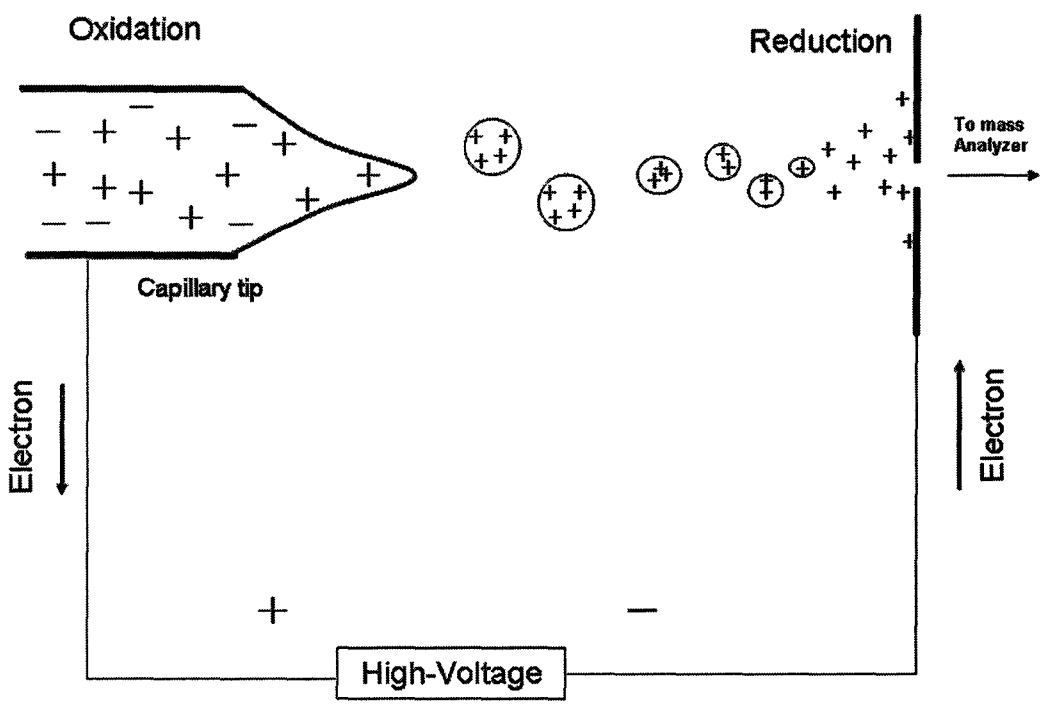


Figure 1.3 ESI ionization scheme.

An electrospray is produced in a strong electric field [67]. In electrospray ionization, the analyte solution is passed through a small diameter capillary tube. A high positive voltage is applied to the solution, inducing the formation of a taylor cone that emits small highly charged droplets from the opening to form highly charged droplets under the high electric field (10^6 V m^{-1}).

The droplets evaporate and shrink to the point where the coulomb repulsion between ions is close to the solvent cohesion force. The droplets explode to form smaller droplets. The whole procedure continues until the charged gas phase analytes are formed.

Electrospray ionization can be coupled directly with low flow reversed-phase liquid chromatography to separate complex peptide mixtures, and has been used in analyses of O-GlcNAc modifications. Chalkey and Burlingame have reported the possibility of direct O-GlcNAc peptide analysis using ESI-Q-Tof experiments [68]. In addition, BEMAD derivatization for more complex samples has been used [69]. Others have used more complicated fragmentation analyses: multi-stage dissociation with an ion-trap [70] and electron-capture induced dissociation on a Fourier transform ion cyclotron resonance mass spectrometer [71].

1.7.2.2. Matrix assisted laser desorption / ionization (MALDI)

The matrix-assisted laser desorption ionization (MALDI) of Karas and Hillenkamp has developed into an effective mass spectrometric technique for analyzing protein and peptide biopolymers [64], [72]. The ionization mechanism of MALDI is not fully understood, but a general ion formation process has been proposed.

The mechanism of MALDI is believed to consist of two steps as shown in Figure 1.4. In the first step, the analyte is mixed with a matrix compound. The matrix consists of small molecules which have a strong absorption at the laser (usually UV) wavelength. The analyte molecules are embedded in matrix crystals to form a so called "solid solution". The second step involves the exposing "solid solution" to short laser pulse in vacuum. The pulsed photon absorption causes the ablation of the matrix and solutes into the gas phase. The origin of the ions produced in MALDI is not fully understood [73]. The MALDI process has been proposed to consist of two parts; excitation of the matrix and proton transfer from the photo-excited matrix molecules to the analyte in the gas phase, see Figure 1.4.

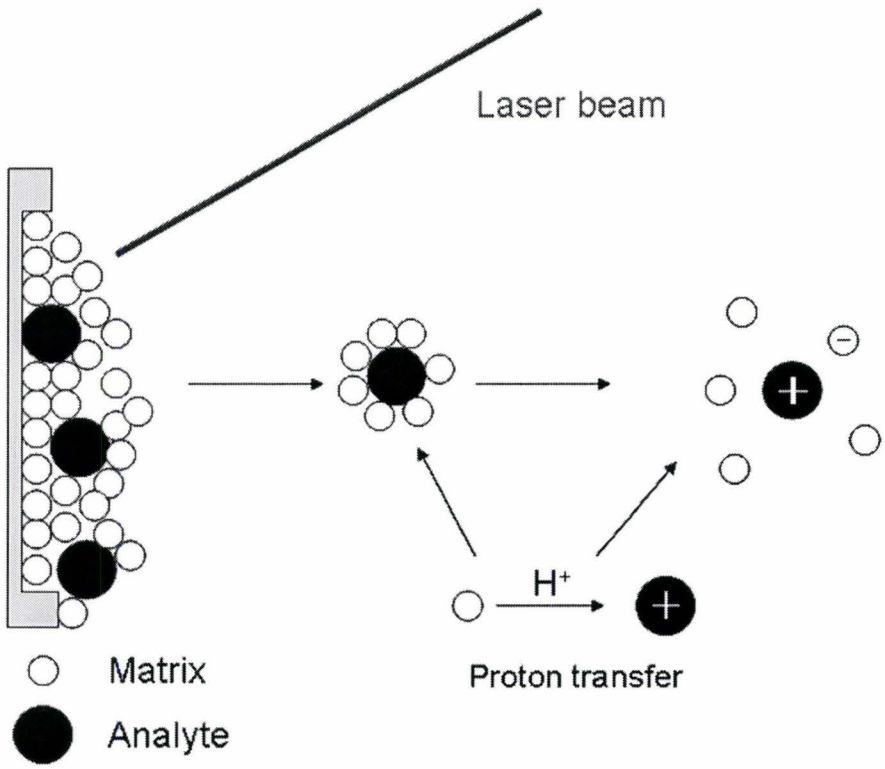


Figure 1.4 MALDI ionization scheme.

The sample is mixed with a large excess of matrix material that strongly absorbs the laser light. Thermal relaxation of the excited matrix molecules leads to evaporation, transfering the analyte molecules into gas phase. Reactions lead to analyte ions with one or two charges, typically $[M+H]^+$ or, if salt is present, $[M+Na]^+$ or $[M+K]^+$. Deprotonation of analyte by the matrix can form $[M-H]^-$ negative ions. The use of a matrix prevents fragmentation of the thermally labile molecules, thereby obtaining intact ionized molecules. MALDI does have some limitations: these include the challenge of reproducibility in crystallization, a not strictly quantitative response and so called "signal suppression" especially in complex samples because some of the analytes are more efficiently being ionized and prevent others from ionizing. However, it is possible to couple capillary scale chromatography with a system that deposits separated peptides onto a target plate for MALDI mass spectrometry analysis. The remarkable efficiency in producing intact ions of large biological compounds and its extraordinary sensitivity (fmol or less) made the MALDI technique widely used for studying O-GlcNAc modified peptides

1.7.2.3. Time-of-flight analyzer

In a time-of-flight analyzer, mass-to-charge ratios are determined by measuring the time that ions take to move through a region between the source pusher-puller stack and the detector. Ideally, when leaving the stack, the ions are all accelerated with the same voltage (V) so that they have a kinetic energy

$$q V = \tfrac{1}{2} m v^2 \qquad 1$$

where v is the ion velocity after acceleration, m is the mass of the ion, q is the charge of the ion, and V is the acceleration voltage. Since v equal to the length of the region L over time t

$$q V = \tfrac{1}{2} m (L / t)^2 \qquad 2$$

The time of flight, t, for an ion is

$$t = (m / 2 qV)^{1/2} L \qquad 3$$

In a time of flight analyzer, L and V are constants and the equation could be modified

$$t = a (m / q)^{1/2} \text{ or } t = a (m / z)^{1/2} \qquad 4$$

This equation shows that the *m/z* can be calculated from measurement of t^2.

The time-of-flight analyzer operates in two modes: the linear and the reflectron mode as shown in Figure 1.5.

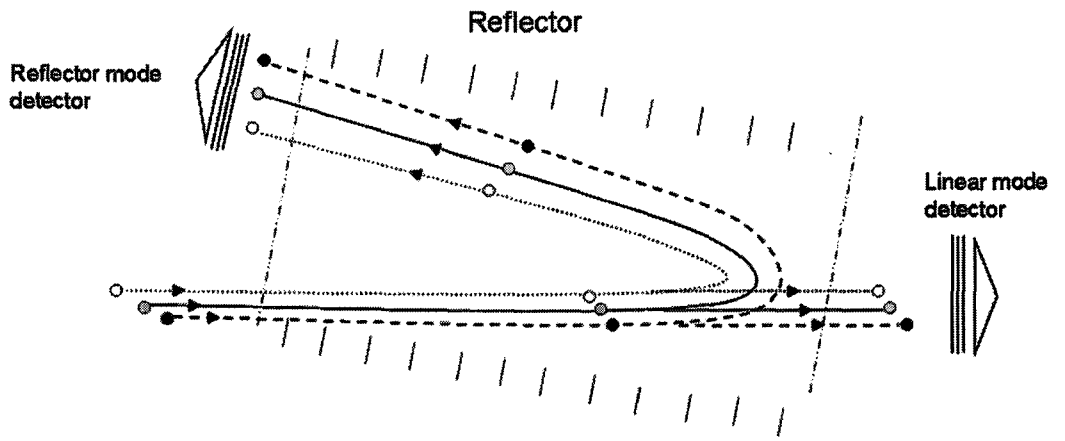


Figure 1.5 Reflector and linear mode of MS detection.

In the reflector mode, the reflector not only increases the length of the flight path, but also compensates ions with the same mass but different initial velocity. Ions with a higher velocity will penetrate the reflectron field further, so it takes a longer time compared to ions with lower velocity. The reflector thus minimizes the spread in time of flight for ions with same m/z.

1.7.2.4. Peptide sequencing by collision induced dissociation

Peptide sequencing is one of the most important applications for mass spectrometry in biological field. The peptide of interest could be selected and introduced into collision cell to collide with inter gas, such as nitrogen or argon. The fragments of the peptide are analyzed to produce a MS/MS spectrum.

As shown in, a series of ions can be generated by the collision induced fragmentation along several bonds of the peptide backbone. The b and y ions are the most common ion types which are formed by the fragmentation at the amide bond. The peptide sequence information could be obtained by comparison of a serious of y and b ions or matched against database directly.

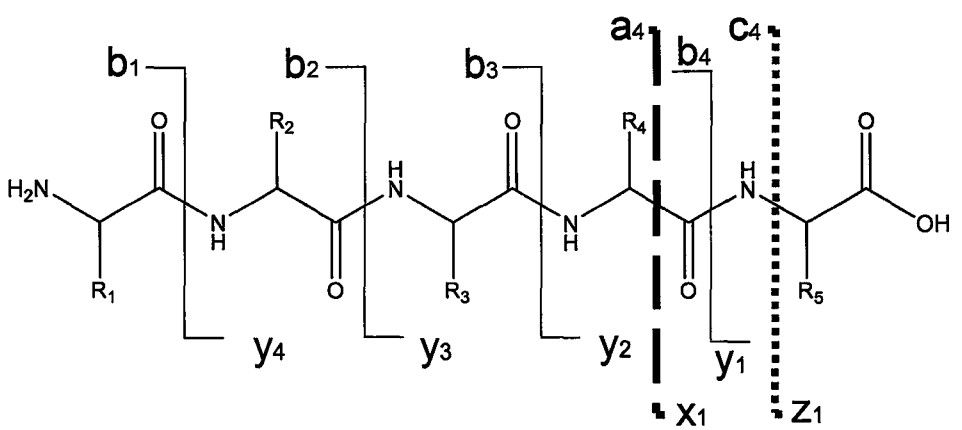


Figure 1.6 Formation of a, b, c, x, y, z ions by collision induced dissociation

1.7.2.5. Capillary LC-MALDI MS

One of the techniques for the detection of glycosylated protein is based on a comparison of the protein digest before and after the release of the oligosaccharides. This requires the release of the glycan from the protein of peptide. There are two ways to achieve this process: chemical release of glycans using anhydrous hydrazine or enzymatic release by a specific glycosidase. However, these methods require that the glycopeptide be clearly detectable as a resolved peak by MALDI. This is not easy to achieve in practice, because of the low amount of the modified peptide in the sample and possible suppression effects in MALDI. MS/MS could provide the sequence information of the modified peptide, but it is difficult to assign the modification position if the sequence contains multiple possible modification sites.

Separation and purification of the glycan pool is essential to investigate the complex mixture of glycoproteins. During the last decades, chromatographic separation of glycans has been developed rapidly. The most common separation technique for glycans is liquid chromatography (LC).

Capillary LC (flow rate around 1μL/min) coupled with nanoelectrospray has been widely applied to peptide sequencing and biomolecular interaction analysis. The online LC-MS/MS technique make it possible to analyze hundreds of samples a day. But the online method makes it hard to approach repeat

analysis of the same sample because the sample after detection by MS detector could not longer be used for further analysis.

The coupling of liquid chromatography with off line MALDI-MS surmounted this. The LC-MALDI-MS generate the matrix /analyte spot on the MALDI plate, therefore multiple analyses of one sample became possible. Also the LC-MALDI-MS allowed sensitive analysis of the biological samples, especially for protein modification studies.

1.8. Objective of this research

The objective of this research was to develop a mass spectrometry based method to detect in vitro O-GlcNAc modification. Such a method could provide a relatively simple system for investigating O-GlcNAc modification and together with an in-vitro glycosylation system to study the relationship between glycosylation and phosphorylation.

Expression, purification and characterization of NCOAT and OGT, the two enzyme system regulating O-GlcNAc modification was necessary. The substrate used for in vitro glycosylation, a nine amino acid peptide YSDSPSTST, was also needed. To optimize the mass spectrometry parameters for detection of O-GlcNAc modified peptide, we used a synthetic glycopeptide YSDSPSgTST as a standard.

2 Experimental Methods

2.1 Bacterial culture condition

Plasmids were propagated in the *E.coli.* top 10 cells before extraction. Bacterial cultures were grown in LB broth media that contained 8 g/L bacto-tryptone, 5 g/L bacto-yeast extract and 5 g/L NaCl. Where applicable, ampicillin or kanamycin was added to a final concentration of 100 μg/mL or 25 μg/mL respectively. A single colony was inoculated into 50 mL of cultures and grown at 37 °C for 10 hours. The cultures were then inoculated to 1000 mL LB broth and grown at 37 °C for 14 to 16 hours under continuous agitation. To pick up single colonies, bacteria were spread over LB agar which composed of 8 g/L bacto-tryptone, 5 g/L bacto-yeast extract, and 5 g/L NaCl, 15 g/L agar. Where appropriate, the ampicillin or kanamycin was added as antibiotic to a final concentration of 100 μg/mL or 25 μg/mL respectively. Bacterial plates were incubated at 37 °C overnight.

2.2 Preparation of competent cells

Electrocompetent BL21 (DE3) Rosetta and Top10 *E.coli.* cells were made as follows. Cells were harvested from 500 mL cell cultures grown to an OD600 of between 0.4-0.5. The cells were then incubated on ice for 15 minutes and then harvested by centrifugation at 4 °C for 20 minutes at 3,000 rpm, washed twice with 250 mL of ice-cold sterile water. Then the cells were diluted and

resuspended with 5 mL SOC media. The final cell suspension was collected as 100 μL aliquots, flash frozen with liquid N_2 and stored at -80 °C. The Top10 electrocompetent E.coli cells were determined to have a transformation efficiency of 5.4 x 10^8 colonies forming unit (cfu) per μg, with vector, pET32a(+)-pPROEX-Htb-TEV.

BL21 (DE3) Rosseta and Top10 E.coli competent cells were also prepared by the calcium chloride method. Cells were collected from 500 mL cell cultures grown to an OD600 of 0.4 and then incubated on ice for 10 minutes. The cells were subsequently harvested by centrifugation at 4 °C for 20 minutes at 3,000 rpm. The pellet was resuspended in 75 mL of an ice-cold 100 mM $CaCl_2$, 20% glycerol and collected as 0.5 mL aliquots which were frozen in liquid nitrogen and stored at -80 °C.

2.3 Transformations

For transformations by electroporation, one aliquot of electrocompetent cells were thawed on ice. 1 ng of plasmid DNA or 15 ng of DNA from ligation was mixed with 50 μL of competent cells. The cells were pelleted by spinning in a Beckman Coulter Allegra 25R benchtop centrifuge at 5000 rpm for 1 min at 4 °C. Then the cells were resuspended in 200 μL of ice cold water twice to get rid of the ionic contaminants. Finally the cells were resuspended in 100 μL cold water and mixed with DNA. The mixture was then transferred to an ice cold

electroporation cuvette (Bio Rad, gap size 1mm). Transformations were done by an Eppendorf 2510 Electroporator using a setting of 1250 V. The sample was quickly transferred the mixture to 1 mL of SOC media (20 g/L bacto-tryptone, 5 g/L bacto-yeast extract, 0.5 g/L NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$ and 20 mM glucose) at room temperature and incubated at 37 °C for 30 minutes. Transformed bacteria were then spread over an LB agar plate containing the appropriate antibiotic.

For transformations of competent cell prepared by the calcium chloride method, the DNA product was mixed with 100 μL of the competent bacteria and incubated on ice for 30 minutes. The mixture was then incubated at 42 °C for 45 seconds. 500 μL of warm LB was added and the mixture was incubated at 37 °C for 1 hour. The transformed bacteria were then spread evenly over the LB agar containing the appropriate antibiotic. The plates were incubated at 37 °C overnight to allow the colonies to grow.

2.4 Purification of plasmid DNA

Plasmid preparations were made using QIAprep Spin Miniprep Kit from Qiagen. Colonies were picked up from LB agar plates and grown overnight in 2 mL LB broth containing the appropriate antibiotic. 1.5 mL of culture was transferred to an eppendorf tube and centrifuged for 1 minute at 13,000 rpm. The supernatant was discarded and the pellet was resuspended in 250 μL RNase added buffer P1. 250 μL of buffer P2 was added into the tube, mixed thoroughly

by inverting the tube 5 times. 350 μL of buffer N3 was added immediately, mixed by inverting the tube for 5 times. The tube was centrifuged at 13,000 rpm for 10 minutes. The supernatant was transferred to a QIAprep spin column using a pipette. After centrifugation at 13,000 rpm for 60 seconds, the flow through was discarded and 750 μL of PE buffer was added to wash the plasmid free of trace nuclease activity. 50 μL of water was added, and the plasmid was collected into a new microcentrifuge tube by centrifuging for 60 seconds.

2.5 Gel electrophoresis

Two different electrophoretic techniques were applied during the work of the thesis; agarose gel electrophoresis and SDS-polyacrylamide gel electrophoresis.

Agarose gels were used to separate and analyze the DNA fragments. Agarose gels were made using 1×TAE buffer (40 mM Tris, 0.11% glacial acetic acid, 1mM EDTA, pH=8.0 and containing 1.0% agarose). Samples for electrophoresis were mixed with one-fourth sample volume of 5×loading buffer (40 mM Tris, 50% glycerol, 0.25% bromophenol blue in water). The electrophoresis was performed at 100V in 1×TAE buffer. After the electrophoresis was done, the gels was stained with 1 μg/mL of ethidium bromide for an hour, and destained with water for half and hour. The DNA fragments were visualized under UV light.

SDS-polyacrylamide gel electrophoresis (SDS-PAGE) was used as a separation technique for protein analysis. The stock acrylamide mixture consisted of 30% acrylamide and 0.8% N,N'-methylene-bis-acrylamide and was stored at 4 °C. The separating gel consisted of 10% or 15% acrylamide (depending on the required resolution), 375 mM Tris-HCl (pH 8.8), 0.1% SDS, 0.1% ammonium persulfate and 0.01% N,N,N',N'-Tetramethyl ethylenediamine (TEMED). The separating gel was poured, covered with water, and allowed to polymerize for an hour at room temperature. After removal of the water cover, a stacking gel contain 4% acrylamide, 125 mM Tris-HCl (pH 6.8) 0.1% SDS, 0.1% ammonium persulfate and 0.01% TEMED was poured and allowed to polymerize for 30 minutes. Samples were mixed with 3×SDS loading buffer (250 mM Tris-HCl (pH 6.8), 33% glycerol, 10% SDS, 16% β-mercaptoethanol, 0.06% bromophenol blue) and boiled for five minutes before loading. Electrophoresis was performed using 1×running buffer (380 mM glycine, 50 mM Tris, 0.1% SDS). SDS-polyacrylamide gels were stained with Comassie blue solution (46% ethanol, 8.4% acetic acid, 0.23% Coomassie blue) for 30 minutes. The gels were then destained in a destain solution (23% ethanol, 8.4% acetic acid) until protein bands were observable.

2.6 PCR

Pfu DNA polymerase (Fermentas, Canada) was used for PCR reaction. All primers used were synthesized by MOBIX laboratories, McMaster University.

The template DNA used to amplify the MGEA5 gene was pcDNA3.1HisC/O-GlcNAcase. For each PCR tube, 5 ng template DNA and 0.5 unit polymerase was used. All PCR reactions were performed using an Eppendorf Mastercycler PCR machine. Temperature and the time setting varied between different trials and the primers.

2.7 Restriction digests

All restriction endonucleases were purchased from Fermentas. All digests were performed with a minimum of 1 μL of each enzyme for 500 μg of DNA to be digested. The reaction volumes were typically 20-50 μL. For double digests, Fermentas double Tango buffer was used. All digest were incubated at 37 °C overnight.

2.8 Agarose gel extractions

DNA was extracted from agarose gel by using a Qiaquick gel extraction kit (Qiagen), following the procedures described in the manual for extraction. For standard gel extractions, DNA was eluted with 40 μL of the elution buffer. To obtain more concentrated DNA, elution was done with 8-30 μL elution buffers.

2.9 Ligation

T4 ligase was obtain from Fermentas, and used in all ligation reactions. Rapid ligase buffer and standard buffer provided by the T4 ligation kit were used in the reaction. For rapid ligation the temperature was 37 °C and the duration an hour while for standard ligation the temperature was 22 °C for 3 hours.

2.10 Subcloning of MGEA5 gene from pCDNA3.1 to PET32 vector.

Conventional methods were selected to subclone the gene encoding the NCOAT protein from a pcDNA 3.1 mammalian expression vector into PET32a based bacterial expression vector. The primers are listed in table 2-1.

