INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UMI®

Université d'Ottawa • University of Ottawa

CHARACTERIZATION OF PROTEIN GLYCOFORMS IN CELLOBIOHYDROLASES AND ENDOGLUCANASES FROM *TRICHODERMA REESEI* RUT-C30 AND MUTANT STRAINS USING CAPILLARY ISOELECTRIC FOCUSING AND MASS SPECTROMETRY

Joseph P.M. Hui

Thesis submitted to the
School of Graduate Studies and Research
University of Ottawa
in partial fulfillment of the requirements for the
M.Sc. degree in the
Ottawa-Carleton Chemistry Institute

Dedicated to

my Parent, Alan and Maria,

and my Sister, Cecilia,

with Love

Abstract

Trichoderma reesei is a filamentous fungus heavily used in the biotechnology industry due to its efficient secretion of cellulases. The enzymatic system of *T. reesei* consists primarily of four glycoproteins referred to as cellobiohydrolases (CBH I, CBH II) and endoglucanases (EG I, EG II). They exhibit microheterogeneity both in the N- and O-linked glycans. This thesis focuses on the method development to characterize the glycosylation profile and post-translational modifications present in these glycoproteins. Crude cellulase fermentation extracts RUT-C30 and its two mutant strains Iogen-M4, Iogen-B13 were initially analyzed by capillary isoelectric focusing (CIEF) to determine the cellulase composition in them. The major cellulase CBH I was purified from each strain and electrospray mass spectrometry (ESMS) was used to reveal the extent of overall glycosylation. To characterize the N-linked glycans and their attachment sites, CBH I from these strains were subjected to tryptic digest with and without PNGase F incubations followed by mass spectrometric detection. The O-linked glycans were released chemically by hydrazinolysis and were analyzed by high performance anion-exchange chromatography with pulsed amperometric detection (HPAEC-PAD). The majority of O-linked glycans was di- and tri-saccharides. Two unusual posttranslational modification from strain RUT-C30 were observed: (1) both high mannose (predominantly $Man_8GlcNAc_2$) and single GlcNAc in putative N-linked sites and (2) mannosylphosphorylation in a O-linked di-saccharide. Heterogeneity in putative N-linked sites was found consistently in CBH II, EG I and EG II from RUT-C30, however, no mannosylphosphorylation was observed at least on the proteins in the purified fractions. These results have led to the proposal of endogenous

endoglycosidase H as well as mannosylphosphorylation activities possibly induced during fermentation.

Acknowledgements

I would first like to acknowledge my supervisor Dr. Pierre Thibault for his patience, guidance and expertise in mass spectrometry which led to the successful completion of this work. His continued support, persistence and work ethic have inevitably kept me motivated throughout this master program. Thank you for giving me the opportunity to carry out the research in the National Research Council (NRC) where I was able to work with state-of-the-art instrumentation and exchange ideas with remarkable scientists. I am also grateful to Dr. René Roy for setting up this joint graduate research opportunity between University of Ottawa and NRC at an expanding research frontier.

Many thanks to Iogen Corporation for providing the materials and financial support that allows me to undertake this challenging project. In particular, I wish to express my appreciation to Dr. Theresa White for her knowledge of *Trichoderma* genetics and cellulases as well as for valuable discussions. Thanks also to Heather Pikor, Dr. Sylvia McHugh and Genny Giroux for support and understanding. A special thanks to Dr. Makoto Yaguchi for his continuous "supply of research papers on cellulases" and his rich knowledge in this field. Thanks to Dr. Thibault's research team, Dr. John Kelly, Dr. Jianjun Li, Tammy LeRiche, Patricia Lanthier, Don Krajcarski and Ken Chan for their kindness and guidance. I am grateful to Dr. Roger MacKenzie and Tomoko Hirama for their technical assistance in protein purification. The generous student grant from Fonds pour la Formation de Chercheurs et l'Aide á la Recherche (FCAR) is greatly acknowledged.

Throughout the course of this project, the uncountable "evening discussion" of science with Dr. Thibault and late working hours which have practically turned the lab into my

domicile were worthwhile. The hard work and experiences have undoubtedly transformed into this thesis, but more importantly, they have taught me many valuable life lessons.

Table of Contents

Chapter 2 Experimental Procedure

Chapter 3 Capillary Electrophoresis Analysis of Cellulase Extracts from RUT-C30 and Two Mutant Strains from *T. reesei*

Chapter 4 Characterization of CBH I Glycoforms from *T. reesei* RUT-C30 and Two Mutant Strains

Chapter 5 Characterization of Posttranslational Modification in CBH II, EG I and EG II from *T. reesei* strain RUT-C30

List of Figures

List of Tables

List of Abbreviations

ac	Alternating current
API	Atmospheric pressure ionization
ß-G	ß-Glucosidase
BLAST	Basic local alignment search tool
BPC	Base peak chromatogram
BSA	Bovine serum albumin
CBD	Cellulose-binding domain
CBH	Cellobiohydrolase
CID	Collision-induced dissociation
CIEF	Capillary isoelectric focusing
CIP	Calf intestinal alkaline phosphatase
cLC	Capillary liquid chromatography
CM Sepharose	Carboxymethyl weak cation exchanger
CMD	Carbohydrate membrane desalter
CZE	Capillary zone electrophoresis
Da	Dalton
dc	Direct current
DEAE Sepharose	Diethylaminoethyl weak anion exchange
2,5-DHB	2,5-Dihydroxybenzoic acid
DTT	DL-Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid

EG	Endoglucanase
EMBL	European Molecular Biology Laboratory
Endo H	Endoglycosidase H
EOF	Electroosmotic flow
ES	Electrospray
eV	Electron volt
F	Response factor
FIA	Flow-injection analysis
FPLC	Fast protein liquid chromatography
FWHM	Full width half maximum
GC	Gas chromatography
GlcNAc	N-Acetylglucosamine
GPC	Gel permeation chromatography
H. insolen	*Humicola insolen*
HPAEC-PAD	High performance anion-exchange chromatography with pulsed amperometric detection
HPLC	High performance liquid chromatography
ID	Inner diameter
LOD	Limit of detection
MALDI	Matrix-assisted laser desorption ionization
Man	Mannose
MH^+	Protonated molecules
MOWSE	Molecular Weight search

MS	Mass spectrometry
MS-MS	Tandem mass spectrometry
m/z	mass-to-charge ratio
NanoES	Nanoelectrospray
OD	Outer diameter
PBE	Polybuffer exchanger
PNGase F	N-glycosidase F
PVA	Poly(vinyl alcohol)
q	Quadrupole collision cell
Q	Quadrupole mass filter
Q-TOF	Quadrupole time-of-flight
rf	Radio frequency
RSD	Relative standard deviation
SAXS	Small angle X-ray scattering
S/N	Signal-to-noise ratio
SP Sepharose	Methyl sulfonate strong cation exchanger
TFA	Trifluoroacetic acid
TIC	Total ion chromatogram
T. reesei	*Trichoderma reesei*
TOF	Time-of-flight
UCSF	University of California, San Francisco
UV	Ultraviolet

A theory is the more impressive the greater the simplicity its premises is, the more different kinds of things it relates, and the more extended is its area of applicability.

Albert Einstein (1879-1955)

"Notes for an Autobiography," *Saturday Review*, 26 November 1949

Chapter 1

Theory

1.1 Introduction

Biological mass spectrometry has gained significant recognition in the past two decades. The rapid development of electrospray and matrix-assisted laser adsorption ionization techniques which allows for the detection of intact large bio-molecules has enhanced the role of mass spectrometry in analyzing proteins and peptides. Electrospray combined with low energy collision-induced dissociation tandem mass spectrometry has become a valuable tool for structure elucidation.

This thesis focuses on the use of biological mass spectrometry and capillary isoelectric focusing (CIEF) for the analyses of cellulolytic enzymes purified from *Trichoderma reesei.* Electrospray mass spectrometry (ESMS) was used to study the protein glycoforms of cellulases. Further structural characterization was accomplished using targeted proteolytic digest followed by mass spectrometric detection on both a quadrupole time-of-flight (Q-TOF) equipped with electrospray source and a matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF). This approach provided a mean to assess the complexity of post-translational modifications such as glycosylation and phosphorylation. CIEF was also used to quantitate cellulolytic enzymes present in crude cellulase fermentation extracts.

The current availability of Q-TOF and MALDI-TOF instruments has enabled the acquisition of a number of mass spectrometry experiments with little user intervention. The continuing evolution of instrumental designs aiming for higher sensitivity, increased resolution, better mass accuracy, ease of operation and ruggedness have made mass spectrometers a standard tool in most pharmaceutical companies. This thesis presents the

analytical capability of mass spectrometry in cellulase analyses which serves as an example to illustrate its power and applications in glycoprotein studies.

1.2 Cellulolytic Enzyme System of *Trichoderma reesei*

Cellulose is composed of repeating ß-1,4 linked D-glucose units predominantly found in plant cell walls and constitutes the most abundant organic polymer in nature. The role of cellulase is thus very important in recycling of the rich carbon source available in plant material in order to maintain the ecological balance. *Trichoderma reesei* is a filamentous fungus capable of effectively hydrolyzing crystalline cellulose to glucose. Its efficient secretory ability also makes it a useful organism for the large-scale production of enzymes for a variety of applications in the pulp and paper, textiles and animal feeds industries as well as a target for the development of recombinant strains for heterologous protein production.[1,2,3]

The cellulolytic enzyme system of *T. reesei* consists of two major cellobiohydrolases (CBHs) and several endoglucanases (EGs), together with small amounts of β-glucosidase and several hemicellulases.[4] In general, the cellobiohydrolases and endoglucanases show significant N- and/or O-linked glycosylation. With the exception of EG III, these cellulases comprise a catalytic domain (core) and a cellulose-binding domain (CBD) connected via a linker peptide which is heavily glycosylated at serine and threonine residues. The CBD domain serves as an anchor on the amorphous surface of crystalline cellulose and facilitates the catalytic action of the cellulolytic enzymes. The role of the O-linked glycosylation in the linker region is believed to be related to the secretion of these extracellular enzymes, to provide sufficient spatial separation between the core and CBD domains and to protect the

linker peptide against proteolysis.[5,6] Synergism exits between the cellobiohydrolases and endoglucanases to carry out the degradation of cellulose.[7,8]

1.2.1 Synergism

The most widely accepted hypothesis (International Union of Biochemistry and Molecular Biology, 1992)[9] for the synergistic action of cellulolytic enzymes starts with endoglucanases that randomly hydrolyse the β-(1→4) glycosidic linkages in the amorphous region of the cellulose. This significantly reduces the degree of polymerization of the substrate and creates new chain ends that facilitate the accessibility of cellobiohydrolases to split off cellobiose units. Subsequently, β-glucosidases hydrolyze cellobiose into glucose units. As an example, the synergistic action of cellulases on cellulose is illustrated in Figure 1.1 below.

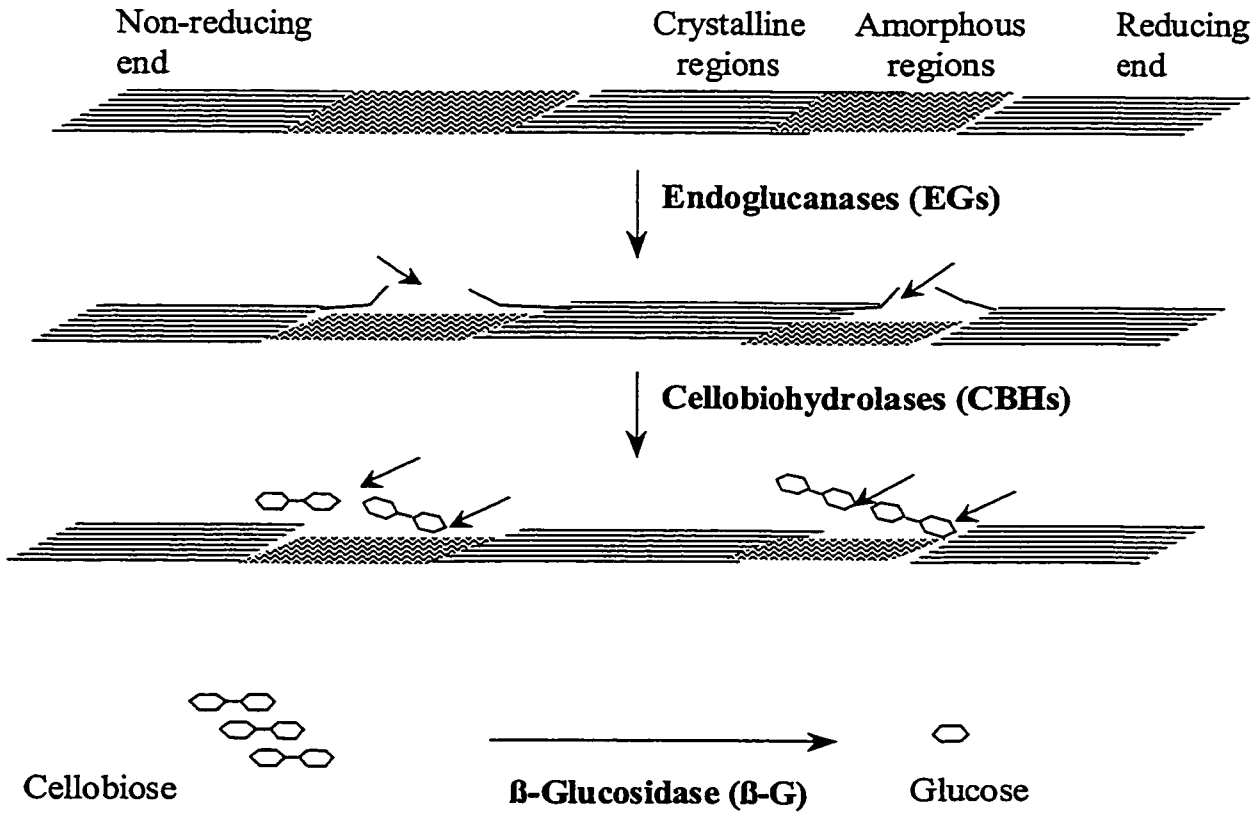


Figure 1.1 Enzymatic hydrolysis of cellulose into glucose.

1.2.2 Domain structures – from Gene to Proteins

The characteristic domain structures of the studied cellulases, as predicted from the gene sequences, are organized as shown in Figure 1.2.[10]

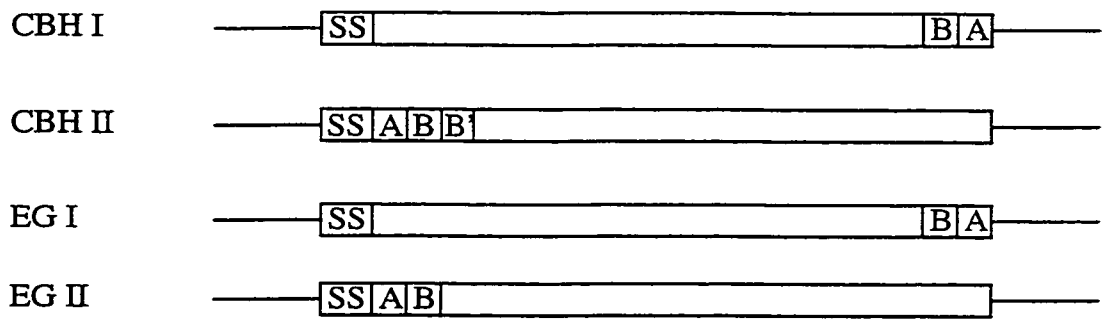


Figure 1.2 Domain structures highlighted on genes from major cellulases from *T. reesei.* The introns were omitted for simplicity. The homology domains are shown in labeled box (A for cellulose-binding domain; B and B' for linker region; SS refers to the signal sequence).

Block A in Figure 1.2 represents the CBD and is located at the C-terminus for CBH I and EG I and at the N-terminus for CBH II and EG II. Up to 70% of the amino acids are conserved in this domain, especially the cysteine and glycine residues (Figures 3.1a-d). The linker domain denoted by B is rich in serine and threonine residues, which are the potential sites for O-linked glycosylation, together with proline and arginine. CBH II contains two homology linker domains and is expected to be more glycosylated. The structure of both CBH I and CBH II have been studied by small angle X-ray scattering (SAXS) and they resemble that of a tadpole,[11] with the core forming the head and the protruding tail from CBD connected by the linker domain. The core can be cleaved from the linker and CBD using papain which specifically hydrolyzes peptide bond at the hinge between these domains.

1.2.3 Cellobiohydrolase I

Among the cellulolytic enzymes, CBH I comprises up to 60% of the total protein secreted by *T. reesei*.[12] Due to its abundance and facile purification, the structure and enzymatic action of CBH I have been intensely studied in the literature. Divne and co-workers have first described the three-dimensional crystal structure of CBH I core in 1994.[13] Subsequently, in a detailed paper published in 1998, the same group has investigated further the functional relationships between CBH I and its substrate.[14] An illustration of the crystal structure of CBH I core is shown below in Figure 1.3:

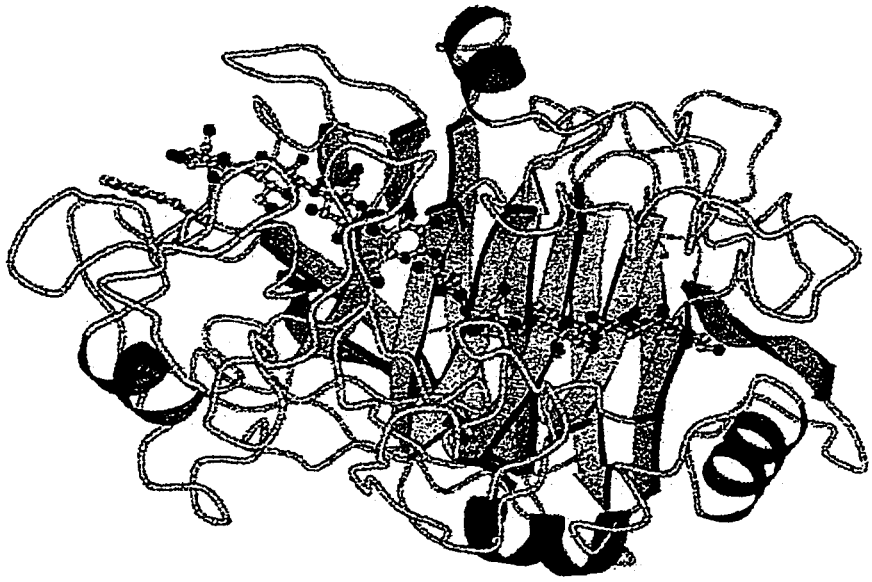


Figure 1.3 Schematic representation of CBH I catalytic domain. The cellulose chain with ten oligomers is depicted as a ball-and-stick assembled molecule being bound inside the tunnel formed by β strands shown as blue arrows. Other secondary structures include the α helices illustrated as red spirals and loop regions seen as yellow coils.

The most striking feature of the proposed structure is the tunnel formed by β strands that span around 50 Å and contains ten glucosyl binding sites (-7 to +4, with hydrolysis occurring between –1/+1). The presence of this tunnel structure allows for the close interaction

between the catalytic domain and the loose chain ends appearing on the crystalline cellulose surface. It was also identified that the two acidic residues, Glu212 and Glu217, are the catalytic residues involved in the hydrolysis of cellulose. Catalytic actions proceed in an acid/base catalysis fashion, releasing cellobiose from the reducing end with an overall retention of configuration at the anomeric C1 atom. The double-displacement nucleophilic substitution suggested at the anomeric carbon is similar to a S_N1- type mechanism involving an enzyme-glycosyl intermediate (Figures 1.3a-d).[15]

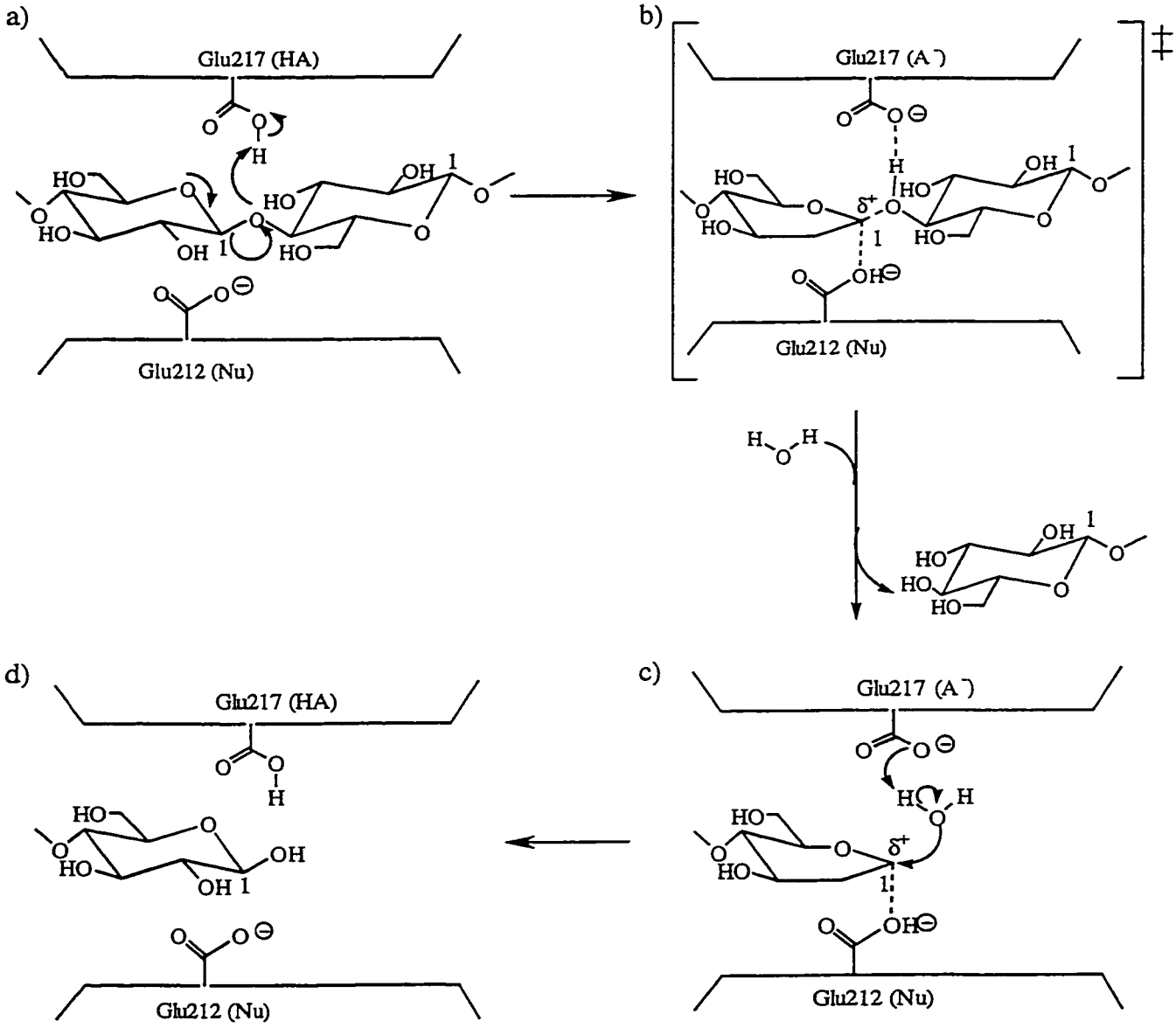


Figure 1.4 Hydrolysis of a glycosidic bond by CBH I via a double displacement mechanism with retention of configuration.

The reaction is initiated with a proton donation from an acidic group (Glu217) to the oxygen atom at the glycosidic linkage (Figure 1.4a), followed by the nucleophilic attack from Glu212 on opposite side of the substrate and this results in the formation of an oxocarbonium transition state with covalent or strong electrostatic interactions (Figure 1.4b).[16] Subsequently, the carboxylate on Glu217 obtains a proton from a water molecule, creating a hydroxyl ion which attacks the anomeric carbon (Figure 1.4c). This leads to the regeneration of the acid group in Glu217 and the β configuration of hydroxyl group at C1 is retained (Figure 1.4d).

1.2.4 Cellobiohydrolase II

In contrast to CBH I, the catalytic action of CBH II proceeds via a single displacement step. This is similar to a S_N2-type mechanism with an inversion of configuration at the anomeric carbon (Figures 1.5a,b).

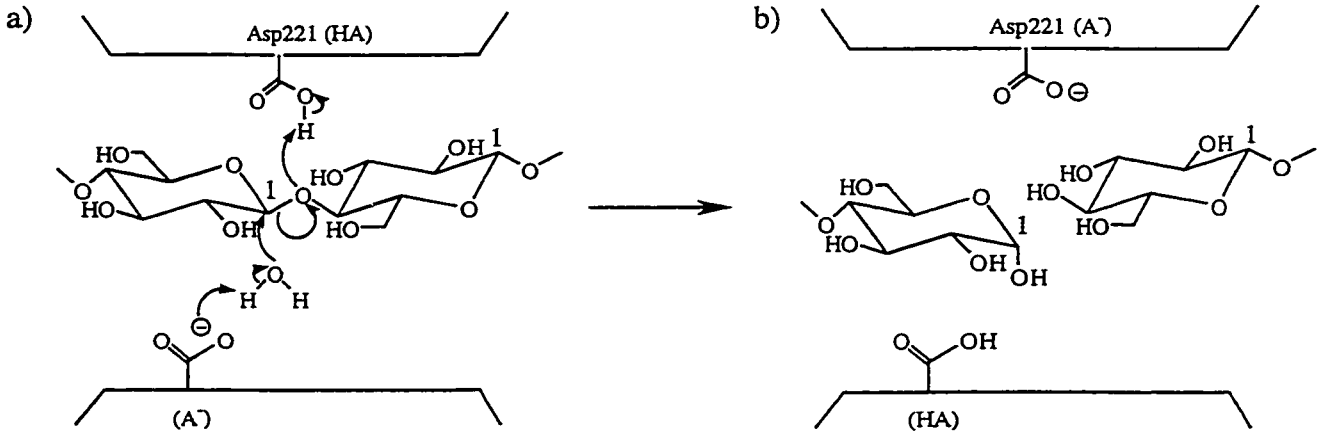


Figure 1.5 Hydrolysis of a glycosidic bond by CBH II via a single displacement mechanism with inversion of configuration.

Hydrolysis of cellulose by CBH II also involves a general acid/base catalysis with an aspartic acid (Asp221) acting as a proton donor, but the catalytic base has not been identified yet (Figures 1.5 a,b).[17] The crystal structure of CBH II suggests that tyrosine 169 is also important by helping to distort the substrate into a more reactive conformation to affect the protonation state of Asp221. The amino acid Asp221 protonates the glycosidic oxygen and

results in the formation of an oxocarbonium transition state. The carboxylate from the base then deprotonates a water molecule and forms a hydroxyl ion to attack the anomeric carbon. The conformation of the C1 hydroxyl group thus changes from β to α position.

CBH II hydrolyses cellobiose from the non-reducing end of the substrate. Its 3-D structure was reported to be consisted of seven parallel β-strands which forms an enclosed tunnel about 20 Å long.[18] Similar to CBH I, this tunnel structure enables close enzyme-substrate interactions through hydrogen bondings and stacking of planar aromatic rings from tryptophan residues. The tunnel has four binding sites for glycosyl units (-2 to +2, with hydrolyses occurring at –1/+1. Thus, the cleavage site is centrally located in the CBH II tunnel but offset in the CBH I tunnel.

1.2.5 Endoglucanase I

In contrast to cellobiohydrolases, the structure of EG I catalytic domain resembles an open substrate-binding cleft rather than a tunnel.[19] This enables EG I to interact with the amorphous or disordered crystalline cellulose more effectively. The crystal structure of EG I core has been compared to CBH I core in Figure 1.6 to illustrate their structural differences. Two stereo-drawings of EG I core are shown in Figure 1.6 for comparison purposes and the *H. insolens* (another commonly cellulase producing fungus used in industry) EG I is included for references. Selected amino acid positions of EG I have been numbered. Close examination of the structures in Figure 1.6 indicates that the loops forming the tunnel in CBH I are deleted in EG 1 and these include residues 191-204, 234-254, 317-321, 329-341 and 381-392.[19] These loops are essential to maintain the tunnel structure in cellobiohydrolases. In addition, EG I loop regions from residue 94-107 and 321-327 located on the right side of both *T. reesei* and *H. insolen* fungi are pulled further away from the top of the substrate-binding cleft which allows for easier accessibility of the catalytic active sites.

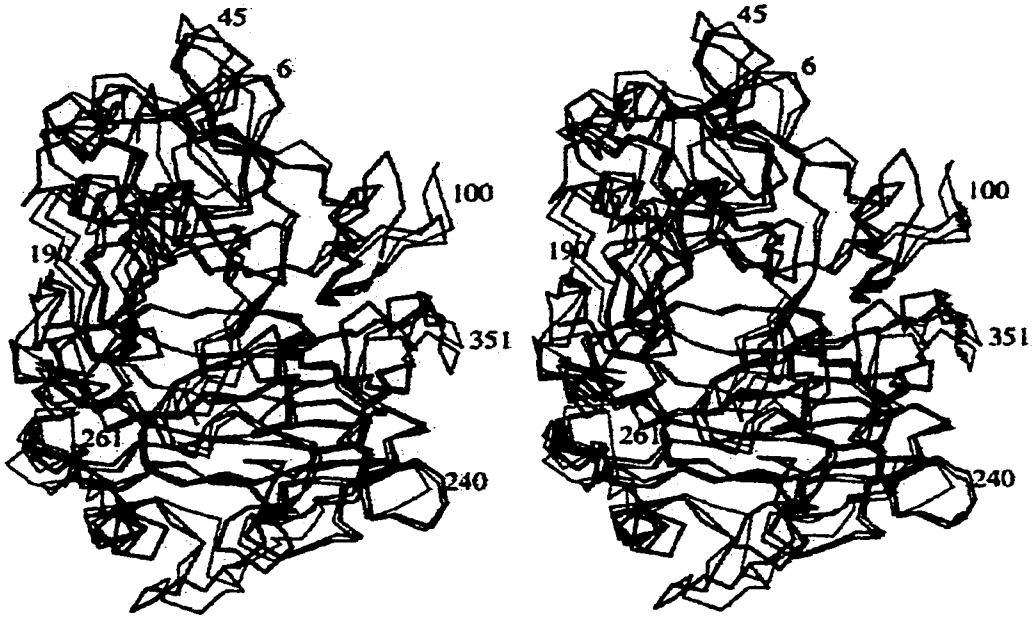


Figure 1.6 Three-dimensional superimposed structures of *T. reesei* EG I (red), *H. insolen* EG I (green) and *T. reesei* CBH I (blue).[19]

EG I is a retaining glycosylase and its catalytic mechanism is similar to that of CBH I illustrated in Figure 1.4. The active sites involve Glu197 acting as nucleophile to mediate the formation of the glycosyl-enzyme intermediate while Glu202 functions as the Brønsted acid/base. These two glutamic acids are located on the opposite side of the substrate separated by ~5.5 Å. Except for Trp38, the four tryptophan residues (38, 40, 367, 376) distributed in the CBH I tunnel to stabilize the cellulose substrate are conserved in EG I.