Number	Description	Sequence 5' to 3'	Tm(°C)
GAM25	NotI Forward	CATTATTCAAAGTTCATAGTAGCGGCCG CTATGGTGCAGAAGGAGAGTCAAG	70
GAM26	SalI Forward	CATCATTCAAAGTTCGTAGTACAGTCGA CGATGGTGCAGAAGGAGAGTCAAG	71
GAM27	StuI Forward	CATCCACTAGCAAAGTTCGTCGTCAGGC CTATGGTGCAGAAGGAGAGTCAAG	73
GAM28	XhoI Reverse	CATCATCATCATCATCATCATCTCGAGTC ACAGGCTCCGACCAAGTATAACC	72
GAM29	NotI Reverse	CATAATTATCATCAATGCGGCCGCTCAC AGGCTCCGACCAAGTATAACC	71

Table 2-1 Primers of PCR for MGEA5 gene subcloning.

The amplification of the MGEA5 gene was done under 4 different primer pairs (NotI / XhoI, SalI/ NotI, StuI/XhoI, StuI/NotI). After amplifying the gene of interest using the PCR reaction, the insert and the vector were agarose purified (Qiagene) to measure their concentration for further ligation study. The ligation was done using an insert to vector ratio of 1:4, 1:2, 1:1, 2:1, 4:1, and 8:1. The ligation product was transformed into BL21 cells and single colony was selected

from the plate that contained the appropriate concentration of antibiotic. The plasmid was purified and double digested with the original restriction enzyme.

2.11 Preparation of NCOAT protein.

Purified NCOAT protein was prepared as followed. The gene encoding the NCOAT protein was subcloned from a pcDNA3.1 mammalian expression vector into PET32a based bacterial expression vector. BL21 (DE3), a protease deficient strain of E. coli was transformed with plasmid encoding the fusion protein and grown overnight in 100 mL ampicillin containing LB media. 100 mL of ampicillin containing medium was inoculated with a single colony and was grown overnight at 37 °C. 1 L of the ampicillin containing medium was inoculated with the overnight culture and allowed to grow at 37 °C until the OD_{600} reached 0.6. Different isopropyl-β-D-thiogalactopyranoside (IPTG) concentrations were used to optimize the best expression conditions. The bacteria were harvested by centrifugation at 4,000 rpm for 15 minutes. Cells were then resuspended in 10 mL of French Press lysis buffer (50 mM Tris-HCl pH 7.5, 100 mM NaCl, 1 mM EDTA, 5 mM DTT) and lysed by French Press. The suspension was centrifuged at 15,000 rpm for 30 minutes. Imidazole was added to the soluble fraction to reach a final concentration of 20 mM. The soluble fraction was then loaded onto a 1 mL Hi-Trap Nickel column (Amersham) pre-equilibrated with wash buffer (20 mM of imidazole, pH=7.5 , 250 mM of NaCl). The column was washed with 50 mM of imidazole and the bound fraction was eluted with 250 mM of imidazole.

After buffer exchange into 50mM sodium phosphate, pH=7.0, using a Amicon stirred cell (Millipore, Corp) over a 10 kDa cutoff membrane (YM10 ultrafiltration membrane, Millipore, Corp). The fraction was loaded on a Mono Q column (Amersham Bioscience, Corp) and eluted with a linear of 500 mM of NaCl, 50mM sodium phosphate. The elution was analyzed by SDS-PAGE and stored at -80 °C.

2.12 Preparation of sOGT protein

The expression of a soluble truncated variant of OGT was made possible using a pET28 based plasmid obtained as a kind gift from Dr. Walker at Harvard University. BL21 (DE3) cells were transformed and grown to an O.D. between 1.1 and 1.3 at 37 °C. The culture was cooled to 16 °C, induced with 0.2 mM IPTG for 24 hours; the cells were pelleted and frozen at –80 °C. The pellets were resuspended in 50 mL of lysis buffer (50 mM Tris-HCl pH 7.5, 200 mM NaCl, 1 mM EDTA, 5 mM DTT, and 1 mM PMSF) and a French press was used to lyse the cells. The lysate was then clarified by centrifugation at 15,000 rpm and then loaded on a Hi-Trap nickel column (GE/Amersham Bioscience) for affinity purification relying on the encoded His-tag. The column was washed with ten column volumes of 50 mM of imidazole, pH=7.5, 250 mM of NaCl, and the protein was eluted with 250 mM imidazole, pH=7.5, 250 mM of NaCl. About 15 mL of the fraction was collected and a Amicon Ultra-15 centrifuge with a 10kDa cut off was used to concentrated the sOGT protein and exchange the buffer to 25

mM Hepes, 10 mM $MgCl_2$, 1mM EDTA, and 500 μM of tris(hydroxypropyl)phosphate. The aliquots were stored at –80 °C for further use.

In subsequent expression attempts, another cell lysis method was used to achieve a better yield of sOGT protein. For this, the cell pellets (1L) were lysed with 25 mL BugBuster (Novagen) Protein extraction reagent complementary with 50 KU of rLysozyme Solution (Novagen) and 600 U Benzonase Nuclease (Novagen). The cell suspension was incubated at room temperature for 30 minutes. The purification using Novagen lysis kit is the same as described above.

2.13 Mass spectrometry.

All experiments were done on a Q-TOF Global Ultima (Waters, Micromass) with a nanoES source. Samples were dissolved in 50% aqueous acetonitrile containing 0.2% formic acid at a concentration of 100 ng/μL, and directly transferred into nanoES glass capillaries (Proxeon, Borosilicate emitters) using a pipette tip. Capillary voltage was typically 1.2-1.6 kV at a and cone voltage of 50-100 V. Mass spectra in Tof-MS and MS/MS mode were obtained in a mass range *m/z* 50-1800 with a resolution of 8000 FWHM (full width at half maximum height). Argon was used as collision gas for MS/MS.

2.14 In-Gel tryptic digestion and nanoES Q-TOF MS analysis.

The gel plugs, containing the protein spots of interest, were destained with 50 mM ammonium bicarbonate, containing 50% acetonitrile and then air dried. The proteins were reduced by adding 30 μL of 10 mM dithiotreitol (DTT) in 25 mM ammonium bicarbonate to each gel plug followed by incubating for 1 h at 56 °C. After cooling to room temperature, the DTT solution was removed and the gel plugs treated with the same volume of 100 mM iodoacetamide in 50 mM ammonium bicarbonate. After 60 min incubation at ambient temperature, in the dark, the gel plugs were washed with 30 μL of 25 mM ammonium bicarbonate for 15 min and then dehydrated with 100% acetonitrile. After 10 minutes the liquid phase was removed, and the gel plugs were completely dried in air. Next the proteins were subjected to in-gel digestion; 0.015 mg of trypsin in 30 μL of 50 mM ammonium bicarbonate solution containing 10 % acetonitrile was added to each gel plug and the resulting mixture was incubated at 37 °C overnight. The digested proteins were desalted and concentrated using a Millipore C18 ZipTip prior to mass spectrometry (MS) analysis eluting the peptides in a final step using 8 μL of 50% aqueous acetonitrile containing 0.2% formic acid. All protein digests were analyzed by a Q-TOF Global Ultima (Waters, Micromass) with a nanoES source.

2.15 Nano-electrospray mass spectrometry

After loading the capillary with 2 μL of sample solution (make sure that the sample is in the tip of the needle without air bubbles), the needle was placed in the holder and pushed forward to a position about 1-2 mm in front of the orifice. The needle position was manipulated in x, y, and z directions via the micrometer screws to ensure the needle tip was centered in front of the orifice.

To open the needle tip, the needle was briefly touched against the interface plate with all potentials at ground and then centered in front of the mass spectrometer's orifice. An air pressure of 0.8 bar was applied on the capillary holder via an air-filled syringe which caused a small droplet to appear at the tip of the needle. A potential of 600-700 V at the needle started the spray. The necessary pressure was adjusted according to the mass spectrometric signal so that a constant stable flow was achieved. A new capillary was used for each experiment.

2.16 HPLC analysis

The peptide substrate used for enzyme activity assay was YSDSPSTST, which is a known peptide substrate of the OGT protein. This substrate peptide is derived from nuclear pore protein p62. Purified peptide substrate for OGT assays was prepared in our laboratory using reversed phase HPLC on crude material obtained either from a commercial source (Elim Biopharmaceuticals, CA) or by

solid phase peptide synthesis with an Advanced ChemTech 348 synthesizer (Dr. John Valliant's lab, McMaster University). Peptide samples were purified using a Delta 600 HPLC system (Waters Canada) to elute with a 30 minute linear gradient of 0-60% acetonitrile with aqueous 0.1% v:v formic acid at 1.4 mL /min at room temperature. A SunFire Prep C18 5 μm (4.5x50 mm) column (Waters Canada) was used to develop the method on an analytical scale and a SunFire Prep C18 5 μm (10x100 mm) column (Waters Canada) was used as a semi-prep scale column to purify the peptide. Fractions from the analytical scale were analyzed by ES-MS and those with a peak at the expected *m/z* 944.4 $[M+H]^+$ were targeted for collection in the corresponding semi-prep purification. Reinjection of these fractions using an analytical column confirmed the purification and the desired fractions were combined, lyophilized and stored at -20 °C.

2.17 NCOAT activity assay

The NCOAT enzyme activity assay was measured using Spectrophotomeric Stop Rate Determination method. β-N-acetylglucosaminidase was purchased from Sigma (A-7708) and used as a positive control. The assay was performed in a 100 μL solution of 50 mM NaH_2PO_4, 100 mM NaCl , 0.1% BSA , pH6.5 for NCOAT enzyme, and 100mM citric acid , 200 mM NaCl, 0.1%BSA, pH 5.0 (β-N-acetylglucosaminidase) for 30 minutes. The reaction was

stopped by addition of 900 μL of 200 mM glycine pH 10.75. Subsequently the absorbance at 400 nM was measured.

The reaction is shown in Figure 2.1. One unit is the amount of enzyme required to catalyze the release of para-nitrophenol from 1uM of pNP-O-GlcNAc, per minute at 30°C.

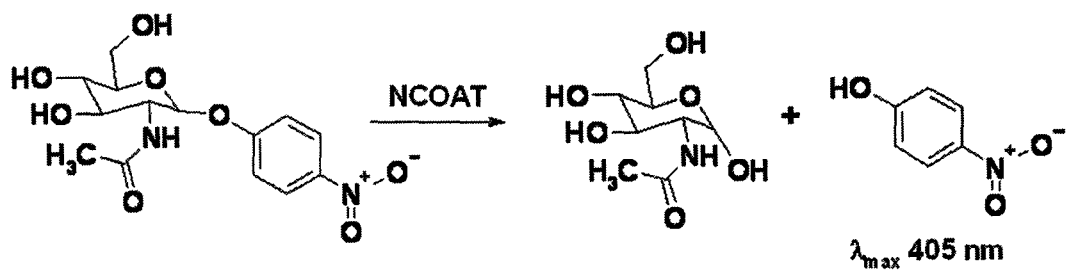


Figure 2.1 NCOAT enzyme cleaves pNP-O-GlcNAc to release pNP.

2.18 Solid phase peptide synthesis (SPPS)

Solid phase peptide synthesis was carried on a peptide synthesizer (Dr. Valliant's group) with the help by Ms. Leslie. Fmoc-Thr(tBu)-Wang resin (100-200 mesh, 0.52-0.60 mmol/g, 2200 mg) was used as the solid support resin and amino acid building blocks (Tyr 0.92 g in 4 mL DMF, Ser 1.92 g in 10 mL DMF, Pro 0.67 g in 4 mL DMF, Asp 0.82 g in 4 mL DMF, Thr 1.19 g in 6 mL DMF) as the starting materials. A t-Bu side-chain protection was used for all amino acids. Fmoc removal was achieved with piperidine in DMF (15 mL in 85 mL DMF). Fmoc-amino acids were coupled for total 30 h in the presence of HBTU (3.03 g in 16 mL DMF)/ DIPEA (4.14 g in 16 mL DMF) at room temperature. After finishing

SPPS, the peptide resin (2 x 200 mg) was transferred to a glass bottle. Peptide YSDSPSTST was cleaved from the resin with concurrent removal of the tBu side-chain protecting groups by treatment with 2 x 2 mL TFA-TIS-H_2O (95%TFA, 2.5%TIS, 2.5%H_2O) for 24 h at 25 °C. The resulting peptide solution was added to a centrifuge tube filled with about 7 mL of cold diethyl ether. After shaking this tube for half a minute, some precipitate formed and the solution was centrifuged for 2 min. The upper solution was removed and the solid was washed with 4x7 mL cold diethyl ether. This solid was lyophilized overnight and then purified by reversed phase HPLC. The collected fractions were pooled and lyophilized to get about 40 mg (yield: 36%) of acetylated glycopeptide.

2.19 sOGT enzyme activity assay

Initially the sOGT enzyme activity assay was performed using the following conditions. 1 μg of the purified sOGT, 100 ng of purified peptide substrate YSDSPSTST and 0.1 μCi of UDP-[^{3}H]-GlcNAc (36.5Ci/mmol, Sigma) were put in 50 μL reaction buffer containing 50mM Tris-HCl, 1mM DTT, and 12.5 mM $MgCl_2$. After incubation at 37 °C for 30 minutes, the reaction was stopped by addition of 450 μL of 50 mM formic acid. Next the reaction mixture was then loaded onto a 0.5 mL Sephadex (SP-25) column equilibrated in 50 mM formic acid. The column was washed with 50 mM formic acid and eluted with 2 mL of 0.5M NaCl. 10 mL of Liquiscint scintillation fluid (DiaMed Canada) was added

and the radioactivity was counted using a LS 6500 scintillation counter (Beckman Coulter Co.).

Subsequently the assay was modified in several ways in an attempt to obtain significant levels of [^{3}H]-GlcNAc labeled peptide product. The assay was performed in 50 μL of 50mM Tris-HCl, 1mM DTT, and 12.5 mM $MgCl_2$. Aliquots (0.1 μCi ,1 μL) of the UDP-[^{3}H]-GlcNAc stock solution (36.5Ci/mmol, Sigma) were dried by speedvac and reconstituted in 1 μL of nanopure water. 0.1 μCi of [^{3}H] UDP-GlcNAc (36.5 Ci/mmol), 100 ng of peptide substrate YSDSPSTST and 1 μg of purified sOGT protein were added in a 0.5 mL microcentrifuge tube. After incubation for 2 hours at 24 °C, the reaction was stopped by adding 100 μL of 15% TCA to precipitate the peptide and sOGT protein. The mixture was centrifuged at 20,000 g for 20 s to pellet both peptide and sOGT protein. The supernatant was discarded and the pellet was washed with 200 μL of 15% TCA and centrifuged again. After repeating the wash step three times, the pellet was resuspended in 100 μL of 0.5% SDS and transferred to a scintillation vial. 5 mL of Liquiscint scintillation fluid was added and the radioactivity was counted in a LS 6500 scintillation counter (Beckman Coulter Co.).

Successful measurement of [^{3}H]-GlcNAc labeled peptide product was achieved using Sep-Pak C18 column (Waters, Canada) to separate the peptide product from the [^{3}H]-GlcNAc. The reaction buffer and starting materials are the same as described above. After incubation at 24 °C for 2 hours, the reaction was

stopped by adding 500 μL of 0.5% formic acid. The Sep-Pak column was activated by 2 mL 60% methanol and equilibrated with 5 mL of 0.5% formic acid. The reaction mixture was then loaded on to Sep-Pak column, washed with 10 mL 0.5% formic acid followed by 5 mL of 60% methanol. 5 mL of Liquiscint scintillation fluid was added to the fractions and the radioactivity was counted in a LS 6500 scintillation counter (Beckman Coulter Co.).

2.20 β–elimination of glycopeptide

To detect the sites of O-GlcNAc modification, the product obtained from in vitro glycosylation was dried in a speed-vac and subjected to β–elimination in 0.5 mL 1% triethylamine, 0.1% NaOH. The reaction was incubated at 45 °C for 2 hours and the peptides were desalted using C18 spin column. The reaction was shown in Figure 2.2.

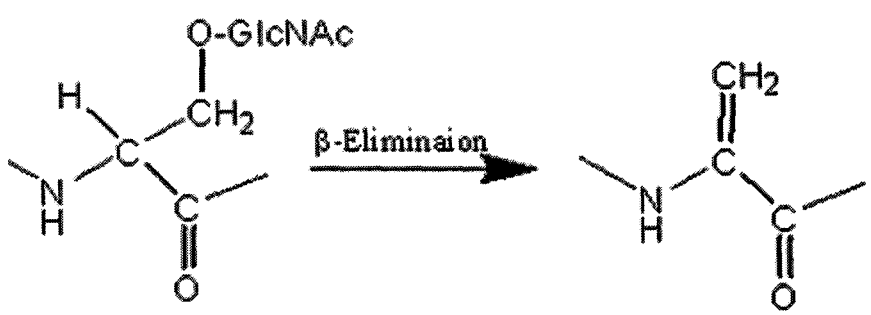


Figure 2.2 β-elimination of glycopeptide

3 Results and Discussion

3.1 Subcloning of MGEA5 gene from pCDNA3.1 to PET32 vector.

PCR amplification of the pcDNA3.1 using primer GAM25 and GAM 28 (Table 2-1) as templates produced DNA of the expected size (8000 bp as shown in Fig 3.1 lane 4). The gene was ligated into pET32a vector successfully at a vector: insert ratio of 1:4 to provide the desired plasmid.

To confirm that the reconstructed plasmids contain the MGEA5 gene, the plasmids were subjected to restriction digests using the same restriction enzyme pair. The gel bands indicate that the subcloning of the MGEA5 from pcDNA3.1HisC to pET32a(+)-pPROEX-Htb-TEV was successful. Clone 1 from lane 4 was selected for sequencing and along with clones from the other three groups, SalI/ NotI, StuI/XhoI, StuI/NotI, which all worked well too (data not shown). The clones that had the right pattern after double digest were submitted to sequencing, which confirmed the ligation result.

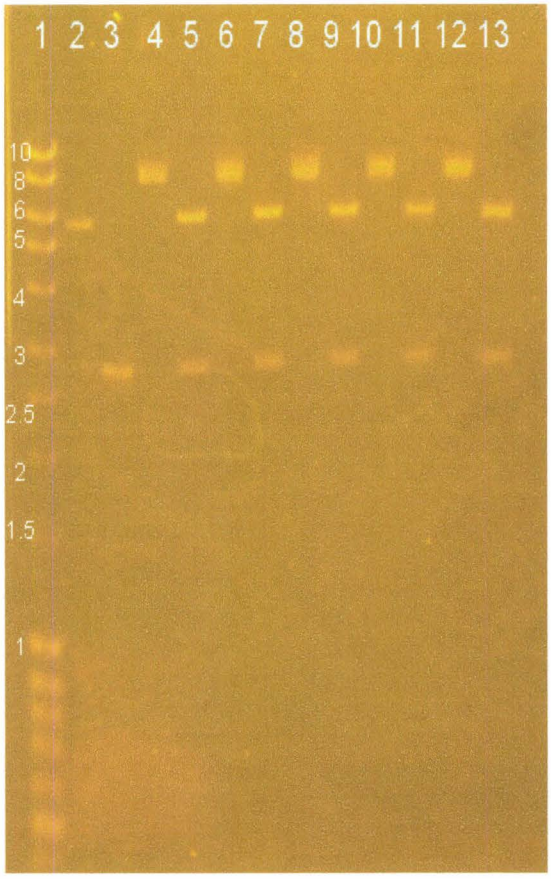


Figure 3.1 Subcloning of MGEA5 from pCDNA3.1 to pET32a(+)pPROEX-Htb-TEV plasmid and mini prep double digests.
Lane1: DNA marker.
Lane2: pET32a(+)-pPROEX-Htb-TEV NotI degest.
Lane3: Standard PCR product
Lane4: Clone 1 plasmid mini prep.
Lane5: Clone 1 plasmid mini prep double digest with NotI/Xho
Lane6: Clone 2 plasmid mini prep.
Lane7: Clone 2 plasmid mini prep double digest with NotI/XhoI
Lane8: Clone 3 plasmid mini prep
Lane9: Clone 3 plasmid mini prep double digest with NotI/XhoI
Lane10: Clone 4 plasmid mini prep
Lane11: Clone 4 plasmid mini prep double digest with NotI/XhoI
Lane12: Clone 5 plasmid mini prep
Lane13: Clone 5 plasmid mini prep double digest with NotI/XhoI

3.2 Expression and purification of NCOAT protein

The expression of NCOAT protein was carried out as described in the experimental section. Interestingly, there was no significant change of the protein expression level before and after induction. To further demonstrate this, we expressed the NCOAT protein with different IPTG concentrations and temperatures. As shown in Figure 3.2, the band at 140 kDa indicated NCOAT protein (red arrow position). The relative amount of NCOAT protein was observed to be less dependent on conditions of both temperature and IPTG concentration.

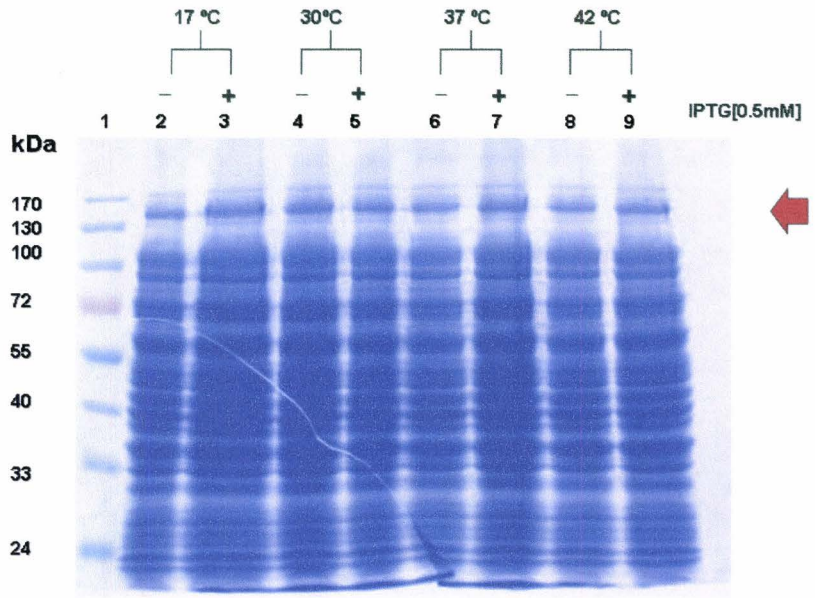


Figure 3.2 NCOAT expression at different conditions.
Lane1 protein marker.
Lane2, 3: -/+ 0.5 mM IPTG at 17°C.
Lane4, 5: -/+ 0.5 mM IPTG at 30°C.
Lane6, 7: -/+ 0.5 mM IPTG at 37°C.
Lane8, 9: -/+ 0.5 mM IPTG at 42°C.

NCOAT protein was purified as described in the Experimental Methods. As shown in Figure 3.3A, pure protein at MW 140 kDa was obtained after His-tag purification followed by ion-exchange purification. The protein samples were saved from each step of purification to determine the enzyme activity.

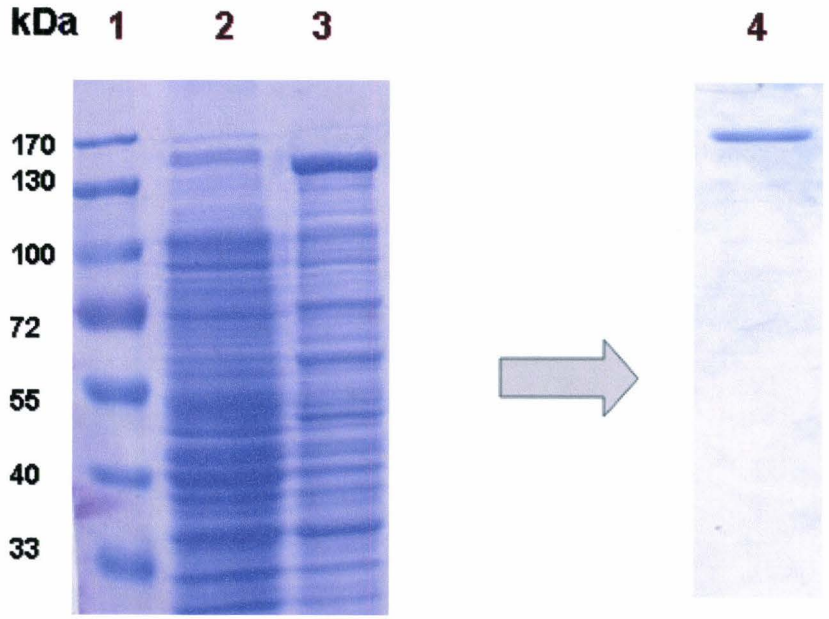


Figure 3.3 The purification of NCOAT protein.
Lane 1: protein marker.
Lane 2: flow through after nickel column purification.
Lane 3: elution from nickel column.
Lane 4: elution from Ion-exchange column.

3.3 Protein identification of the NCOAT protein by in gel digestion

The results from Figure 3.3 indicated that the purification of NCOAT enzyme was successful. The purified protein mass was about 140 kD according to the protein molecular weight marker which was compatible with what we expected. Further evidences are required to fully confirm the identity of the protein of interest. The general approaches to achieve this are either by immunoblotting against specific antibody or protein identification using mass spectrometry. Taking advantage of the instrument available in our lab, a quadrupole-time-of-flight (Q-Tof) mass spectrometer, I decided to carry out the protein identification using mass spectrometry followed by in-gel digestion.

The gel band from Figure 3.3 lane 4 was cut and in gel digested by trypsin to conduct the protein identification study using ESI-MS/MS. Figure 3.4 is the mass spectrum of the tryptic digest of the protein of interest. This mass spectrum displays the ions between *m/z* 200 and *m/z* 1000 from a spectrum acquired with a quadrupole-time-of-flight (Q-Tof) mass spectrometer. The data were acquired with an *m/z* resolution of ~7000. Because of the resolution of the time-of-flight mass analysis, the charge state of each ion can be determined from the spacing in the isotope clusters. The doubly charged *m/z* 679.85, 736.88 and triply charged 794.76 were selected for MS/MS to obtain the sequence information of the selected peptides indicated in Figure 3.4.

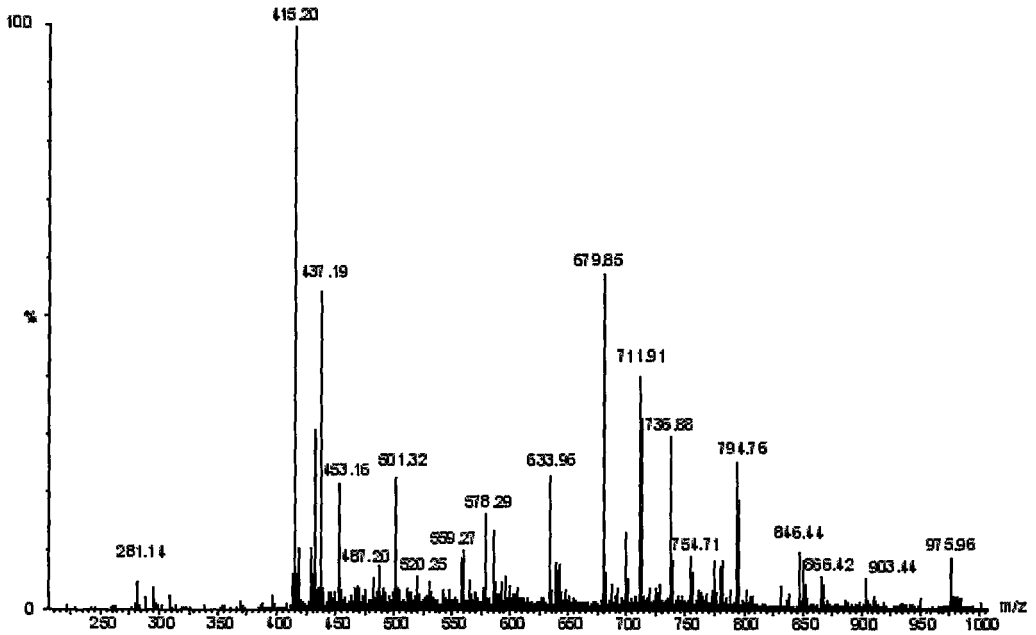


Figure 3.4 NanoESI Mass Spectrum of NCOAT protein after trypsin digestion

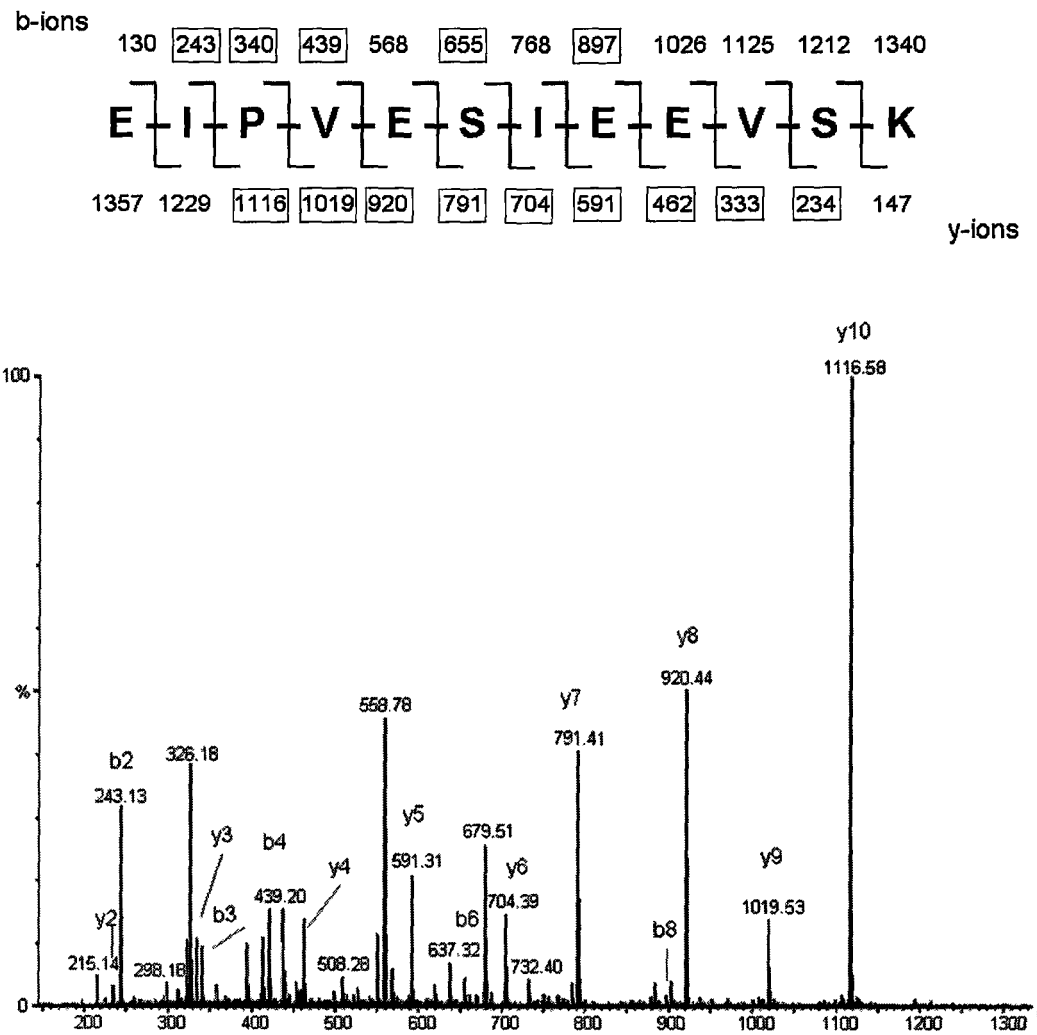


Figure 3.5 MS-MS of *m/z* 679.85 peak (EIPVESIEEVSK) of NCOAT protein trypsin digestion. The corresponded y series and b series of ions are labeled in the spectrum. The sequence of the peptide was obtained by matching the mass spectrum with Mascot database. The numbers listed on top and button of the sequence are theoretical calculation of the b and y series of ion. The experimental data which matched the theoretical calculation is highlighted with a box.

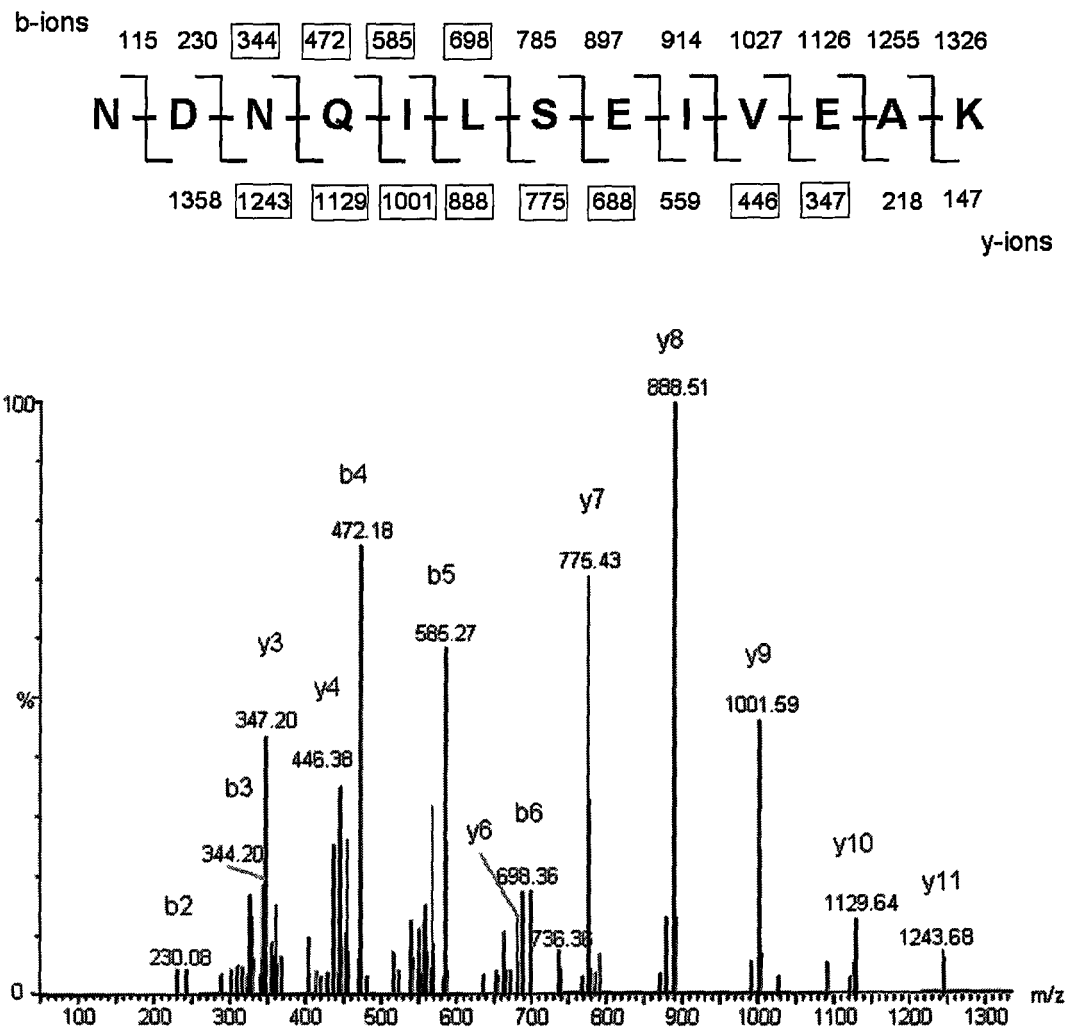


Figure 3.6 MS-MS of *m/z* 736.88 peak (NDNQILSEIVEAK) of NCOAT protein trypsin digestion

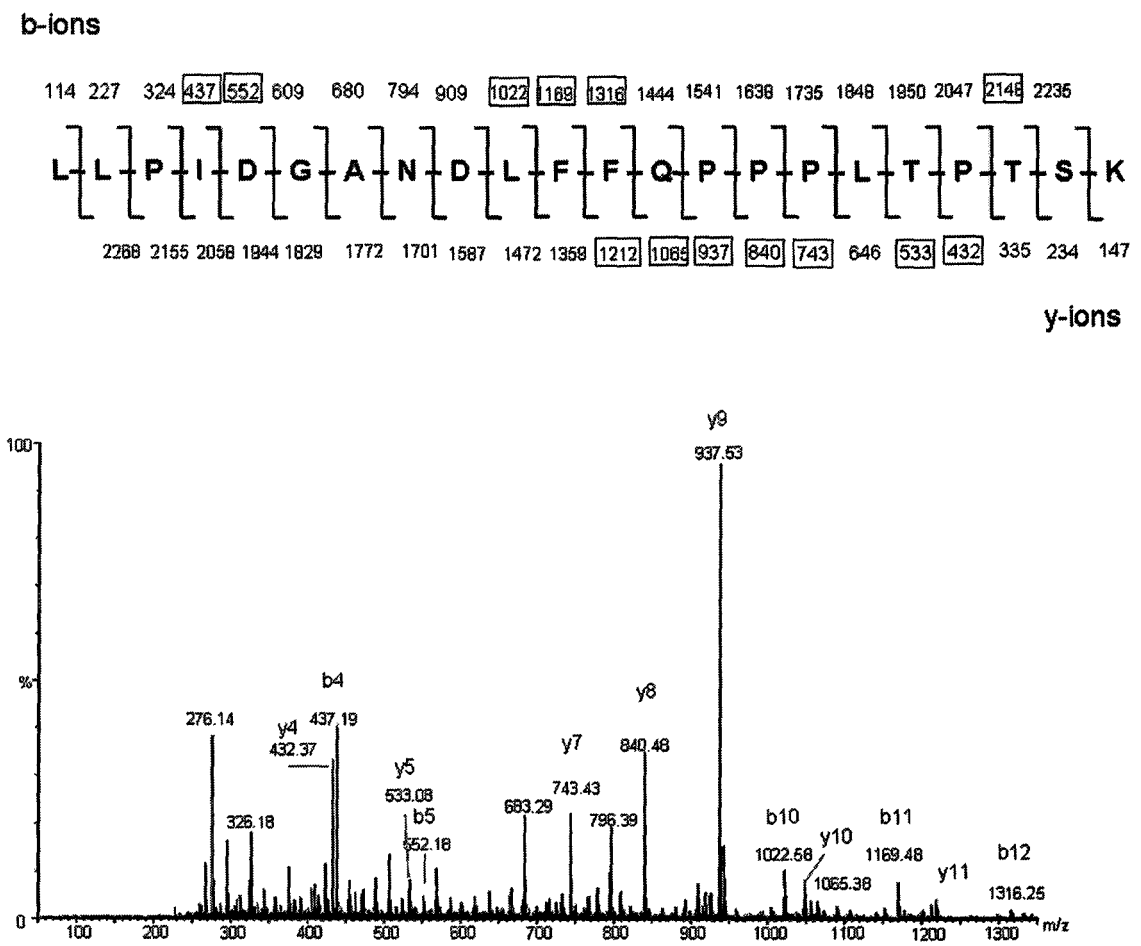


Figure 3.7 MS-MS of the triply charged ion *m/z* 794.76.

Figure 3.5 is a product ion spectrum of the doubly charged peptide ion at *m/z* 679.85. The product ion spectrum of *m/z* 679.85 was acquired by using an electrospray ionization-Q-Tof system. The precursor ion was selected by using the quardupole mass analyzer, and the product ion spectrum was detected by using time-of-flight *m/z* analysis.

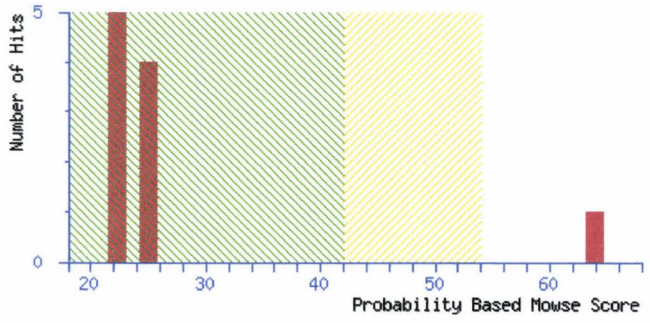
Figure 3.6 is a product ion spectrum of a doubly charged peptide ion *m/z* 736.88. Figure 3.7 is a product ion spectrum of a triply charged peptide ion *m/z* 794.76. Doubly charged ions are ideal for MSMS protein identification from a trypsin digested peptide. It produces a series of y and b-ions because it has basic group at both C and N terminal of the peptides. The ions listed in a box are the ions which matched the predicted series of y and b-ions.

The database search for the peptide mass fingerprint and the product ion spectra was performed using Mascot, an internet-based search engine. The search parameters used for this analysis were as follows: the minimum mass intensity cut off used is 1%, the peptide mass tolerance was kept at 0.2 Da, only one missed cleavage was allowed and to account for the alkylation reaction performed with iodoacetamide, carbamidomethylation of cysteine was chosen as the only variable modification. The sequence list at the top of the figure is the predicted peptide sequence generated from the database. The ions listed in brackets are the experiment ions which matched the predicted y and b-ion series.

With the development of the "soft ionization" technique (ESI and MALDI), the increase of instrument resolution, mass spectrometry is now one of the most efficient ways to identify proteins [74,75]. Figure 3.8 shows an example of the generated Mascot result report of the doubly charged *m/z* 736.88 from the in-gel tryptic digestion of NCOAT protein.

The development of several algorithms and computer programs facilitate the protein identification process [76]. The Mascot is a probability-based search engine developed in 1999 [77]. The out put from Mascot score is $-10\text{Log}_{10}(P)$, where P is the probability that the observed match between the experimental data and searched sequence database entry is a random even [77]. The program used significance threshold, where the probability of the observed match occurring by chance is less than 0.05 ($p<0.05$). Whether the observed match is significant depends on the database size for each entry. As shown in Figure 3.8 the peptide score of 64 is more than the 54 score that is considered as identity or extensive homology.

Probability Based Mowse Score
Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 42 indicate peptides with significant homology.
Individual ions scores > 54 indicate identity or extensive homology ($p<0.05$).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



Query	Observed	Mr(expt)	Mr(calc)	Delta	Miss	Score	Expect	Rank	Peptide
1	736.88	1471.7454	1471.752	-0.0064	0	64	0.0053	1	K.NDNQILSEIVEAK.M

Proteins matching the same set of peptides:

Q9HAR0_HUMAN **Mass**: 76829 **Score**: 64 **Queries matched**: 1
Meningioma-expressed antigen 5s splice variant.- Homo sapiens (Human).
Q86WV0_HUMAN **Mass**: 68322 **Score**: 64 **Queries matched**: 1
MGEA5 protein (Fragment).- Homo sapiens (Human).
T00360 **Mass**: 87364 **Score**: 64 **Queries matched**: 1
Hypothetical protein KIAA0679 - human (fragment)
O60502_HUMAN **Mass**: 103932 **Score**: 64 **Queries matched**: 1
Meningioma-expressed antigen 5.- Homo sapiens (Human).
O75166_HUMAN **Mass**: 104088 **Score**: 64 **Queries matched**: 1
KIAA0679 protein (Fragment).- Homo sapiens (Human).
CAE90035 **Mass**: 54835 **Score**: 64 **Queries matched**: 1

Figure 3.8 Example of the main result report from the MS/MS of *m/z* 736.88 from NCOAT protein in-gel tryptic digest.

Table 3-1 lists all the analyzed tryptic peptides as selected from the in-gel digestion, and Figure 3.9 shows the sequence coverage map of the NCOAT protein. 16 tryptic peptides were used with a signal to noise ratio > 3. Protein identification using at least two significant peptides is considered acceptable in the proteome analysis [78]. I used three peptides as shown in Figure 3.5, Figure 3.6, Figure 3.7 to identify the NOCAT protein and all of them obtained scores higher than the significant match thresholds.

The sequence coverage of NCOAT protein is about 30%. Sequence coverage is an important factor especially in the protein posttranslational modification studies [79]. It provides a tool for mapping the modification sites. Nowadays, less than 50 O-GlcNAc sites have been identified and there is a growing need to improve the current methods for O-GlcNAc site mappings [80]. One of the commonly used methods to map the O-GlcNAc modification site is by comparison of the peptide fingerprint map generated before and after releasing the sugar from the protein. This required a high sequence coverage rate to target the modified peptides. There are several methods to increased the sequence coverage such as using a combination of both MALDI and ESI spectrum of the same proteolysis fragments, using different enzymatic cleavage approach (such as Lys-C or Asp-N) or combine separation technique before mass spectrometry analysis [81].

One of the drawbacks of using MALDI or ESI MS to analyze peptides mixture is their lower efficiency when characterizing the hydrophobic peptides. The discrimination based on the peptide hydrophobicity makes hydrophobic peptides more challenging to detect. This is mainly due to their limitation of the solubility in aqueous solutions [82].

It has been reported that the presence of low amounts of surfactants such as SDS, CHAPS (3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate), Triton X-100 could improve the signal of hydrophobic peptides and increase the sequence recovery rate [83]. This is quite interesting because the surfactants are generally considered to be problematic and should be avoid before MS analysis. Considering the fact that proteins expressed in *E.coli* system are lack of proper post translational modification, experiments to improve the sequence recovery rate were not been done in this thesis. However this should be considered when analyzing the samples obtained from mammalian expression system.

Table 3-1 Observed fragments in NCOAT tryptic digest and comparison with the Database Output (SWISS-PROT).