1.2.6 Endoglucanase II

At present, knowledge on the structure and catalytic mechanism of EG II is limited.[20,21] This endoglucanase is also a retaining glycosyl and the catalysis thus proceeds via a double displacement mechanism similar to CBH I. It has been proposed that the active site in EG II involves Glu329 acting as nucleophile while Glu218 functions as the Brønsted acid/base.[20] The catalytic activity of EG II is slightly lower than that of EG I on certain substrates.[21]

1.2.7 Glycosylation

Cellulolytic enzymes are glycoproteins containing both N- and O-linked glycan and exhibiting microheterogeneity in glycosylation.[22] There are four putative N-linked sites in the CBH I core (Asn45, Asn64, Asn270 and Asn384). Mammalian high-mannose-type N-linked glycans have been reported in CBH I from *T. reesei* RUT-C30 strain by Maras and co-workers.[23] They have found that these high mannoses are heterogeneous, ranging from the simple $Man_5GlcNAc_2$ to the most complex configuration $ManP\text{-}GlcMAn_7GlcNAc_2$ (P = phosphodiester linkage). Klarskov et al.[24] and Harrison et al.[25] have found independently the presence of single GlcNAc in Asn45, Asn64 and Asn270 of the CBH I catalytic domains from strains QM9414 and ALKO2877, respectively (see section 1.2.8 for genealogy of *T. reesei*). The nature of glycosylation in CBH II, EG I and EG II is not well documented in the literature, probably due to the difficulties in their purification. The putative N-linked sites, Asn56 and Asn182 in EG I core from VTT mutant strains have been reported to bear a single GlcNAc residue.[19] The Asn289 and Asn310 in CBH II have also been reported to be glycosylated, although the nature of the attached oligosaccharide was not determined.[18]

All the linker domains in these major cellulases are rich in serine and threonine residues and are heavily O-linked glycosylated. However, the heterogeneity and the proximities of these O-linked glycans make it difficult to assign their nature and attachment sites on the protein backbone. The O-linked glycans in CBH I from ALKO2877 have been reported to consist primarily of short-chain mono-, di- and tri-saccharides in the linker domain.[25] Differences in biological roles between N- and O-linked glycosylation in *T. reesei* have been studied in some extent, and it has been reported that only the O-linked glycosylation is necessary for the secretion of EG I and EG II.[26]

1.2.8 Genealogy of *T. reesei* Mutant Strains

All mutant strains of *T. reesei* are originated from the wild-type QM6a (Figure 1.7).[27]

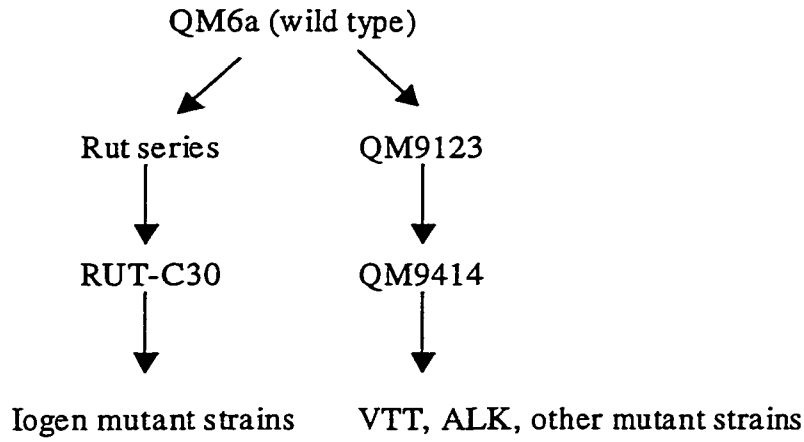


Figure 1.7 Partial family tree of *T. reesei* mutant strains.

The major high cellulase-producing *T. reesei* strains used in industry and for research purposes are shown in Figure 1.7. These mutant strains were produced using ultraviolet light and gamma irradiation in combination with chemical treatment. The wild-type QM6a was introduced by Mandels and Reese in 1957.[28] The VTT and ALK mutants were developed in the Technical Research Centre in Finland.[29] The Rut series were pioneered in the Rutgers University in the late 1970s.[30] RUT-C30 strain is major mutant strain in the Rut series and has been the parental strain from which the Iogen mutant strains (Iogen-M4 and Iogen-B13) were produced.

1.3 Capillary Electrophoresis (CE)

The technique of capillary electrophoresis involved the separations of charged molecules in a small fused silica capillary under the influence of an electric field. Compared to the slab-gel electrophoresis, CE is superior in terms of speed, quantitation and ease of automation. Since its introduction, CE has demonstrated versatile applications in the separations of small organic and inorganic molecules, as well as in biomolecules such as peptides, proteins, oligonucleotides, DNA, carbohydrates, and oligosaccharides.[31,32] In this project, two well-established modes of CE, capillary zone electrophoresis and capillary isoelectric focusing, were evaluated and the latter was found to be more suitable for cellulase analysis.

1.3.1 Capillary Zone Electrophoresis (CZE)

Capillary zone electrophoresis is among the most general applied mode in CE separations. A simple set-up showing the separation of charged and neutral solutes is shown in Figure 1.8 below.[33]

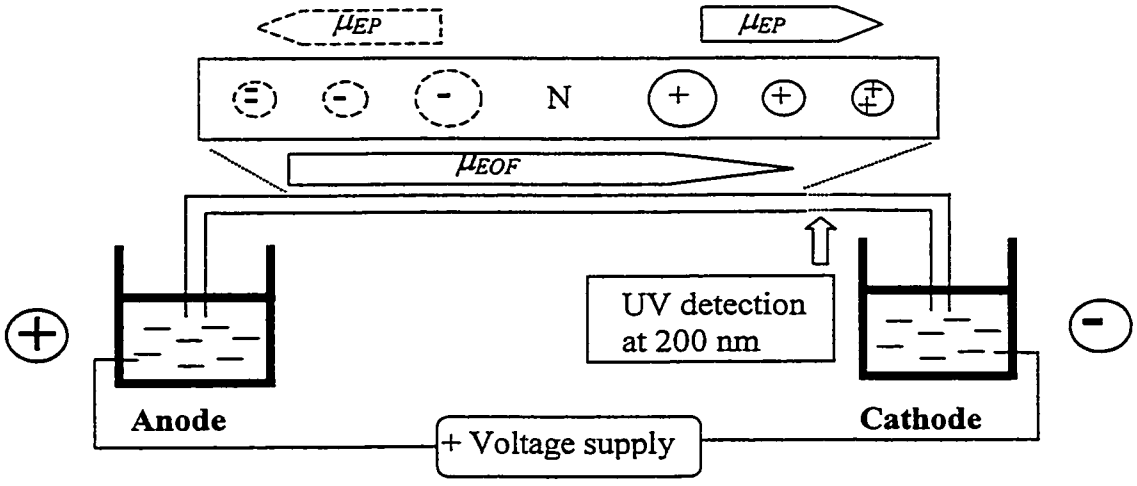


Figure 1.8 Schematic representation of CZE separation. (μ_{EP} = electrophoretic mobility; μ_{EOF} = electroosmosis mobility; circled +'s with solid line and -'s with dotted line represent cations and anions, respectively; N = neutral components).

Both ends of the capillary are inserted into reservoirs containing the same electrolyte solution, which are connected to the power supply to maintain electrical conductivity. The UV detection is generally monitored at 200 nm for protein and peptide analyses. Separation of charged molecules and neutrals are achieved by electrophoretic mobility. Under an electric field, positively charged cations migrate towards the cathode while negatively charged anions migrate towards the anode. The migration rates depend on their charge-to-size ratio (Figure 1.8) and ions with more charges and smaller sizes have a higher mobility. This movement is governed by the electrophoretic mobility (μ_{EP}) and is described by the following equation:[34]

$$\mu_{EP} = \frac{Q}{6\pi\eta r} \tag{1}$$

where Q is the charge on the ion, r is the hydrodynamic radius and η is the viscosity of the buffer. As expected, the electrophoretic mobility is proportional to Q while it is inversely related to r and η.

The overall mobility of the analyte is also influenced by the electroosmotic flow which propels the neutrals and anions towards the detector. The surface of the inner fused silica capillary is composed of silanol group (Si-OH) and are ionized to silanoate ($Si\text{-}O^-$) at pH above three.[35] These silanoate groups attract the positive ions from the buffer and form an inner fixed layer of cations on the capillary wall. On top of this fixed layer is a mobile layer of cations that is not tightly held. The potential drop across these two layers is described by the zeta potential (ζ):[35]

$$\zeta = \frac{4\pi\delta e}{\varepsilon} \tag{2}$$

where δ is the thickness of the double layer, e is the charge per unit surface area and ε is the dielectric constant of the electrolyte buffer. Under the influence of an electric field, the mobile layer of cations is pulled towards the cathode and bring with them solvated electrolyte

molecules, causing both the neutral and anions to migrate towards the detector. In a bare fused silica capillary, the combination of electrophoretic and electroosmotic mobility thus lead to the migration order of cations, neutral and lastly anions, depending on their charge-to-size ratios, as shown in Figure 1.8. The magnitude of electroosmotic mobility (μ_{EOF}) is proportional to the zeta potential and is given by:

$$\mu_{EOF} = \frac{\varepsilon\zeta}{4\pi\eta} \qquad (3)$$

The inner surface of the capillary is often coated with ionic phases to prevent the absorption of analytes to the capillary wall and to reverse the direction of EOF in order to enhance resolution.[36] Analyses of proteins and peptides at pH less than 7 are improved by introducing a cationic coating which imparts a positive charged layer on the capillary surface. The use of cationic polymer hexadimethrine bromide was adopted in all CZE experiments in this project. An acidic buffer and a reversal of polarity with cathode at the inlet and anode at the outlet, were required. Under these conditions, proteins and peptides become positively charged and the electrophoretic mobility is in the opposite direction to the electroosmosis mobility, resulting in migration after the neutral components.

1.3.2 Capillary Isoelectric Focusing (CIEF)

The technique of CIEF was first introduced by Hjertén and Zhu in 1985.[37] It combines the high resolving power of conventional IEF performed in slab gels and the efficient dissipation of Joule heat due to the separation in small diameter capillaries.[38] Separation is performed usually in two steps: focusing of analytes according to their pI values in a pH gradient created by carrier ampholyte, followed by hydrodynamic or chemical mobilization. The applied voltage remains the same during focusing and mobilization. This is illustrated in Figure 1.9:

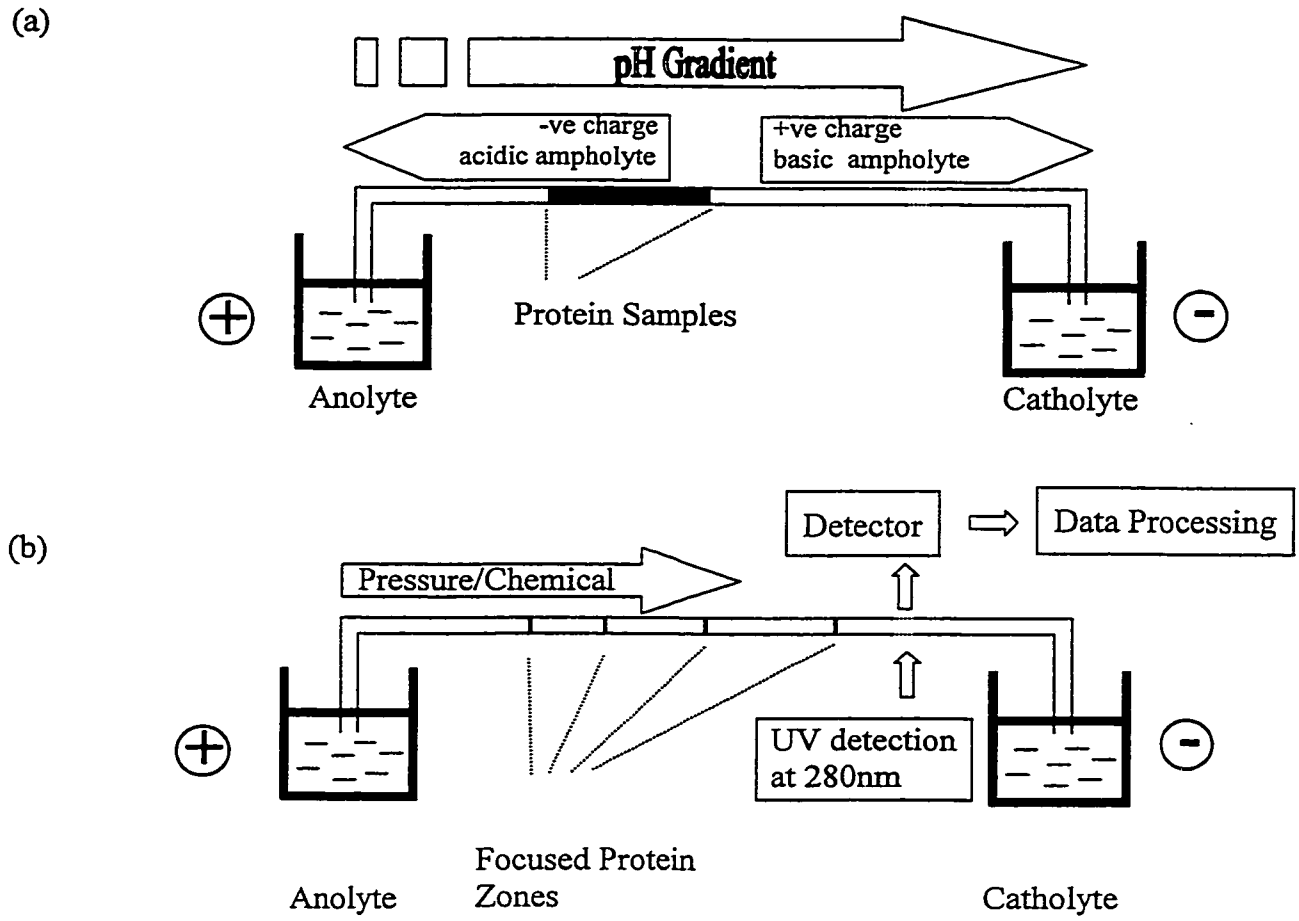


Figure 1.9 Schematic representation of two-step CIEF in neutral coated fused-silica capillary. (a) Focusing; (b) hydrodynamic or chemical mobilization.

In contrast to the low nL sample consumption in CZE, the entire capillary in CIEF is filled with protein samples dissolved in ampholyte solution. The buffer reservoir containing the anode and cathode is termed anolyte and catholyte, respectively. The role of anolyte and catholyte is to prevent the ampholyte molecules from migrating out of the capillary.

Ampholytes consist of polyamino-polycarboxylic acids with slightly different isoelectric (pI) values and have excellent buffering capacity. However, they absorb at 200 nm and CIEF is generally monitored at 280 nm in order to reduce the chemical background noise. When a voltage is applied, the positively charged basic ampholyte migrate towards the cathode until their isoelectric point have been reached and creates the high pH end. Simultaneously, the negatively charge acidic ampholyte migrate towards the anode, resulting in the low pH

extreme. This established a pH gradient and the protein samples are focused at their isoelectric point accordingly. Mobilization is accomplished by simply using low pressure typically of 0.5 psi. It is also achieved by replacing either the catholyte with a more acidic solution (cathodic mobilization) or the anolyte with a more basic solution (anodic mobilization).

Two-step CIEF is performed in neutral coated capillary to eliminate electroosmotic flow (EOF) during focusing. In general, this coating is created using methylcellulose or non-cross-linked acrylamide.[39,40] Recently, the use of poly(vinyl alcohol) (PVA) has grown popularity due to its simplicity in the coating process.[41,42] Single-step CIEF has also been used, in which dynamic focusing is performed simultaneously with mobilization. This mode of CIEF is performed in both uncoated and dynamically coated capillaries.[43,44]

The theoretical basis of CIEF was thoroughly described in a paper published by Hjertén and co-workers.[45] The main principal behind CIEF is the electroneutrality observed when focusing is completed. Such condition can be formulated as following:

$$C_{H+} + \Sigma C_{NH3+} = C_{OH-} + \Sigma C_{COO-} \qquad (4)$$

where C_{H+}, C_{OH-}, ΣC_{NH3+} and ΣC_{COO-} represent the concentrations in coulomb per cm^3 of capillary volume of protons, hydroxyl ions and total positive and negative groups in the carrier ampholytes, respectively. In pressure mobilization, the condition expressed in equation (4) remains the same. Chemical mobilization involves perturbing the established electroneutrality to initiate a shift in the pH gradient. Anodic mobilization is accomplished by introducing a non-proton cation X^{n+} in the left side of equation (4) such that:

$$C_{Xn+} + C_{H+} + \Sigma C_{NH3+} = C_{OH-} + \Sigma C_{COO-} \qquad (5)$$

where n represents the valency of charge state. Migration of this non-proton cation into the capillary results in a decrease in C_{H+}, or an increase in pH, which causes a progressive pH shift along the capillary and the focused analyte will thus mobilize towards the detector window. Typically, anodic mobilization is performed experimentally by replacing the anolyte with solutions containing cations that can enter the separation capillary electrophoretically (Figure 1.9). Conversely, cathodic mobilization is accomplished by introducing a non-hydroxyl anion C_{Ym-} to equation (4), resulting in the following equation:

$$C_{H+} + \Sigma C_{NH3+} = C_{OH-} + \Sigma C_{COO-} + C_{Ym-} \qquad (6)$$

Introduction of this non-hydroxyl anion results in a decrease in the concentration of hydroxyl ions, or a decrease in pH. To maintain electroneutrality, the ampholyte solution will slowly migrate towards the capillary outlet (Figure 1.9).

CIEF provides resolution of proteins differing only by 0.05 pI units with an enhancement of concentration detection limit ranging from 50-100 fold compared to conventional zone electrophoresis.[46] The application of CIEF has been demonstrated previously for clinical samples[47,48] and for the quantitation of hemoglobins and its variants.[49,50] However, precise quantitation measurements often require the use of internal standard to account for variation in migration time and sample buffers. The difficulty in analyzing crude cellulase samples is further compounded by the natural heterogeneity of these glycoproteins as recently described by Medve and co-workers for CBH I, CBH II and EG II using fast protein liquid chromatography (FPLC) and IEF separation of the resulting fractions.[51]

In this project, all CIEF experiments were performed using two step conditions with cathodic mobilization, in which the catholyte is replaced by an acidic solution comprised of methanol:water:acetic acid (50:49:1 v/v/v). The non-hydroxyl anion in this case is the acetate ion (CH_3COO^-).

1.4 Formation of Ions in Mass Spectrometry

The rapid development of MALDI and electrospray ionization techniques in the last decade have dominated the applications of mass spectrometry for biological samples such as protein characterization, peptide sequencing, carbohydrate analysis as well as in drug discovery.[52] These two techniques are sometimes referred to as "soft" ionization, which defines the relatively low internal energy imparted on the analyte compared to other existing ionization techniques for formation of gas-phase ions such as electron impact. The theory behind these ionization methods will be discussed and their analytical performance will be demonstrated and compared in chapters 3 and 5.

1.4.1 Matrix-Assisted Laser Desorption Ionization

The role of matrix is crucial to the successful formation of ions in this mode of ionization.[53] All matrices must absorb the UV-laser light and they all possess at least one aromatic residue. The most common laser source is the nitrogen laser operating at 337 nm. Sinapinic acid and 2,5-dihydroxy benzoic acid (2,5-DHB) were the matrices used for protein and peptide analyses in this project and their structures are illustrated in Figure 1.10.

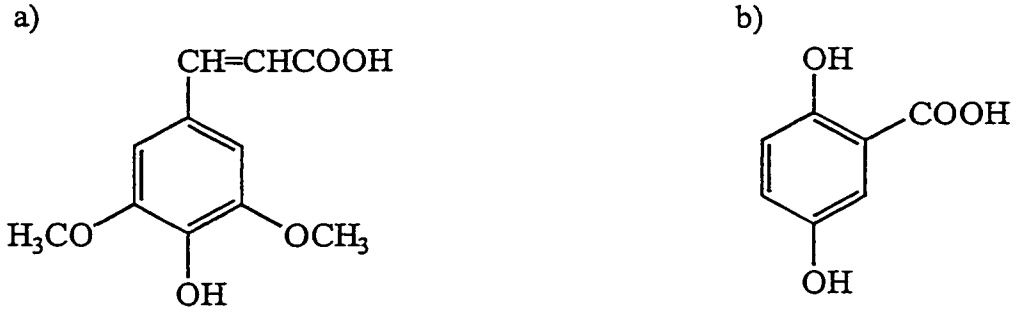


Figure 1.10 Chemical structures of selected MALDI matrices. (a) Sinapinic acid and (b) 2,5-DHB.

The mechanism of MALDI ionization is far from well understood.[54] Briefly, the samples are co-crystallized with a molar excess of matrix on a stainless steel surface, and upon absorbing

the laser light causes the matrix molecules to rapidly dissociate into cations and anions which act like an ion exchange resin to protonate the sample molecules. At the surface of the matrix/sample crystal, a high pressure region (supersonic molecular beam) is formed from the rapid sublimation of matrix/crystal. This subsequently produces a supersonic expansion which provides a mean for the formation of ions in the gas phase.

1.4.2 Electrospray

Electrospray ionization was first demonstrated by Fenn et al. in the early 1980s[55] and it involves the formation of gas-phase ions from solution in the presence of a strong electric field (Figure 1.11).[56] This ionization process occurs in the atmospheric pressure and has largely facilitated the coupling of separation techniques such as liquid chromatography and capillary electrophoresis to mass spectrometry. Sample effluent is typically introduced to the mass spectrometer at a flow-rate of 1-100 μL/min. Kebarle and Tang have proposed a detailed mechanism for this ionization.[57] When a high electric tension is applied to the electrospray needle (+ 5 kV), positive electrolyte in solution is pulled downfield (towards the

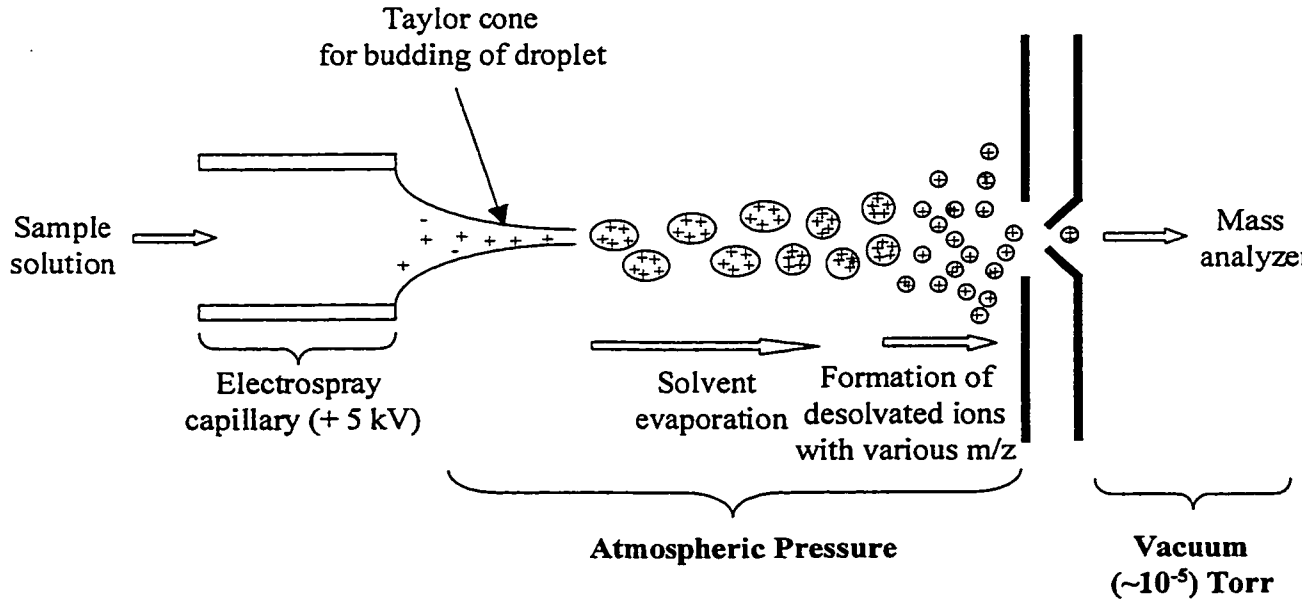


Figure 1.11 Schematic representation for the formation of ions in electrospray ionization process.

mass spectrometer), leading to the formation of a Taylor cone from which charged droplets in a fine mist are emitted. As the droplets traverse towards the sampling cone of the mass spectrometer, solvent evaporates from the surface and continuously shrink in size until they approach Rayleigh limit where they undergo fission into even droplets. These continuous fission and solvent evaporation processes eventually lead to ejection of ions.

Another theory proposed by Iribarne and Thomson suggested the formation of highly charged droplets, including both the analyte and other charged electrolytes at the tip of a fine capillary typically held at high electric field.[58,59] As the solvent evaporates, these droplets further reduce in size. The charges in the droplets become more compact and concentrated to a point when the repulsion between the same charged ions become greater than the surface tension of the droplets. At this point, ions with various mass-to-charge begin to eject from the droplets.

An unique characteristic of electrospray mass spectrometry is its ability to form multiply-charge ions. This is particular useful for protein and peptide analyses that have multiple charged sites and their molecular weights are easily calculated from two consecutive mass-to-charge states, $(m/z)_a$ and $(m/z)_b$ on the spectrum:

$$(m/z)_a = \frac{M + nH}{n} \qquad (m/z)_b = \frac{M + nH + 1}{n+1} \qquad (7)$$

Where M is the molecular mass of the analyte and n is the number of charges. This results in two equations with two unknowns (M, n) from which the molecular mass can be undoubtedly obtained. Examples of electrospray mass spectra are shown throughout chapters 3 to 5.

Compared to MALDI ionization, electrospray has higher mass accuracy of ± 0.01% as opposed to ± 0.1% of MALDI in protein analysis. It required only weekly calibration in

routine usage and most importantly, it can be interfaced on-line to HPLC and CE. However, electrospray has low salt tolerance from buffer solution and the presence of complex mixtures often complicates the deconvolution process to obtain the molecular masses. The limit of detection (LOD) is typically 10 ten-fold more sensitive in MALDI than in electrospray ionization.

1.4.3 Nanoelectrospray

The rapid expansion of biological mass spectrometry requires ionization techniques with minimal sample consumption and high sensitivity. The development of nanoelectrospray was pioneered by Matthias Wilm and Matthias Mann in the mid 90s.[60] Typically, a 1 μL sample solution is introduced into a gold-coated silica capillary emitter with pulled tip of about 1 μm inner diameter. The analyte is delivered to the mass spectrometer at a flow-rate of 20-40 nL/min by the electrospray process (1.5 – 2 kV), with miminal amount of nebulizing gas or hydrodynamic flow assistance. This low flow-rate regime permits prolonged infusion of sample solution (1 μL of sample lasts for ~ 30 min) and thus provides sufficient time to perform low collision-induced dissociation on selected precursor ions (section 1.5), especially from mixtures such as proteolytic digest. Compared to the conventional electrospray ionization process, nanoelectrospray has shown improvement in desolvation, ionization and transfer efficiencies.[60] This is attributed to the small size of droplets emitted from nanoelectrospray which are about two to three order of magnitude smaller than conventional electrospray thereby preventing the formation of clusters and facilitating rapid fission and solvent evaporation.

1.5 Mass Analyzers

Mass spectrometry involves the formation of gas-phase ions and their separation based on mass-to-charge ratios (m/z). Separation can be achieved using for example electric or magnetic sector instruments or combination of them to improve selectivity. Magnetic and electrostatic analyzers are used conventionally for small molecule analysis. Recently, the development and refinement of time-of-flight and quadrupole analyzers have revolutionized the role of mass spectrometry in biological studies.

1.5.1 Time-of-Flight Mass Analyzer

Time-of-flight mass analyzers first emerged in the 1960s but were not popular at that time due to the relatively poor resolution and sensitivity compared to other types of mass spectrometers.[61] With the improvement in detector electronics, TOF analyzer has now been recognized as a powerful tool for detection of large bio-molecules. TOF analyzer measures the time of arrival of ions to the detector in a drift tube at an applied potential. The principle of time-of-flight mass analyzer is based on the kinetic energy of ions:[61]

$$qV = \frac{1}{2} m u_d^{\,2} \tag{8}$$

where q is the charge in coulombs, qV is the translational kinetic energy of ions experienced across a potential drop V applied along the flight tube and u_d is the drift velocity. Substituting the simple expression of time equals to distance divided by velocity into equation (8), an expression for the time required (t_d) for ions to travel through the drift tube with distance d can be derived:

$$t_d = d\sqrt{\frac{m}{2qV}} \tag{9}$$

Since t_d is proportional to the square root of mass, the flight time for ions at higher masses becomes closer and thus reduces the resolution.

TOF analyzer is compatible with pulsed ionization techniques and the coupling of MALDI to TOF has become an excellent choice for analyzing biological macromolecules such as proteins, oligosaccharides, glycoconjugates and polynucleotides.[62,63] A schematic of the Biosystem DE-Elite MALDI-TOF used in this project is illustrated in Figure 1.12.