Peptide Sequence	Theoretical Mass	Experimental Mass	Position	Charge State
IEEWRSRAAK	415.8948	415.8	604-613	3
SNMNGVRK	453.2347	453.16	334-341	2
MEGFPKDVVILGR	487.6025	487.2	902-914	3
APVIWDNIHANDYDQK	633.6392	633.96	274-289	3
EIPVESIEEVSK	679.8561	679.85	258-269	2
LLPGIEVLWTGPK	711.9214	711.91	241-253	2
NDNQILSEIVEAK	736.8832	736.88	479-491	2
ISWIPFMQEKYTKPNGDKELSEAEK	742.8745	742.76	795-819	4
RKLDQVSQFGCR	754.3935	754.71	156-167	2
QPDEEPMDMVVEK	789.8367	789.69	459-471	2
LLPIDGANDLFFQPPPLTPTSK	794.76	794.76	692-713	3
VTDPSVAKSMMACLLSSLKANGSR	846.4343	846.44	851-874	3
LSNCANRTILYDMYSYVWDIKSIMSMVK	866.6656	866.42	628-655	4
RFLCGVVEGFYGRPWVMEQRK	876.7824	876.46	59-79	3
GAQMLREFQWLRANSSVVSVNCK	904.1233	904.46	575-597	3
GVLTNPNCEFEANYVAIHTLATWYK	975.8128	975.96	309-333	3

```
1    MVQKESQATL EERESELSSN PAASAGASLE PPAAPAPGED NPAGAGGAAV AGAAGGARRF
61   LCGVVEGFYG RPWVMEQRKE LFRRLQKWEL NTYLYAPKDD YKHRMFWREM YSVEEAEQLM
121  TLISAAREYE IEFIYAISPG LDITFSNPKE VSTLKRKLDQ VSQFGCRSFA LLFDDIDHNM
181  CAADKEVFSS FAHAQVSITN EIYQYLGEPE TFLFCPTEYC GTFCYPNVSQ SPYLRTVGEK
241  LLPGIEVLWT GPKVVSKEIP VESIEEVSKI IKRAPVIWDN IHANDYDQKR LFLGPYKGRS
301  TELIPRLKGV LTNPNCEFEA NYVAIHTLAT WYKSNMNGVR KDVVMTDSED STVSIQIKLE
361  NEGSDEDIET DVLYSPQMAL KLALTEWLQE FGVPHQYSSR QVAHSGAKAS VVDGTPLVAA
421  PSLNATTVVT TVYQEPIMSQ GAALSGEPTT LTKEEEKKQP DEEPMDMVVE KQEETDHKND
481  NQILSEIVEA KMAEELKPMD TDKESIAESK SPEMSMQEDC ISDIAPMQTD EQTNKEQFVP
541  GPNEKPLYTA EPVTLEDLQL LADLFYLPYE HGPKGAQMLR EFQWLRANSS VVSVNCKGKD
601  SEKIEEWRSR AAKFEEMCGL VMGMFTRLSN CANRTILYDM YSYVWDIKSI MSMVKSFVQW
661  LGCRSHSSAQ FLIGDQEPWA FRGGLAGEFQ RLLPIDGAND LFFQPPPLTP TSKVYTIRPY
721  FPKDEASVYK ICREMYDDGV GLPFQSQPDL IGDKLVGGLL SLSLDYCFVL EDEDGICGYA
781  LGTVDVTPFI KKCKISWIPF MQEKYTKPNG DKELSEAEKI MLSFHEEQEV LPETFLANFP
841  SLIKMDIHKK VTDPSVAKSM MACLLSSLKA NGSRGAFCEV RPDDKRILEF YSKLGCFEIA
901  KMEGFPKDVV ILGRSL
```

Figure 3.9 Coverage map of NCOAT protein tryptic digestion. 16 proteolytic fragments result are found with S/N>3, the sequence coverage is about 30.6% (280 out of 916).

3.4 Determination of the specific enzyme activity.

The specific enzyme activity of NCOAT protein was determined after each step of purification. The total protein amount was determined by Bradford assay and the yield of active enzyme obtained from each step was calculated and shown in Table 3-2. As shown here, the total protein amount during each step was decreased as well as the total enzyme unit. The specific activity was calculated and increased during each step, indicating a successful purified of active enzyme.

Table 3-2 Determination of specific enzyme activity.

Purification steps	Protein (mg)	Enzyme (unit)	Specific Activity (units/mg)	Fold Purification	% yield
Cell extract	292.0	180	0.61	N/A	N/A
Soluble Fraction	124.8	156	1.25	2	86
Hi trap	3.1	115	37.1	60.82	63
Mono Q	1.0	97	97.0	159	53

One Unit is the amount of enzyme required to catalyze the release of p-nitrophenyl from 1 umole of pNP-O-GlcNAc per minute at 30 °C.

3.5 Purification of sOGT protein

A smaller variant of OGT (identified as sOGT) was reported in the literature and the solubility of the new construct was significantly higher, 10 mg/L culture. To take advantage of the sOGT plasmid obtained from Walker's group [25], expression and purification of sOGT protein was undertaken. French press lysis gave about 2 mg per liter of culture rather than the 10mg per liter reported in the literature, but lysis using the lysozyme kit from Novagen gave a better yield of 5mg protein per liter cells. During the experiment it was observed that the Novagen lysis buffer will precipitate at 4 °C. As a result, the purification using Hi-Trap column had to be done at room temperature. Out of concern for possible effects of this on the enzyme's activity, the French press lysis method was used to purify the active protein, see experimental 2.12.

After purification, SDS-PAGE was run to test the purification results as shown in the Figure 3.10. The yield of the protein was estimated to be 2 mg per liter. The purity was relatively high as well.

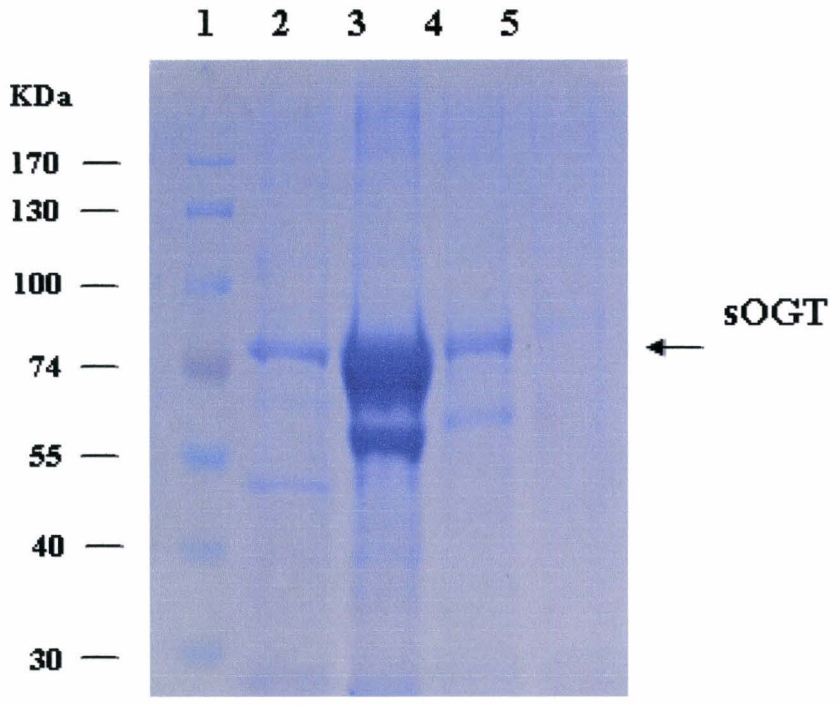


Figure 3.10 sOGT protein purification. Lane 1 is protein marker, lane 2: elution with 50 mM imidazole, lane 3: elution with 150 mM imidazole, lane 4: elution with 175 mM imidazole, lane 5: elution with 250mM Imidazole

3.6 sOGT protein identification by in gel digestion and MS.

The gel band in Figure 3.10 lane three at 74 kDa was cut after electrophoresis and the stain step. The protein identification study was conducted using the experimental ESI-MS/MS protocol.

Figure 3.11 shows the mass spectrum of ions between *m/z* 400 and *m/z* 940 from a spectrum acquired with a Q-Tof mass spectrometer. The doubly charged precursor ions (*m/z* 552.8, *m/z* 696.8 and *m/z* 915.44) were selected for MS/MS, see Figure 3.12, Figure 3.13 and Figure 3.14. The ions listed in a box are the observed ions which matched the predicted series of y and b-ion, using the Mascot search engine.

The protein sequence coverage is about 27.9% for sOGT protein as shown in Figure 3.15, which is sufficient for protein identification [84].

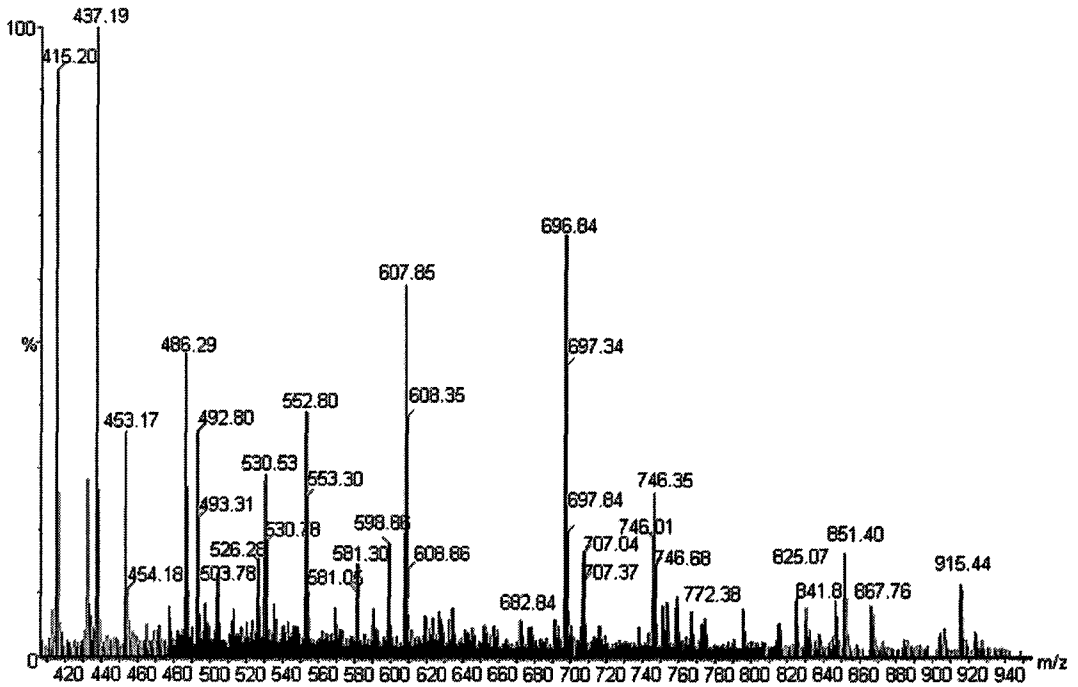


Figure 3.11 ESI-MS of the sOGT protein after trypsin digestion.

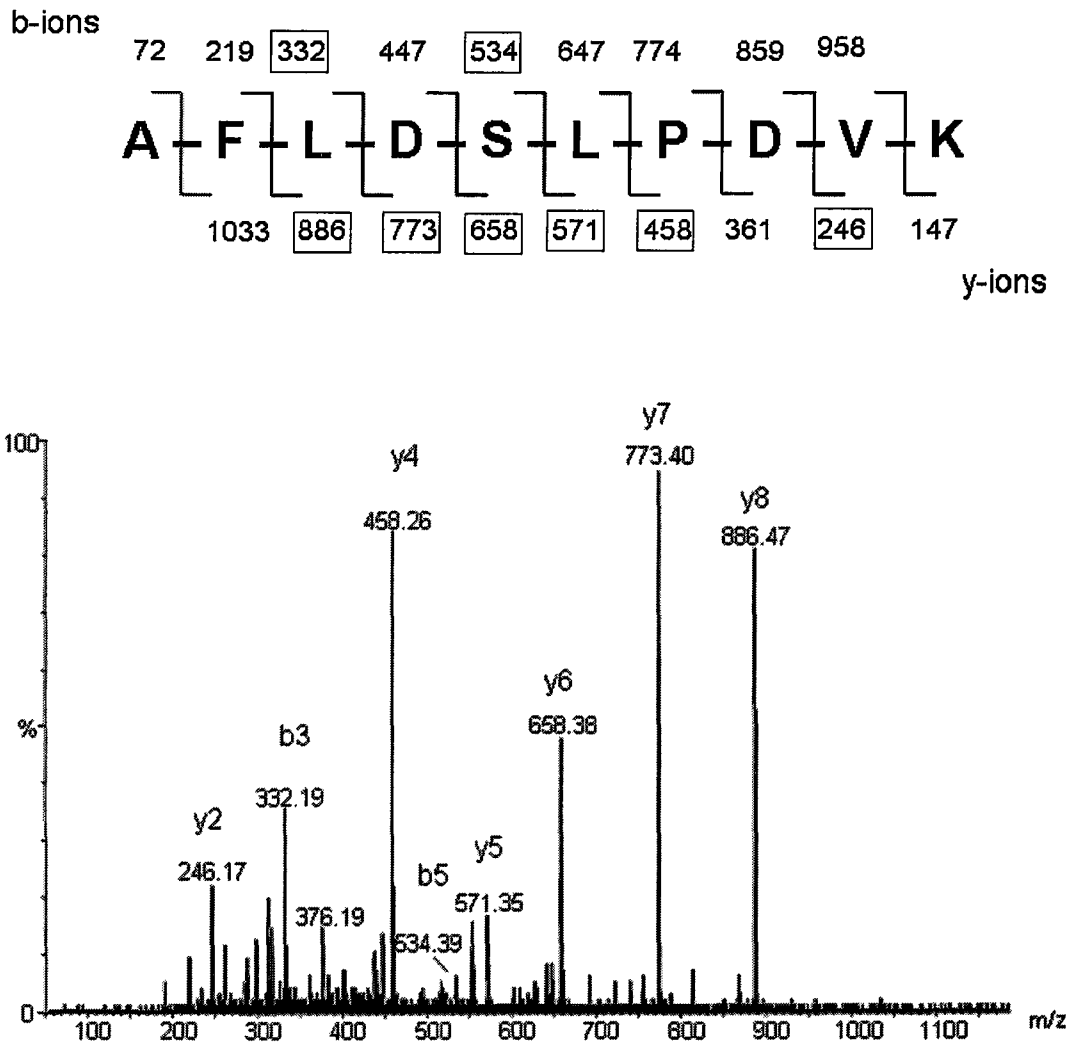
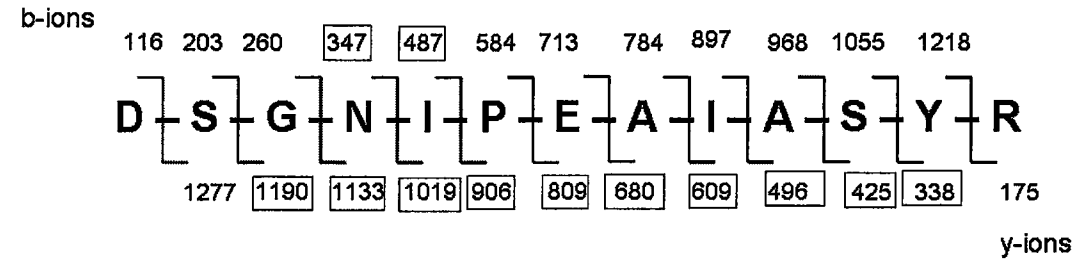


Figure 3.12 MS/MS of doubly charged ion *m/z* 552.8



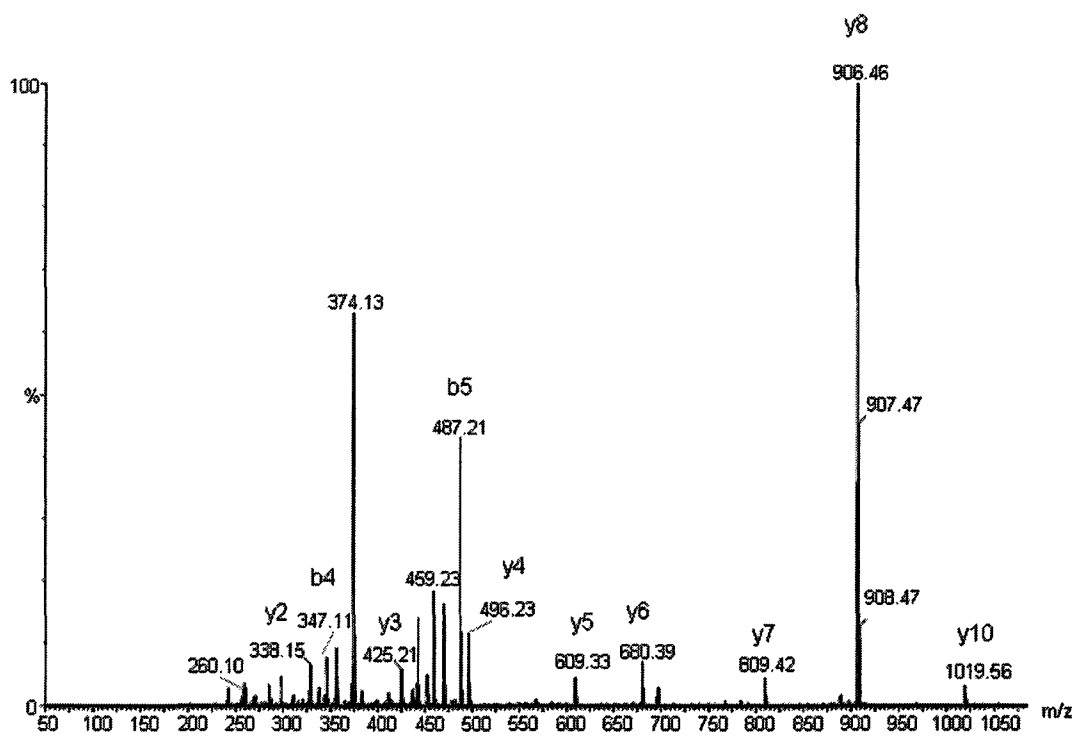


Figure 3.13 MS/MS of doubly charged ion *m/z* 696.80

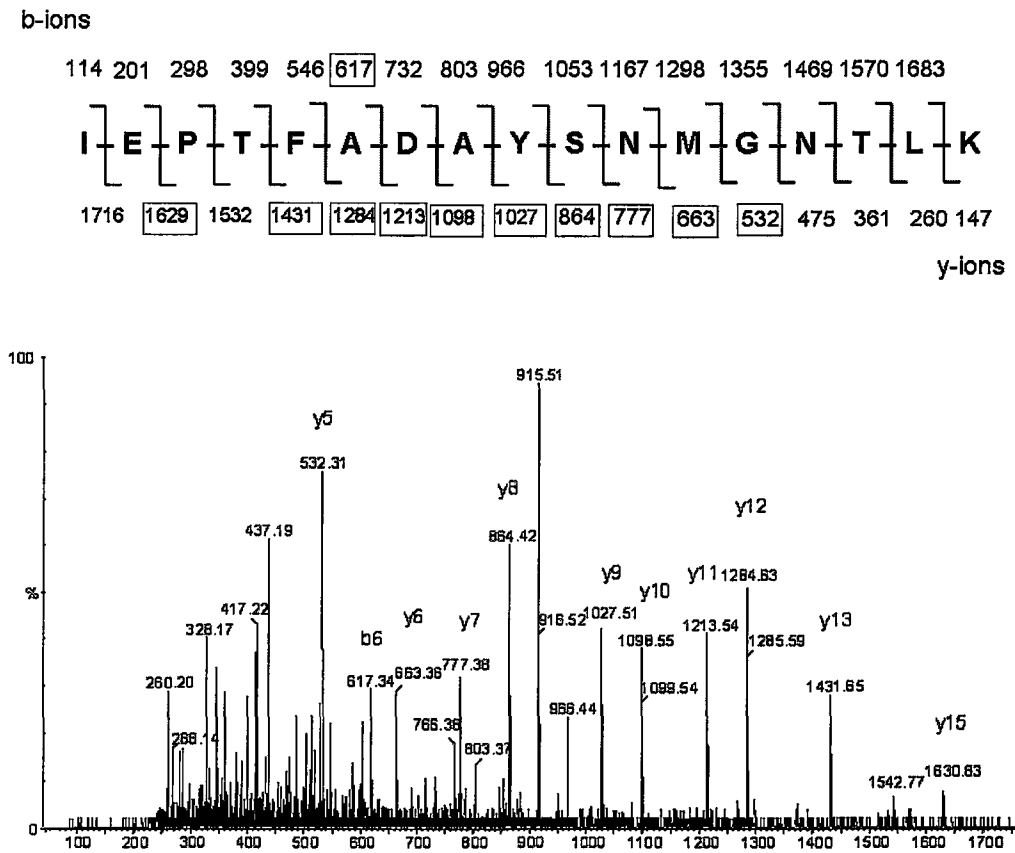


Figure 3.14 MS/MS of the doubly charged ion *m/z* 915.44

```
1    MASMTGGQQM GRDPQQQGKL QEALMHYKEA IRISPTFADA YSNMGNTLKE MQDVQGALQC
61   YTRAIQINPA FADAHSNLAS IHKDSGNIPE AIASYRTALK LKPDFPDAYC NLAHCLQIVC
121  DWTDYDERMK KLVSIVADQL EKNRLPSVHP HHSMLYPLSH GFRKAIAERH GNLCLDKINV
181  LHKPPYEHPK DLKLSDGRLR VGYVSSDFGN HPTSHLMQSI PGMHNPDKFE VFCYALSPDD
241  GTNFRVKVMA EANHFIDLSQ IPCNGKAADR IHQDGIHILV NMNGYTKGAR NELFALRPAP
301  IQAMWLGYPG TSGALFMDYI ITDQETSPAE VAEQYSEKLA YMPHTFFIGD HANMFPHLKK
361  KAVIDFKSNG HIYDNRIVLN GIDLKAFLDS LPDVKIVKMK CPDGGDNADS SNTALNMPVI
421  PMNTIAEAVI EMINRGQIQI TINGFSISNG LATTQINNKA ATGEEVPRTI IVTTRSQYGL
481  PEDAIVYCNF NQLYKIDPST LQMWANILKR VPNSVLWLLR FPAVGEPNIQ QYAQNMGLPQ
541  NRIIFSPVAP KEEHVRRGQL ADVCLDTPLC NGHTTGMDVL WAGTPMVTMP GETLASRVAA
601  SQLTCLGCLE LIAKNRQEYE DIAVKLGTDL EYLKKVRGKV WKQRISSPLF NTKQYTMELE
661  RLYLQMWEHY AAGNKPDHMI KPVEVTESAL EHHHHHHHH
```

Figure 3.15 Coverage map of sOGT protein tryptic digestion. 15 proteolytic fragments result are found with a S/N>3, the sequence coverage is about 27.9% (195 out of 699)

Table 3-3 Observed Fragments in sOGT Tryptic Digest and Comparison with the Database Output (SWISS-PROT).

Peptide Sequence	Theoretical Mass	Experimental Mass	Position	Charge State
IIFSPVAPK	486.2999	486.29	543-551	2
IVLNGIDLK	492.808	492.8	377-385	2
ISSPLFNTK	503.782	503.78	645-653	2
AADRIHQDGIHILVNMNGYTKGAR	530.8784	530.53	267-290	5
AFLDSLPDVK	552.8004	552.8	386-395	2
VPNSVLWLLR	598.8611	598.86	511-520	2
LVSIVADQLEK	607.8532	607.85	132-142	2
NRQEYEDIAVK	682.8439	682.84	615-625	2
DSGNIPEAIASYR	696.8413	696.84	84-96	2
IHQDGIHILVNMNGYTKGAR	746.3918	746.35	271-290	3
EAIRISPTFADAYSNMGNTLK	772.3827	772.38	29-49	3
EEHVRR	825.4326	825.07	552-557	2
HGNLCLDKINVLHKPPYEHPK	841.7812	841.82	170-190	3
EMQDVQGALQCYTRAIQINPAF---ADAHSNLASIHKDSGNIPEAIASYR	867.7589	867.76	50-96	6
ISPTFADAYSNMGNTLK	915.4406	915.44	33-49	2

3.7 Purification and characterization of normal peptide substrate

To determine if the sOGT was active or not, a peptide substrate was used for an enzyme activity assay. The peptide YSDSPSTST is reported as one of the best OGT substrates so it was synthesized by solid phase peptides synthesis. The crude YSDSPSTST peptide samples contained unknown impurities which may influence the OGT enzyme activity. HPLC purification was used to remove the impurities.

Purification of 70 mg of crude peptide, a semi-preparative column with identical packing material was used with the same linear gradient but different flow rate (analytical at 1.4 mL/min, and 3 mL/min for semi-prep).