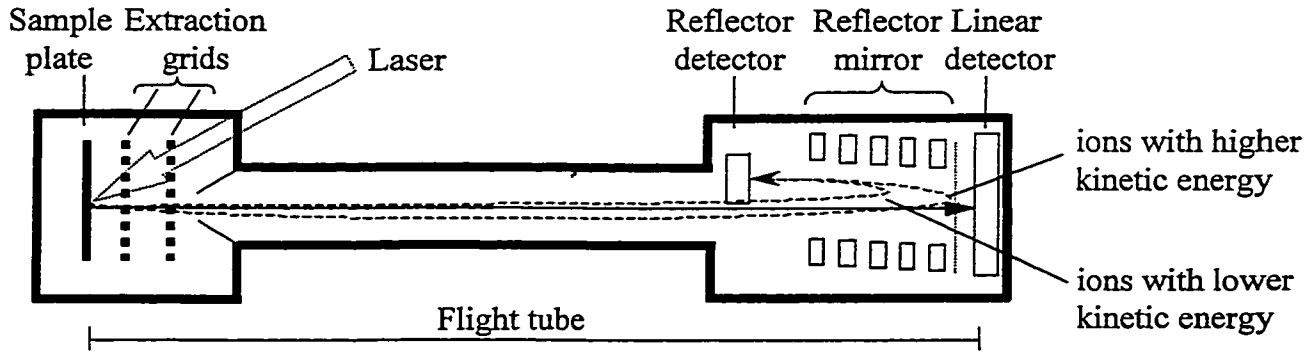


Figure 1.12 Schematic representation of a MALDI-TOF mass spectrometer (arrow with solid line represents ion path in linear mode; arrow with dotted line represents ion path in reflectron mode).

In view of improving the resolution of TOF analyzers, most commercial MALDI-TOF are equipped with delayed extraction technology and are capable of operating in both linear or reflectron modes. Delayed extraction was first proposed by Wiley and MacLauren in 1955.[64] They had introduced a method called "time-lag velocity focusing" to correct for the initial velocity distribution of ions in the source. When ions are produced from the sample plate, their positions are correlated with their initial velocity and application of a fast pulse at a predetermined delay time makes those ions with the same mass-to-charge ratio to reach the detector at the same time. Upon leaving the source, ions travel through the flight tube in a linear pathway or a longer reflectron pathway which further enhances the mass resolution.

Reflectron mode focuses ions of the same m/z with initial spread of kinetic energy. After drifting in the flight tube ions traversed through a series of reflector mirrors which serves to decelerate and eventually reflects their paths (Figure 1.12). Those ions with higher kinetic energy penetrate deeper into the reflector mirrors and thus arriving at the detector simultaneously with others having the same m/z.

Incorporation of both delayed-extraction and reflectron technologies in MALDI-TOF instruments have largely facilitated peptide and protein analyses. Vestal and co-workers have reported nearly a two-fold enhancement of resolution for standard peptides by switching from static to delayed extraction in reflectron mode.[65] The Biosystem DE-Elite MALDI-TOF has resolving power greater than 10,000 (M/ΔM) for peptides up to about 5,000 Da while progressively less than 500 (M/ΔM) for proteins from 10,000 Da upwards. The detection limit is typically 0.001-1 pmol for peptides and 0.01-10 pmol for proteins.

1.5.2 Quadrupole Mass Filter

Quadrupole mass filters also plays an important role in biological mass spectrometry due to their fast scanning capacities and are thus desirable for coupling to gas chromatography (GC) as well as liquid chromatography (LC).[66] Compared to magnetic sector instruments, quadrupole mass spectrometers are more rugged, less expensive, easier to operate and have higher transmission rate.[67] A quadrupole consists of four metal rods arranged symmetrically as shown in Figure 1.11.

Quadrupole mass spectrometer employs a combination of direct current (dc) voltage and alternating current (ac) voltage operating at radio frequency (RF) to carry out mass filtering.[68] The two metal rods from opposite sides are connected electrically and carries the

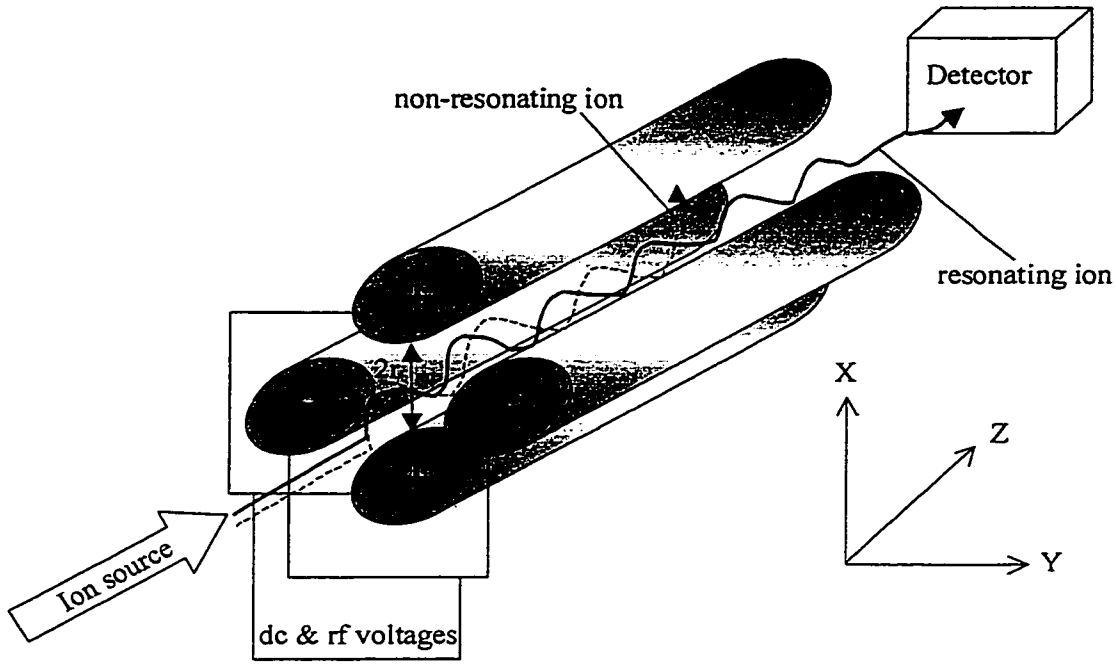


Figure 1.13 Schematic representation of a quadrupole mass filter. Ion path with stable trajectory (resonating) is shown in arrow with solid line while ion path with unstable trajectory (non-resonating) is shown in arrow with dotted line.

dc voltage (U). An oscillating ac voltage ($V \cos \omega t$) rapidly alternates between positive and negative voltages at radio frequency superimposes on the dc voltage. This results in the applied potential (ϕ_o) experienced by ions in the quadrupole field described in the following equation:

$$\phi_o = U + V \cos \omega t \tag{10}$$

where ω is the angular frequency (rads/sec). At a particular combination of ac and dc voltages, only ions possess a certain range of given m/z values have a stable trajectory along the quadrupole while all other ions become unstable and hit the metal rods. The Mathieu equations describe the motion of ions along the x and y axes in the quadrupole field with opposite rods separated by $2r_0$:[69]

$$\frac{d^2x}{d(\omega t/2)^2} + [a + 2q\cos 2(\omega t/2)]x = 0 \tag{11}$$

$$\frac{d^2y}{d(\omega t/2)^2} - [a + 2q\cos 2(\omega t/2)]y = 0 \qquad (12)$$

where

$$a = \frac{8eU}{mr_0^2\omega^2} \qquad q = \frac{4eV}{mr_0^2\omega^2} \qquad (13)$$

Therefore

$$\frac{a}{q} = \frac{2U}{V} \qquad (14)$$

The parameters a, q in equation 13 define a stability diagram showing the boundaries for ions with mathematical stable trajectories (Figure 1.14).

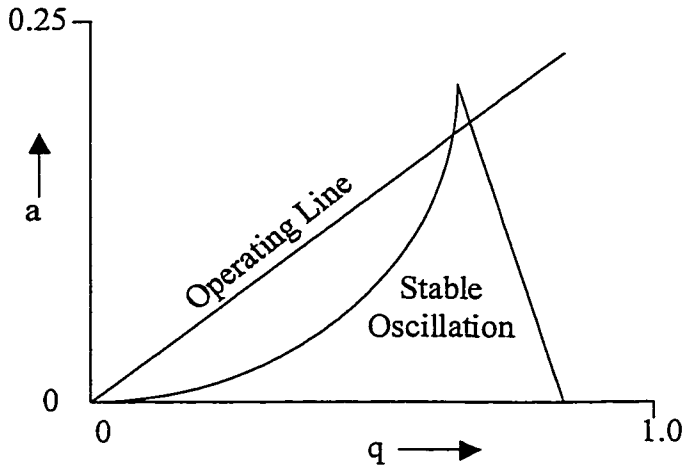


Figure 1.14 Stability diagram for quadrupole mass filter showing the area for stable and unstable oscillation.

The operating line has a constant ratio of a/q ratio and only ions with m/z satisfying the (a,q) coordinates (equation 13) at the apex of the bounded region would have stable trajectories in the quadrupole filter. An increase in the slope of a/q would result in a decrease in the apex area, which enhances the resolving power but at the expense of sensitivity.

Likewise, if the slope of a/q decreases, the operating line travels through a larger area and results in an increase in sensitivity at the expense of resolution.

Quadrupole mass filters are widely used in conjunction with electrospray because they can tolerate high pressure regime (10^{-5} Torr). In addition, ions traversing the quadrupole mass filters are accelerated to less than 15 V. This is ideal for collision activation at low energy regime (< 500 eV) that enables dissociation of selected precursor ions in a triple quadrupole instrument (section 1.6.1 for more details) thereby providing an excellent tool for peptide sequencing and bio-molecule characterization.

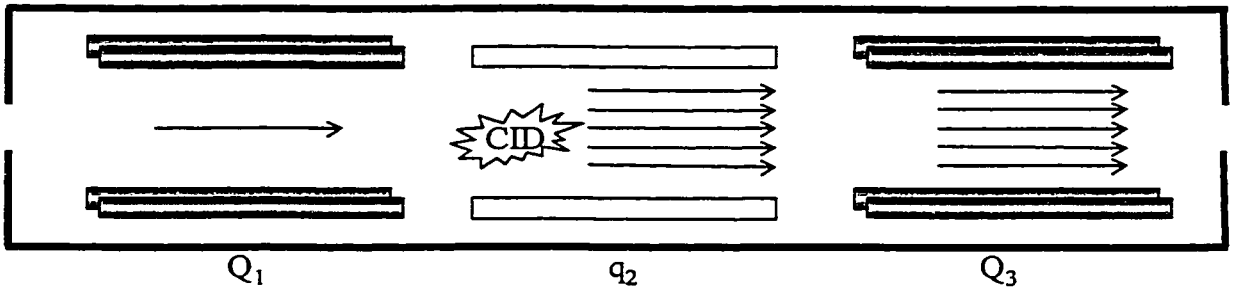
1.6 Tandem Mass Spectrometry (MS-MS)

The basic principle behind tandem mass spectrometer involves the selection of precursor or parent ion(s) in the first mass analyzer followed by separation of product or daughter ion(s) formed in a CID process in the second mass analyzer.[70] Development of triple quadrupoles and hybrid quadrupole time-of-flight tandem mass spectrometers coupled to electrospray ionization have played an important role in structural elucidation of bio-molecules.

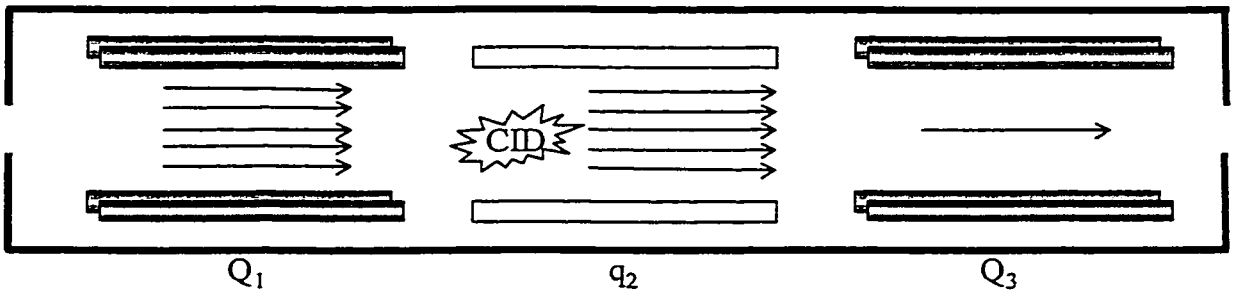
1.6.1 Triple Quadrupoles

This tandem mass spectrometer consists of two quadrupole mass filters (Q_1 and Q_3) separated by a rf-only (radio frequency) quadrupole (q_2) aligned linearly (Figure 1.15).[71] The second quadrupole acts as the collision cell for low energy CID (< 500 eV) to occur while the first and third quadrupoles select or scan independently for precursor and product ions, respectively. Three common strategies are performed by triple quadrupole instruments as

a) Product ion scan



b) Precursor ion scan



c) Constant neutral loss

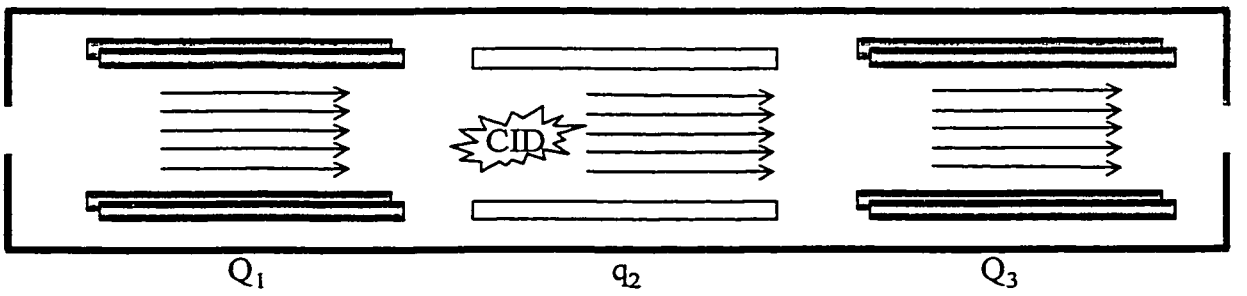


Figure 1.15 Schematic representations of three major experiments performed by triple quadrupoles. (a) Product ion scan; (b) Precursor ion scan and (c) constant neutral loss.

illustrated in Figure 1.15.[72] In product ion scan, a precursor ion is selected in Q_1 which undergoes CID, and the resulting fragment ions are scanned in Q_3. This mode is particular useful for mixture analyses such as the proteolytic digest where a precursor ion of interest is subjected to low energy CID to obtain the peptide sequence information. Conversely, a product ion is selected in Q_3 while all ions are scanned in Q1 in precursor ion mode. This strategy allowed for the determination of homologous classes of compounds such as

assigning the phosphate or sulfate groups transferred to proteins during posttranslational modification. Constant neutral loss involves scanning both Q1 and Q3 in parallel simultaneously while keeping a given offset between Q_1 and Q_3 corresponding to a neutral fragment of a predetermined mass. This approach allows for the determination of related compounds with common fragmentation pathways. Combination of these strategies has resulted a powerful method for structure characterization.[73,74]

Product ion scan was used in some extent to carry out partial peptide sequencing in this project. However, the unit resolution of quadrupole fails to differentiate the charge states of the resulting product ions and has significantly complicated the data interpretation process. The emergence of a new hybrid tandem mass spectrometer, whereby a quadrupole is coupled to a TOF analyzer, has eliminated such problem and is becoming the tool of choice for partial peptide sequencing.

1.6.2 Quadrupole Time-of-Flight

The first commercial hybrid quadrupole time-of-flight (Q-TOF) instrument was pioneered in Howard Morris' group recently in 1996.[75,76] In order to provide a pulsed beam of ions to the TOF analyzer, an orthogonal acceleration first described by Dawson and Guilhaus was introduced.[77] The combination of accelerator TOF after precursor selection and collision in the quadrupole analyzers has significantly improved the resolution and enlarged the applicability of tandem mass spectrometry. Compared to the unit resolution of a triple-quadrupole, this hybrid Q-TOF mass spectrometer allows for the unambiguous sequencing of low femtomole amount of peptides commonly extracted from 2-D gel protein separation.[76] The geometry of a Micromass Q-TOF instrument used during this project is illustrated in Figure 1.16.

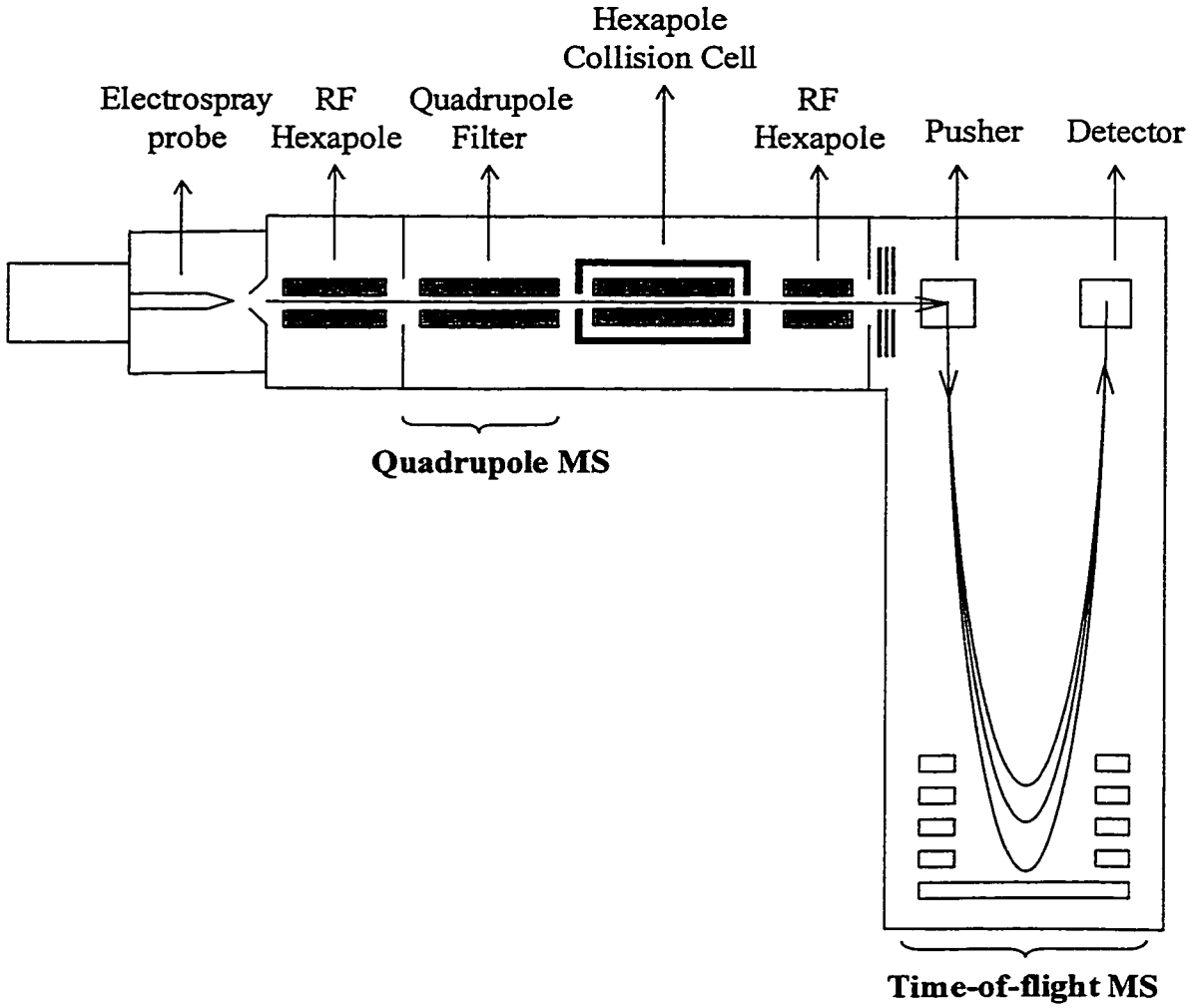


Figure 1.16 Schematic illustration of a Q-TOF instrument and its major components.

The main advantage of constructing the orthogonal geometry is the significant improvement in sensitivity and acquisition speed compared to the scanning mass analyzers. For peptide sequencing, samples are first ionized by electrospray and guided into the first quadrupole through a rf-only hexapole. The selected precursor ion then passes through the hexapole gas cell and is fragmented using low collision energy, following which the product ions are pulsed into TOF mass analyzer by a pusher electrode. In normal MS acquisition, the quadrupole is operated in the RF-only mode and acts as a wide band pass to transmit all ions into the TOF analyzer. The prototype Q-TOF introduced by Morris et al.[75] had an

accelerating potential of 7.8 kV between the ion source and the TOF analyzer, and the pulse rate was operated at 16 kHz for mass range of 1500 u.

The performance of a Q-TOF prototype on peptide sequencing has been evaluated.[76] A product ion mass resolution greater than 3000 FWHM (M/ΔM defined at full width half-maximum) and a mass accuracy of 0.05 Da was reported. With this resolution, the different charge states of product ions appeared on the mass spectra could be unambiguously assigned. Except for the two pairs isoleucine/leucine and glutamine/lysine, the mass accuracy allowed for the precise identification of amino acids based on the space difference between fragment ions in the mass spectra. These performances have alleviated the ambiguities in peptide sequencing posed by triple quadrupole MS-MS experiments.

Commercial hybrid Q-TOF mass spectrometers nowadays typically provide a mass resolution of 10,000 FWHM with attomole sample detection limits. Concurrent with the rapid expansion of proteomic research requiring highly sensitive mass spectrometer and very precise mass measurement in peptide sequencing, the pivotal role of Q-TOF in this research endeavor is rapidly becoming recognized.

1.7 Peptide Sequencing using Tandem Mass Spectrometry

Low CID energy give rise to fragmentation of peptides primarily at the amide bonds (Figure 1.17).[78] Combined with nanoelectrospray and quadrupole time-of-flight mass spectrometer, this technique has become a very sensitive and accurate method for rapid "de novo" peptide sequencing.[79] The linear backbone of peptides composed of 20 naturally occurring amino acids typically generates a ladder of sequence ions.

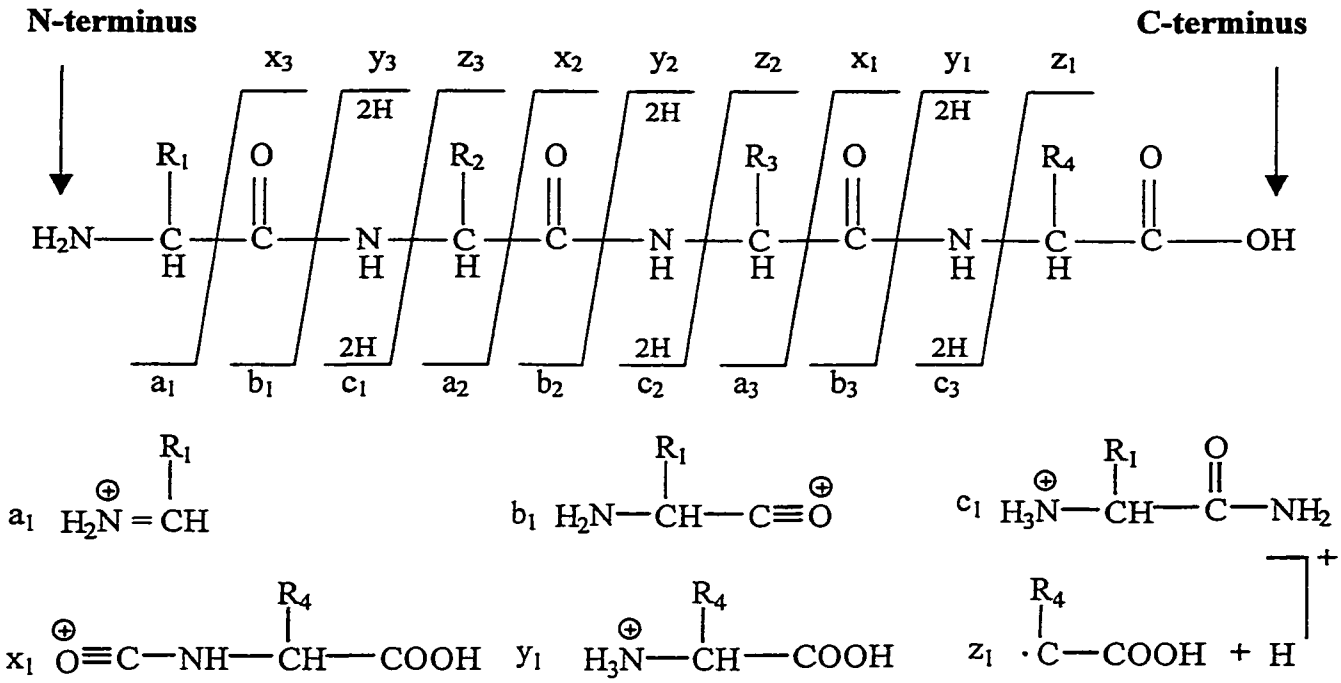


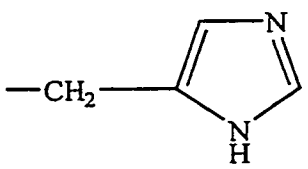
Figure 1.17 Nomenclature of fragment ions produced from low energy CID of a typical protonated peptide. The 2H indicates that an additional proton is obtained from opposite side in c and y ions.[78]

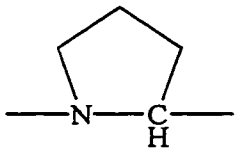
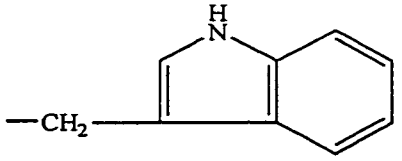
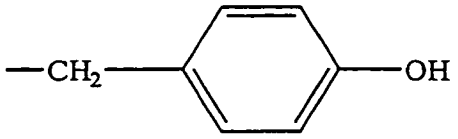
Upon collisional activation, some of the kinetic energy of the original protonated peptides is converted into vibrational energy thereby assisting the fragmentation process to occur. The nomenclature of peptide sequencing is created in accordance to the charge residency upon cleavage. If the charge is retained on the N-terminus, a series of type a_n, b_n and c_n ions are produced. If the charge is retained on the C-terminus, fragment ions belonging to type x_n, y_n and z ions are generated.

Production of type b_n and y_n ions are favoured because they correspond to the cleavage at the amide linkages. In general, the presence of basic residues, arginine and lysine, near the C-terminus of tryptic peptides favours the formation of y-type ions. Despite the numerous number of peaks appearing on the MS-MS spectra, interpretation of the mass spectrum with respect to peptide sequencing is facilitated through the following criteria: (a) the spacing

between adjacent peaks generally correspond to the m/z of one of the 20 amino acids (Table 1.1); (b) the fragment ions form a continuous ladder series with the same charge state. Often time, the use of terminal y-type ions such as arginine (m/z 175) and lysine (m/z 147) are necessary to complete the assignment of partial sequence. The immonium ions from tryptophan (m/z 159), tyrosine (m/z 136) and phenylalanine (m/z 120) are also diagnostic fragment ions.

Table 1.1 Chemical compositions and monoisotopic masses of twenty natural occurring amino acid residues

Name (Composition)	Symbols	Side Chain Structure	Residue Monoisotopic Mass (Da)
1. Alanine (C_3H_5NO)	Ala (A)	$-CH_3$	71.03711
2. Arginine ($C_6H_{12}N_4O$)	Arg (R)	$-CH_2-(CH_2)_2-NH-\underset{\underset{NH}{\|}}{C}-NH_2$	156.10111
3. Asparagine ($C_4H_6N_2O_2$)	Asn (N)	$-CH_2-CONH_2$	114.04293
4. Aspartic Acid ($C_4H_5NO_3$)	Asp (D)	$-CH_2-COOH$	115.02694
5. Cysteine (C_3H_5NOS)	Cys (C)	$-CH_2-SH$	103.00919
6. Glutamic Acid ($C_5H_7NO_3$)	Glu (E)	$-CH_2-CH_2-COOH$	129.04259
7. Glutamine ($C_5H_8N_2O_2$)	Gln (Q)	$-CH_2-CH_2-CONH_2$	128.05858
8. Glycine (C_2H_3NO)	Gly (G)	$-H$	57.02146
9. Histidine ($C_6H_7N_3O$)	His (H)		137.05891

10. Isoleucine ($C_6H_{11}NO$)	Ile (I)	$-CH(CH_3)CH_2-CH_3$	113.08406
11. Leucine ($C_6H_{11}NO$)	Leu (L)	$-CH_2CH(CH_3)_2$	113.08406
12. Lysine ($C_6H_{12}N_2O$)	Lys (K)	$-CH_2-(CH_2)_3-NH_2$	128.09496
13. Methionine (C_5H_9NOS)	Met (M)	$-CH_2-CH_2-S-CH_3$	131.04049
14. Phenylalanine (C_5H_9NO)	Phe (F)	$-CH_2-Ph$	147.06841
15. Proline (C_5H_7NO)	Pro (P)		97.05276
16. Serine ($C_3H_5NO_2$)	Ser (S)	$-CH_2-OH$	87.03203
17. Threonine ($C_4H_7NO_2$)	Thr (T)	$-CH(OH)CH_3$	101.04768
18. Tryptophan ($C_{11}H_{10}N_2O$)	Try (W)		186.07931
19. Tyrosine ($C_9H_9NO_2$)	Tyr (Y)		163.06333
20. Valine (C_5H_9NO)	Val (V)	$-CH(CH_3)_2$	99.06841

1.8 Peptide Mapping and Database Searching

The large amount of peptide fragments produced from enzymatic digestion provides a fingerprinting identity to determine the protein of origin. Peptide mapping is a process

through which the peptide masses of an unknown protein resulting from proteolytic digest are compared to the theoretical digest from known protein candidates available in the database systems.[80,81] Search engines such as ProteinProspector from the University of California, San Francisco (UCSF), PeptideSearch from the European Molecular Biology Laboratory (EMBL), MOlecular Weight SEarch (MOWSE) from Daresbury laboratory in Cheshire can all be assessed from the internet. They are generally developed from public databases like SwissProt and NCBI (National Center for Biotechnology Information) and the user has to define constraint parameters such as the protein mass range, the proteolytic enzymes, modifications in cysteine residue after reduction and alkylation, peptide mass accuracy and a list of peptide masses.

The sequence coverage from peptide mapping sometimes is low due to unidentified peptide masses or unexpected cleavages. In these cases, the unknown peptides are further subjected to MS-MS for sequencing and a partial sequence consisting of a few amino acid forms a peptide tag is also used for the database search to match potential protein candidates. If no match is found at this stage, the identified peptide tag is subjected to a BLAST (Basic Local Alignment Search Tool) search which explores all the available databases to match for sequence homology. In contrast to the previous mentioned databases, BLAST seeks for local alignment in isolated regions of similarity with no mass constraint.[82] It is therefore very useful to detect remote sequence relationships and the presence of less occurring residues such as tryptophan, histidine and methionine increases the likehood of accurate match.

The grand aim of all science is to cover the greatest number of empirical facts by logical deduction from the smallest number of hypotheses or axioms.

Albert Einstein (1879-1955)

In Lincoln Barnett, "The Meaning of Einstein's New Theory," *Life*, 9 January 1950

Experimental Procedure

2.1 Overall Method

A specific strategy was developed to characterize the structure of cellulase enzymes and thus to enable the identification of post-translational modifications (Figure 2.1). This scheme is similar to the approach undertaken by Harrison et al. for the complete CBH I structural analysis,[25] except that more emphasis is placed on the use of mass spectrometry techniques. Individual cellulase enzymes were purified from crude extracts received from Iogen Corporation using ion-exchange chromatography and chromatofocusing. Three approaches were carried out to study the N- and O-linked glycans. First, papain proteolysis was used to separate the cellulase core from the linker and CBD domains, and the digests were subsequently purified by gel permeation chromatography (GPC). The purified linker peptides were analyzed by MALDI-TOF to establish the extent of O-linked glycan. In the second approach (N-linked glycan), individual cellulases were incubated with trypsin and the resulting peptides were analyzed by MALDI-TOF and LC-ESMS. Glycopeptides with oligosaccharides attached to putative N-linked sites were initially identified using front-end collision-induced dissociation mass spectrometry (front-end CID MS) and incubation with PNGase F was subsequently used to remove the high mannose glycans. The suspected glycopeptides were confirmed by tandem MS with product ion scan which allowed for partial peptide sequencing. In the third approach (N- and O-linked glycans), purified cellulases were introduced into the mass spectrometer via flow-injection analysis (FIA) to determine the microheterogeneity in glycosylation profile. Automated hydrazinolysis was used to chemically release both the N- and O-linked glycans. The released oligosaccharides were

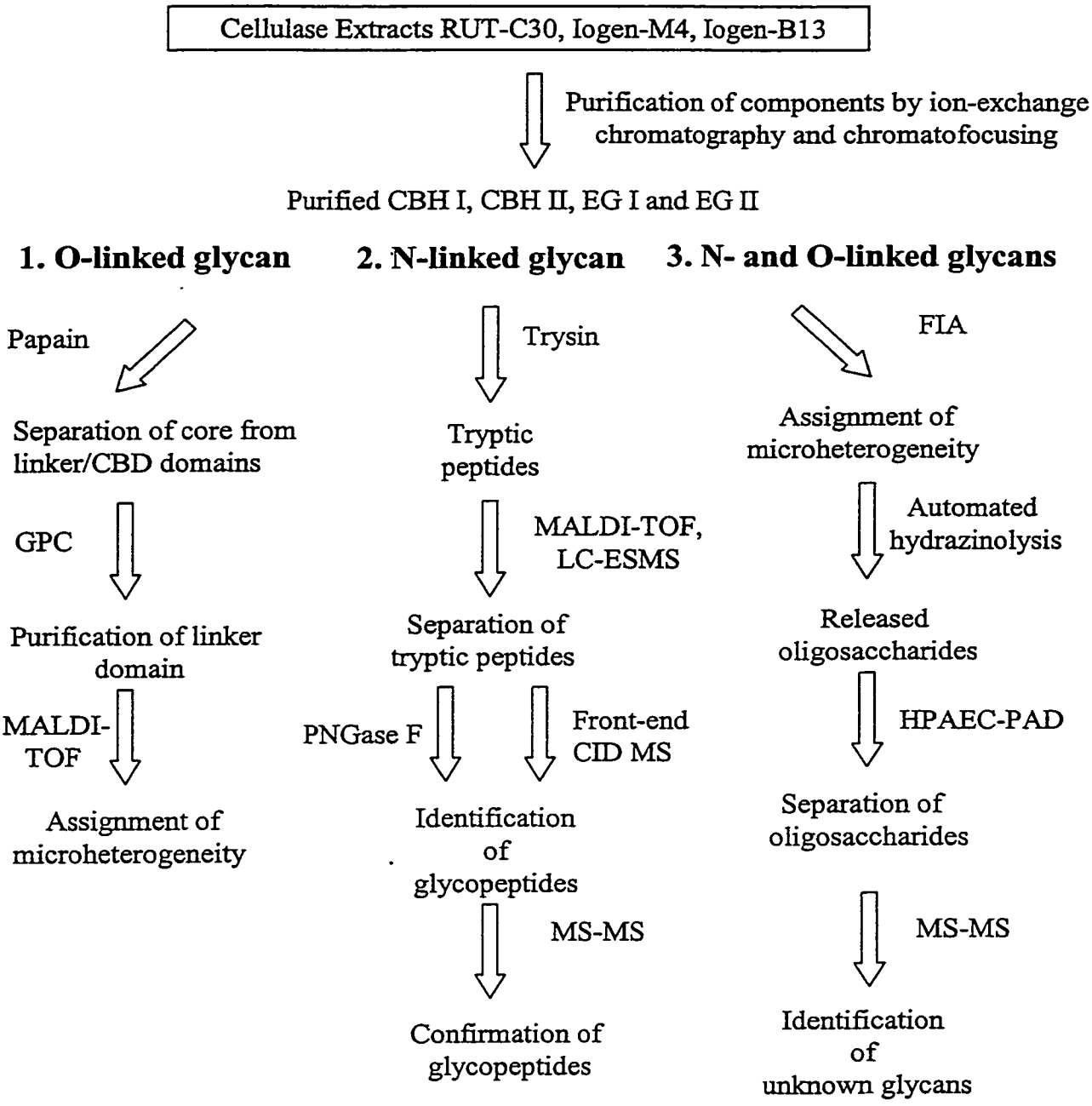


Figure 2.1 Flow chart showing the overall strategy used for the characterization of glycosylation in the purified cellulases from *T. reesei*.

separated by high performance anion-exchange chromatography with pulsed amperometric detection (HPAEC-PAD). Unidentified oligosaccharides were further subjected to tandem MS for structural characterization.

2.2 Reagents and Materials

All cellulase fermentation samples and purified cellulolytic enzyme standards were provided by Iogen Corporation (Ottawa, ON, Canada). For CBH I purification, the DEAE sepharose (diethylaminoethyl weak anion exchanger) was obtained from Amersham Pharmacia Biotech (Amersham Place, Buckinghamshire, UK) and packed into a poly-prep chromatography column purchased from Bio-Rad (Hercules, CA, USA). Protein concentration and desalting was achieved using Ultrafree-4 centrifugal filter units, Biomax-5 (5,000 Da MWCO) and Biomax-10 (10,000 Da MWCO) from Millipore (Bedford, MA, USA). Micro-dialysis was carried out using 2,000 Da and 5,000 Da molecular weight cut-off (MWCO) membranes from AmiKa Corporation (Columbia, MD, USA). Sodium phosphate monobasic $NaH_2PO_4(H_2O)$ and dibasic $Na_2HPO_4(7H_2O)$ were purchased from BDH (Toronto, ON, Canada). The DL-dithiothreitol and iodoacetamide were from Sigma (St. Louis, MO, USA) while β-mercaptoethanol and the ammonium bicarbonate were obtained from BDH (Toronto, ON, Canada) and from Fisher Scientific (New Jersey, NY, USA), respectively. Sequencing-grade trypsin was purchased from Promega (Madison, WI, USA), lyophilized papain powder was from Sigma (St. Louis, MO, USA), and the calf intestinal alkaline phosphatase (CIP), PNGase F and Endo H were from New England BioLabs (Beverly, MA, USA). Wide range pH 3-10 and narrow range pH 3-6 carrier ampholytes

were obtained from Beckman (Fullerton, CA, USA) and from SERVA Electrophoresis GmbH (Heidelberg, Germany), respectively.

For polybrene and neutral coating, the fused silica capillary (ID 50 μm or 76 μm x OD 360 μm) was purchased from Polymicro Technologies (Phoenix, AZ, USA). Hexadimethrine bromide (Polybrene®), ethylene glycol and poly(vinyl alcohol) with average M_w of 89,000 – 98,000 were obtained from Aldrich (Milwaukee, WI, USA). Ultra high pure (UHP) grade nitrogen gas was used for purging and pressurizing the coating solution through the capillary. A column rinse kit was used for capillary coating and was purchased from J & W Scientific (Folsom, California, USA).

2.3 Culture and Growth Conditions

T. reesei strains RUT-C30, Iogen-M4 and Iogen-B13 were grown in 10L fermentation vessels at 28°C for 5-6 days essentially as described by Mandels and Reese for optimal growth and production of cellulase by QM6a.[83] The growth medium typically contains: 1.4 g/L $(NH_4)_2SO_4$, 2 g/L KH_2PO_4, 0.3 g/L $CaCl_2$, 0.3 g/L $MgSO_4$, 5 mg/L $FeSO_4 \cdot 7H_2O$, 1.6 mg/L $MnSO_4 \cdot H_2O$, 1.4 g/L $ZnSO_4 \cdot 7H_2O$, 2g/L $CoCl_2$, 4.2% corn steep liquor adjusted to pH 4.0 or pH 5.0. The carbon source for the fermentations comprises 5 g/L glucose plus 10 g/L Solka flocor or other carbon source which can be sterilized separately as an aqueous solution at pH 2 to 7 and added to the remaining media.

2.4 Sample Preparation

Purification of cellulolytic components from culture filtrate was carried out using similar procedures reported by Pettersson et al.,[84] except chromatofocusing was incorporated to purify the endoglucanases.

2.4.1 CBH I Purification

The intact CBH I from fermentation samples produced by RUT-C30 and two derivative strains, Iogen-M4 and Iogen-B13, was fractionated in several steps using ion-exchange chromatography with diethylaminoethyl weak anion exchanger (DEAE) sepharose gel.[85] A low salt phosphate buffer (50 mM sodium phosphate, pH 6.0, 25 mM NaCl) was used for column equilibration. Approximately 10 mg of cellulase extract was loaded onto the column packed with 3 mL bed volume. After elution of the unretained components, CBH I was eluted in a phosphate high salt buffer (50 mM sodium phosphate, pH 6.0, 300 mM NaCl). The DEAE sepharose was then washed with five column volumes of 1.0 M NaOH, followed by ten column volumes of de-ionized water and stored in 20% ethanol. The enriched CBH I fraction was desalted and concentrated by using the Ultrafree-4 centrifugal filter unit, on a refrigerated swinging-bucket centrifuge (Sorvall RT6000B, Du Pont Instruments, Newtown, Connecticut, USA).

2.4.2 Purification of CBH I Papain Digestion Products

The proteolytic fragments obtained from papain digestion of CBH I (section 2.7.2) were separated using gel-filtration chromatography on a Pharmacia Superdex 75 HR 10/30 (high resolution, ID 10 mm x 30 cm in length) column as described previously.[86] The column was initially equilibrated with 50 mM ammonium carbonate pH 5.0, 150 mM NaCl and the flow-rate was maintained at 0.5 mL/min using a Pharmacia fast protein liquid chromatographic

(FPLC) system. Approximately 1 mg of CBH 1 papain digest in 200 μL final volume was introduced into the column via an injection port. The elution of individual components was monitored by an on-line UV detector with wavelength set at 280 nm. Initially, a sharp peak of CBH I core was eluted, followed by a broad peak of the combined CBH I linker and CBD domains of weaker absorbance. The collected components were concentrated and desalted by the Ultrafree-4 centrifugal filter unit, and subsequently diluted with deionized water to the desired concentration.

2.4.3 CBH II Purification

Purification of CBH II from RUT-C30 was carried out using two cation exchangers: methyl sulfonate (SP) sepharose followed by carboxymethyl (CM) sepharose. CBH I was initially removed following the procedure described in section 2.3.1, except that the starting material was increased to approximately 50 mg of protein and 5 mL of DEAE sepharose volume bed was used. The unretained fraction enriched in CBH II and endoglucanases was collected, concentrated and desalted with Biomax-5 filter unit. This fraction was loaded onto a column packed with 5 mL volume bed of S sepharose previously equilibrated with 3 mM acetate buffer, pH 4.0. Detection of cellulolytic components was monitored on-line by a UV detector at 280 nm wavelength. After elution of the unretained components, the buffer was changed to a high salt acetate buffer (50 mM sodium acetate, pH 4.0, 1M NaCl) to elute the CBH II enriched fraction. This fraction was again concentrated and desalted using the Biomax-5 filter unit, and was further purified using CM Sepharose (2 mL volume bed) equilibrated with 6 mM sodium acetate, pH 4. The unretained components were first removed, followed by elution of CBH II using 6 mM sodium acetate buffer with 40 mM NaCl, pH 4. Similarly, the collected CBH II fraction was concentrated and desalted. Both

anion-exchangers were washed and stored using the same procedure as described in section 2.3.1 for DEAE sepharose.

2.4.4 Chromatofocusing Purification of EG I and EG II

Chromatofocusing was performed using the Pharmacia Polybuffer exchanger PBE 94 and Polybuffer 74 (pH 4-7) adjusted to pH 4 as elution buffer. The PBE 94 was mixed with small amount of starting buffer (0.025 imidazole-HCl, pH 7.4), degassed and slowly introduced into a FPLC column (20 cm length, 4 mL volume capacity). Prior to the purification, the PBE 94 was equilibrated with 10-15 columns of the starting buffer flowing at 0.2 mL/min, followed by 1 column volume of the elution buffer. The unretained fraction from S Sepharose (enriched in endoglucanases) was slowly pumped into the column and was separated into different bands of components of varying isoelectric points (low to high pI) resulting from the pH gradient created by the Polybuffer 74. Collected fractions were concentrated using Biomax-10 filter units, which was necessary to remove the ampholyte present in the Polybuffer.

2.5 Capillary Coatings

The inner surface of bare fused-silica capillary was coated with polybrene and poly(vinyl alcohol) in CZE and CIEF experiments, respectively.

2.5.1 Polybrene

This cationic coating procedure is based on the same method reported previously.[87,88] All the rinsing was carried out using 20 psi in the CE system. The fused-silica capillary (ID 50 μm x OD 350 μm) was initially rinsed with 1M NaOH for 15 min, followed by the same deionized water for 15 min. Polybrene solution consisting of 5% (w/v) dimethrine bromide,

2% (v/v) ethylene glycol was passed through the column for 20 min. The column was then equilibrated for 15 min with the separation buffer prior to the CZE analysis. This dynamic coating was re-conditioned between subsequent runs with 0.1 M NaOH for 2 min, followed by deionized water (3 min), polybrene (4 min) and separation buffer (5 min).

2.5.2 Poly(vinyl alcohol)

This neutral coating procedure is a modified version reported previously.[41,42] A 1m piece of fused silica capillary (ID 50 μm x OD 360 μm or ID 76 μm x OD 360 μm) was rinsed with deionized water for 20 min with nitrogen gas purged at 40 psi. A 5% (v/w) PVA solution in deionized water was prepared by heating in a hot water bath (close to boiling point) for 10 min and the solution was filtered through a Millpore Ultrafree-MC 0.45 μm filter unit. This solution was then passed through the capillary with nitrogen gas at 40 psi for 20 min. To prevent no blockage, the capillary outlet is immersed in a beaker filled with water.

The excess PVA solution was blown out of the capillary outlet with nitrogen pressure. At this point, the nitrogen gas was reduced to a steady flow of 20 psi and left purging for 30 min. This step is crucial especially for capillary with 50 μm inner diameter to initiate a smooth PVA coating prior to the capillary subjected to heating. The capillary was then transferred to the GC oven to allow for the polymerization of PVA at 145°C for 3 hours, with nitrogen gas purged at 30 psi.

The capillary was left to cool to room temperature in the GC oven, after which was removed and both ends were protected with GC septa. To make a window for CIEF-UV detection, a section (~0.3 cm) located approximately 7cm from the capillary outlet was marked and etched with concentrated sulfuric acid heated to 150°C. This section of

polyimide coating would expand, however not enough to remove the polyimide, and it is necessary to gently scrape the peeled section with a scalpel. Through repeated ectching and scraping, this section of polyimide would be easily peeled off by wiping gently with a Kimwipe partly soak in methanol.

2.6 CE

Capillary electrophoresis experiments were performed using either capillary zone electrophoresis (CZE) or capillary isoelectric focusing (CIEF) modes with UV detection.

2.6.1 CZE-UV

CZE-UV experiments were conducted on a Beckman P/ACE 5000 system (Fullerton, CA, USA). Separation was achieved on a 97 cm polybrene coated capillary (ID 50 μm x OD 350 μm) with voltage maintained at 20 kV. The separation buffer was 1.0 M formic acid. Samples were injected using pressure (0.5 PSI) for 20 s, which is equivalent to approximately 15 nL. The detector was set at 200 nm.

2.6.2 CIEF-UV

All CIEF experiments were conducted on the same CE instrument. Separations were carried out using a Beckman eCAP neutral coated capillary (ID 50 μm x OD 365 μm, 27 cm in length). The cellulase standards or cellulase extracts were dissolved to their final concentration in an aqueous solution consisting of 3% blended carrier ampholyte (Beckman ampholyte:Servalyt 3:1, v/v). The anolyte (inlet) was 10 mM phosphoric acid and a catholyte of 20 mM NaOH was used at the capillary outlet. A voltage of 13.5 kV was used to focus the analyte band for a duration of 10 min followed by cathodic mobilization using a buffer of methanol:water:acetic acid (50:49:1 v/v/v).[89] During focusing and mobilization, the

voltage was maintained at a field strength of 13.5 kV (500 V/cm). The detector was set at 280 nm.

2.7 Protein Digestion

The purified cellulases were subjected to selected proteolytic digestion including trypsin, papain, alkaline phosphatase, PNGase F and Endo H enzymes described in the following paragraphs.

2.7.1 Trypsin

Tryptic digests (typically 200-500 μg of cellulase protein) were performed on reduced and alkylated proteins previously denatured in 50 μL of 6.0 M guanidine/HCl, 0.1 M Tris-HCl, 1 mM EDTA, pH 8.5. Argon gas was gently blown over the solution surface for 10 min. Reduction was carried out by the addition of 50 μL of 4 mM DL-dithiothreitol, and the solution was then incubated for 2 hours at 37°C. Alkylation was performed by adding 40 μL of 50 mM iodoacetamide dropwise to the mixture, followed by incubation in the darkness for 1 hour at 37°C. The excess iodoacetamide was quenched by adding 1 μL of β-mercaptoethanol and subsequently dialysed against 500 mL of 50 mM ammonium bicarbonate for a minimum of 24 hours with periodic change of dialysis buffer. Sample solution was then lyophilized and redissolved in 0.1 M ammonium bicarbonate, 0.1 mM calcium chloride, pH 8. Tryptic digestion was carried out at 37°C for 12 hours using a substrate-to-enzyme ratio of 25:1 (w/w).

2.7.2 Papain

The papain digestion of cellulase enzymes was conducted according to a procedure described previously.[25] Briefly, the lyophilized papain powder was re-suspended in 200 mM

K_2HPO_4, 2 mM EDTA, 5 mM L-cysteine, pH 7.0 to a concentration of 2 mg/mL. An aliquot of 8 μL of this enzyme suspension was added to 200 μL of 5 mg/mL CBH I dissolved in 100 mM ammonium acetate, pH 5 equivalent to a substrate-to-enzyme ratio of 60:1 (w/w). The mixture was incubated at 37°C for 16 hours. The resulting digest was purified using gel filtration chromatography described in section 2.4. The substrate-to-enzyme ratio for CBH II was adjusted to 10:1 (w/w) while for EG I and EG II was 1:5 (w/w).

2.7.3 Alkaline Phosphatase

The calf intestinal alkaline phosphatase (CIP) incubation was performed on purified CBH I linker and CBD domains from RUT-C30. Approximately 100 μg of CBH I linker plus CBD was dissolved in 20 μL of 1X NE Buffer 3 also supplied by New England BioLabs, followed by the addition of 10 units of CIP. This 1X NEBuffer 3 consists of 50 mM Tris-HCl, 10 mM $MgCl_2$, 100 mM NaCl, and 1 mM DTT. One unit of CIP activity is defined as the amount of enzyme required to catalyze 1 μmole of *p*-nitrophenylphosphate to *p*-nitrophenol in 1 minute in a reaction volume of 1 mL at 37°C. The mixture was incubated at 37°C and small aliquot was pipetted out periodically to monitor the dephosphorylation using MALDI-TOF mass spectral analysis. No further structural change was detected after 3 hr.

2.7.4. N-Glycosidase F (PNGase F)

All PNGase F incubations were performed on approximately 20 μg of the original tryptic digests of CBH I to which were added 500 U endoglycosidase corresponding to an equivalent of substrate-to-enzyme ratio of 70:1 (w/w). The digestion was allowed to proceed for 24 hours at 37°C in 50 mM ammonium bicarbonate, pH 8.[90] An unit of PNGase F is defined as the amount of enzyme required to remove more than 95% of carbohydrate from 1 0 μg of denatured RNase B in 1 hour in a 10 μL reaction volume at 37°C.

2.7.5 Endoglycosidase H (Endo H)

The Endo H enzymatic activity was performed on CBH I and selected papain digest from RUT-C30. Approximately 50 μg of intact CBH I was freeze-dried, and re-dissolved in 45 μL of 50 mM ammonium acetate at pH = 5.5 and 5 μL of 10X G5 buffer supplied by New England BioLabs. This 10X G5 buffer consists of 0.5 M sodium citrate, pH 5.5. An aliquot of 5 μL of Endo H, equivalent to 2500 units, was added and the mixture was incubated at 37°C for 5 hours. An unit of Endo H is defined as the amount of enzyme required to remove more than 95% of carbohydrate from 10 μg of denatured RNase B in 1 hour in a 10 μL reaction volume at 37°C.

2.8 Mass Spectrometry

Structure characterization of cellulase enzymes were primarily carried out on a MALDI-TOF instrument and a Q-TOF mass spectrometer equipped with electrospray ionization.

2.8.1 MALDI-TOF

All MALDI-TOF mass spectra were acquired on a Voyager Elite STR Biospectrometry Workstation (PerSeptive Biosystems, Framingham, MA) equipped with delayed extraction technology. This instrument is capable of operating in linear or reflector mode. Sinapinic acid was the matrix solution for protein analysis and it was prepared as 10 mg/mL in a mixture of methanol:acetonitrile:water in equal proportion. For peptide detection, the matrix solution was 2,5-dihydroxybenzoic acid (2,5-DHB) and prepared as 0.2 M solution in 50:50 methanol:water. A nitrogen laser operated at 337 nm was used to initiate the ionization. All samples and matrix solution were spotted on polished stainless steel MALDI standard 100-well sample plates and air-dried to allow for crystal formation.

For the intact cellulase and papain digest analyses, a 1 μL of sample with typical concentration ranging from 1-5 mg/mL was mixed with 10 μL of sinapinic acid (SA) matrix solution, and an aliquot of 0.5 μL was deposited on the MALDI plate. These sample spots were analyzed in linear positive ion mode with typical conditions of 25 kV accelerating voltage, 91% grid voltage, 0.3% guide wire voltage and a delay time varying from 200 – 300 ns. External mass calibration was carried out using the singly-protonated ions of insulin (m/z 5,734.59) and apo-myoglobin (m/z 16,952.56) for papain digest, while the singly- and doubly-protonated ions of bovine serum albumin (BSA) with average masses of m/z 66,431 and m/z 33,216, respectively for intact cellulase analyses. For tryptic peptide analysis, an aliquot of 0.5 μL of 2,5-DHB matrix solution was first applied on the MALDI plate, followed by spotting 0.5 μL of sample solution typical of 1 mg/mL, and another 0.5 μL of matrix solution. These sample spots were analyzed in reflectron positive ion mode with 20 kV accelerating voltage, 72% gride voltage, 200 ns delay time and the guide wire voltage varying from 0.02 – 0.05%, depending on the resolution requirement. External calibration was carried out using the monoisotopic masses of protonated molecular ions of des-Arg^1-Bradykinin (m/z 904.4681) and PTH 1-31 Amide (m/z 3716.9845).

2.8.2 ESMS

ESMS experiments were acquired using either a Micromass Q-Tof (Manchester, UK) or a PE/Sciex Q-Star (Thornhill, ON, Canada) hybrid quadrupole/time-of-flight instrument. Flow injection analyses were obtained by dissolving the CBH I preparation in 50 % acetonitrile (0.2 % acetic acid). Approximately 1 μL of this solution was injected into a stream of 50 % acetonitrile (0.2 % acetic acid) introduced at a flow rate of 5 μL/min to the mass spectrometer.

2.8.3 LC-ESMS

An HPLC HP1100 (Hewlett Packard, Palo Alto, CA) was used for on-line LC-ESMS experiments. Chromatographic separations were achieved on a 15 cm x 0.32 mm PepMap capillary column (LC Packings, San Francisco CA) using a linear gradient elution of 5-95 % acetonitrile (0.2 % HCOOH) in 35 min. A 100:1 flow splitter was mounted before the injector such that only 3-4 μL/min flow rate was introduced in the capillary column. The sample was manually injected via a Rheodyne six-port valve (Rohnert Park, California, USA) fitted with a 5 μL sample loop. For experiments involving fraction collection, a Zorbax C8 reversed-phase column (2.1mm x 15 cm) was used with flow-rate at 0.2 mL/min. Conventional mass spectra were obtained by operating the quadrupole in a RF-only mode while a pusher electrode was pulsed to transfer all ions to the time-of-flight analyzer. Mass spectra were acquired using stepped orifice-voltage scanning similar to that described previously.[91,92] In the present work, the sampling orifice voltage was maintained at 100 V during low mass scanning (m/z 150 to 400) and 30 V during high mass scanning (m/z 400 to 2000). In MS-MS experiments, glycopeptide precursor ions were selected by the first quadrupole while a pusher electrode was pulsed to transfer fragment ions formed in the R.F.-only hexapole cell to the time-of-flight analyzer. Mass spectral resolution was typically 4,000-5,000. A scan duration of 1 s and 2 s was set for conventional and MS-MS mass spectral acquisition, respectively. Collisional activation was performed using argon collision gas with a typical 25 V offset between the dc voltage of the entrance quadrupole and the rf-only hexapole cell. Data were acquired and processed in the MassLynx Window NT based data system.

2.8.4 Nanoelectrospray

Nanoelectrospray emitters were purchased from Micromass (Manchester, U.K.) and Protana (Odense, Denmark). These emitters were made from pulled borosilicate tubing and were gold coated on the outside to maintain electrical conductivity for sample ionization. Typically, the emitters were loaded with 1-2 μL of tryptic peptides and slowly introduced into the Micromass Q-Tof instrument at a flow-rate of ~ 20 nL/min. An ionization voltage of 1000 – 1200 V was generally applied.

2.8.5 CIEF-MS

The interface for CIEF-MS developed was similar method reported by Lee et al. and was briefly illustrated in Figures 2.2a-b.[93,94,95] A Sciex API 3000 triple quadrupole mass spectrometer equipped with an electrospray source was used. Protein standards and cellulase samples were dissolved in 1% Beckman ampholyte and were filled into a PVA coated capillary (ID 50 μm x OD 180 μm, 30 cm in length) using a gas-tight syringe. During focusing, the tapered capillary outlet was immersed into an eppendorf vial (1.5 mL) containing 0.05 M ammonium hydroxide (catholyte) close to the microelectrospray probe (Figure 2.2a). The capillary inlet was immersed into another eppendorf vial filled with 1.0 % formic acid (anolyte) and was positioned at the same height as the outlet. A 15 kV focusing voltage generated from a Spellman CZE 1000R High Voltage power supply (Hauppauge, NY, USA) was applied for 10 min. When focusing was completed, the power supply was turned off and the catholyte removed. The microelectrospray probe was carefully turned toward the orifice of the mass spectrometer with the tapered end of the capillary barely protruding from the electropray needle (Figure 2.2b).