As described in 2.16, ACN and H_2O were used for separation. 0.1% formic acid was added to each solvent. This acid is used to improve the chromatographic peak shape and to provide a source of protons for reverse phase LC/MS analysis. Fractions including the peak of interest were collected and detected by ESI-Mass spectrometry. As shown in Figure 3.16, the major peak eluted around 15.43 minutes contained the expected *m/z* 944. The peak area was integrated and the yield of the SPPS was estimated about 45 %.

The fractions eluted around 15.5 minutes were collected into a 15 mL falcon tube, frozen using dry ice for 15 minutes and then lyophilized in a speed

vac for 8 hours. The purity of the peptide was confirmed by re-injection into to same column as shown in Figure 3.17. The homogeneity measured at 280 nm was estimated about 94%, sufficient for my ezymological purposes.

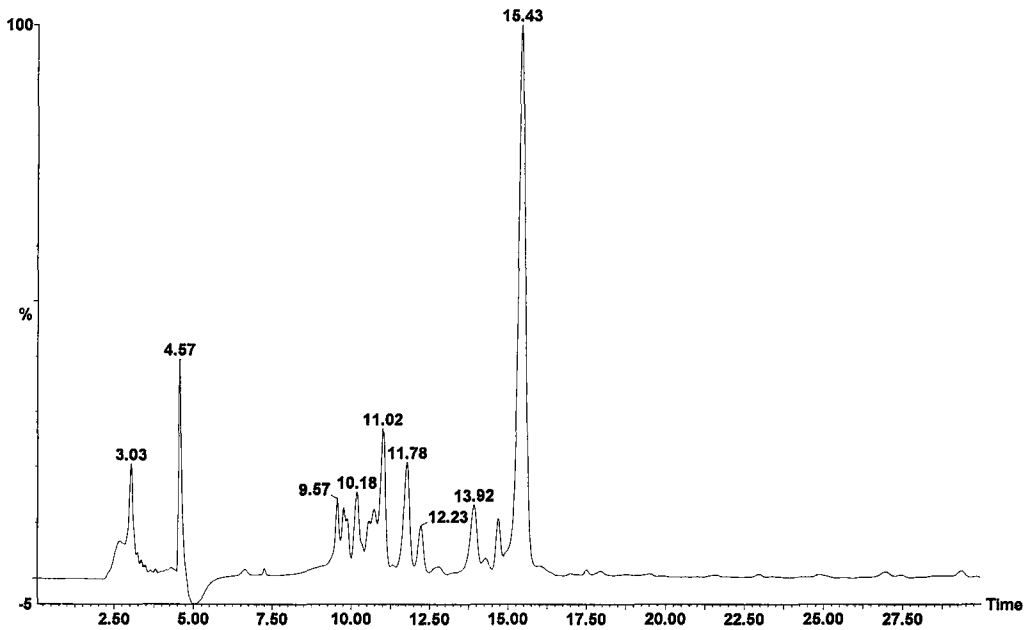


Figure 3.16 HPLC UV chromatogram of peptide substrate before purification

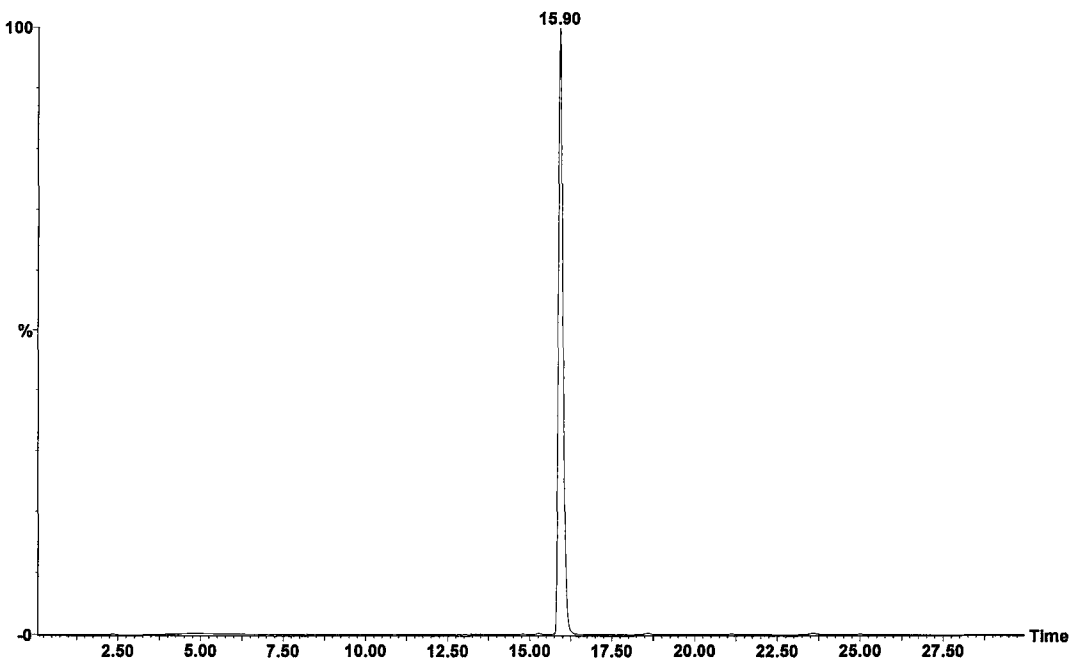


Figure 3.17 HPLC UV chromatogram of peptide substrate after purification

The purified peptide was also tested by ESI-mass spectrometry as shown in figure Figure 3.18. Before purification, there were a lot of unassigned peaks, presumably from impurities, in the spectrum (data not shown). After the purification, the mass spectrum is quite clean. The peak at *m/z* 944.41 represents $[M+H]^+$ and *m/z* 491.61 $[M+H+K]^{2+}$ of YSDSPSTST peptide.

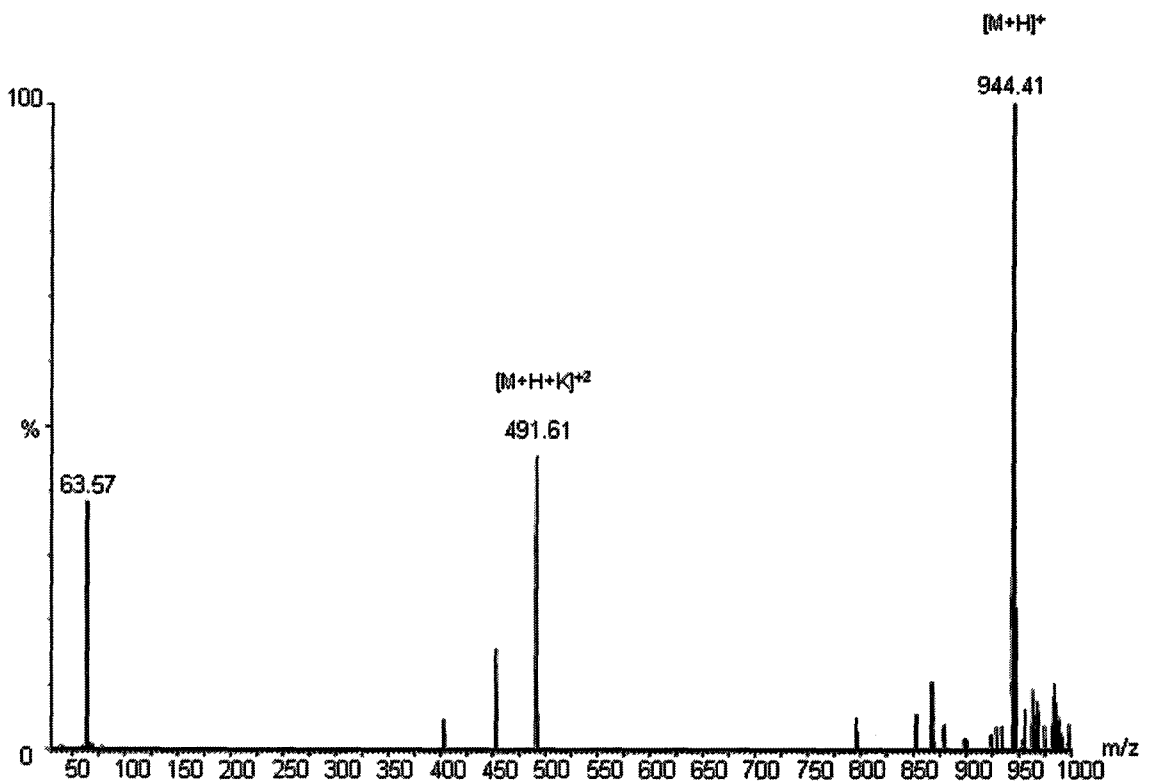


Figure 3.18 ESI-MS of YSDSPSTST peptide after purification.

The purified peptide was then subjected to ESI-MS/MS to generate the sequence information. The singly protonated forms (*m/z* 944.41) was used for CID-ES-MS/MS using argon as collision gas, see Figure 3.20. The spectrum contains a higher intensity fragment ion *m/z* 492.23. This indicated that the major cleavage of C-N bond in the main chain occured between serine and proline, and is a proline characteristic cleavage [85]. The imino group is included in a five-atom ring and has a higher proton affinity than the other amide bonds in the peptide. The protonation and cleavage on the N-terminal side of the proline is favoured to be generated. A series of y and b ions were found except for y8 ions, but the amino acid composition could be obtained by comparison of the mass difference between b2 and b1 ion. The mass difference of the two ions equals 87.03 which correspond to the serine residue at the second position from the N terminal.

Figure 3.19 shows the predicted fragmentation pattern of YSDSPSTST peptide.

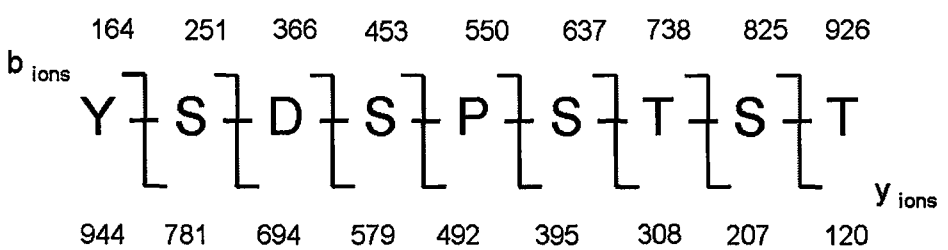


Figure 3.19 Predicted fragmentation pattern of YSDSPSTST peptide by ESI-MS/MS.

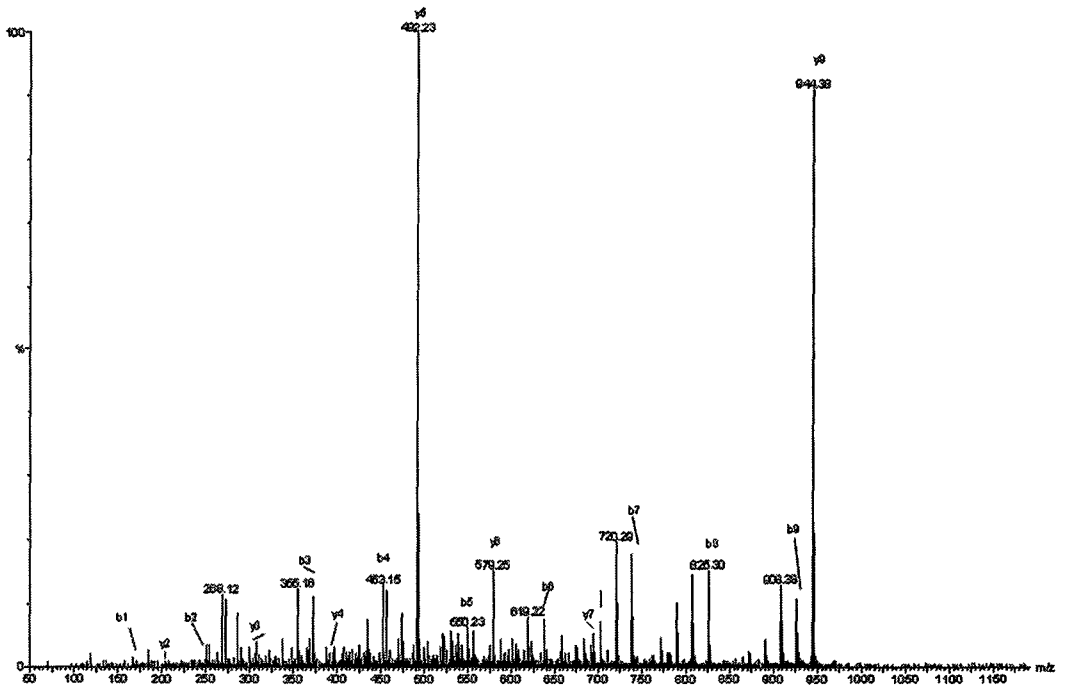


Figure 3.20 MS-MS singly charged ion *m/z* 944.41.

3.8 Purification and characterization of the glycopeptide.

As a standard for use in the in vitro glycosylation study, a glycopeptide was synthesized. This glycopeptide YSDSPgSTST was synthesized by solid phase peptides synthesis as described in experimental section 2.17. The letter "g" represents the N-acetylglucosamine molecule attached to the following amino acid, either serine or threonine. The Fmoc protected building block of modified serine (Fmoc-GlcNAc(OAc3)-Ser-COOH) was synthesized by a graduate student in our group, Mr. Hongchao Zheng. The glycopeptide was purified as described in section 2.16 by collecting HPLC fractions which contained the target ion (*m/z* 1147.41). The purity of the sample was determined by HPLC re-injection followed by ESI–MS analysis. As shown in figure, the purity of the glycopeptide after purification is about 90%. Again, this is sufficient for its use as a standard for glycosylation.

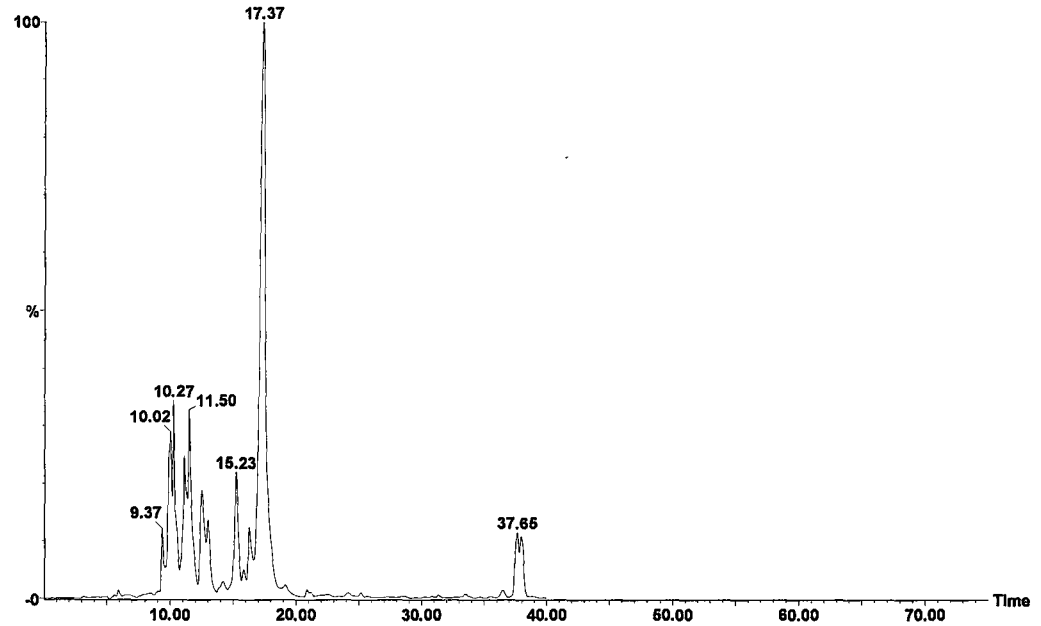


Figure 3.21 HPLC UV chromatogram of glycopeptide before purification.

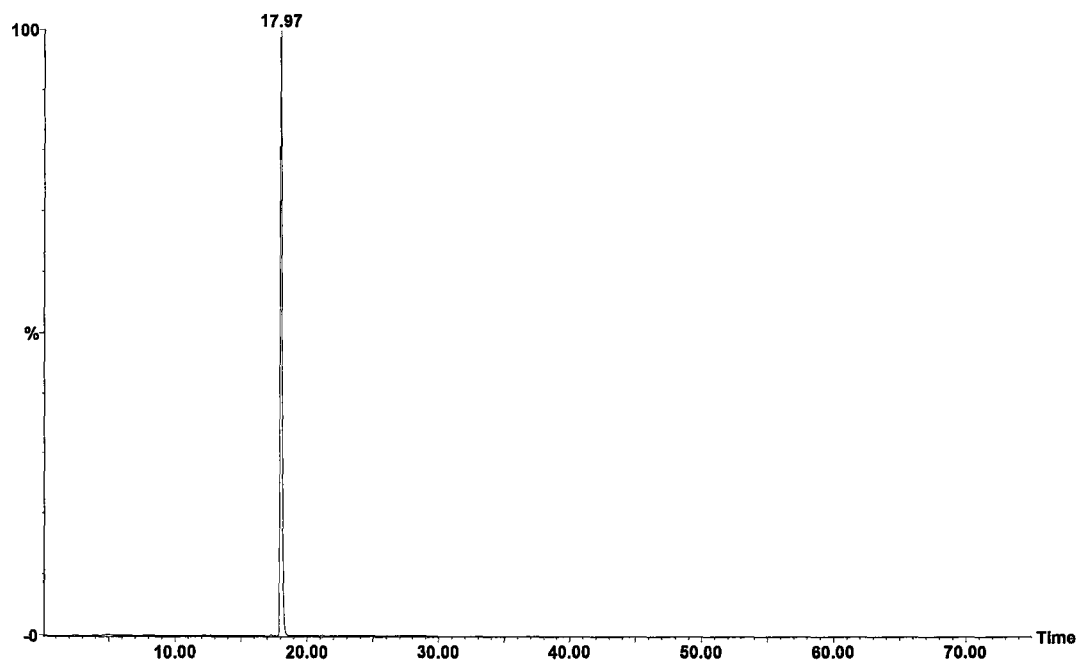


Figure 3.22 HPLC UV chromatogram of glycopeptide after purification

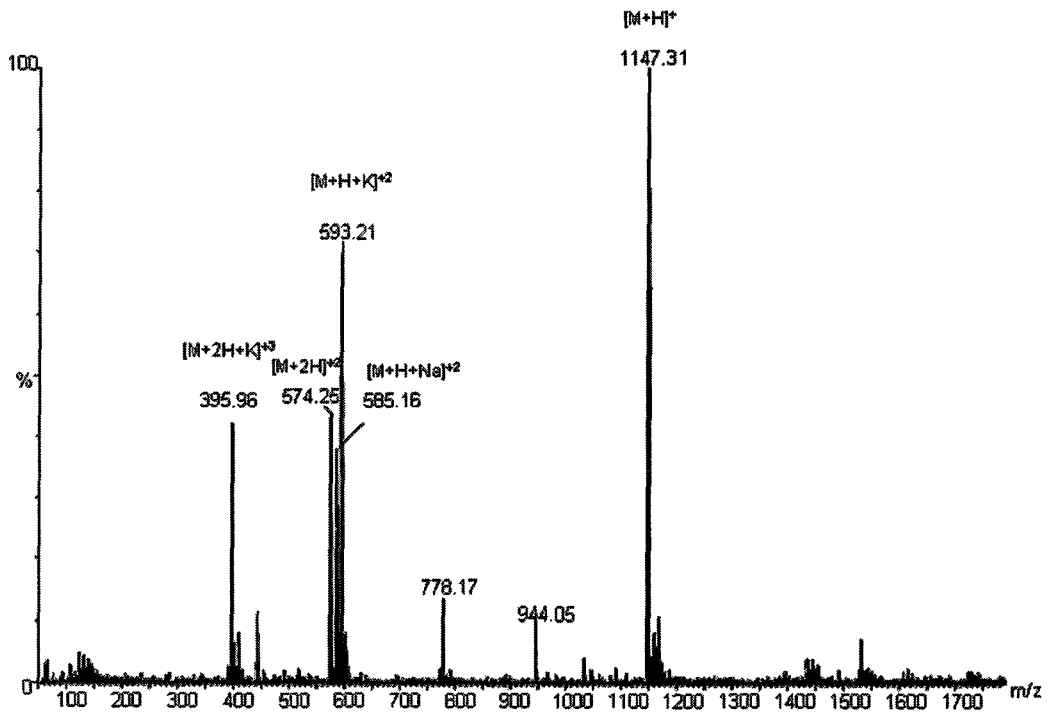


Figure 3.23 ESI-MS of glycopeptide after purification

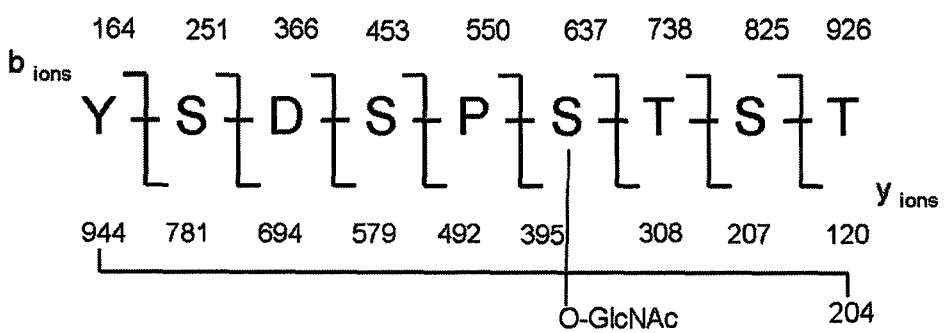


Figure 3.24 Predicted fragmentation pattern of glycopeptide by ESI-MS/MS.

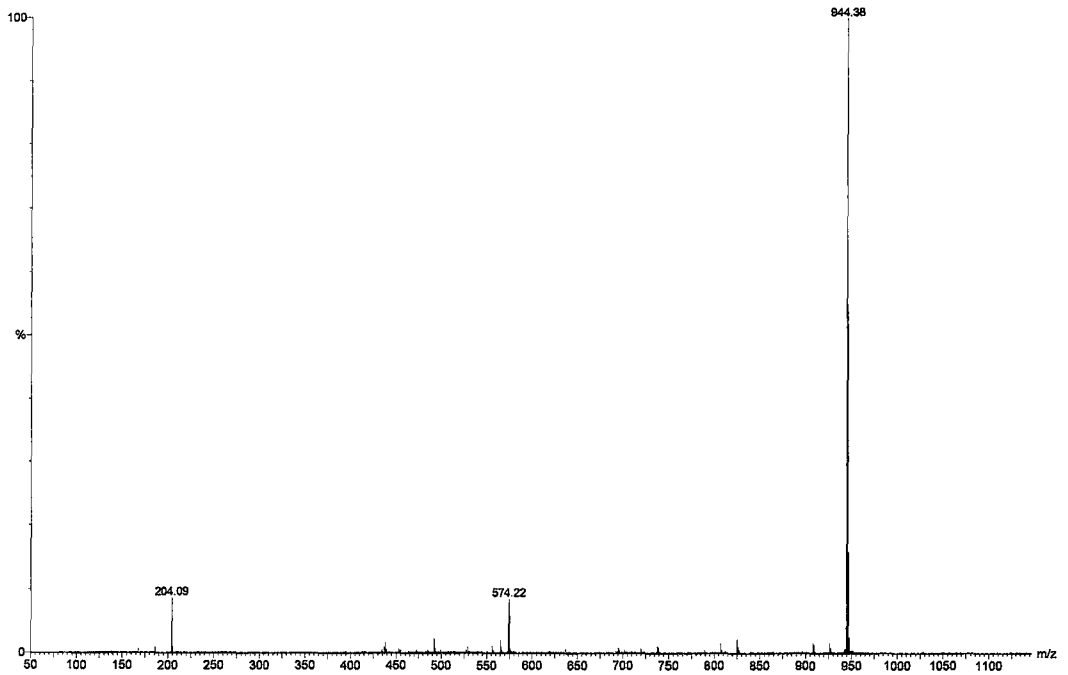


Figure 3.25 MS/MS of the doubly charged ion *m/z* 574.25 at 5eV collision energy.