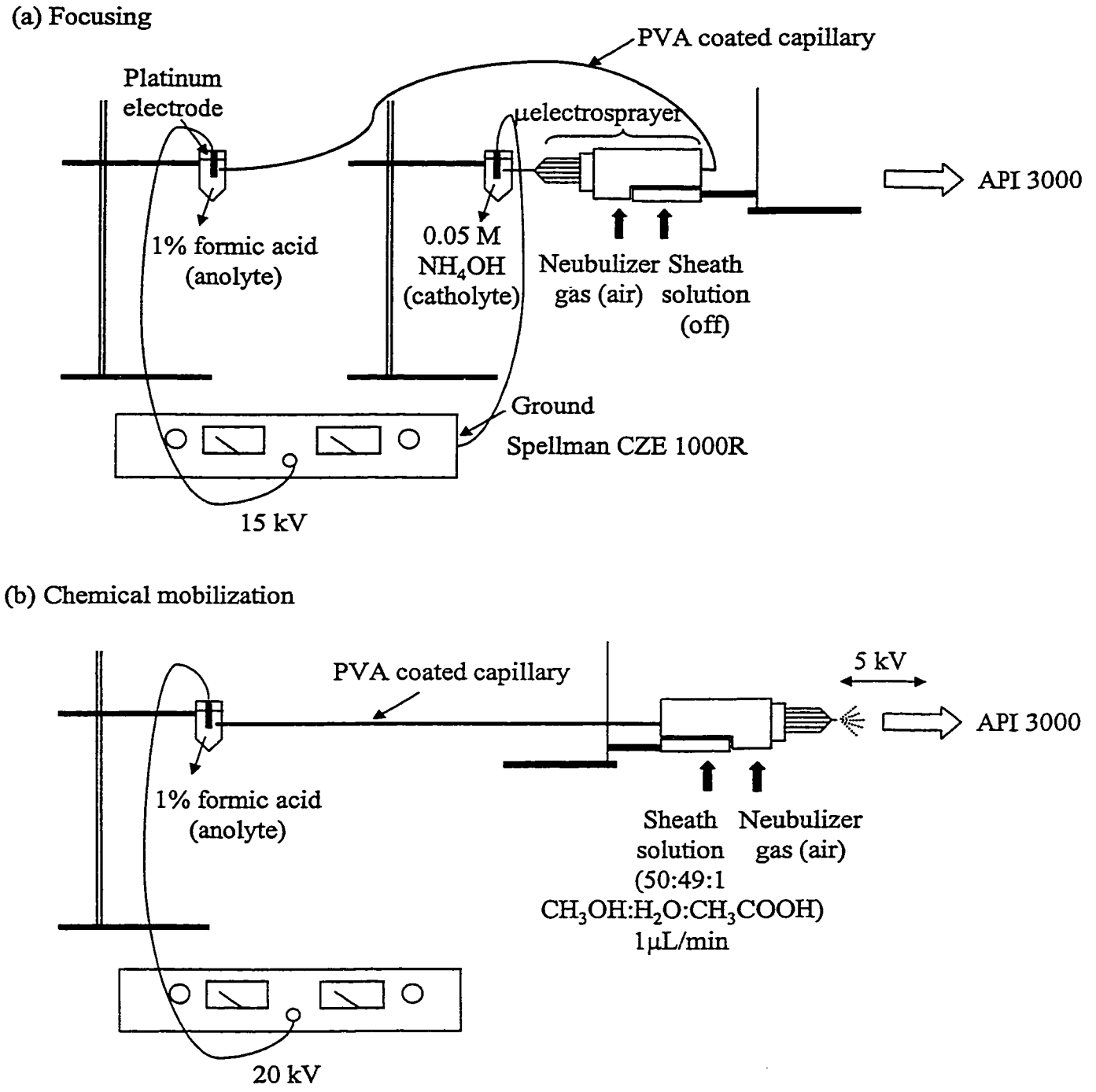


Figure 2.2 Schematic representation of CIEF-MS interface. (a) Focusing of analytes; (b) mobilization by replacing the catholyte with a solution of 50:49:1 CH_3OH:H_2O:CH_3COOH (v/v/v).

Mobilization was initiated by introducing a sheath liquid comprising of methanol:acetonitrile:water (50:49:1 v/v/v) delivered from a syringe pump at a flow-rate of 1 μL/min. The power supply from Spellman was immediately turned on with voltage increased to 20 kV to counteract the 5 kV electrospray ionization voltage. The orifice-to-skimmer voltage in API 3000 for protein analysis and the ring voltage was set at 100 V and 350 V, respectively. Typical mass spectra were acquired using dwell times of 3 ms per 1 Da with a scanning mass range of m/z 800 – 2,000. Proteins appeared in order of their isoelectric points starting from the most basic ones.

2.9 Automated Hydrazinolysis

Pools of glycans were released from 2 mg of each glycoprotein using the Glycoprep 1000 glycan release and preparation system (Oxford GlycoSciences, U.K.) for automated hydrazinolysis. Each glycoprotein was processed using a N+O program (95°C, 5 h.) mode for optimal release and isolation of N- and O-linked oligosaccharides. At the end of each processing run, the intact and unreduced pools of glycans were immediately filtered through a 0.2 μm nylon membrane and dried by centrifugal vacuum evaporation at room temperature.

2.10 High Performance Anion-Exchange Chromatography with Pulsed Amperometric Detection (HPAEC-PAD)

The dried glycans released by automated hydrazinolysis were dissolved in deionized water and aliquots were analyzed by HPAEC-PAD using a DX-300 chromatograph (Dionex Corporation, Sunnyvale, CA) equipped with a PA-100 analytical column (4 x 250 mm). Chromatographic conditions typically used a linear gradient of 5-230 mM sodium acetate

of the glycan pool was injected on the column. All data processing and peak integration was performed using the AI-450 chromatography software (Dionex). To further determine the structure of released oligosaccharides, a carbohydrate membrane desalter (CMD) from Dionex was introduced after the PAD detection cell thereby allowing direct fraction collection. The CMD used 0.15 M trifluoroacetic acid (TFA) to generate H^+ ions required for cation exchange. The desalted fractions were subjected to mass spectral analysis.

2.11 Database Searching

All the amino acid sequence information of *T. reesei* was obtained from Swiss-Prot database (http://www.expasy.ch/sprot). The accession numbers for CBH I, CBH II, EG I and EG II are P00725, P07987, P07981 and P07982, respectively. Theoretical tryptic digest was initially performed on the candidate proteins by using the following criteria: the selected enzyme was trypsin, cysteines were carboxylamidomethylated, methionines were not oxidized, the maximum number of missed cleavage sites was set to one and the monoisotoptic masses of protonated peptide $[M+H]^+$ were displayed. These theoretical masses were then compared to the y-type fragment ions obtained from MS-MS.

Unidentified sequences from MS-MS analyses were further examined using Peptidesearch database developed by EMBL (http://www.rmann.embl-heidelberg.de/Services/PeptideSearch/) or ProteinProspector by UCSF (http://prospector.ucsf.edu/). Typically, the peptide sequence tag approach was used where a partial amino acid sequence together with the molecular mass of the peptide were entered to match for potential proteins available in the database. The searching parameters were limited to tryptic peptides with all cysteine residues converted to carbamidomethylated cysteine. A protein mass range from 0-300 kDa was used, the peptide mass accuracy was less than 0.5 Da and all origins of species were included in the search.

It is a capital mistake to theorize before one has data.

Sir Arthur Conan Doyle (1859-1930)

A Scandal in Bohemia in The Adventures of Sherlock Holmes, 1894

Capillary Electrophoresis Analysis of Cellulase Extracts from RUT-C30 and Two Mutant Strains from *T. reesei*

3.1 Introduction

In order to carry out routine analysis of crude cellulase samples obtained at different fermentation stages, it is necessary to develop a rapid and reliable separation method. Capillary electrophoresis is commonly applied for the separation of bio-molecules and this technique was investigated in the present study. Both capillary zone electrophoresis (CZE) and capillary isoelectric focusing (CIEF) were evaluated as potential screening tools. The latter techniques has the unique characteristic of focusing the analyte of interest according to their isoelectric points (pI) which turns out to be a valuable method for cellulase analyses. Compared to the conventional slab IEF gel, CIEF has the added advantages both in terms of speed and resolution. The coupling of these CE techniques to mass spectrometry offers a second dimension of separation and thus further enhances the overall analytical capability.

Crude cellulase fermentation samples from *T. reesei* RUT-C30 and two mutant strains, Iogen-M4 and Iogen-B13, were provided by Iogen. These crude samples include two cellobiohydrolases (CBH I, II), at least three endoglucanases (EG I, II and III) and small amount of ß-glucosidase. Except for EG III, the cellulolytic enzymes are comprised of a cellulose-binding domain (CBD), a flexible linker peptide and a catalytic domain (core). The putative N-linked glycosylation sites on asparagine are assigned according to the consensus sequence Asn-X-Ser/Thr, where X is any amino acid except proline.[96] The amino acid sequence for the major CBHs and EGs are listed below in Figures 3.1a-e and their structural domains and characteristics are summarized in Table 3.1.

a) CBH I

pESACTLQSET HPPLTWQKCS SGGTCTQQTG SVVIDANWRW THATNSSTNC[50]
YDGNTWSSTL CPDNETCAKN CCLDGAAYAS TYGVTTSGNS LSIGFVTQSA[100]
QKNVGARLYL MASDTTYQEF TLLGNEFSFD VDVSQLPCGL NGALYFVSMD[150]
ADGGVSKYPT NTAGAKYGTG YCDSQCPRDL KFINGQANVE GWEPSSNNAN[200]
TGIGGHGSCC SEMDIWEANS ISEALTPHPC TTVGQEICEG DGCGGTYSDN[250]
RYGGTCDPDG CDWNPYRLGN TSFYGPGSSF TLDTTKKLTV VTQFETSGAI[300]
NRYYVQNGVT FQQPNAELGS YSGNELNDDY CTAEEAEFGG SSFSDKGGLT[350]
QFKKATSGGM VLVMSLWDDY YANMLWLDST YPTNETSSTP GAVRGSCSTS[400]
SGVPAQVESQ SPNAKVTFSN IKFGPIGSTG NPSGGN*PPGG* *NPPGTTTTRR*[450]
PATTTGSSPG ***PTQSHYGQCG*** ***GIGYSGPTVC*** ***ASGTTCQVLN*** ***PYYSQCL***[497]

b) CBH II

pEACSSVWGQC **GGQNWSGPTC** **CASGSTCVYS** **NDYYSQCLPG** *AASSSSSTRA*[50]
ASTTSRVSPT *TSRSSSATPP* *PGSTTTRVPP* *VG*SGTATYSG NPFVGVTPWA[100]
NAYYASEVSS LAIPSLTGAM ATAAAAVAKV PSFMWLDTLD KTPLMEQTLA[150]
DIRTANKNGG NYAGQFVVYD LPDRDCAALA SNGEYSIADG GVAKYKNYID[200]
TIRQIVVEYS DIRTLLVIEP DSLANLVTNL GTPKCANAQS AYLECINYAV[250]
TQLNLPNVAM YLDAGHAGWL GWPANQDPAA QLFANVYKNA SSPRALRGLA[300]
TNVANYNGWN ITSPPSYTQG NAVYNEKLYI HAIGPLLANH GWSNAFFITD[350]
QGRSGKQPTG QQQWGDWCNV IGTGFGIRPS ANTGDSLLDS FVWVKPGGEC[400]
DGTSDSSAPR FDSHCALPDA LQPAPQAGAW FQAYFVQLLT NANPSFL[447]

c) EG I

pEQPGTSTPEV HPKLTTYKCT KSGGCVAQDT SVVLDWNYRW MHDANYNSCT[50]
VNGGVNTTLC PDEATCGKNC FIEGVDYAAS GVTTSGSSLT MNQYMPSSSG[100]
GYSSVSPRLY LLDSDGEYVM LKLNGQELSF DVDLSALPCG ENGSLYLSQM[150]
DENGGANQYN TAGANYGSGY CDAQCPVQTW RNGTLNTSHQ GFCCNEMDIL[200]
EGNSRANALT PHSCTATACD SAGCGFNPYG SGYKSYYGPG DTVDTSKTFT[250]
IITQFNTDNG SPSGNLVSIT RKYQQNGVDI PSAQPGGDTI SSCPSASAYG[300]
GLATMGKALS SGMVLVFSIW NDNSQYMNWL DSGNAGPCSS TEGNPSNILA[350]
NNPNTHVVFS NIRWGDIGST TNSTA*PPPPP* *ASSTTFSTTR* *RSSTTSSSPS*[400]
***C*TQTHWGQCG** **GIGYSGCKTC** **TSGTTCQYSN** **DYYSQCL**[437]

d) EG II

pEQTVWGQCGG **IGWSGPTNCA** **PGSACSTLNP** **YYAQCI***PGAT* *TITTSTRPPS*[50]
GPTTTTRATS *TSSSTPPTSS* GVRFAGVNIA GFDFGCTTDG TCVTSKVYPP[100]
LKNFTGSNNY PDGIGQMQHF VNEDGMTIFR LPVGWQYLVN NNLGGNLDST[150]
SISKYDQLVQ GCLSLGAYCI VDIHNYARWN GGIIGQGGPT NAQFTSLWSQ[200]
LASKYASQSR VWFGIMNEPH DVNINTWAAT VQEVVTAIRN AGATSQFISL[250]
PGNDWQSAGA FISDGSAAAL SQVTNPDGST TNLIFDVHKY LDSDNSGTHA[300]
ECTTNNIDGA FSPLATWLRQ NNRQAILTET GGGNVQSCIQ DMCQQIQYLN[350]
QNSDVYLGYV GWGAGSFDST YVLTETPTSS GNSWTDTSLV SSCLARK[397]

e) EG III

pETSCDQWATF TGNGYTVSNN LWGASAGSGF GCVTAVSLSG GASWHADWQW[50]
SGGQNNVKSY QNSQIAIPQK RTVNSISSMP TTASWSYSGS NIRANVAYDL[100]
FTAANPNHVT YSGDYELMIW LGKYGDIGPI GSSQGTVNVG GQSWTLYYGY[150]
NGAMQVYSFV AQTNTTNYSG DVKNFFNYLR DNKGYNAAGQ YVLSYQFGTE[200]
PFTGSGTLNV ASWTASIN[218]

Figure 3.1 Amino acid sequence of cellulolytic enzymes from *T. reesei*. (a) CBH I; (b) CBH II; (c) EG I; (d) EG II and (e) EG III. The CBD domain is highlighted in bold type, the linker domain in italics and the rest of the sequence constitutes the catalytic core. Putative N-glycosylation sites are circled. See Table 3.1 for other pertinent information (pE represents pyroglutamic acid).

Table 3.1 Summary of structural features in major cellulolytic enzymes from *T. reesei*

			Amino acid sequence			
Cellulases	Molecular mass (Da)	No. of disulfide bonds	Putative N-glycan sites	Core	Linker	CBD
CBH I[a]	52205.8	12	45, 64, 270, 384	1-436	437-461	462-497
CBH II[a]	47162.4	4	14, 289, 310	83-447	42-82	1-41
EG I[a]	46011.3	2	56, 142,182, 186, 259, 372	1-375	376-401	402-437
EG II[a]	42164.5	2	103	71-397	37-70	1-36
EG III[b]	23463.6	n.a.	164, 167	1-218	N/A	N/A

[a] information obtained from Swiss-Prot database (http://www.expasy.ch/sprot)
[b] information obtained from Ward et al.[97]
n.a. = not available; N/A = not applicable

3.2 Capillary Zone Electrophoresis with UV Detection (CZE-UV)

Separation of cellulase standards was initially carried out using the most common CE mode – capillary zone electrophoresis (Figures 3.2a-e). These cellulases were originally purified from strains Iogen-M4. A high concentration buffer solution (1.0 M formic acid) and a capillary 97 cm in length were necessary to enhance resolution of the separation individual components (Figure 3.2e). The initial drastic decrease in absorbance corresponds to the EOF. Although CZE has been reported to separate different glycoforms in carbohydrate analysis,[98] the microheterogeneity of cellulases was not successfully resolved under the current conditions.

Interestingly, both CBH II and EG I displayed two peaks of similar intensity (Figures 3.2b,c), but of different electrophoretic mobilities. Such occurrence may reflect the presence of two conformers for each glycoproteins. In CZE, separation is achieved based on the charge-to-size ratio. The polarity is reversed in experiments involving polybrene coated capillaries. All positive ions are attracted to the cathode located at the capillary inlet side, but are migrating towards the detector window due to overwhelming EOF (see introduction section 1.3.1 for more details). Cations with fewer positive charges or higher masses (smaller electrophoretic mobility) have a shorter migration time. Thus, the first components in Figures 3.2b and c would have a smaller charge-to-mass ratio than the second one. The separation of a four cellulase mixture is shown in Figure 3.2e and the migration time observed correlates well with those individual components.

Crude fermentation samples from two mutant strains of RUT-C30 were also separated by CZE (Figures 3.3a,b). The cellulolytic components were identified based on the migration times appearing on each CZE electropherograms. The two peaks appearing in both CBH II

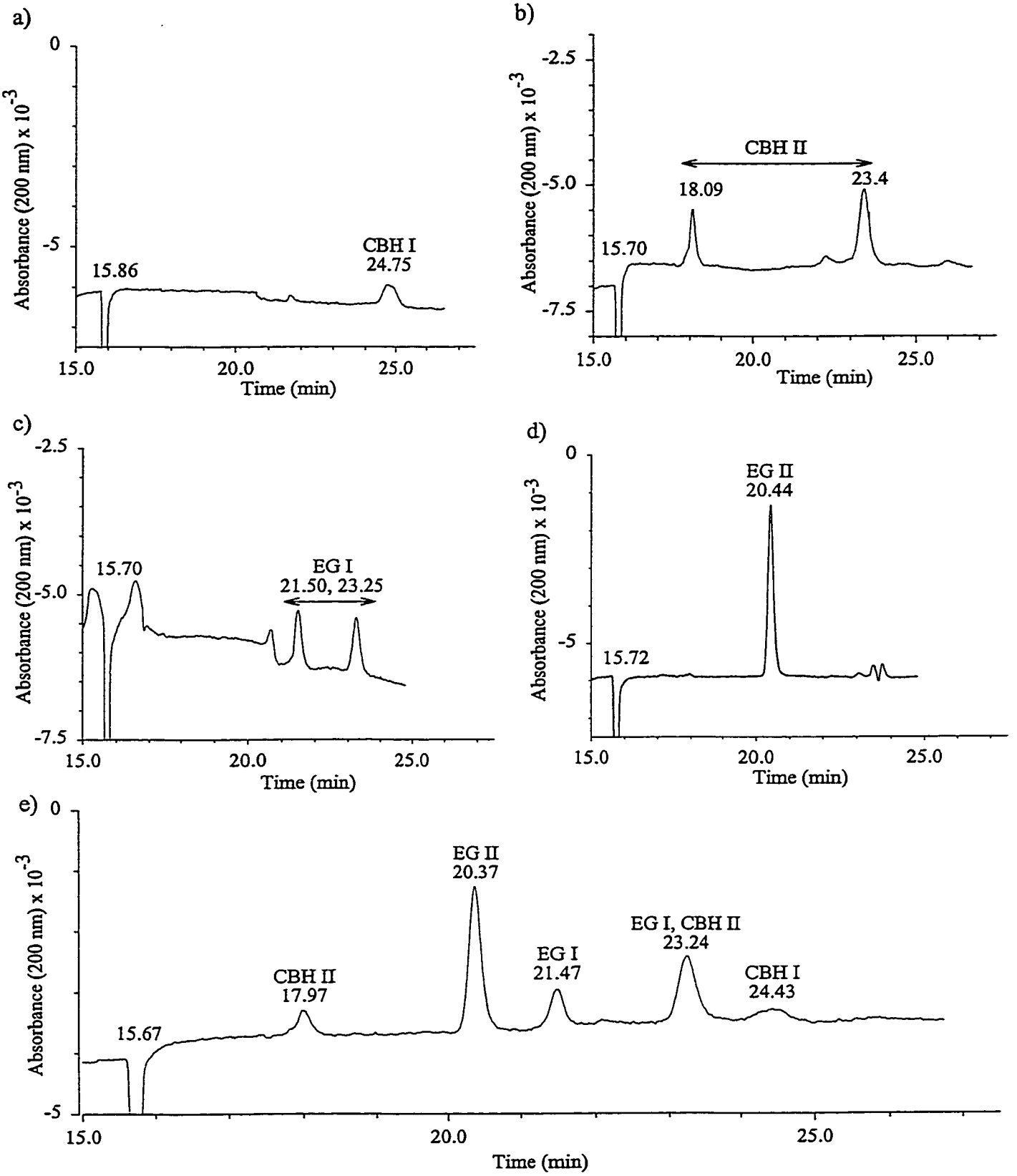


Figure 3.2 CZE analysis of cellulase enzyme standards purified from Iogen-M4. (a) CBH I (183.5 μg/mL); (b) CBH II (212.5 μg/mL); (c) EG I (227.5 μg/mL); (d) EG II (3 23.5 μg/mL); (e) mixture (same concentrations as individual standards). Conditions as described in section 2.6.1 with a 97 cm polybrene coated capillary, 1.0M formic acid as separation buffer and a voltage of 20 kV was used.

and EG I enzymes were matched and tentatively assigned. In both cases, CBH I represents the major cellulase enzymes accounting for more than 50% of total secreted cellulases based on peak area. In Iogen-B13, the peak at 23.43 min corresponds to the combined EG I and CBH II components and it is significantly more abundant in this strain compared to that of Iogen-M4 (23.43 min). Extra shoulder peaks were observed in CBH I, EG I and EG II from Iogen-B13 strain (Figure 3.3b). It is not known if these peaks were caused by buffer salts in the crude fermentation samples or glycoform variations between strains. Despite the fact that CZE offers a reasonable separation method for crude cellulase extracts, the lack of baseline resolution of certain components in Iogen-B13 and the co-migration of EG 1 and CBH II failed to provide a valid method for quantitative analysis.

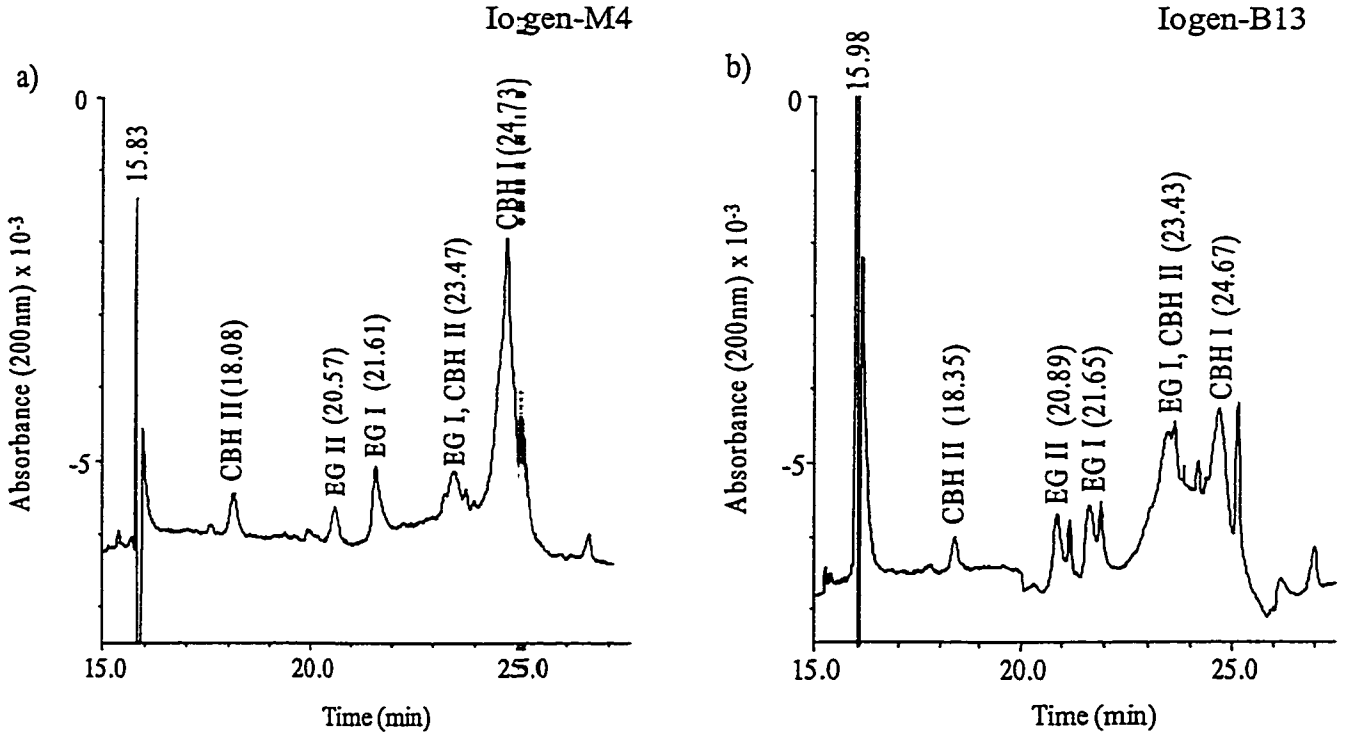


Figure 3.3 CZE analysis of crude fermentation samples from *T. reesei*. (a) Iogen-M4 [3.3 mg/mL] and (b) Iogen-B13 [2.12 mg/mL]. Same conditions as described in Figure 3.2.

3.3 Capillary Isoelectric Focusing with UV Detection (CIEF-UV)

The main cellulases from *T. reesei* are acidic glycoproteins (pI < 5.5) and this property is exploited advantageously for their purification using ion-exchange chromatography.[51,83] The most abundant cellobiohydrolases and endoglucanases have distinct pI values which make them ideal analytes for separation based on isoelectric focusing. All experiments reported in this section were carried out using the Beckman neutral coated capillary. The CIEF electropherograms for individual cellulase components separated on a wide range pH 3-10 are shown in Figures 3.4a-e. These components were concentrated and focused into narrow peaks. The enlarged peaks are shown as inset and in most cases, shoulder peaks were visible reflecting the heterogeneous nature of the cellulases. As expected, the order of migration time corresponded to the calculated pI values (Table 3.2).

The analysis of cellulase standards together with crude secreted protein samples from three different *T. reesei* strains (RUT-C30, Iogen-M4 and Iogen-B13) are shown in Figures 3.5a-d. In this application, a wide range ampholyte 3-10 was blended with a Servalyt 3-7 buffer to enhance the resolution of the cellulases in the acidic region of the pH gradient. The electropherogram shown in Figure 3.5a represents the separation of six cellulase standards at a concentration level of 120 μg/mL each. The limit of detection of the CIEF technique was investigated using serial dilutions ranging from 1 to 500 μg/mL of each cellulase (Table 3.2). At concentrations above 200 μg/mL, the capillary was overloaded with analytes, resulting in high current (> 100 μA) with reduced band focusing.

The calibration plots for cellobiohydrolases and endoglucanases are shown in Figures 3.6a and b, respectively. In most cases good linearity, with a correlation coefficient r^2 = 0.982-0.999 was found over analyte concentrations ranging from 10 μg/mL to 200 μg/mL. A

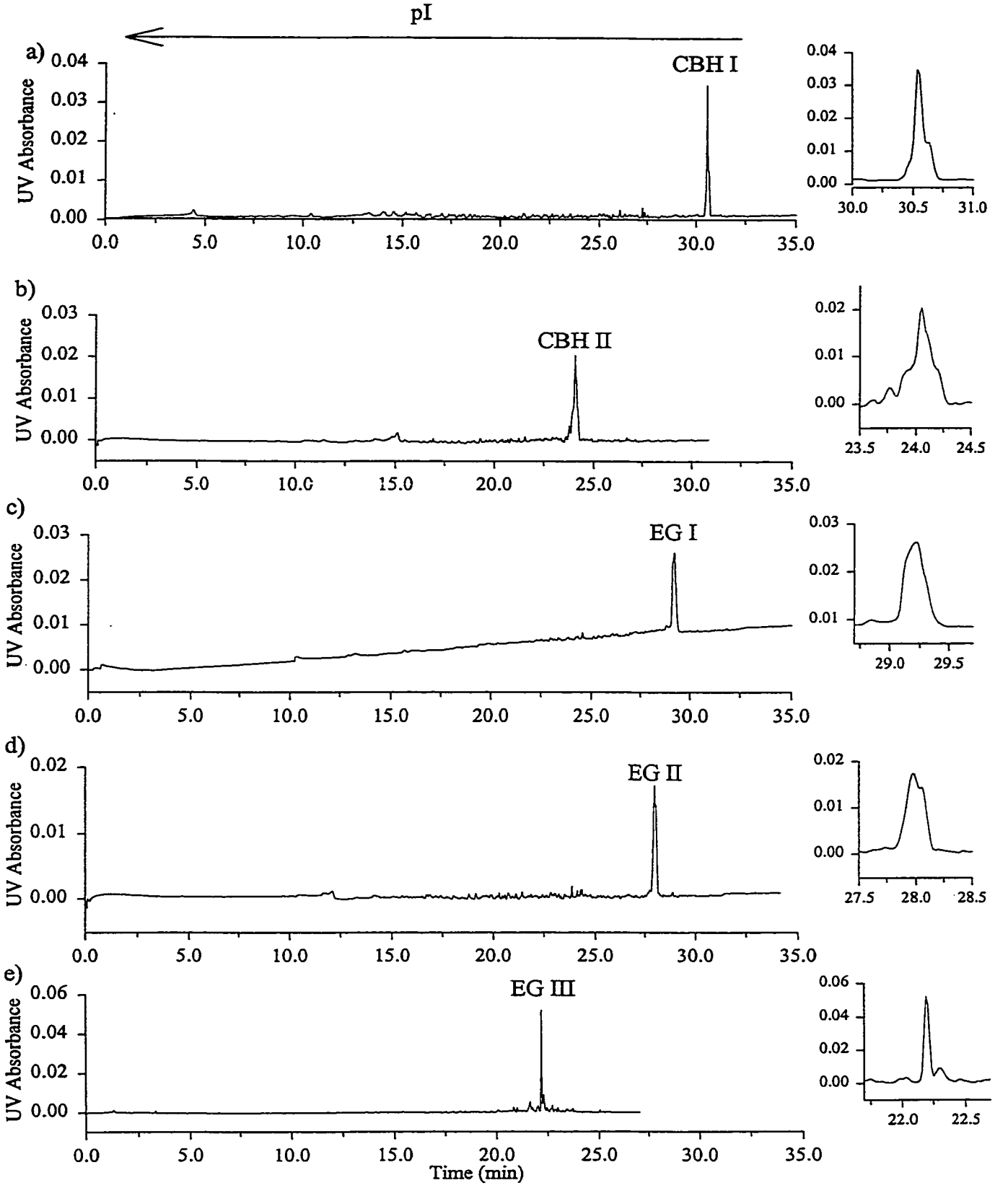


Figure 3.4 CIEF electropherograms of cellulase standards using Beckman neutral capillary. (a) CBH I; (b) CBH II; (c) EG I; (d) EG II and (e) EG III. Peaks are enlarged and shown in the inset. (All standards were purified from Iogen-M4). Conditions as described in section 2.6.2 except only Beckman ampholyte 3-10 was used.

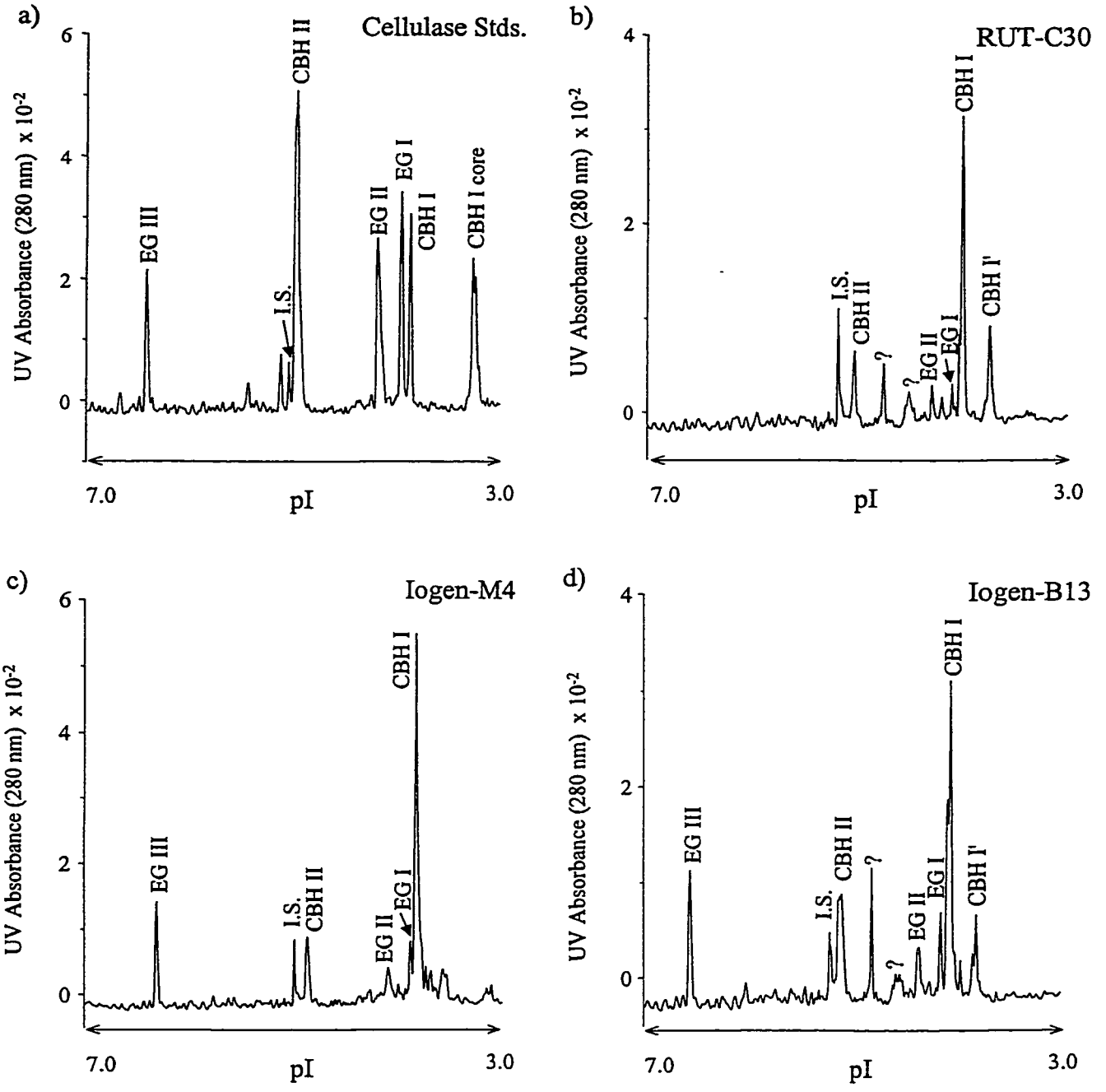


Figure 3.5 CIEF analysis of cellulase standards and crude fermentation samples from *T. reesei* using Beckman neutral capillary. (a) 6 cellulase standards [120 μg/mL except EG III = 30 μg/mL]; (b) RUT-C30 [480 μg/mL]; (c) Iogen-M4 [578 μg/mL] and (d) Iogen-B13 [531 μg/mL]. All samples were spiked with internal standard β-Lactoglobulin A [50 μg/mL]. The CBH I' indicated on (b) and (d) corresponds to the phosphorylated isoform of CBH I. See experimental section 2.6.2 for conditions.

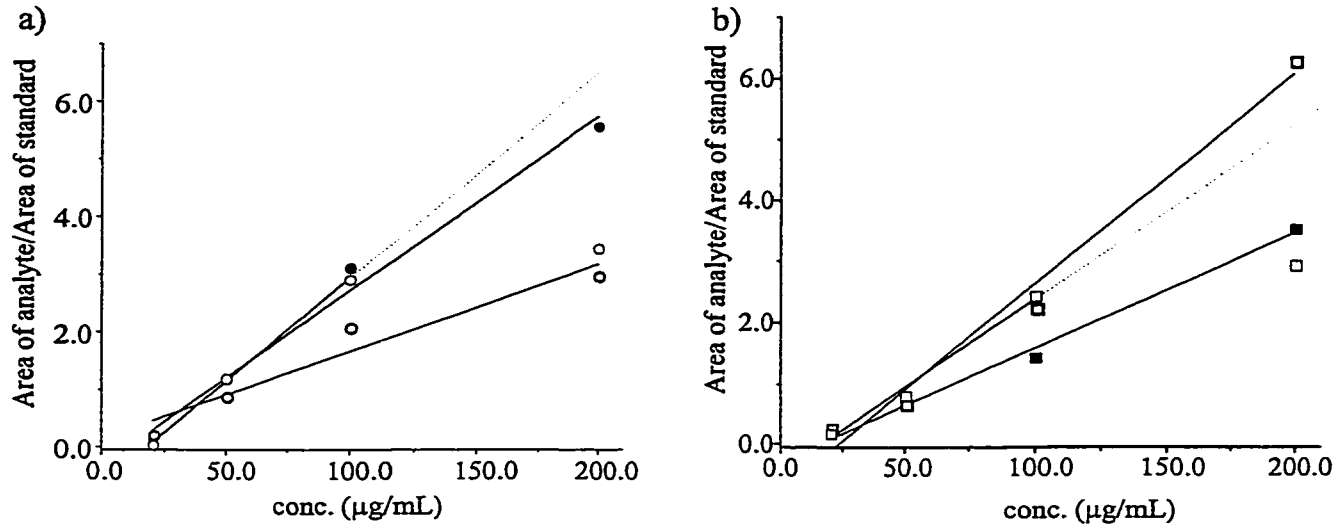


Figure 3.6 Calibration curves for CIEF analyses of cellulase enzymes from 20 μg/mL to 200 μg/mL. (a) cellobiohydrolases (○ CBH I, • CBH II, ○ CBH I core); (b) endoglucanases (□EG I, □EG II, ■EG III).

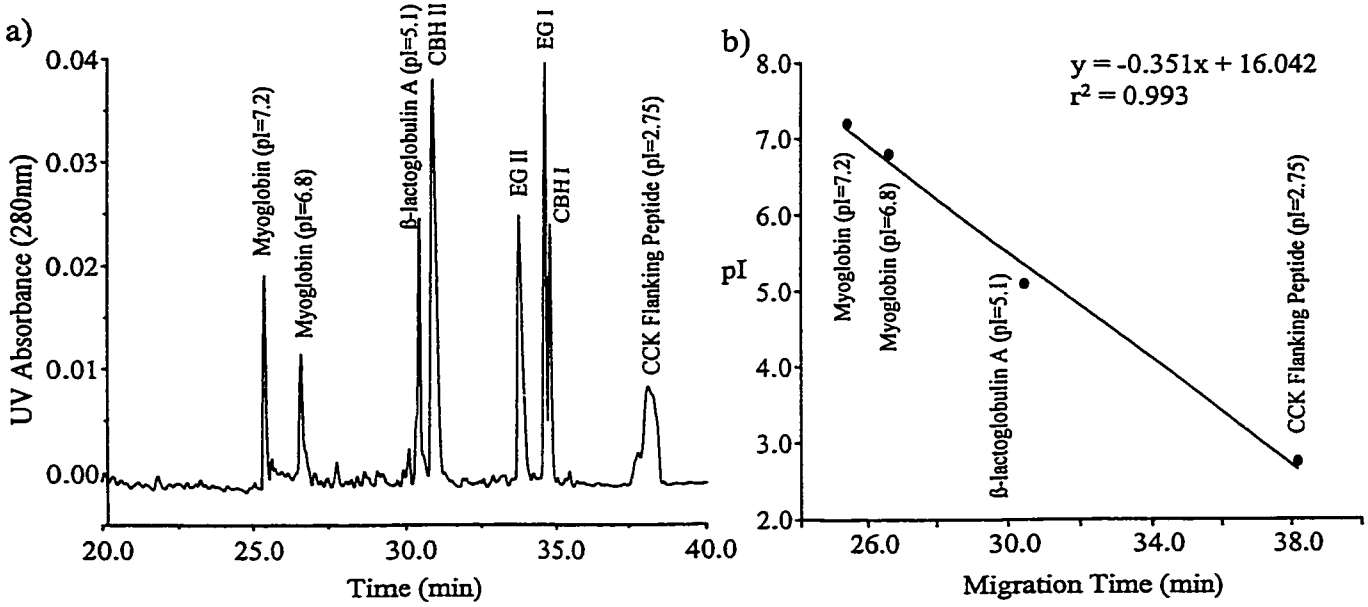


Figure 3.7 CIEF analysis of a mixture of protein standards and cellulase enzymes for pI determination. (a) CIEF electropherogram; (b) calibration plot for pI versus migration time. Conditions as described in section 2.6.2 except Beckman ampholyte 3-10 was used as carrier ampholyte.

departure from linearity was observed in CBH I core and EG I for concentrations above 100 μg/mL. Low μg/mL detection limits using signal-to-noise ratio (S/N) of 3/1 were achieved for all cellulase standards (Table 3.2).

CIEF analyses of the same samples spiked with chemical markers obtained separately also provided a convenient mean of determining their pI values (Figures 3.7a,b). A mixture of protein standards mixed with the cellulase standards was analyzed in a single CIEF run (Figure 3.7a). The protein and peptide standards selected were two isoforms of myoglobin (pI = 6.8, 7.2), ß-lactoglobulin A (pI = 5.1) and CCK flanking peptide (pI = 2.75) providing four points for preparing the calibration curve (Figure 3.7b). The observed pI values are listed in Table 3.2 below.

Table 3.2 Calibration data, limit of detection and pI values for cellulase enzymes determined using CIEF-UV

Cellulases	a^{II}	b^{II}	r^2	LOD (μg/mL)	$pI_{Obs.}$	$pI_{Calc.}^{III}$
CBH I	0.015	0.150	0.941	6	3.82	4.30
CBH I core[I]	0.036	-0.643	0.999	5	<3.0	4.17
CBH II	0.030	-0.315	0.987	24	5.20	5.12
EG I[I]	0.029	-0.474	0.989	9	3.88	4.51
EG II	0.035	-0.820	0.982	9	4.19	4.73
EG III	0.038	-0.292	0.995	4	6.79	6.49

[I] only three data points, 20 μg/ml, 50 μg/ml and 100 μg/ml, were used for regression analysis (N=3).

[II] linear calibration: y = ax + b where y = [area of analyte/area of int. std.]; a = slope; x = concentration; b = y-intercept.

[III] pI values obtained from ProMaC 1.1.1 program (urs.roethlisberger@roche.com).

It is noteworthy that the observed pI for most cellulases are generally lower than that calculated from their corresponding amino acid sequences (Table 3.2). This suggests possible changes in acidity due to the effects of tertiary structure and protein glycosylation. One source for the relatively high concentration detection limit observed in CIEF analysis is the instability of the chemical noise at 280 nm presumably associated with the carrier ampholyte.

The performance of several carrier ampholytes including ampholyte and pharmalyte was evaluated (Figures 3.8a,b). The Beckman ampholyte 3-10 (2% v/v) provided a reasonable separation for the cellulases, although EG I and CBH I were not baseline resolved (Figure 3.8a). When the analogous experiment was performed using pharmalyte 3-10 (2% v/v) instead, the resolution was drastically decreased and an artifact peak was also observed (Figure 3.8b). A blended ampholyte mixture (3% v/v) of the Beckman ampholyte 3-10 with the Servalyt 3-7 was found to yield a reasonable separation compromise in terms of resolution and sensitivity, and was thus used in most CIEF experiments.

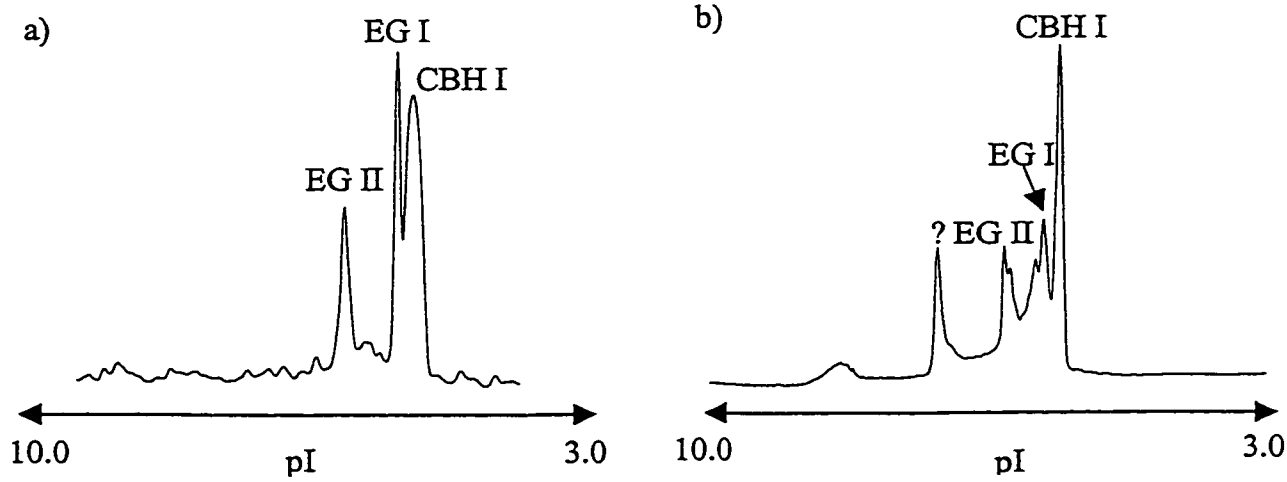


Figure 3.8 Comparison between CIEF separation of three cellulase standards using different carrier ampholyte. (a) Beckman ampholyte 3-10 (2% v/v); (b) pharmalyte 3-10 (2% v/v). See experimental section 2.6.2 for conditions.

In the present case, CIEF analyses using the optimized blended ampholyte system provided a rapid analytical tool to determine the relative distribution of individual enzymes in crude cellulase fermentation samples and offered a viable alternative to slab IEF gel. In order to obtain good reproducibility, an internal standard (I.S. = β-lactoglobulin A) was spiked in all samples to correct for instrumental variations between consecutive runs. Individual response factors were first calculated in a preliminary CIEF experiment. These values were then used to determine the concentration of each cellulase in their respective extracts (Table 3.3). As an example, the concentration of CBH II in Iogen-M4 (Figure 3.5c) was demonstrated.

In method of internal calibration, the response factor (F) is calculated using:

[concentration of analyte/concentration of I.S.] = F [area of analyte/area of I.S.] (15)[99]

In Figure 3.5a, the arbitrary peak area for CBH II is 16.7 (120 μg/mL) and for I.S. is 1.1 (50 μg/mL)

Thus, F = (120/50)*(1.1/16.7) = 0.16

In Figure 3.5c, the arbitrary peak area for CBH II is 1.9 and for I.S. is 1.6 (50 μg/mL).

Using equation 3.1, concentration of CBH II = 0.16*(1.9/1.6)*50 μg/mL = 9.5 μg/mL

The concentrations obtained from five CIEF runs are 8.6, 9.0, 9.5, 9.8 and 11.0 μg/mL

This gives an average concentration of 9.6 μg/mL with a standard deviation of 0.8 μg/mL for CBH II

In all cases data presented in Table 3.3 correspond to normalized cellulase concentrations based on peak area ratios. The relative concentrations shown in Table I represent average values taken from five consecutive injections. For CBH I, the intra-day % RSD for relative percentage ranged from 0.3% to 3%. This is in agreement with the reproducibility evaluation of CIEF reported by Tang and co-workers.[100] However, the % RSD becomes larger for other cellulases of lower concentration thus reflecting the larger influence of the noise level on the corresponding measurement.

Table 3.3: Distribution of cellulases in secreted protein extract from different strains of *T. reesei*

Cellulases	RUT-C30[a] (μg/mL)	Iogen-M4[a] (μg/mL)	Iogen-B13[a] (μg/mL)
CBH I	169.3 ± 12.1 (87.9 ± 0.3)	135.5 ± 16.3 (76.8 ± 1.0)	174.5 ± 21.9 (74.5 ± 2.3)
CBH I core	n.o.	n.o.	n.o.
CBH II	8.1 ± 0.2 (4.2 ± 0.2)	9.6 ± 0.8 (5.5 ± 0.2)	21.2 ± 2.5 (9.1 ± 1.4)
EG I	7.0 ± 0.6 (3.6 ± 0.2)	9.5 ± 1.1 (5.4 ± 0.3)	14.5 ± 1.7 (6.2 ± 0.2)
EG II	8.2 ± 0.6 (4.2 ± 0.07)	10.8 ± 0.9 (6.2 ± 0.4)	13.8 ± 2.4 (5.9 ± 0.7)
EG III	n.o.	10.8 ± 0.4 (6.2 ± 0.5)	9.9 ± 1.7 (4.3 ± 0.8)

[a] Total protein concentration was 480, 578 and 531 μg/mL for RUT-C30, Iogen-M4 and Iogen-B13, respectively (Based on Bradford assay). Numbers in parantheses correspond to relative percentage of the different cellulases in the corresponding protein extract (n=5) using ß-lactoglobulin A as an internal standard. n.o. = not observed.

Consistent with previous investigations,[12,101] the results displayed in Table 3.3 clearly indicate that CBH I is the major cellulase observed in the secreted protein from *T. reesei*. On a relative weight basis, its presence accounts for more than 50 % of the total protein secreted by three strains examined. RUT-C30 showed the highest relative proportion of CBH I with a value of 87.9% of the total secreted cellulase complement. In the fermentation samples from Iogen-B13 and RUT-C30 a component migrating between CBH I and CBH I core was observed at approximately pI 3.6. Separate experiments suggested that this protein corresponded to a phosphorylated isoform of CBH I (section 4.6). CBH II and EG II were the next most prominent enzymes found in these extracts though their relative proportions were generally below 10 % of the corresponding total protein. Interestingly, no significant level of EG III was observed in RUT-C30 in contrast to Iogen-M4 and Iogen-B13 which showed a relative abundance of at least 5 % in each case. Other peaks of unknown identity were also observed (Figure 3.5) but could not be assigned to any of the known cellulases.

3.4 CIEF with PVA Coating

CIEF analysis using the Beckman neutral capillary supplied commercially has provided a rapid way of screening the crude samples and allows for quantitative assessment of variations in enzyme concentrations during fermentation stages. However, it would be beneficial to develop a neutral coating method in order to reduce the cost of analysis. PVA is a synthetic polymer generally used as a binder in ceramic applications, an additive in paper and pulp industries, a film former as well as in cosmetic formulations as a viscosity increasing agent.[102,103] It forms a transparent thin film with non-covalent binding to the inner capillary wall to reduce or eliminate EOF.

A larger size capillary (ID 75 μm X OD 350 μm) was used to minimize capillary occlusion and non-uniform derivatization during the PVA coating. The analyte concentration was decreased to account for the increase in capillary volume. Analogous experiments performed using the Beckman neutral capillary shown in Figure 3.5 were repeated in PVA coated column (Figures 3.9a-d). The order of migration for individual cellulase standards with respect to the internal standard was similar (Figure 3.9a). Except for EG I and EG II in strain Iogen-M4 (Figure 3.9c), both the identified and the unknown components observed previously were assigned in the crude extracts analyzed using PVA coated capillary. Consistently, the cellulase EG III was not observed in RUT-C30 while CBH I' was seen in both RUT-C30 and Iogen-B13 (Figures 3.9b,d). However, inspection of these CIEF electropherograms suggested that the peak resolution decreased from the basic end to the acidic end. In all electropherograms displayed, the internal standard and CBH II were focused into very narrow peaks while the peak shape for CBH I became broader (Figures 3.9a-d). This is likely arising from the non-linearity of the pH gradient in local area. The non-

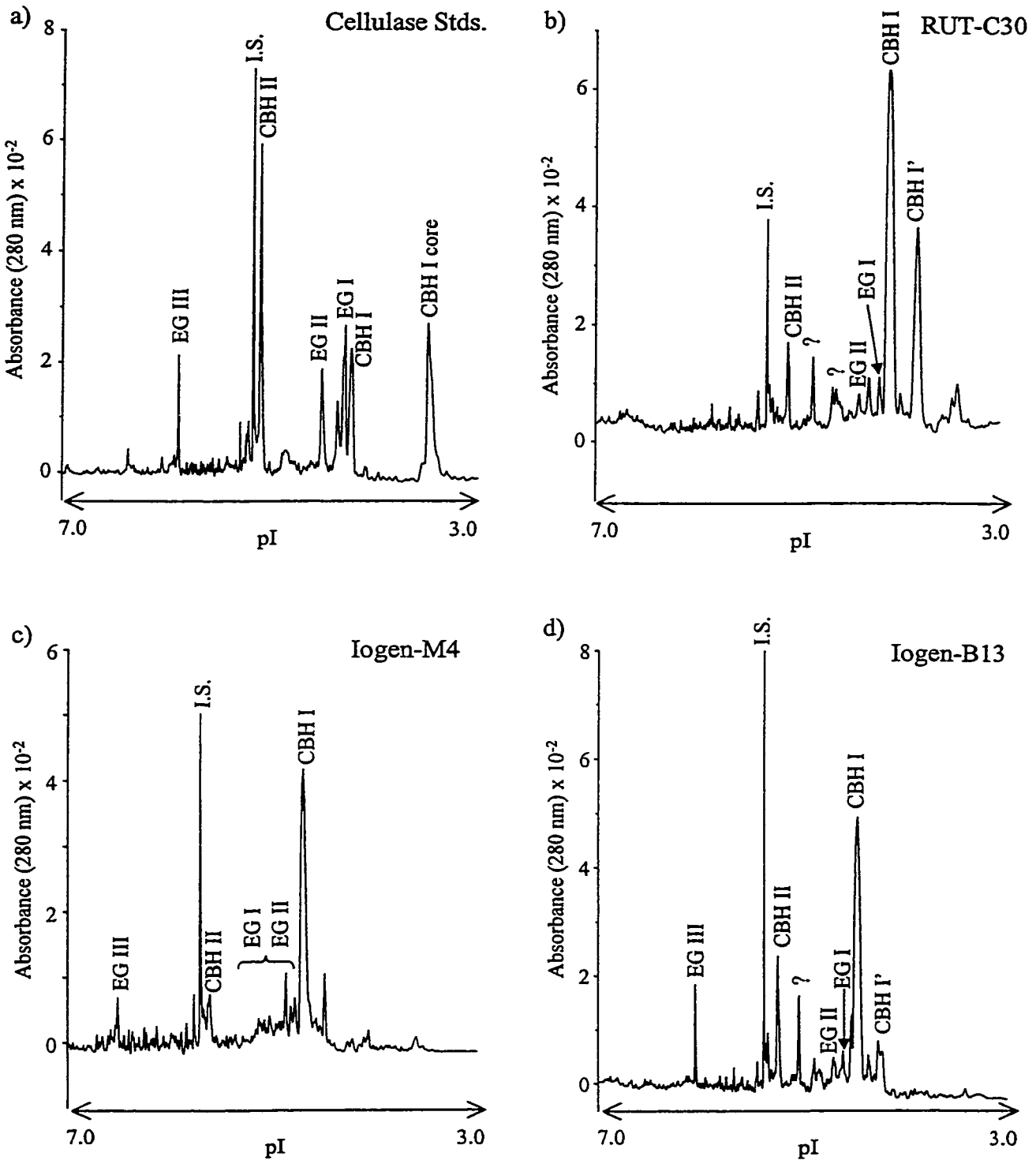


Figure 3.9 CIEF analysis of cellulase standards and crude fermentation samples from *T. reesei* using PVA coated capillary. (a) 6 cellulase standards [50 μg/mL except EG III = 15 μg/mL]; (b) RUT-C30 [400 μg/mL]; (c) Iogen-M4 [413 μg/mL] and (d) Iogen-B13 [354 μg/mL]. All samples were spiked with internal standard β-Lactoglobulin A [50 μg/mL]. See experimental section 2.6.2 for conditions.

uniformity of PVA coating may lead to the presence of EOF in certain section of the capillary thus causing this effect. The resulting peak areas would not truly reflect the actual intensity of each component. In a separate experiment using another batch of PVA coated capillary, the resolution in the acidic end was enhanced. These results demonstrated the neutral coating was not uniformly deposited as a thin film. Thus, further works on refinement of the coating procedure is needed to incorporate PVA for routine CIEF analysis.

3.5 CIEF-MS

To determine the unknown components in the crude extracts, CIEF was coupled on-line to mass spectrometry. At an applied voltage, analytes were first focused to their respective pI values and were then chemically mobilized to the mass spectrometry with electrospray ionization (see section 2.8.5 for more details). A mixture containing two protein standards, ribonuclease A and myoglobin were analyzed initially (Figures 3.10a,b). The corresponding base peak chromatogram shown in Figure 3.10b displayed a good separation between the two standards according to their pI values. The two isoelectric points exhibited in myoglobin (7.2, 6.8) were also observed. In this analysis, only a 1 % (v/v) Beckman ampholyte 3-10 was used, which would result in a decrease in resolution compared to the 3% (v/v) applied in CIEF-UV. However, the carrier ampholyte molecules would cause a large chemical noise at the low mass range. Thus, the poorer resolution in CIEF-MS compared to CIEF-UV is compromised by the improvement in signal-to-noise ratio.

The extracted mass spectra for ribonuclease A and myoglobin were displayed in Figures 3.11a and b, respectively. In both cases, the multiply-charged ion envelope showing the

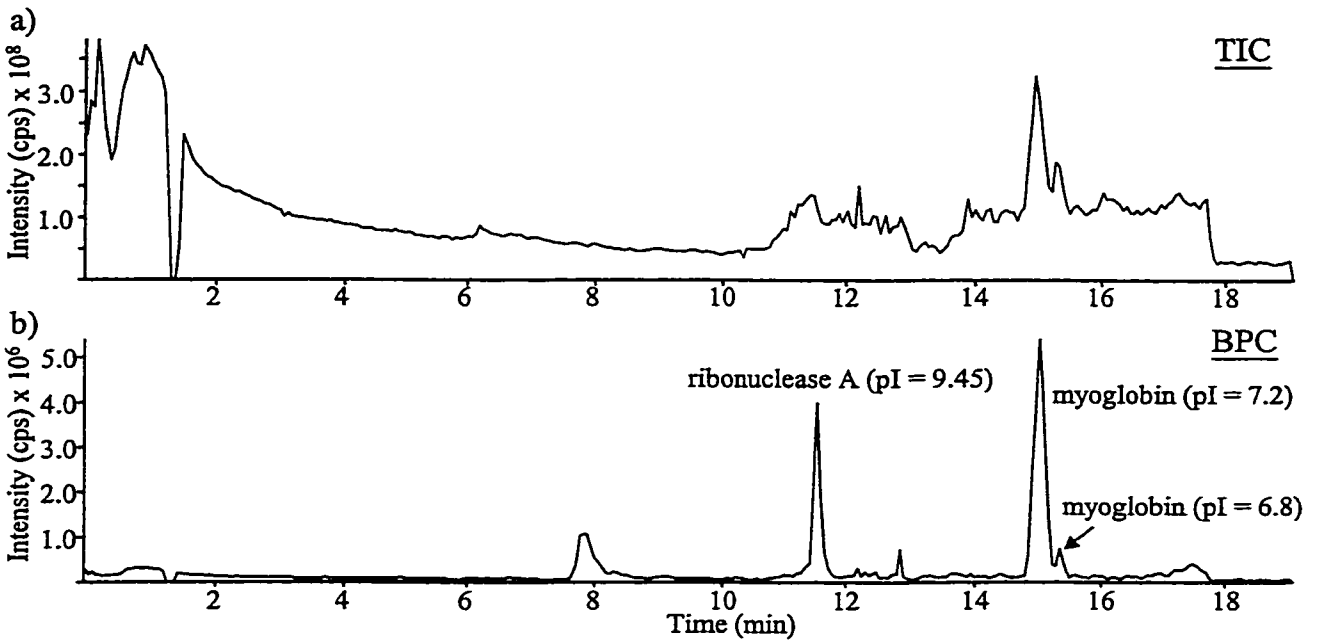


Figure 3.10 CIEF-MS analysis of a mixture of ribonuclease A and myoglobin each at 50 μg/mL. (a) Total ion chromatogram (TIC); (b) base peak chromatogram (BPC). The protein standards were dissolved in a 1% (v/v) Beckman ampholyte pH 3-10 solution.