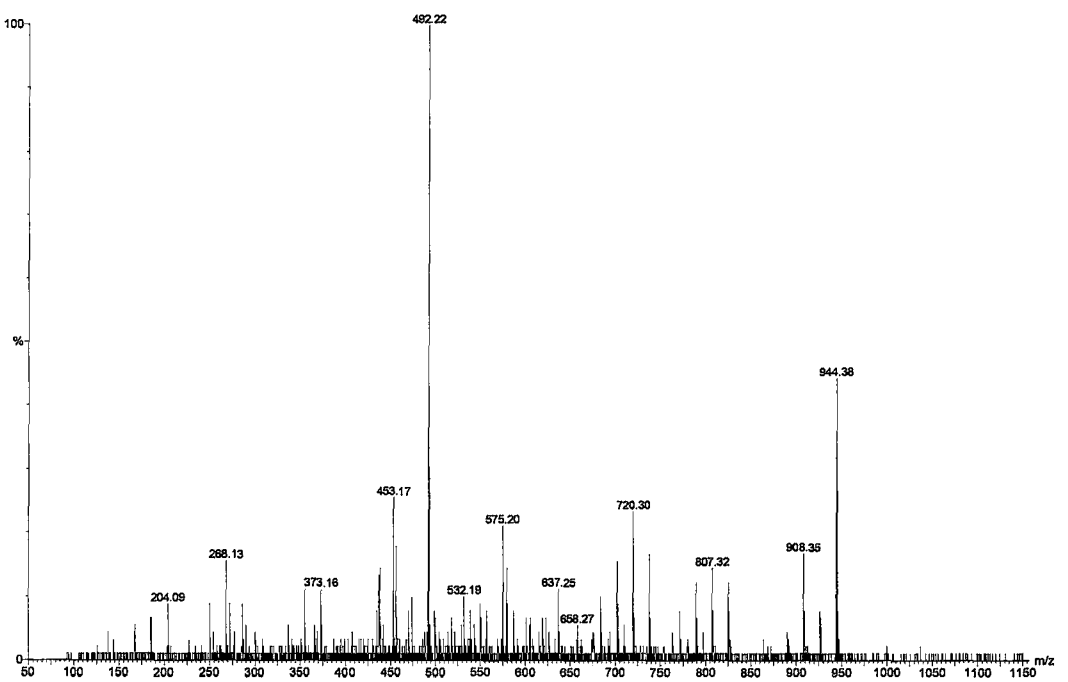


Figure 3.26 MS/MS of the doubly charged ion *m/z* 574.25 at 15eV collision energy.

Figure 3.23 shows the ESI-MS spectrum after purification. The major peaks are the singly charged $[M+H]^+$. Interestingly, the glycopeptide formed the doubly charged spaces, $[M+2H]^{2+}$, $[M+H+Na]^{2+}$, $[M+H+K]^{2+}$, the first two doubly charged ions were not found in the non-glycopeptide spectrum Figure 3.18. Triply charged ion $[M+2H+K]^{3+}$ was also detected.

Figure 3.25 shows the MS/MS spectrum generated from the doubly charged ion (*m/z* 574.25) of the standard glycopeptide at a collision energy of 5 eV. The observed mass shift from 1147.31 to 944.38 corresponds to the loss of a single GlcNAc molecule. The ion *m/z* 204.09 represented the protonated GlcNAc molecule. More fragmentation of the glycopeptide were observed at higher collision energy, see Figure 3.26

When increasing the collision energy, more fragmentations of the glycopeptide occurred.

Figure 3.24 is the theoretical calculation of the b- and y-ions mass fragments of standard glycopeptide.

3.9 Assay development of sOGT enzyme activity

The sOGT enzyme was purified and the enzyme activity assay was carried out. The measurement of the enzyme activity was necessary because the enzyme would be used for further glycosylation study. The details were described in section 2.19. Figure 3.27 shows the scheme of the enzyme activity assay.

The most challenging step in this in vitro glycosylation assay is to separate the tritium labeled glycopeptide from [H^3]-UDP-GlcNAc. As described in section 2.19, three different methods (SP column separation. TCA precipitation and Sep-Pak C18 column separation) were carried out to improve the separation.

The Sep-Pak C18 column gave the best separation. The pre-packed column performed better separation compared to the self packed SP column, and provide more constant result compared to the TCA precipitation.

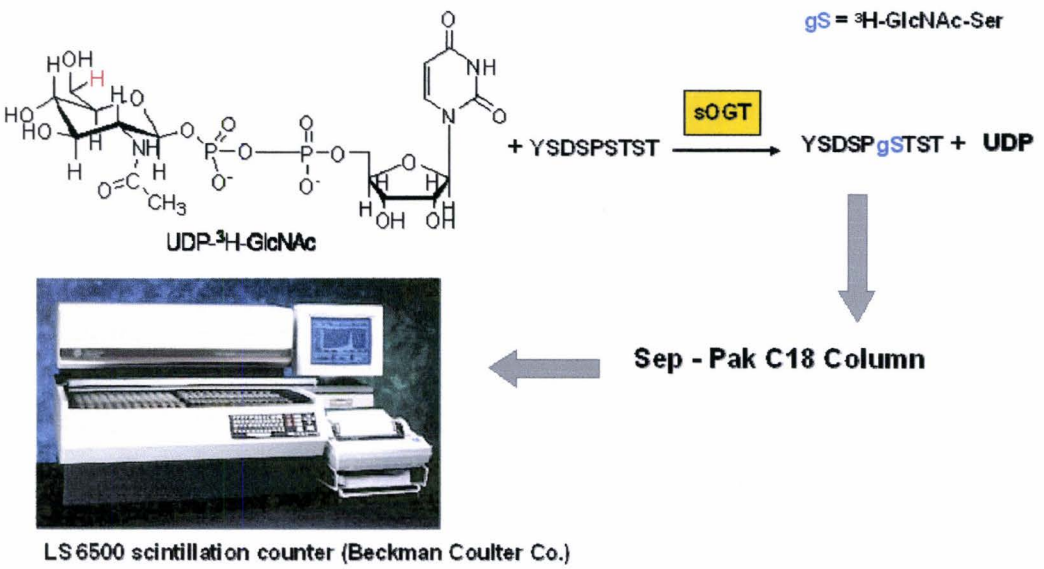


Figure 3.27 sOGT activity measurement.

3.9.1 Standard curve of [H^3]-UDP-GlcNAc

To convert the cpm from the scintillation counter to concentration units, the counting efficiency of the scintillation counter was calculated. From a plot of cpm (count per minutes) values obtained from the machine against dpm value of [^{3}H]-UDP-GlcNAc at different concentrations, the counting efficiency (cpm/dpm) was found to be 0.47 as shown in Figure 3.28.

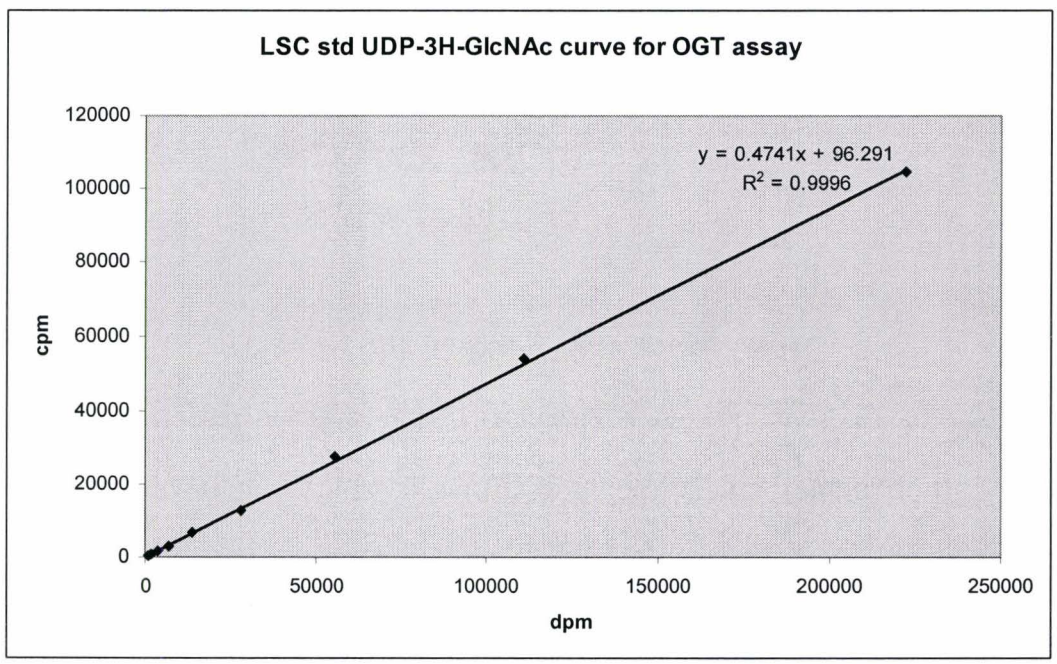


Figure 3.28 Counting efficiency of the liquid scintillation counter.

3.9.2 Elution of reaction mixture from Sep-Pak column

The elution profile of Sep-Pak column was performed and the fractions from elution are collected and counted using the liquid scintillation counter.

Figure 3.29 shows the elution profile. Tubes 1-13 were obtained during column wash and the elution buffer was loaded from tube 14. Tubes 15-17 showed increased cpm in Figure 3.29 a, whereas no significant increase in counts was found from corresponding tubes in negative controls Figure 3.29 b and c. These data indicated the glycosylated product, which would be the peptide YSDSPgSTST, eluted mostly in tubes 15 and 16 after the unretained UDP-[^{3}H]-GlcNAc was removed by the wash step mainly into Tubes 1-4. The amount of [^{3}H]-GlcNAc modified peptide was calculated using the sum of counts Tubes 15 and Tube 16.

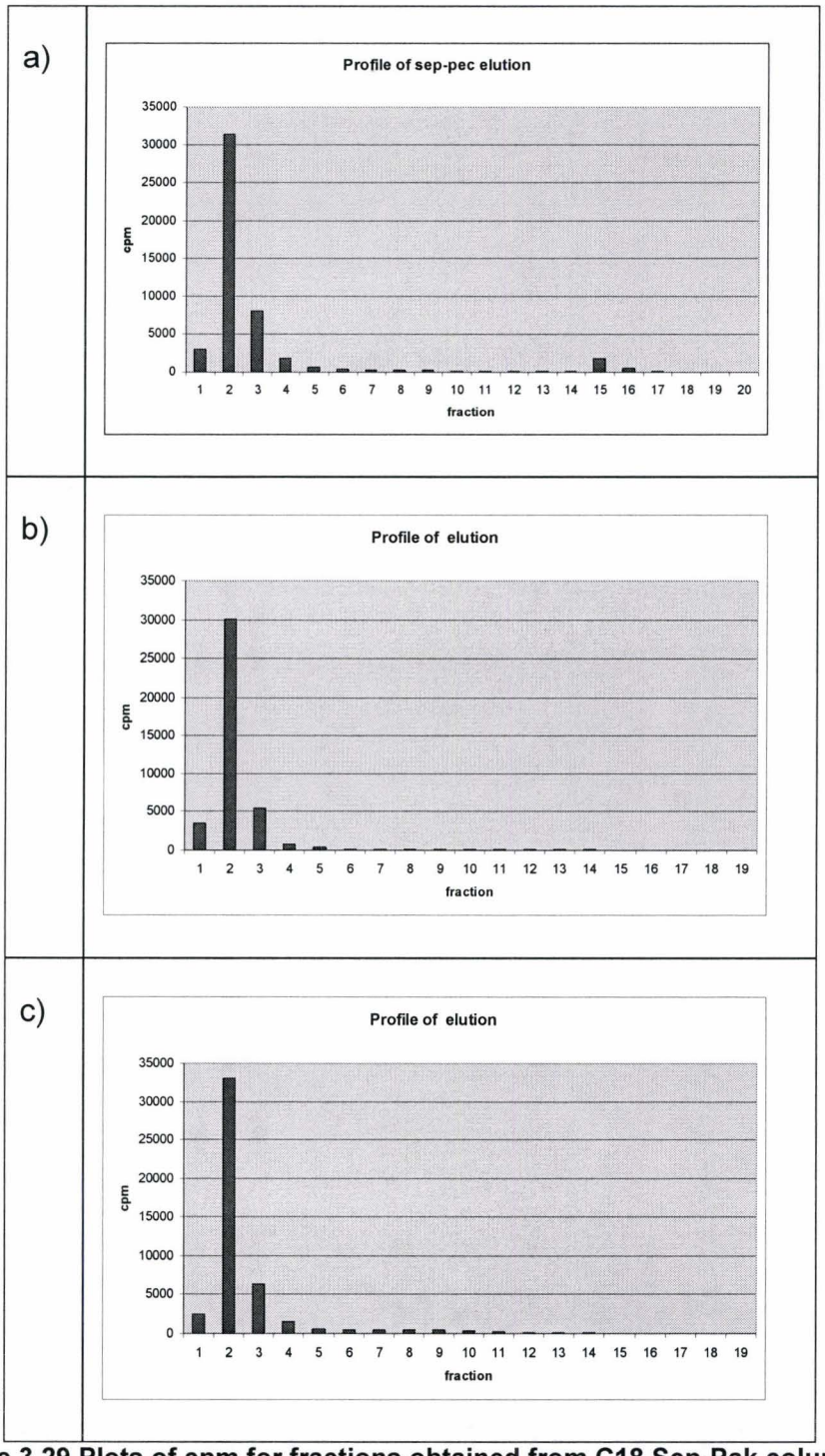


Figure 3.29 Plots of cpm for fractions obtained from C18 Sep-Pak columns. In vitro glycosylation reactions of 0.1 µCi of UDP-[^{3}H]-GlcNAc incubated in 50 µl reaction buffer for 30 min at 37 °C with: a) 20ug of peptide substrate and 1 µg of sOGT protein, b) 20ug of peptide substrate but without sOGT enzyme. c) 1 µg of sOGT protein.

3.9.3 Temperature effect on enzyme activity

The sensitivity of the radioactive assay is higher than that of mass spectrometry. Optimization of the enzyme activity was performed so that more glycosylation product could be generated to further facilitate the development of MS based glycosylation analysis. To test the temperature effect of the enzyme activity, the assay was performed at three different temperatures. Figure 3.30 shows a sOGT assay measured at 24 °C (A), 30 °C (B), 37 °C (C). The highest enzyme activity was achieved after incubation at 24 °C for three hours. To test the stability of the sOGT enzyme, pre-incubation of sOGT enzyme before the in vitro assay was done. The result in Figure 3.30C shows that the enzyme lost its activity rapidly at 37 °C after incubation for 40 minutes. The reason for the inactivation is not obvious, which is similar to the reported full length OGT. But the purification process of sOGT protein should be, and was therefore, done at low temperature to maintain the enzyme activity. Based on these results, we choose to perform the sOGT assay at 24 °C and use 2 hours incubation as our assay condition.

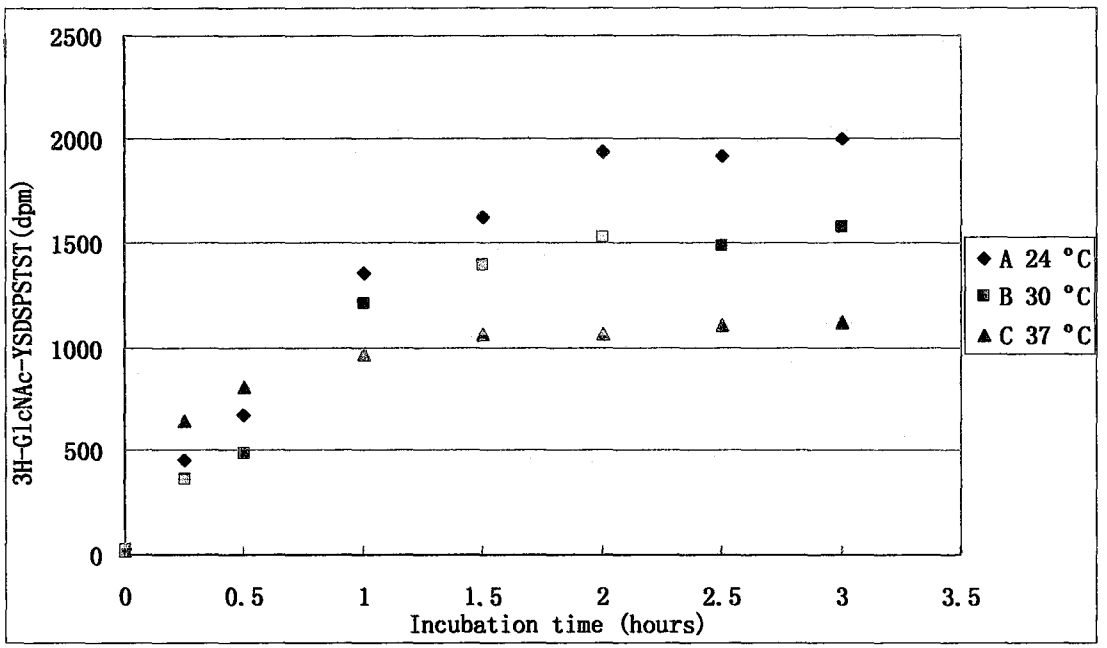


Figure 3.30 sOGT Activity assay at different temperatures. The sOGT assay was done at 24 °C (A), 30 °C (B), 37 °C (C).

3.9.4 Salt effect on enzyme activity

The in vitro glycosylation assay using sOGT enzyme and peptide substrate could be extended to protein substrates. One of the factors that challenge this is the buffers needed for most proteins, which typically have salt for stabilization and high salt concentration such as NaCl or KCl are known to inhibit full length OGT enzyme activity. Of course, a desalting step could be used before assaying to solve this problem. But many proteins will not tolerate desalting and this additional step is both time and material requiring. Therefore, we tested the effect of salt on the sOGT enzyme using NaCl and KCl. The results are shown in figure 3.31. The IC_{50} of NaCl and KCl are about the same, 45 mM, and at 20 mM of salt, the enzyme still maintains about 90% activity, which means the in vitro assay could be extended to protein substrate.

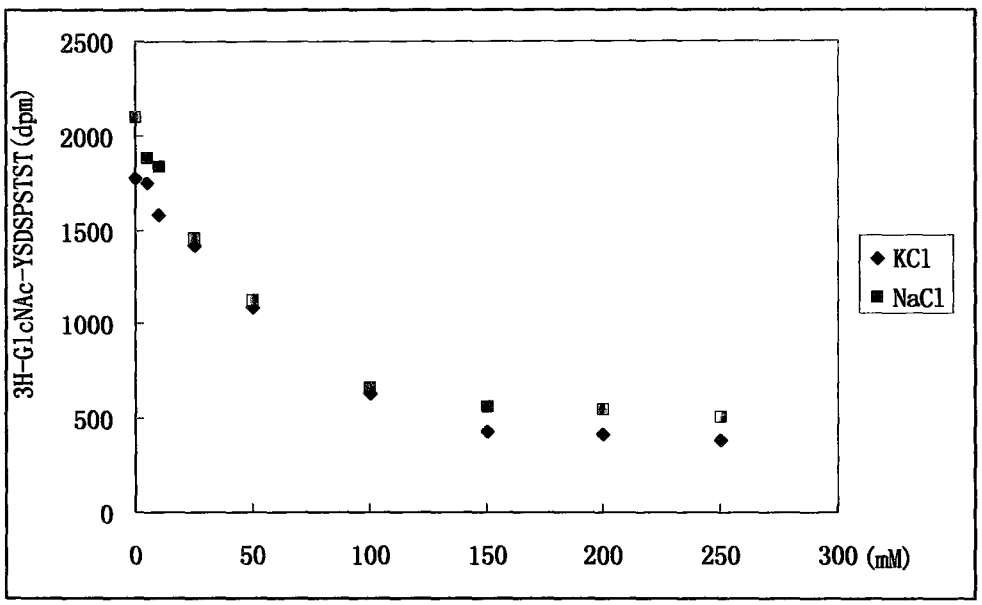


Figure 3.31 Salt effect of NaCl and KCl on sOGT activity.

3.10 MS method development of glycopeptide detection

3.10.1 Sensitivity determination of ESI and MALDI.

A mass spectrometry-based assay would be useful because it removes the need to use a radioactive substrate, and might also mapping of the modification sites. But the sensitivity of mass spectrometry is lower compared to the radioactivity assay. The developed in vitro glycosylation assay could yield about 5% of glycosylation product according to the obtained data from the radioactivity-based assay. Therefore I determined the sensitivity of the ESI and MALDI MS using synthesis peptide mixture.

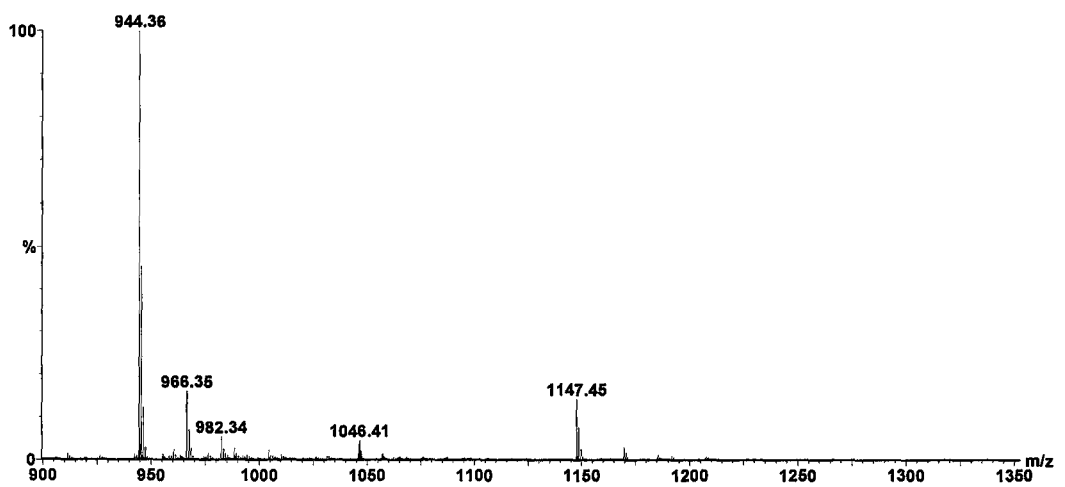


Figure 3.32 ESI of glycopeptide and normal peptide mixture. The ratio of normal peptide : glycopeptide is 1 : 1 , and the glycopeptide concentration is 1pmol/ul.