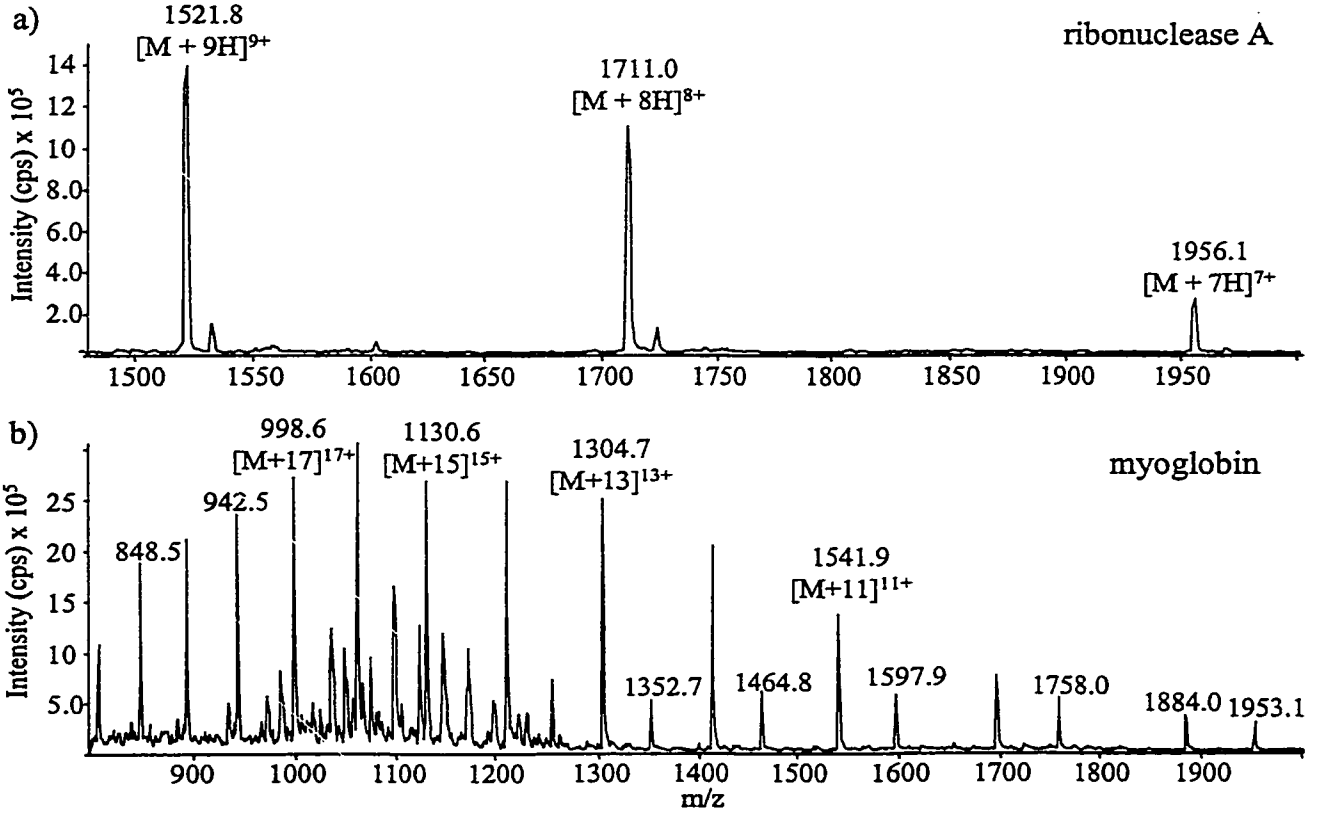


Figure 3.11 Extracted mass spectra from BPC of CIEF-MS analysis shown in Figure 3.10b for (a) ribonuclease A; (b) myoglobin.

various charge states of the ionized molecule was clearly observed. Interestingly, an extra series of peaks (m/z 1352.7, 1464.8, 1597.9, 1758.0) appeared in myoglobin and this may be associated with the ampholyte adduct ions. Reconstructed mass spectra revealed a molecular mass of 16951.7 Da (± 3.6 Da) for myoglobin and a mass of 13684.3 Da (± 4.1 Da) for ribonuclease A. Despite the fact that CIEF-MS with electrospray ionization was shown here to be a successful hyphenated technique, it failed to provide adequate detection of the cellulase enzymes. In a separate experiment, selected cellulase standards were mixed with the protein standards and only the latter were observed in the resulting CIEF-MS mass spectra. This experiment demonstrated that the analytes were well focused at their pI values, however, the cellulase enzymes were not successfully ionized under the current conditions. It is thus postulated that a certain amount of organic solvent was required to assist the ionization of these highly heterogeneous cellulolytic enzymes.

3.6 Solvent Effect on Ionization

Ionization of cellulases has posed a problem in CIEF-MS analysis and the effect of organic solvent addition was further investigated. Individual cellulases were mixed with a protein standard and infused into the mass spectrometer by pressure (Figures 3.12a-d). All the analytes were arbitrary dissolved in a sample buffer consisting of acetonitrile (50% v/v), 0.05 M formic acid and introduced into the mass spectrometer at 300 mbar. A sheath buffer composed of acetonitrile:deionized water:acetic acid (50:49:1 v/v/v) to assist ionization was operated at a flow-rate of 1.5 μL/min. Among the four selected cellulases, CBH I from RUT-C30 showed a relatively strong signal compared to the protein standard, although the glycoforms within each charge state were not well resolved (Figure 3.12b). Ionization of

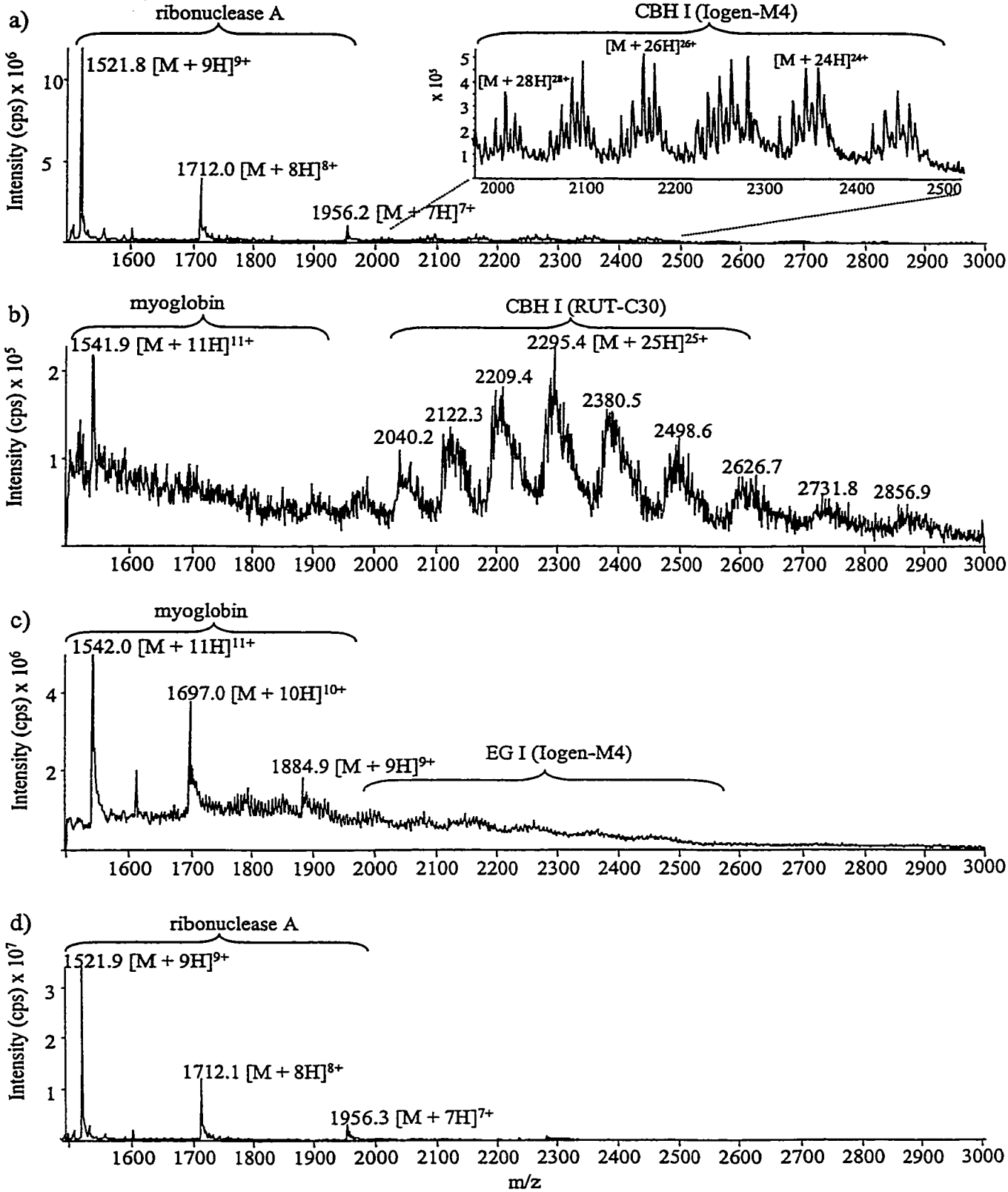


Figure 3.12 Infusion ESMS analysis of selected cellulase and protein standard mixtures. (a) CBH I (Iogen-M4) + ribonuclease A; (b) CBH I (RUT-C30) + myoglobin; (c) EG I (Iogen-M4) + myoglobin and (d) EG II (Iogen-M4) + ribonuclease A. The concentration for all protein standards is 70 µg/mL while for cellulases is 0.5 mg/mL except for (b) where it is 1 mg/mL.

CBH I from Iogen-M4 was relatively poor compared to the protein standard ribonuclease A (Figure 3.12a), and yet its glycoforms were resolved (see inset diagram).

In contrast to the cellobiohydrolases, the endoglucanases were not ionized at all under the current experimental conditions. EG I from Iogen-M4 was barely observed (Figure 3.12c) while EG II from the same strain was not ionized (Figure 3.12d). These results clearly suggested that cellulases do not exhibit the same ionization efficiency at equal organic solvent composition.

The effect of varying the organic composition in the sample solution on ionization was subsequently studied (Figures 3.13a-f). Strong signal observed in CBH I from RUT-C30 (Figure 3.12b) has prompted its use as a model cellulase for this experiment. The percent composition of acetonitrile in the sample solution was increased progressively from 30% to 70%. No ionization was observed for sample solution containing 30% acetonitrile (Figure 3.13a). As the organic solvent increased to 35 % (Figure 3.12b), the multiply-charged envelope of CBH I was observed. The ionization efficiency was maintained from 35% to 60% (Figure 3.12b-e), and gradually weakened at above 70% acetonitrile with no apparent change in the multiply-protonated patterns (Figure 3.12e). These observations lead to the proposal that ionization of cellulases, in particular CBH I from RUT-C30 studied presently, is strongly dependent on the composition of organic solvent present in the sample. The conformation of CBH I may play a role in which certain percent of organic solvent is necessary to unfold the protein and to expose basic amino acids leading to its efficient multiply-charged state. Even though a sheath buffer consist of 50% methanol was used in CIEF-MS experiment, the failure to detect the cellulases may indicate the insufficient amount of time for sample exposure to the organic solvent. It is also observed above 70%

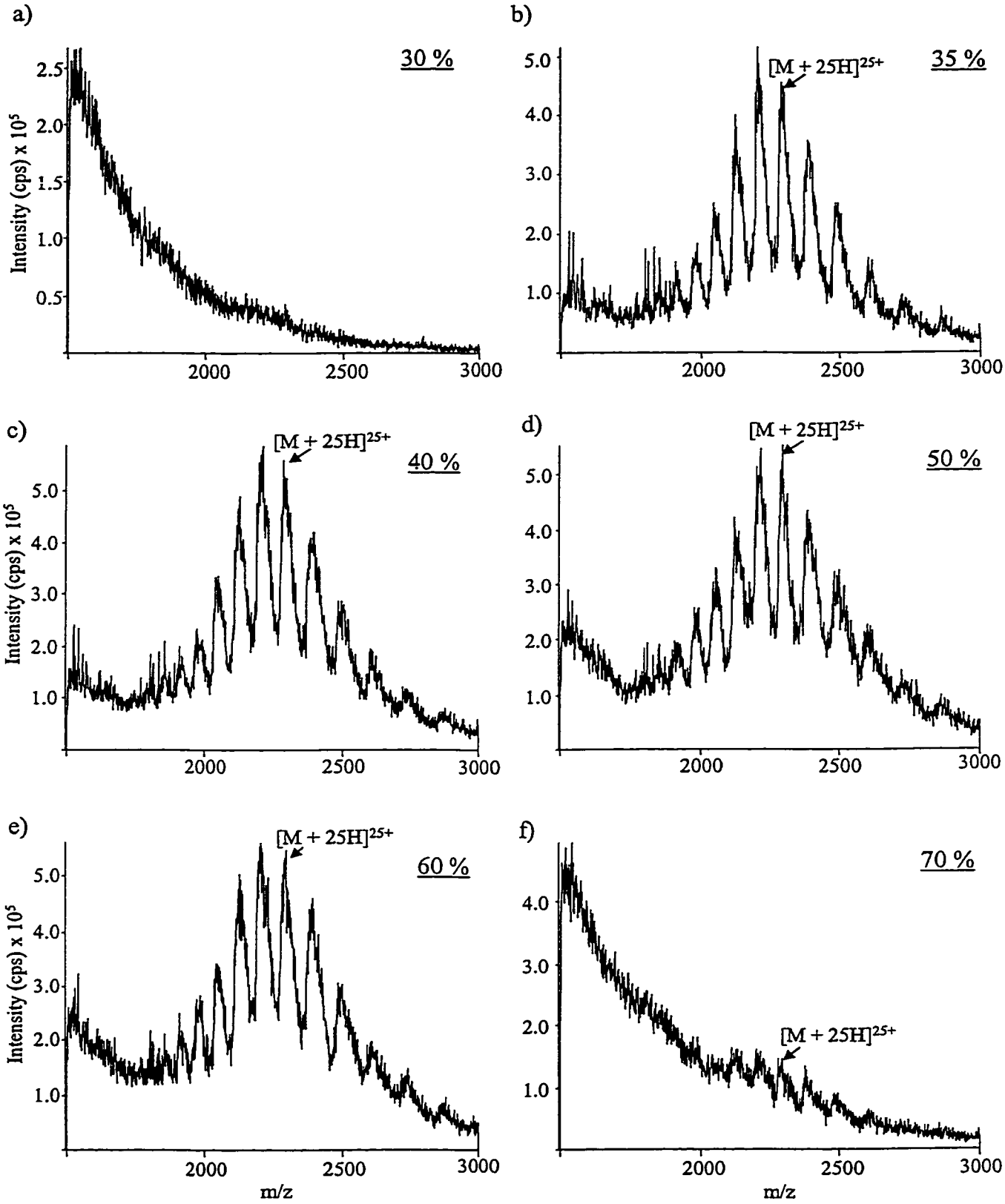


Figure 3.13 Effect of varying percent composition of organic solvent in the sample buffer on ionization of CBH I (RUT-C30). The % of acetonitrile (v/v) is 30 in (a); 35 in (b); 40 in (c); 50 in (d); 60 in (e) and 70 in (f). Concentration of CBH I is 0.5 mg/mL in all cases.

acetonitrile, the ionization is hampered and thus suggesting the formation an unfavourable conformation.

In the next chapter, an alternative approach was used to analyze the cellulase enzymes by ESMS. Using flow injection analysis (FIA), individual cellulolytic enzymes (1 μL) were injected into a stream of 50 % acetonitrile acidified with 0.2% acetic acid and delivered to the mass spectrometer at a higher flow-rate of 5-10 μL/min. All cellulases were ionized successfully using this method.

3.7 Conclusion

Capillary electrophoresis was evaluated as a potential analytical tool to analyze the crude cellulase fermentation samples from *T. reesei*. The analyses were carried out using both CZE and CIEF modes. In CZE, broad peaks and non-baseline resolution were generally observed despite the use of a long capillary and high concentration buffer. In contrast, CIEF is capable of focusing and separating the cellulase components according to their pI values. To further increase the resolution at the acidic region of the pH gradient featuring the pI of the cellulases, a broad range pH 3-10 carrier ampholyte was blended with a narrow range pH 3-7 ampholyte. This allows for baseline resolution between cellulase components with close pI values such as EG I and CBH I. The technique of CIEF not only offers a rapid screening method for crude fermentation samples, but also provides a valid method to determine the concentration and relative distribution of individual enzymes in these samples. By using the internal standardization, the instrumental variations between consecutive runs were corrected. In all the investigated strains, the relative abundance of CBH I accounts for more than 50% of the total protein secreted.

An interface for coupling CIEF on-line to mass spectrometry was developed to determine the unknown components in crude cellulase samples. However, CIEF-MS was only proven successful for the detection of non-glycosylated protein standards and was not suitable for cellulase analysis. This has prompted us to investigate the effect of organic solvent on ionization efficiency of cellulase enzymes. After a series of experiment, it was concluded that a certain percentage composition of organic solvent in the cellulase sample solution was crucial to enhance the formation of gas-phase ions. The use of LC-MS, which will be discussed in the next chapter, turned out to be a better method because of a higher content of organic solvent to assist ionization.

When you have eliminated the impossible, whatever remains, *however improbable*, must be the truth.

Sir Arthur Conan Doyle (1859-1930)

The Sign of Four (1890), Ch. 6

Characterization of CBH I Glycoforms from *T. reesei* RUT-C30 and Two Mutant Strains.

4.1 Introduction

The analysis of crude fermentation extracts from *T. reesei* by CIEF has led to the conclusion that CBH I is the major cellulolytic components. To further understand the variation in structural aspect of this glycoprotein between different mutant strains, it is necessary to examine the changes in post-translational modifications in more details. M. Maras et al. had reported the presence of mammalian high-mannose-type N-linked glycans in CBH I from *T. reesei* RUT-C30. [23] Klarskov et al.[24] and Harrison et al.[25] have characterized the sites of N-linked glycosylation in CBH I from strain QM9414 and strain ALKO2877, a derivative of strain QM9414 from which the genes for the major endoglucanases have been deleted. In each study, selective identification of glycosylated residues revealed the presence of single GlcNAc in three putative N-linked sites located in the catalytic domain. In addition, the linker region of this CBH I from ALKO2877 was found to be partially sulfated.[25] These two glycosylation modifications have not been reported previously for CBH I from other strains from *T. reesei*. The O-linked glycans, on the other hand, are located mainly in the linker domain, and the glycosylation consists of one to three mannoses.[25]

In the present chapter, electrospray mass spectrometry (ESMS) was used to monitor the extent of glycosylation in the native CBH I purified by ion-exchange chromatography from three related *T. reesei* strains. The identification of N-linked glycosylation sites in tryptic peptides from CBH I was achieved using LC-ESMS on a quadrupole/time-of-flight instrument with a mixed scan function enabling alternate acquisition of $m/z < 400$ at high orifice voltage and $m/z > 400$ at low orifice voltage.[91,92] The structures of the

oligosaccharides and the peptide backbone were determined by on-line tandem mass spectrometry. Papain digestion of CBHI followed by MALDI-TOF analysis of the resulting proteolytic fragment revealed an extensively O-linked glycosylated linker peptide. The extent of this heterogeneity was monitored by HPAEC-PAD analysis of the hydrazinolysis products with subsequent characterization using ESMS.

4.2 CBH I Sequence Information

The *cbh 1* gene from *T. reesei* strain L27 has been sequenced by Shoemaker et al.[104] and the corresponding amino acid sequence is available in the Swiss-Prot database (http://www.expasy.ch/sprot). The protein sequence of CBH I from the strains used in the present investigation differs from that of CBH I from strain L27 by replacement of an arginine at position 442 for two prolines (Figure 3.1a). As a result, the present CBH I has 497 amino acid residues and an average molecular mass of 52205.8 Da accounting for an N-terminal pyroglutamic acid and twelve disulfide bridges. Interestingly, the same mutation at position 442 has also been reported for the CBH I sequence from strain ALKO2877.[25]

4.3 Intact CBH I Analysis

Individual CBH I was analyzed by both MALDI-TOF and electrospray mass spectrometry of which the latter has resolution capable of separating glycoforms.

4.3.1 Analysis of CBH I by MALDI-TOF

MALDI-TOF analysis has provided a rapid mean to determine the purity and the average mass of the purified CBH I from the three investigated strains (Figures 4.1 a-c). The purity was also verified using CIEF, which further confirms the presence of a single enzyme.

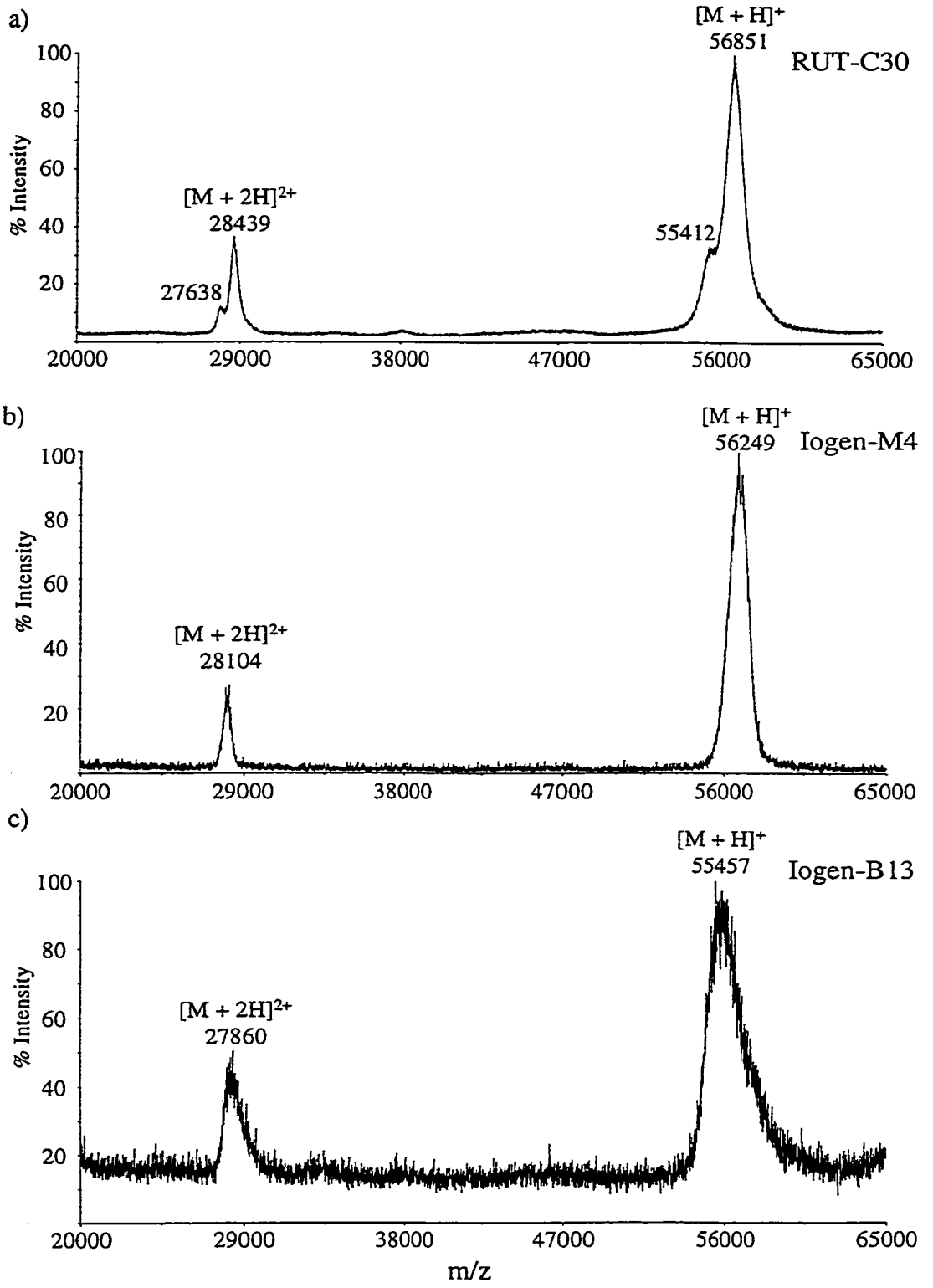


Figure 4.1 MALDI-TOF analysis of CBH I purified from three different strains. (a) RUT-C30; (b) Iogen-M4 and (c) Iogen-B13. Masses of the doubly-charged ions indicated do not exactly correspond to half of the single-charged ions due to the relatively large range of uncertainty (± 0.1 %) of MALDI-TOF in protein analysis.

The relative proportion of glycans based on the average mass with respect to the protein molecular weight is summarized in Table 4.1. Among these purified CBH 1, the one from RUT-C30 displayed the highest average molecular mass (Figure 4.1a) and this mass increment is attributed to the high mannose at Asn270 as shown later in section 4.4. Interestingly, a small shoulder peak at m/z 55412 is observed in Figure 4.2a, and was suspected to arise from a truncated glycan chain at Asn270. Incubations with Endo H completely removed this peak, resulting in a smaller mass at m/z 55468 (Figure 4.2b), and thus support the proposal of a heterogeneous nature of oligosaccharides at this N-linked glycan position.

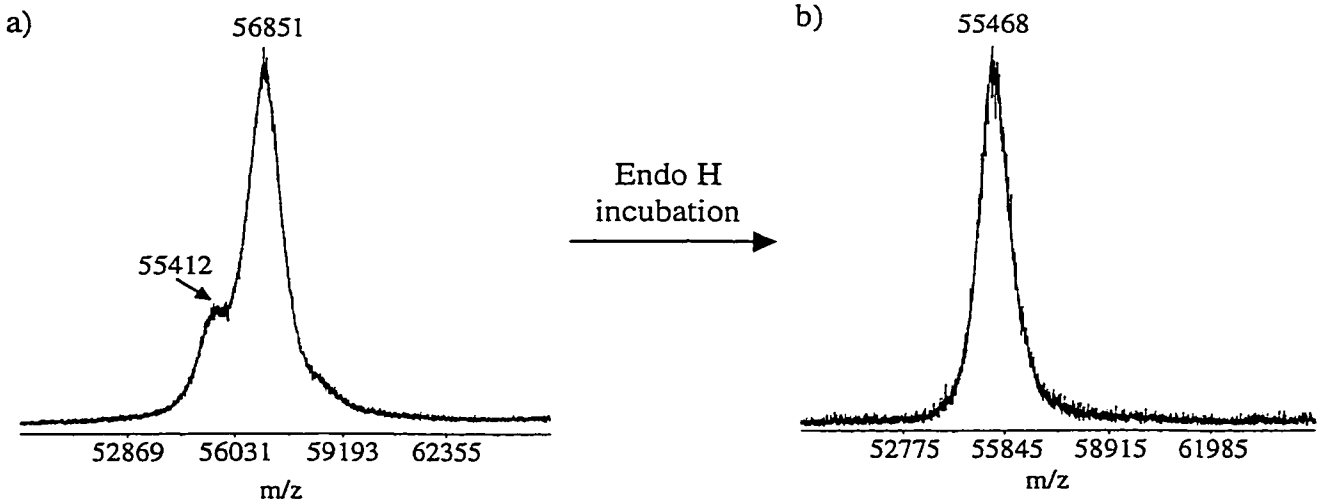


Figure 4.2 MALDI-TOF analysis of CBH I from RUT-C30. (a) Intact cellulase; (b) after Endo H incubation. See experimental section 2.7.5 for digestion conditions.

A relatively broad peak was observed in CBH I from the mutant strain Iogen-B13 (Figure 4.1c) which was later found to be related to the partial deglycosylation at the same N-linked site, Asn270 (section 4.4). These results illustrate that the resolution of MALDI-TOF is relatively limited (± 0.1% mass accuracy for protein > 10,000 Da), and determination of glycoform patterns of the intact cellulases is thus difficult to obtain. However, as indicated in section 4.5 the highly heterogeneous O-linked glycan in the linker/CBD regions at the mass

range m/z 8,000-10,000 was successfully separated by MALDI-TOF. The low resolution of MALDI-TOF in cellulase analysis has prompted the use of electrospray mass spectrometry from a Q-TOF instrument in order to enhance the separation of glycoforms.

4.3.2 Profiling Microheterogeneity in CBH I using ESMS

Compared to MALDI-TOF analysis, flow-injection analysis of intact CBH I using ESMS provided a unique method to assess their corresponding glycoform distributions. The reconstructed molecular mass profiles of CBH I from RUT-C30, Iogen-M4 and Iogen-B13 and are shown in Figures 4.3a-c. As an example, the mass spectrum of CBH I from Iogen-M4 (Figure 4.4) shows a multiply-charged ion envelope extending from m/z 3000-4000 from which the molecular mass profile can be obtained (Figure 4.3a). The CBH I from Iogen-M4 (Figure 4.3b) also showed a relatively narrow distribution of glycoforms each spaced by 162 Da (hexose) extending from 55,000-57,000 Da. Based on these molecular distributions, the relative proportion of glycan in CBH I (with respect to the protein mass) from RUT-C30, Iogen-M4 and Iogen-B13 was calculated and listed in Table 4.1. In contrast, CBH I from Iogen-B13 showed more significant heterogeneity as reflected by spacing of 162 Da (hexose) for the lower mass glycoforms (<56,000 Da) whereas a series of doublet spaced by 162 and 203 Da (hexose and N-acetyl-hexosamine) was noted for the higher molecular mass glycoforms (56,000-57,000 Da). This heterogeneity is accounted for by the glycan population at Asn_{270}. Indeed, tryptic digest analysis has revealed the ratio of single GlcNAc to high mannose $Man_8GlcNAc_2$ at Asn270 of CBH I from Iogen-B13 to be 2:1 (section 4.4), and the overlap between these two series of glycoforms has significantly broadened the glycosylation profile. As indicated in Figure 4.3c, the spacing between the glycoform with high mannose attachment (56593.0 Da) and that of the glycoprotein with a single GlcNAc (55094.0 Da)

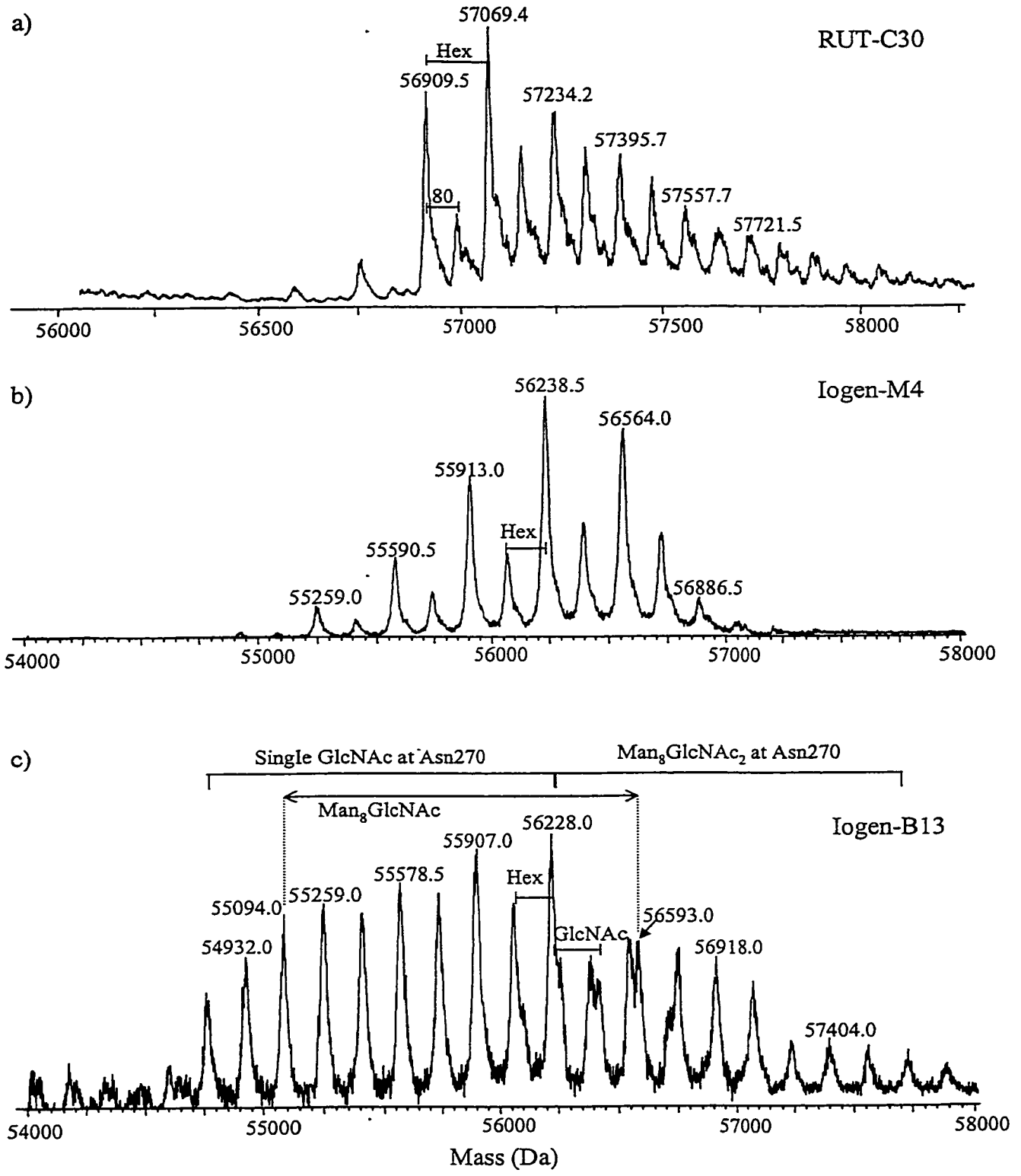


Figure 4.3 Reconstructed molecular mass profiles of CBH I for (a) RUT-C30; (b) Iogen-M4 and (c) Iogen-B13. See experimental section 2.7.2 for conditions.