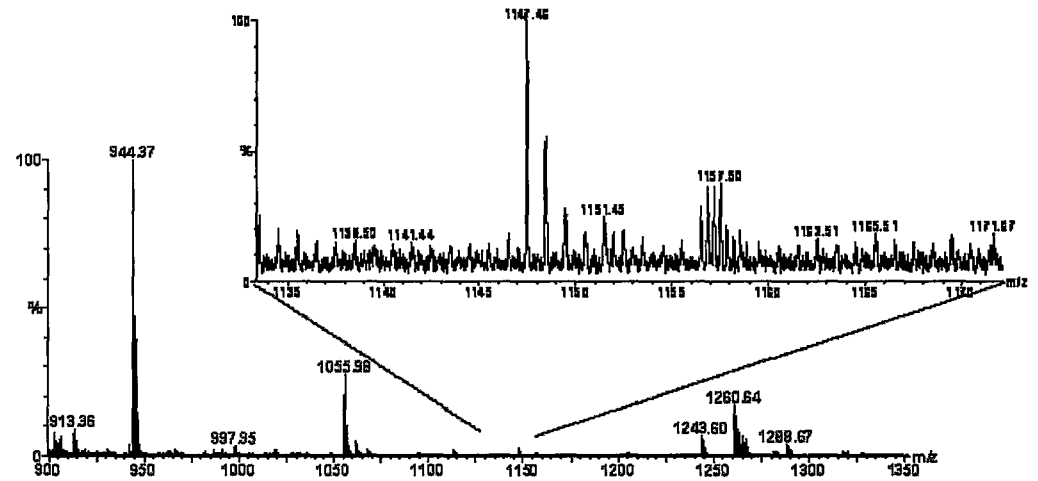


Figure 3.33 ESI of glycopeptide and normal peptide mixture. The ratio of normal peptide:glycopeptide is 10:1, and the glycopeptide concentration is 1pmol/ul.

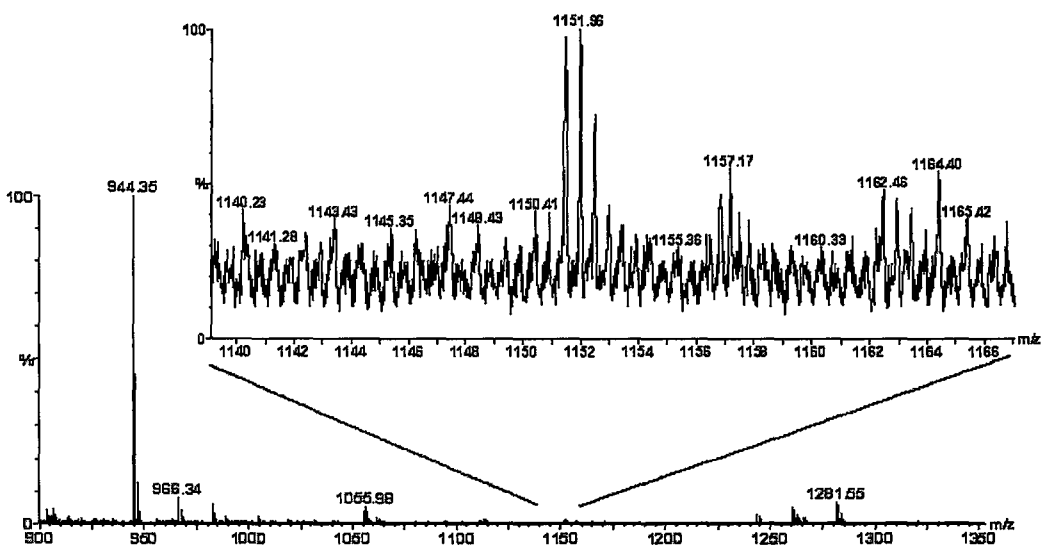


Figure 3.34 ESI of glycopeptide and normal peptide mixture. The ratio of normal peptide:glycopeptide is 100:1, and the glycopeptide concentration is 1pmol/ul.

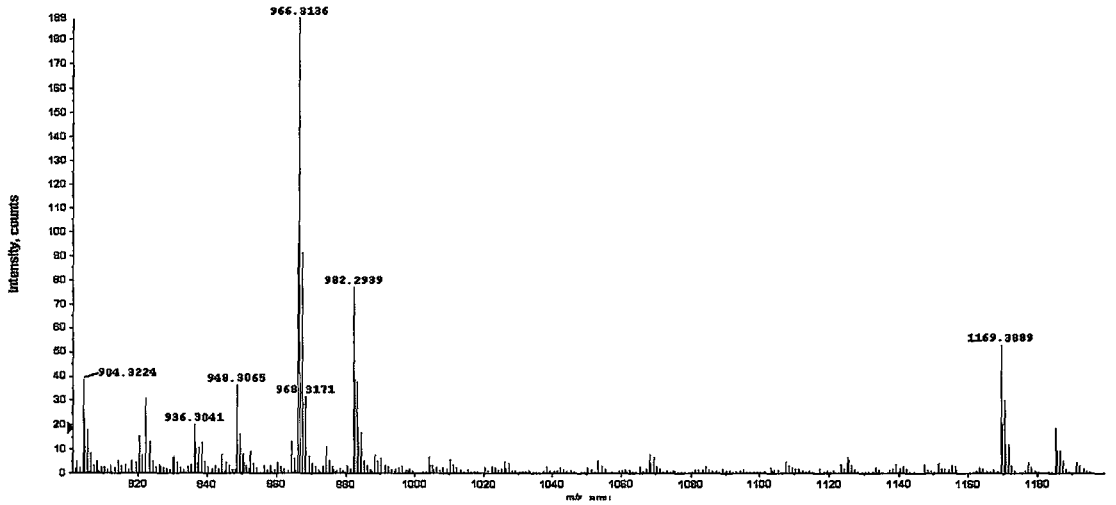


Figure 3.35 MALDI-MS of glycopeptide and normal peptide mixture. The ratio of normal peptide:glycopeptide is 1:1, and the glycopeptide concentration is 500 fmol/spot.

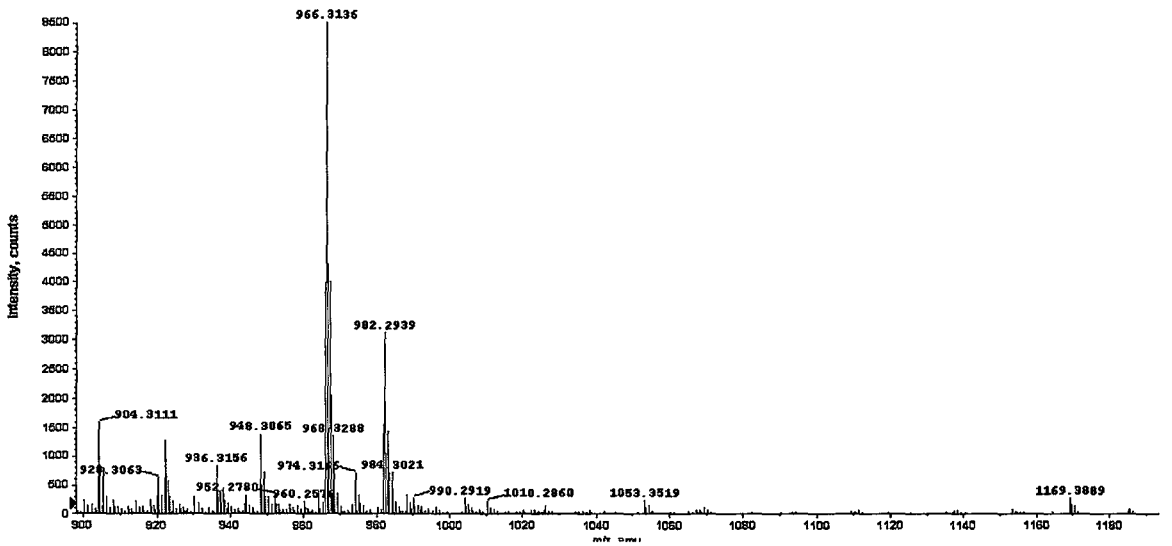


Figure 3.36 MALDI-MS of glycopeptide and normal peptide mixture. The ratio of normal peptide:glycopeptide is 10:1, and the glycopeptide concentration is 500 fmol/spot.

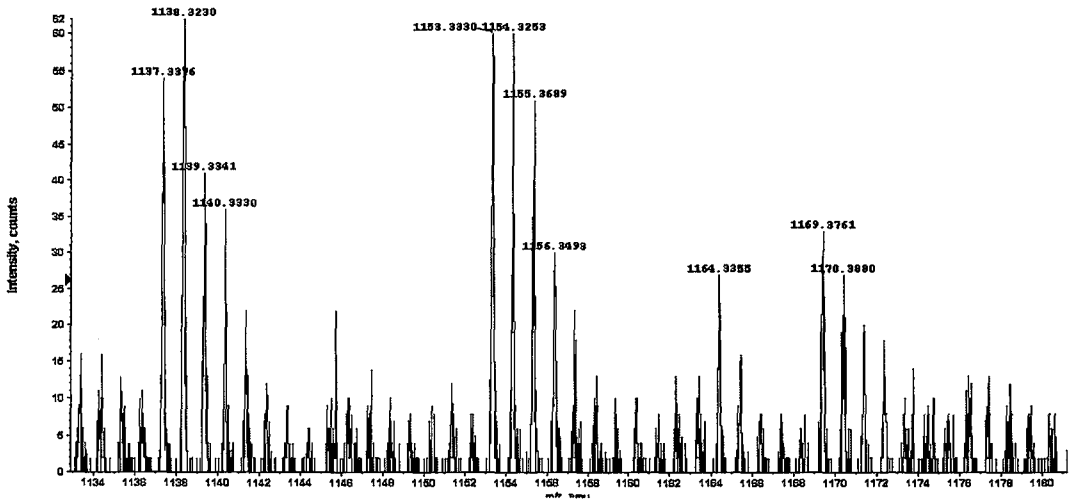


Figure 3.38 MALDI-MS of glycopeptide and normal peptide mixture. The ratio of normal peptide:glycopeptide is 100:1, and the glycopeptide concentration is 500 fmol/spot.

3.10.2 Optimization of MALDI MS TOF

When the glycopeptide is analysed using MALDI MS, we observed that at high focusing potential, dissociation of the GlcNAc residue from the glycopeptides takes place. The detection of the release of sugar residues could be quite helpful when the primary sequence of the glycopeptides is known. The information provided by the mass shift from glycopeptide to deglycosylated peptide could aid the characterization of the modification pattern. But in most cases, the sequence of the peptide from a protein digestion is unknown. In addition, most of the proteins are only partially modified. If the modified protein releases its modification during the analysis by mass spectrometry, it will be hard to identify the modification pattern and also the modification sites. Thus optimized the MALDI MS instrument parameters are essential for glycopeptide analysis. Two different parameters were found to affect the mass spectrum most; laser power and focusing potential. As shown below, at focusing potential 110eV most of the glycopeptides maintain its sodium ion form, and give the strongest signal for detection.

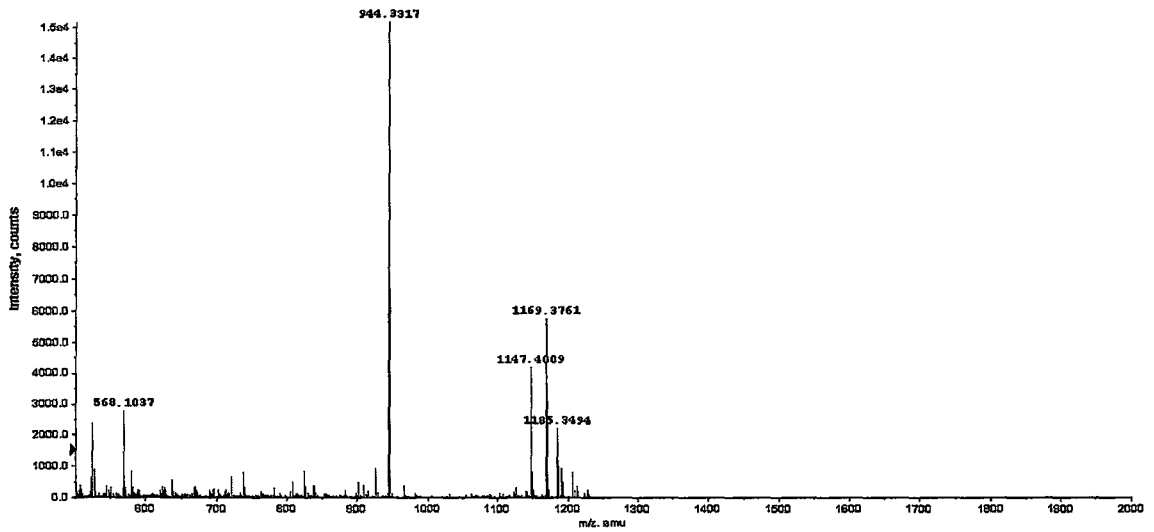


Figure 3.37 MALDI-MS of glycopeptide at focusing potential 60 eV. The concentration is of 2pmol/spot.

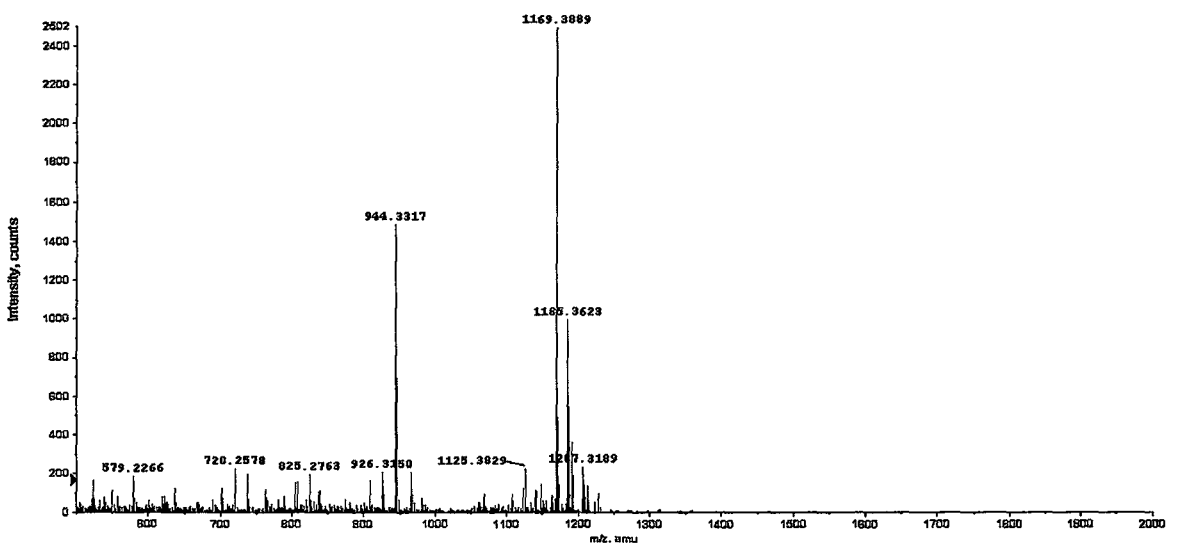


Figure 3.38 MALDI-MS of glycopeptide at focusing potential 90 eV. The concentration is of 2pmol/spot.

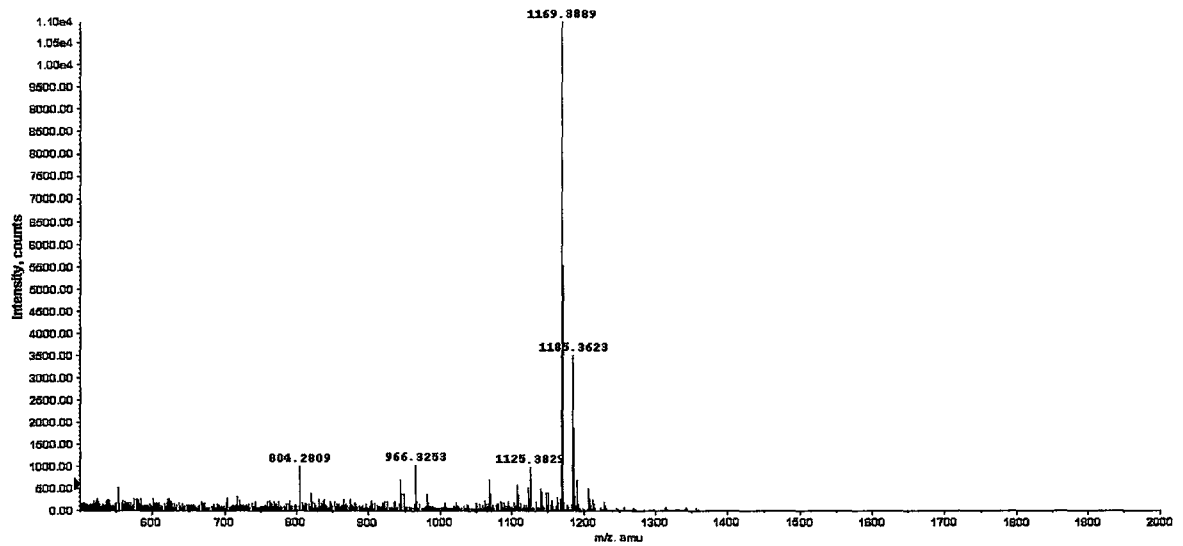


Figure 3.39 MALDI-MS of glycopeptide at focusing potential 110 eV. The concentration is of 2pmol/spot.

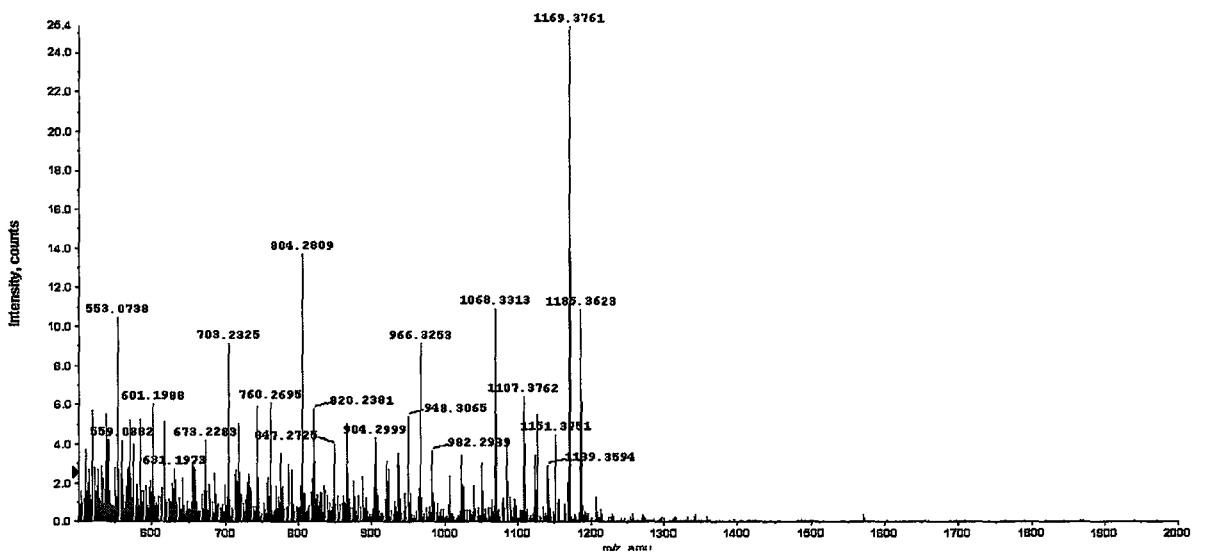


Figure 3.40 MALDI-MS of glycopeptide at focusing potential 140 eV. The concentration is of 2pmol/spot.

3.10.3 Detection of in vitro glycosylation by Cap-LC MALDI MS.

3.10.3.1 Separation of a glycopeptide and normal peptide mixture.

A separation technique combined with mass spectrometry provides a quite useful tool to study the protein modification. We used synthesized glycopeptide (YSDSPSgTST) and normal peptide (YSDSPSTST) to optimize the separation conditions. Because of the comparable retention times of the used peptides, a series of conditions were tried to optimize the separation. The used gradients are shown in Figure 3.41.

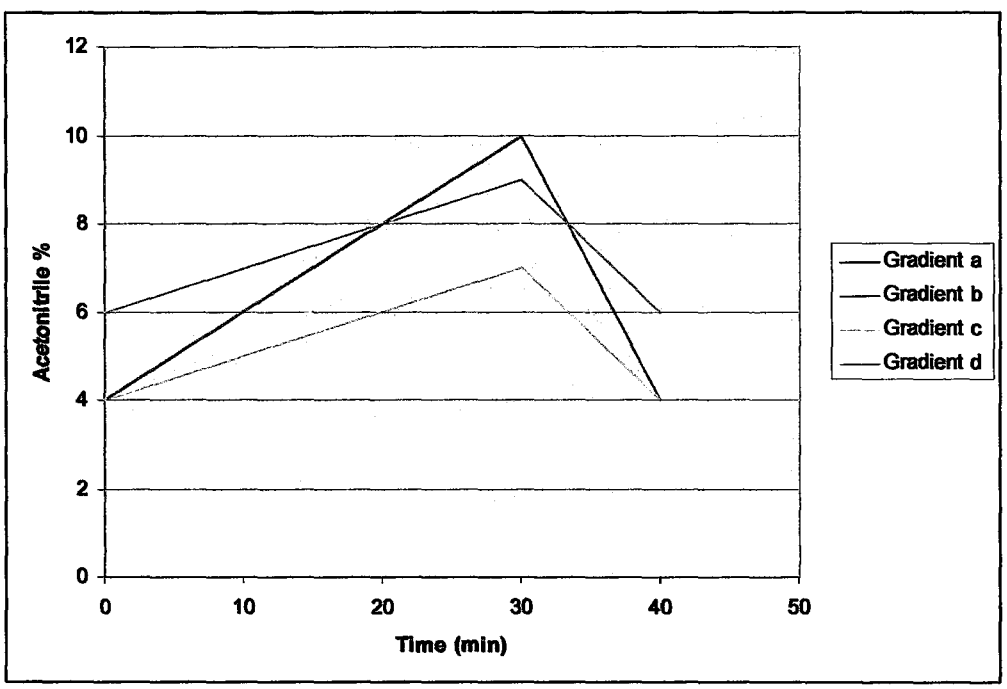


Figure 3.41 Optimizing of separation condition of glycol and normal peptide mixture using Cap-LC.

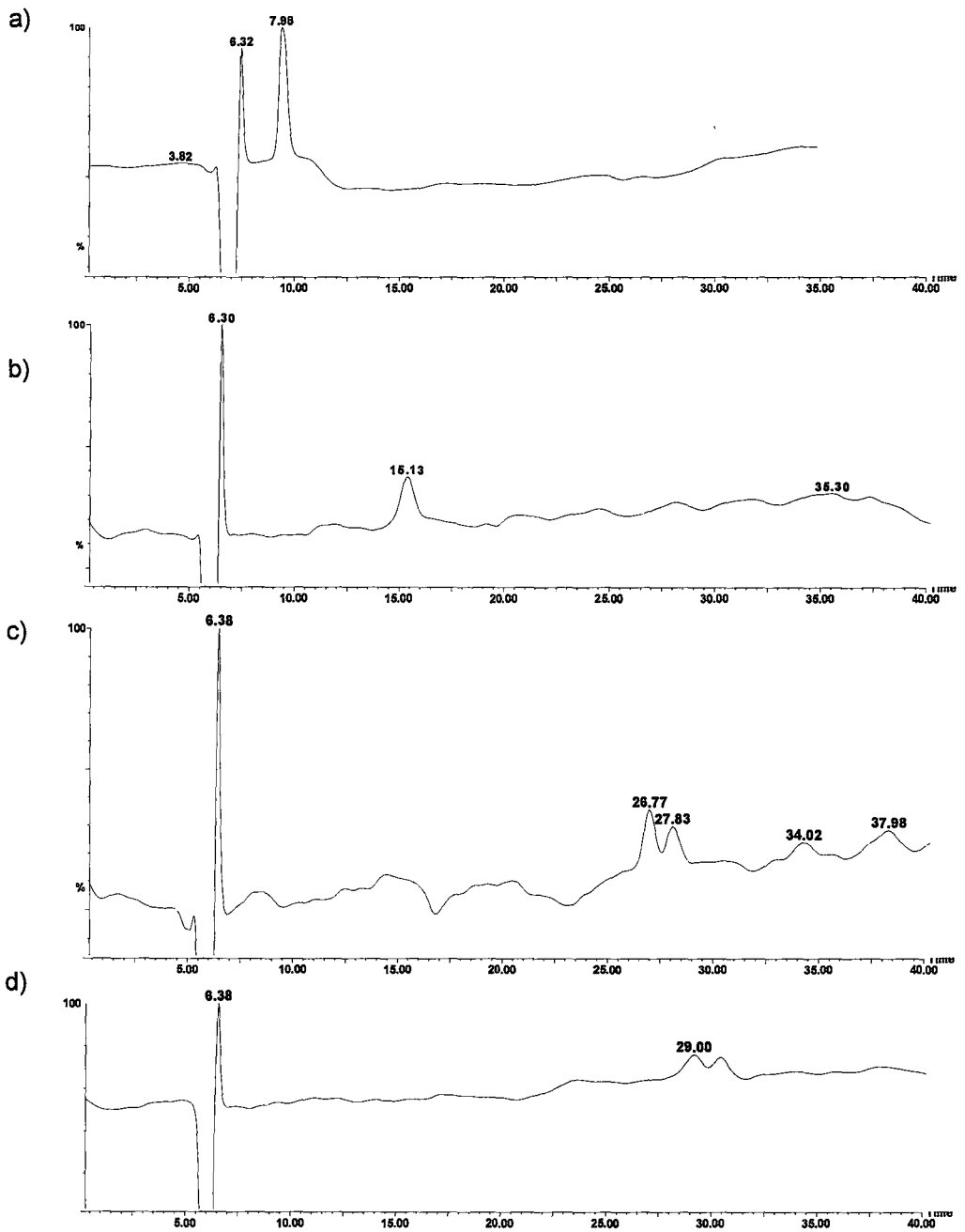


Figure 3.42 UV-Cap-LC chromatogram of different gradient separation of the glyco-normal peptide mixture (1:1).

As shown in Figure 3.42. The glycol and normal peptide mixture eluted as one single peak at 8 minutes as shown in figure a). A later elution of the peptide mixture was achieved by increasing the initial acetonitrile percentage and a decrease of acetonitrile percentage at 30 minutes. The peptide mixture eluted at 15 minutes but still migrate as a single peak as shown in figure b). Best separation of the peptide mixture was achieved in Figure 3.42 c with the gradient shown in Figure 3.41 gradient c. Decrease of the gradient slop results in greater peak broadening effect. This elution condition was used in further study.

3.10.3.2 Detection of glycopeptide generated from in vitro glycosylation.

We then performed the in vitro glycosylation study using peptide substrate and unlabeled UDP-GlcNAc. The in vitro glycosylation reaction condition was described in the experimental procedure section 2.19 with little modification. 10 μg of peptide substrate was added to obtain enough modified peptide after in vitro glycosylation. After incubation for 2 hours at 24 °C, the reaction was stopped by adding 500 μL of 50 mM formic acid. The solution was then filtered though a membrane with a 10 kD cut off before the capillary LC and spotted on MALDI plate. The MALDI MS parameter was set as optimized condition. Figure 3.43 presents the MALDI MS/MS spectrum of the $(M+H)^+$ ion of the glycopeptide at 1147.3 Da. During the CID, most of the glycopeptide was deglycosylated. The

fragmentation ion came from the deglycosylated peptide was bracketed as shown in figure.

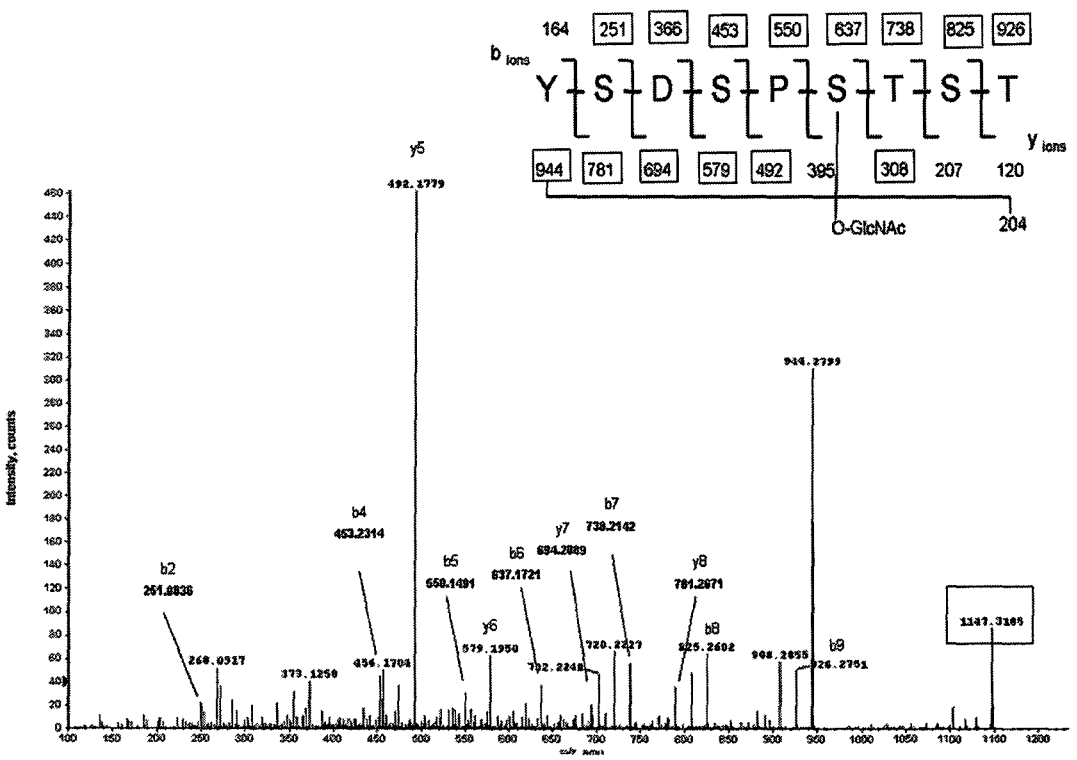


Figure 3.43 CID-MALDI MS/MS spectrum of in vitro glycosylated peptide. The glycopeptide is singly charged, *m/z* 1147.3. The predicted b- and y-ion mass fragments are shown on top. The *m/z* of the unfragmented precursor ion in the CID spectrum is highlighted with a box.

3.11 Identification of glycosylation sites using alkaline β-elimination.

Identification of the glycosylation site was done by β-elimination as described in the section 2.20. CID-MALDI-MS/MS of singly charged ion was shown in Figure 3.44. Using alkaline β-elimination converts the GlcNAc modified serine to 2-aminopropenoic acid (2-ap). The y-ion series and b-series was calculated and shown below, the *m/z* ratio which corresponded Experimental results are shown in a box. The position of the 2-aminopropenoic position was confirmed in the b-ion series which contain ions at *m/z* 550 and 619. The mass shift of 69 amu indicated the 2-aminopropenoic position.

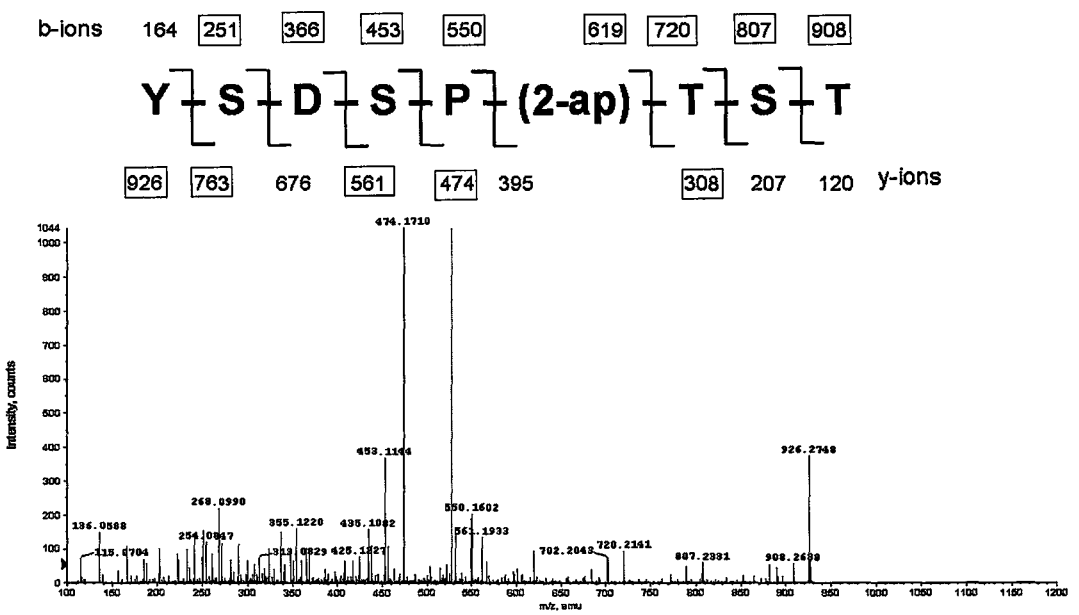


Figure 3.44 Identification of the glycosylation site by MALDI-CID-MS/MS. The predicted b- and y-ion series are shown above the CID mass spectrum

Base hydrolysis β-elimination has been used to release sugar conjugate from glycoprotein for a long time [86]. It is important to point out, the β here referred to the β-carbon atoms in the Ser/Thr unit, completely different from the β position of C1 carbon on the sugar residue. Using this method, we could identify the glycosylation site on pmol level of modified peptides.

4 Conclusion and future directions

Site mapping of O-GlcNAc modification is one of the most important tasks in post translational modification study. In this study, I have purified and characterized two enzymes (NCOAT and sOGT) which catalyse the removal and addition of a single N-acetylglucosamine residue to the serine or threonine residue of a substrate.

In order to develop a new mass spectrometry based assay for O-GlcNAc modification detection. I used synthetic glycopeptides as a standard to optimize the separation and detections of O-Glycosylated peptides. I also determined the limitation of detection of ESI and MALDI mass spectrometry by mixing the two standard peptides (non-glycosylated and glycosylated) at different ratio.

After optimization of the Capillary LC separation conditions for the glycopeptide and unmodified peptide, I successfully detected the glycopeptide using MALDI-TOF after in vitro glycosylation by purified sOGT enzyme. The site of in vitro glycosylation was mapped using alkaline β-elimination.

The application of Mass spectrometry in protein study especially in the study of post-translational modification has cause a revolution in molecular biology field. The main focus of this dissertation is to develop a new mass spectrometry based method to detect the in-vitro O-GlcNAc modification. The current substrate used in the in vitro glycosylation is a known peptide substrate

which has its own limitation. Future studies could extend the usefulness of the developed method by including detection of O-GlcNAc modification of protein substrate by in vitro glycosylation.

Further more, the synthesized glycopeptide has intriguing feature. The multiple charged glycopeptide spaces indicate its unique geometry in the gas phase of mass spectrometry. The formation of the triply charged K^+ may indicate the non-covalent interaction between the K^+ with O-GlcNAc residue in the gas phase.

Another Graduate student in our lab Mr.Zheng has characterized the synthetic sOGT enzyme inhibitor property using the current developed in vitro glycosylation method coupling with Cap-LC-UV detection. Quantitive analysis of the O-GlcNAc product from in vitro glycosylation assay could carry on for further glycosylation study. This provides an alternate tool to determine enzyme activity.

5 References

1. Yang, X. J. (2004) *Oncogene.* **24**, 1653-1662.

2. Jensen, O. N. (2006) *Nat. Rev. Mol. Cell Biol.* **7**, 391-403.

3. Kornfield, R., Kornfield, S. (1985) *Annu. Rev.Biochem.* **54**, 631-664.

4. Moens, S., Vanderleyden, J. (1997) *J. Arch. Microbiol.* **168**, 169–175.

5. Hofsteenge, J., Muller, D. R., de Beer, T., Loffler, A. (1994) *Biochemistry.* **33,** 13524–13530.

6. Lennarz, W. J. (1987) *Biochemistry.* **26**, 7205-7210.

7. Hart, G. W., Holt, G. D., Haltiwanger, R. S. (1988) *Trends Biochem. Science.* **13**, 380-384.

8. Torres, C. R., Hart, G. W. (1984*) J.Biol.Chem.* **259**, 3308-3317.

9. Holt, G. D., Hart,G. W., (1986*) J.Biol.Chem.* **261**, 8049-8057.

10. Wells L, Vosseller K, Hart G.W. (2001) *Science.* **291**, 2376 – 2378.

11. Chou, C. F., Omary, M. B., (1993*) J.Biol.Chem.* **268**, 4465-4472.

12. Holt, G. D., Snow, C. M., Senior, A., Haltiwanger R. S., Hart, G. W., (1987) *J.Cell.Biol.* **104**, 1157-1164.

13. Kearse, K.P., Hart, G. W. (1991*) Proc. Natl. Acad. Sci. U. S. A.* **88**, 1701-1705.

14. Lefebvre, T., Baert, F., Bodart J. F., Flament, S, Michalki, J. C., Vilain J. P. (2004*) J.Cell.Biochem.* **93**, 999-1010.

15. Haltiwanger, R. S., Holt, G. D., Hart, G . W. (1990) *J.Biol.Chem.* **265**, 2563-

2568.

16. Whelan, S.A., Hart, G.W. (2003) *Circulation Research.* **93**(17), 1047-1058.

17. Haltiwanger, R. S., Blomberg, M. A., Hart, G . W. (1992*) J.Biol.Chem.* **267**, 9005-9013.

18. Kreppel, L. K., Blomberg, M. A., Hart, G . W. (1997*) J.Biol.Chem.* **272**, 9308-9315.

19. Lubas, W. A., Frank, D. W., Krause, M., Hanover, J. A. (1997) *J.Biol.Chem.* **272**, 9316-9324.

20. Wrabl, J. O., Grishin, K. V. (2001) *J. Mol. Biol.* **314**, 365–374.

21. Kreppel, L. K., Hart, G . W. (1999*) J.Biol.Chem.* **274**, 32015-32022.

22. Zhang, F., Su, K., Yang, X., Bowe, D. B., Paterson, A. J., Kudlow, J. E. (2003) *Cell.* **115**, 715-725.

23. Nolte, D., Muller, U. (2002) *Mamm. Genome.* **13**, 62-64.

24. Hanover, J. A., Yu, S., Lubas, W. B., Shin, S. H., Ragano-Caracciola, M., Kochran, J., Love, D. C. (2003) *Arch. Biochem. Biophys.* **409**, 287-297.

25. Gross, B. J., Kraybill, B. C., Walker, S. (2005) *J. Am. Chem. Soc.* **127**, 14588-14589.

26. Lubas, W. B., Hanover, J. A. (2000) *J.Biol.Chem.* **272**, 10983-10988.

27. Jinek, M., Rehwinkel, J., Lazarus, B. D., Izaurralde, E., Hanover, J. A., Conti, E. *Nat Struct Mol Biol.* (2004), **11**, 1001-1007.

28. Haltiwanger, R. S., Blomberg, M. A., Hart, G. W. (1991) *Glycoconjugate.J.* **8**, 210-212.

29. Okuyama, R., Marshall, S., (2003) *J. Neurochemistry.* **86**, 1271-1280.

30. Konrad, R. J., Zhang, F., Hale, J. E., Knierman, M. D., Becker, G. W., Kudlow, J.E. (2002) *Biochem. Biophys. Res. Commun.* **293**, 207-212.

31. Gao, Y., Wells, L., Comer, F. I., Parker, G. J., Hart, G. W. (2001) *J.Biol.Chem.* **276**, 9838-9845.

32. Wells, L., Gao, Y., Mahoney, J. A., Vosseller, K., Chen, C., Rosen, A., Hart, G.W. (2002) *J.Biol.Chem.* **277**, 1755-1761.

33. Bertram, L., Blacker, D., Mullin, K., Keeney, D., Jones, J., Basu, S., Yhu, S., McInnis, M. G., Go, R. C., Vekrellis, K., Selkoe, D. J., Saunders, A. J., Tanzi, R. E. (2000) *Science.* **290**, 2302–2303.

34. Comtesse, N., Maldener, E., Meese E., (2001) *Biochem. Biophys. Res. Commun.* **283**, 634-640.

35. Schultz, J., Pils, B. *FEBS Lett.* (2002), **529**, 179-182.

36. Dong, D. L., Hart, G. W. (1994) *J. Biol. Chem.* **269**, 19321-19330.

37. Whisenhunt, T. R., Yang, X., Bowe, D. B., Paterson, A. J., Van Tine, B. A., Kudlow, J. E. (2006) *Glycobiology.* **16**, 551-563.

38. Macauley, M. S., Vocadlo, D. J. (2005) *J. Biol. Chem.* **280**(27), 25313–25322.

39. Dennis, R. J., Taylor, E. J., Macauley, M. S., Stubbs, K.A., Turkenburg, J. P., Hart, S. J., Black, G. N., Vocadlo, D. J., Davies, G. J., (2006) *Nat. Struct. Mol.*

Biol. **13**, 365-371.

40. Rao, F. V., Dorfmueller, H. C., Villa, F., Allwood, M., Eggleston, I. M., van Aalten, D.M. (2006) *EMBO J.* **25**,1569-1578.
41. Cetinbas, N., Macauley, M. S., Stubbs, K. A., Drapala, R., Vocadlo, D.J. (2006) *Biochemistry.* **45**, 3835-3844.
42. Love, D. C., Hart, G. W. (2005) *Science STKE.* 312, 1-15.
43. Hart, G.W. (1997) *Annu. Rev. Biochem.* **66**, 315-335.
44. Kelly, W. G., Dahmus, M. E., Hart, G.W. (1993) *J. Biol. Chem.* **268**, 10416–10424.
45. Chou, T. Y., Hart, G. W., Dang, C.V. (1995) *J. Biol. Chem.* **270**, 18961–18965.
46. Comer, F. I., Hart, G.W., (2001) *Biochemistry.* **40**, 7845-7852.
47. Hershko, A., Ciechanover, A., Heller, H., Haas, A. L., Rose, I. A. (1980) *Proc. Natl. Acad. Sci. U S A.* **77**, 1783-1786.
48. Wells, R., Rechsteiner, M. (1986) *Science.* **234**, 364-368.
49. Rechsteiner, M., Rogers, S. W. (1996) *Trends. Biochem. Sci.* **21**, 267-271.
50. Cheng, X., Hart, G. W. (2001) *J. Biol. Chem.* **276**, 10570-10575.
51. Zhang, F., Su, K., Yang, X., Bowe, D. B., Paterson, A. J., Kudlow, J. E. (2003) *Cell.* **115**, 715-725.
52. Holt, G. D., Snow, C. M., Senior, A., Haltiwanger, R. S., Hart, G. W. (1987) *J.Cell.Biol.* **104**, 1157-1164.
53. Starr, C. M., Hanover, J. A. (1990) *J. Biol. Chem.* **265**, 6868-6873.

54. Han, I., Kudlow, J. E. (1997) *Mol. Cell. Biol.* **17**, 2550-2558.

55. Chou, T. Y., Hart, G. W., Dang, C. V. (1995) *J. Biol. Chem.* **270**, 18961-18965.

56. Lefebvre, T., Ferreira, S., Dupont-Wallois, L., Bussiere, T., Dupire, M.J., Delacourte, A., Michalski J.C., Caillet-Boudin, M.L. (2003) *Biochim. Biophys. Acta* **1619**, 167–176.

57. Zhu, W., Leber, B., Andrews, D. W. (2001) *EMBO J.* **20,** 5999-6007.

58. Marshall, S., Okuyama. R., (2004) *Biochem. Biophys. Res. Commun.* **318,** 911-915.

59. Vosseller, K., Trinidad, J.C., Chalkley, R.J., Specht, C.G. (2006) *Mol. Cell. Proteomics.* 923-934.

60. Hanover, J. A., Cohen, C. K., Willingham, M. C., Park, M. K, (1987) *J. Biol. Chem.* **262,** 9887-9894.

61. Sprung, R., Nandi, A., Chen, Y., Kim, S. C., Barma, D., Falck, J. R., Zhao, Y. J. (2005) *Proteome Res.* **4**, 950-957.

62. Haynes, P. A., Aebersold, R. (2000) *Anal. Chem.* **72**, 5402–5410.

63. Wells, L. (2002) *Mol. Cell. Proteomics.* **1**, 791–804.

64. Karas, M., Hillenkamp, F. (1988) *Anal. Chem.* **60**, 2299–2301.

65. Whitehouse, C.M., Dreyer, R.N., Yamashita, M., Fenn, J.B. (1985) *Anal. Chem.* **57**, 675-679.

66. Fenn, J.B., Mann, M., Meng, C.K., Whitehouse, C.M. (1989) *Science.* 246.

67. Fenn, J.B., Yamashita, M. (1984) *Phys. Chem.* **88**, 4451-4459.

68. Robert. J. Chalkley., A. L. Burlingame. (2003) *Mol. Cell. Proteomics*, **2**(3), 182–190.

69. Robert. J. Chalkley., A. L. Burlingame. (2001) *J. Am. Soc. Mass. Spectrom.* **12**(10), 1106–1113.

70. Wilson J., Vachet R.W. (2004) *Anal. Chem.* **15**, 7346–7353.

71. O'Connor P.B., Costello C.E., Vachet R.W. (2000) *Anal. Chem.* **130**, 363–370.

72. Tanaka, K., Akita, S. (1987) *Second Japan-China Joint Symposium on Mass Spectrometry.* 185-188.

73. Zenobi, R., Knochenmuss, M. (1999) *Mass. Spectrom. Rev.* **17**, 337-366.

74. Roepstorff P. (1997) *Curr.Opin.Biotechnol.* **8**, 6-13.

75. Yates J.R. (1998) *J.Mass Spectrom.* **33**, 1-19.

76. Yates J.R. (1998) *Electrophoresis.* **19**, 893-900.

77. Perkins D. N., Pappin D. J., Creasy D. M., Cottrell J. S. (1999) *Electrophoresis.* **20**, 3551-3567.

78. Aggarwal K., Lee K. H. (2005) *Proteomics.* **5**, 2297–2308.

79. Cantin G. T., Yates J. R. (2004) *J. Chromatogr. A.* **1053**, 7-14.

80. Wells L., Vosseller K., Hart G. W. (2002) *Mol. Cell. Proteomics*, **1**, 791–804.

81. Haynes P.A., Aebersold R. (1998) *Electrophoresis.* **19**, 939-945.

82. Breaux G.A., France A., Limbach P.A., (2000) *Anal. Chem.* **72**, 1169–1174.

83. Amado F.M., Tomer K.B. (1997) *Anal. Chem.* **69**, 1102–1106.

84. Hiroyuki Katayama., Yoshiya Oda. (2001) *Rapid. Commun. Mass. Spectrom.*

15, 1416-1421.

85. Breci L.A., Wysocki V.H. (2003) *Anal. Chem.* **75**, 1963–1971.

86. Zinn, A.B., Carlson, D.M. (1977) *The Glycoconjugates*. 69-85.