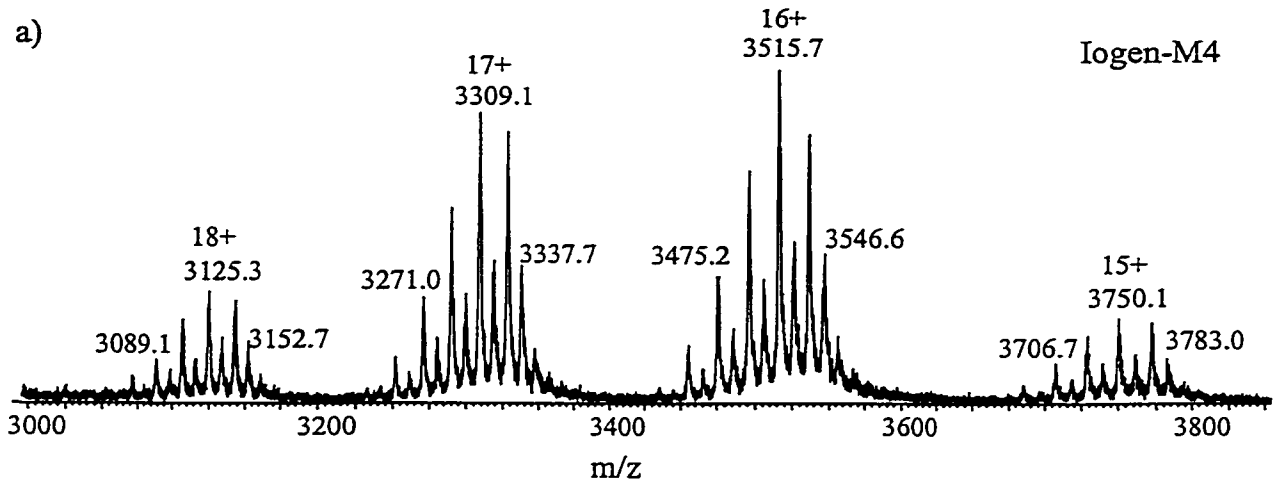


Figure 4.4 Electrospray mass spectrum of CBH I purified from Iogen-M4 strain.

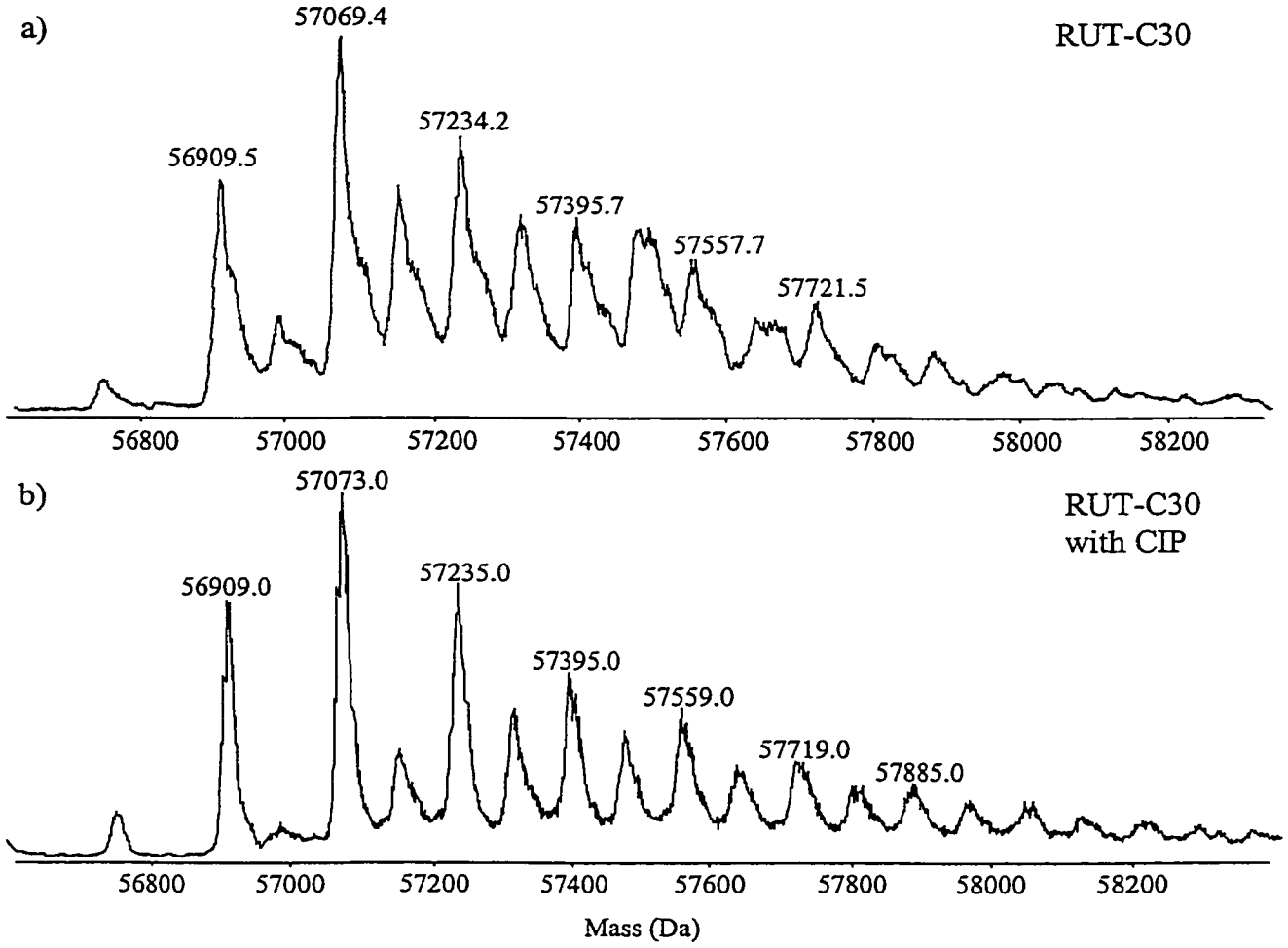


Figure 4.5 Reconstructed molecular mass profiles of CBH I (a) before and (b) after alkaline phosphatase incubation.

corresponds precisely to the mass difference of $Man_8GlcNAc$. The mass accuracy of ESMS is generally ± 0.01 % for protein analysis. In contrast, CBH I from strain Iogen-M4 has single GlcNAc at Asn270 with no high mannose thus narrowing the glycosylation profile (section 4.4).

Table 4.1 Comparison of molecular weight and % glycosylation of CBH I purified from three investigated strains detected by MALDI-TOF and ESMS

		MALDI-TOF		ESMS	
Strains	$Mass_{avg.}$*(Da)	$Mass_{avg.}$ (Da)	% glycosylation	$Mass_{avg.}$ (Da)	% glycosylation
RUT-C30	52205.8	56850	8.9	56500-58000	8.2-11.1
Iogen-M4		56248	7.7	55000-57000	5.4-9.2
Iogen-B13		55456	6.2	54500-58000	4.4-11.1

* calculated from the protein sequence

The CBH I from RUT-C30 showed the highest molecular mass distribution of all three strains examined. This will be more evident in the following discussion on cLC-ESMS analysis of tryptic digest, which revealed the presence of predominantly high mannose $Man_8GlcNAc_2$ at Asn270 with minor occupancy of single GlcNAc in CBH I from RUT-C30. Interestingly, an extra series of peaks spaced by 80 Da was found embedded between the hexose spacing in the RUT-C30 sample. This unusual glycoform pattern, which was previously reported to be associated with sulfation of the linker peptide of CBH I from strain ALKO2877,[25] was subsequently identified as a phosphorylated saccharide from HPAEC-PAD analyses described in section 4.6. Separate experiment has shown that incubation of this cellulase with alkaline phosphatase removed the phosphate group on the first two glycoforms except for the larger oligomer series above 57,200 Da (Figures 4.5a-b). This observation possibly suggests that extensive glycosylation of the linker might impart steric hindrance

preventing the catalytic activity of the phosphatase. Similar results were obtained when the analogous experiment was repeated on the purified CBH I linker/CBD domains (section 4.5).

4.4 Tryptic Digestion

Tryptic digest of CBH I were analyzed by ESMS and tandem mass spectrometry to characterize the posttranslational modification in CBH I.

4.4.1 LC-ESMS Analysis of Tryptic Peptides from CBH I

The tryptic digest of the reduced and alkylated CBH I from the three *T. reesei* strains were subjected to on-line cLC-ESMS. In order to facilitate the identification of tryptic glycopeptides on the quadrupole/time-of-flight instrument, a mixed scan function was used to promote the in-source formation of selected oxonium ions under high orifice voltage conditions (100 V), while enabling detection of multiply-protonated ions using low orifice voltage (30 V).[91,92] The extracted ion electropherogram for the characteristic oxonium disaccharide ion Hex-HexNAc at m/z 366 is shown in Figure 4.6 together with the total ion chromatogram (TIC) for m/z 400-2000.

The peptide masses derived from the extracted mass spectra were matched against those predicted from the theoretical digest of CBH I (Figure 3.1a) as calculated by the ProMaC 1.1.1 program (urs.roethlisberger@roche.com). The observed tryptic peptides are presented in Table 4.2 in ascending order of molecular masses together with their peak number and sequence assignment. It is noteworthy that hydrophobic peptides of masses above 4000 Da were, in general, not eluted from the C18 reverse-phase column under the present chromatographic conditions or were not efficiently solubilized prior to injection. This was the case for a number of tryptic peptides from the cellulose binding domain (462-497) and

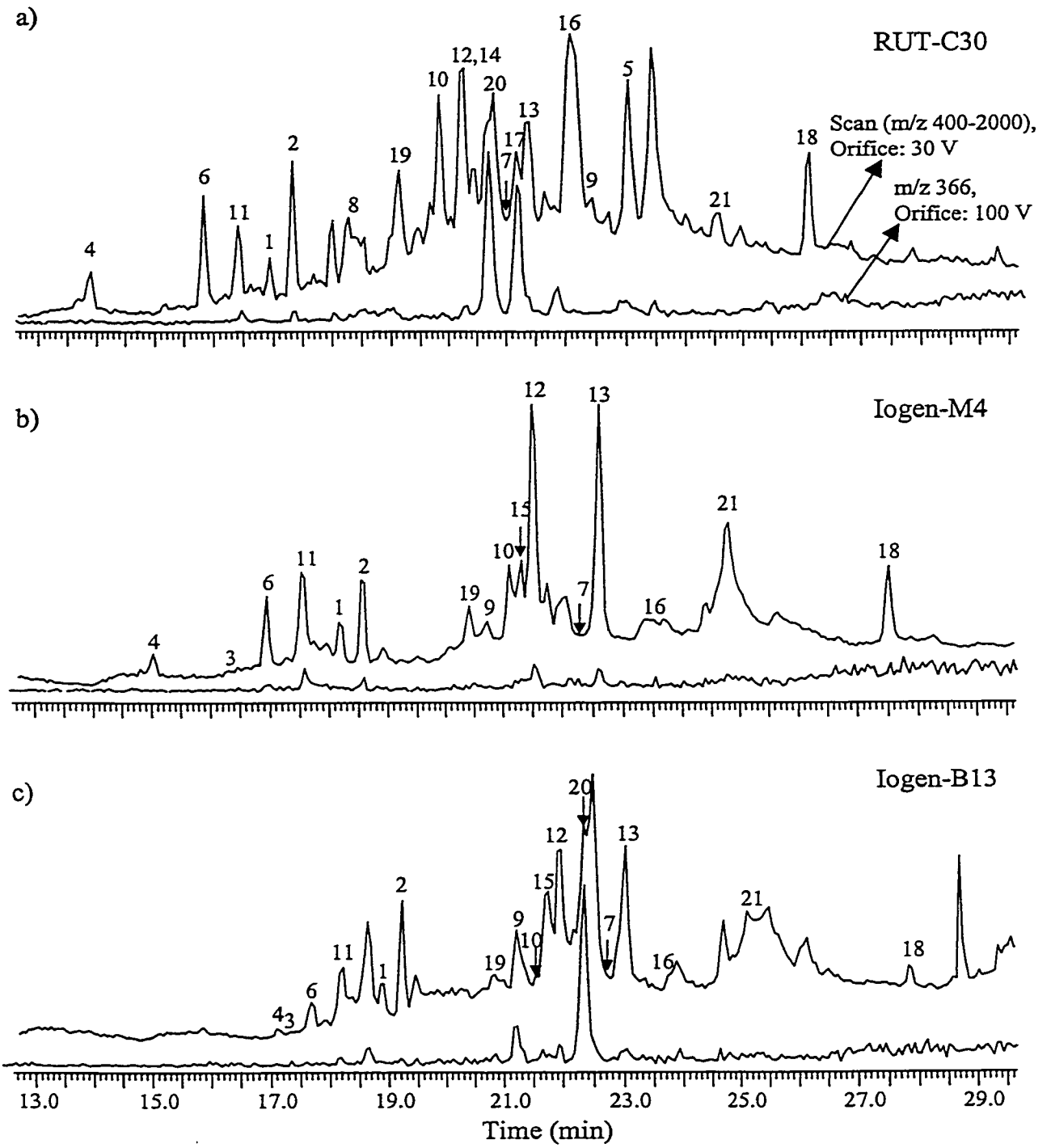


Figure 4.6 cLC-ESMS analysis of tryptic digests of CBH I for the three investigated strains. (a) RUTC30; (b) Iogen-M4; (c) Iogen-B13. The number and corresponding amino acid sequences are shown in Table 4.2. Separation conditions as described in section 2.8.3.

Table 4.2: Assignment of tryptic peptides of CBH I from *T. reesei*

Peak #	Residues	Assignment	$Mass_{mono.}$(Da)	CBH I (tryptic masses)[a]		
				RUT-C30	Iogen-M4	Iogen-B13
1	347-353	GGLTQFK	749.4	749.4	749.3	749.3
2	416-422	VTFSNIK	807.4	807.4	807.3	807.3
3	347-354	GGLTQFKK	877.5	-	877.4	877.4
4	158-166	YPTNTAGAK	921.5	921.4	921.4	921.4
5	355-364	ATSGGMVLVM	964.5	964.4	-	-
6	167-178	YGTGYCDSQCPR	1462.6	1462.4	1462.5	1462.4
7	288-302	LTVVTQFETSGAINR	1634.9	1634.6	1634.6	1634.6
8*	5-18	TLQSETHPPLTWQK	1664.9	1664.7	-	-
9	287-302	KLTVVTQFETSGAINR	1763.0	1762.8	1762.8	1762.8
10	252-267	YGGTCDPDGCDWNPYR	1931.7	1931.6	1931.4	1931.4
11	395-415	GSCSTSSGVPAQVESQSPNAK	2076.9	2076.9	2076.6	2076.9
12*	1-18	pESACTLQSETHPPLTWQK	2094.0	2093.8	2093.7	2093.7
13*	268-286	LGNTSFYGPGSSFTLDTTK + GlcNAc	2195.0	2194.8	2194.8	2194.8
14	19-39	CSSGGTCTQQTGSVVIDANWR	2283.0	2283.4	-	-
15*	268-287	LGNTSFYGPGSSFTLDTTKK + GlcNAc	2323.1	-	2322.9	2322.9
16*	374-394	MLWLDSTYPTNETSSTPGAVR + GlcNAc	2528.2	2528.1	2528.1	2528.1
17*	268-280	LGNTSFYGPGSSF + $Man_8GlcNAc_2$	3035.2	3034.8	-	-
18*	365-394	SLWDDYYANMLWL DSTYPTNETSSTPGAVR + GlcNAc	3655.6	3655.2	3654.9	3655.2
19*	40-69	WTHATNSSTNCYDGNTWSSTLC PDNETCAK + GlcNAc	3680.5	3680.4	3680.0	3680.4
20*	268-286	LGNTSFYGPGSSFTLDTTK + $Man_8GlcNAc_2$	3694.5	3694.2	-	3693.9
21	303-346	YYVQNGVTFQQPNAELGSYSGNEL NDDYCTAEEAEFGGSSFSDK	4887.1	4887.6	4888.0	4887.6

[a] Refers to ^{12}C monoisotopic component

* Tryptic peptides subjected to MS-MS analyses

Cysteine is carbamidomethylated

linker (437-461) regions. As a result, the sequence coverage obtained by cLC-ESMS was only 53 %. However, MALDI-TOF analysis of the corresponding tryptic digest confirmed the presence of additional peptides including those corresponding to amino acids 108-157, 182-251 and a relatively large and heterogenous peptide at approximately 10 kDa corresponding to the combined linker and CBD domains. The combined results from both MALDI-TOF and cLC-ESMS analyses provided close to 92 % coverage of the entire CBH I sequence.

More importantly, the present cLC-ESMS analysis enabled the identification of potential glycopeptides from the RUT-C30 and Iogen-B13 CBH I samples such as peak 20 in Table 4.2. The extracted mass spectrum of the corresponding peak showed an abundant $[M+3H]^{3+}$ ion at m/z 1232.3 consistent with a tryptic peptide of 3693.9 Da (Figure 4.7a). The cLC-ESMS analysis of this tryptic fragment incubated with PNGase F (Figure 4.7b) resulted in a later eluting doubly-charged tryptic peptide of m/z 997.3 (Figure 4.8). The molecular mass of this deglycosylated peptide (1992.6 Da) was consistent with the mass predicted for the tryptic fragment 268-286.

4.4.2 LC-ESMS-MS Analysis of Tryptic Peptides from CBH I

The product ion of m/z 1232.3 corresponding to peak 20 for Iogen-B13 CBH I (Figure 4.7a) is presented in Figure 4.9a and shows fragment ions associated with cleavages of glycosidic and peptide bonds. For example, a series of doubly-charged fragment ions at m/z 996.9, 1098.4, 1200.0, 1281.5, 1363.0 and 1443.4 corresponded to loss of hexose (Man) and acetylhexosamine (GlcNAc) residues supporting the proposal of a high mannose N-linked glycan. The identity of the corresponding tryptic peptide was assigned to 268-286 based on the observation of a series of y-type fragment ions (Figure 4.9a). The site of N-linked

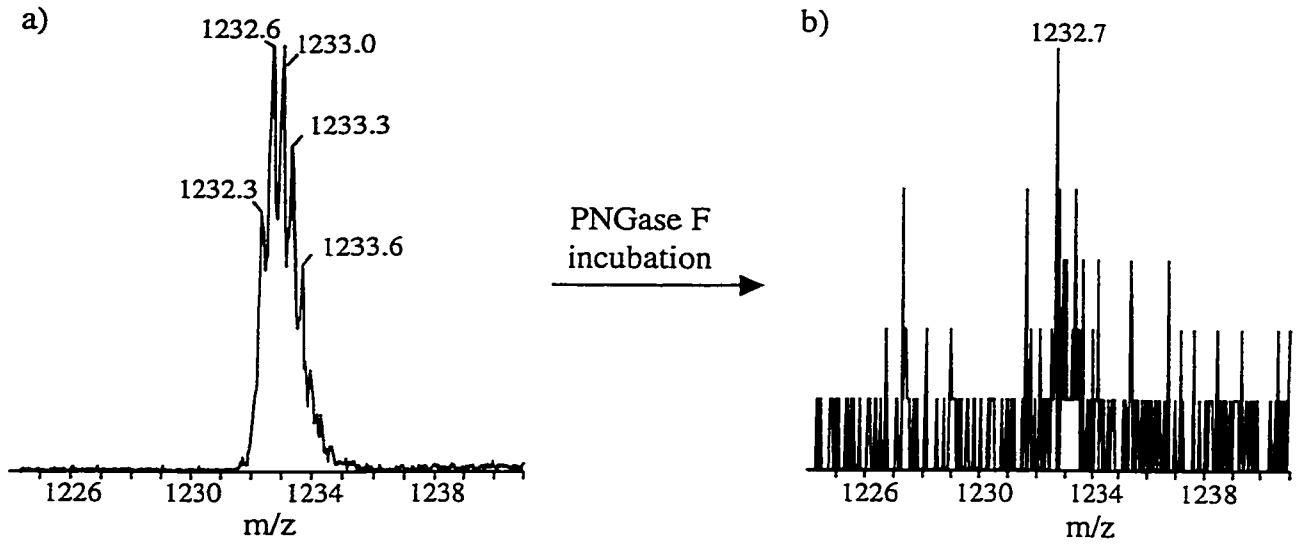


Figure 4.7 The expanded region of the triply-charged ion with $Man_8GlcNAc_2$ attached to Asn270 corresponding to peak 20 from CBH I (RUT-C30 or Iogen-B13) (a) before and (b) after PNGase F incubations.

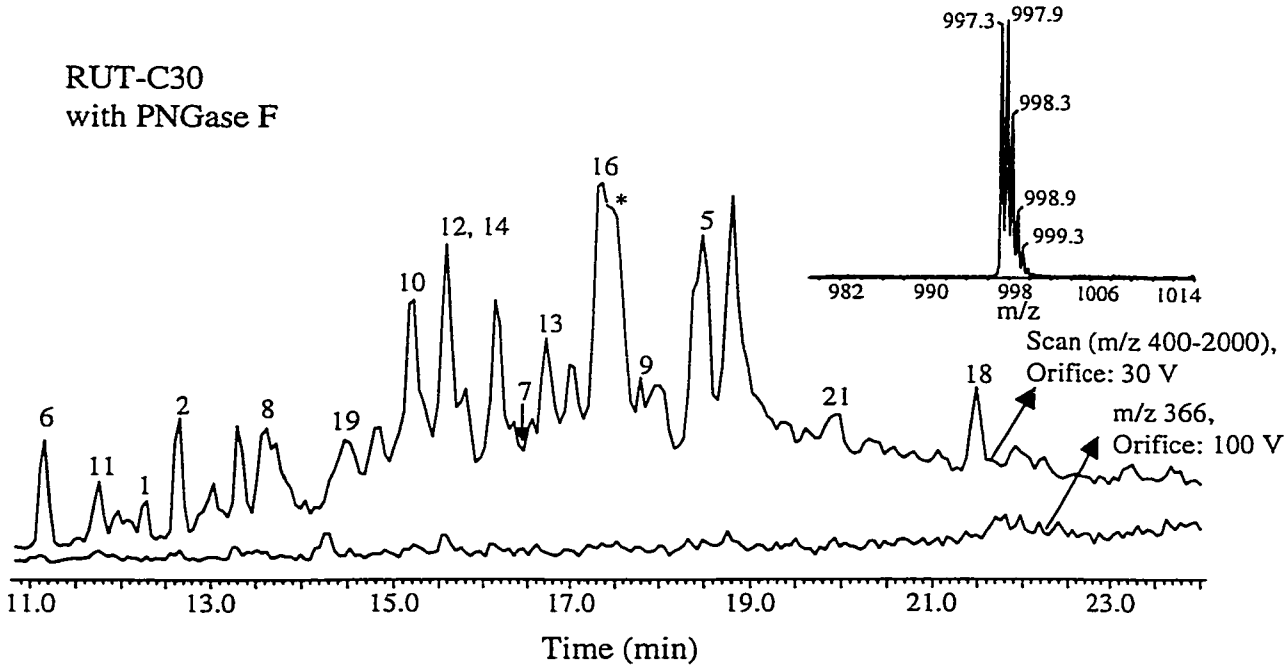


Figure 4.8 cLC-ESMS analysis of tryptic digests of CBH I from RUT-C30 with PNGase F incubations. Inset corresponds to the extracted mass spectrum of the peak with asterisk which appeared after PNGAse F incubations.

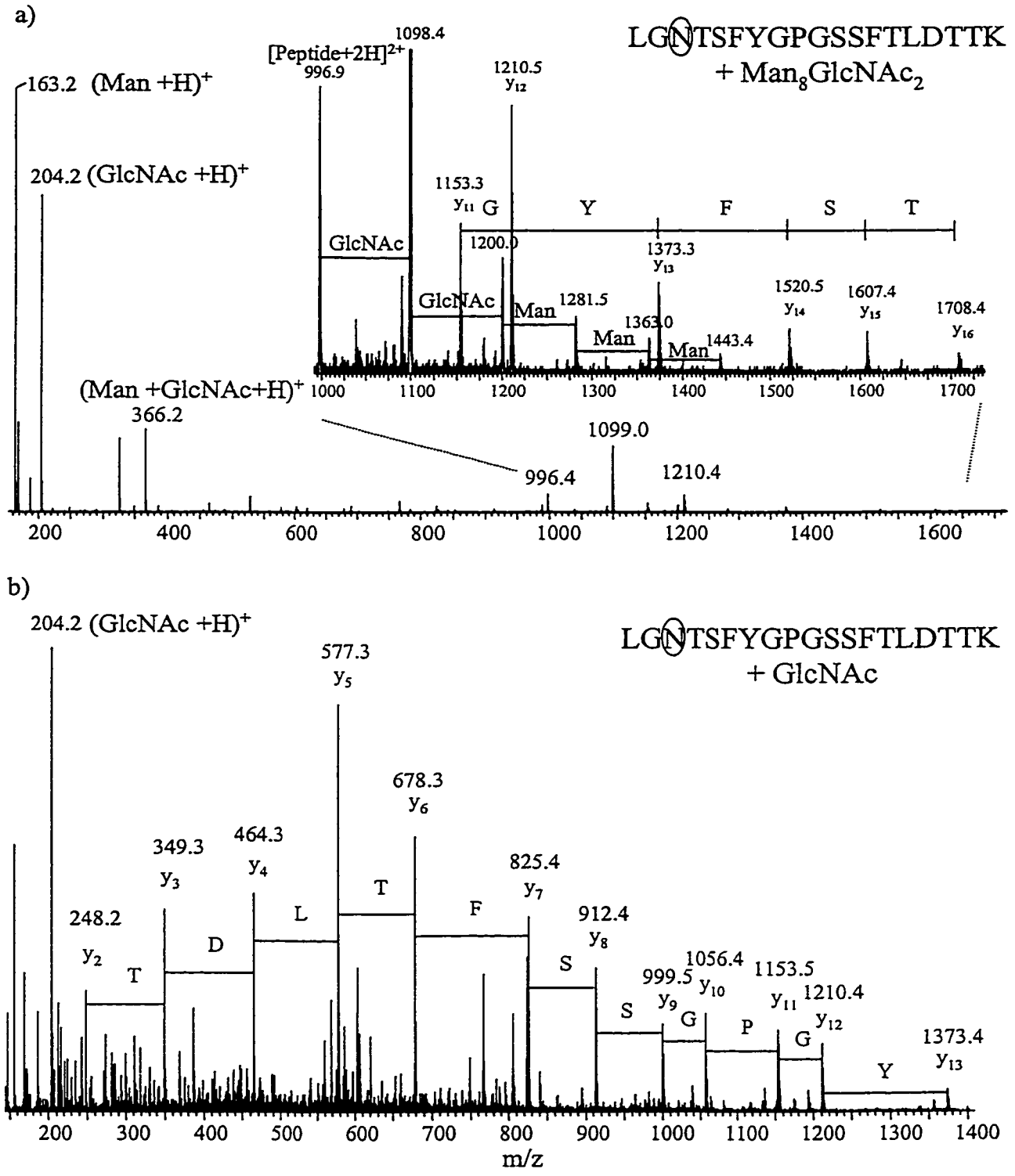


Figure 4.9 MS-MS spectra of N-linked glycopeptides of CBH I tryptic digest. (a) Product ion scan of m/z 1232.3 (peak 20 in Table II) confirming the presence of high mannose $Man_8GlcNAc_2$ on Asn270; (b) Product ion scan of m/z 732.6 (peak 13 in Table II) confirming the presence of a single GlcNAc on Asn270. Conditions as described in 2.7.3 except collision energy (laboratory frame of reference) of 90 eV for (a) and 75 eV for (b).

attachment was assigned to Asn270 (consensus sequence Asn-Thr-Ser) and comprised a predominant $Man_8GlcNAc_2$ glycan with minor Man_5-Man_9 glycoforms as displayed in Figure 4.10. This result is consistent with previous reports of the release of these other high mannose-type N-linked glycans from the CBH I from RUT-C30.[23]

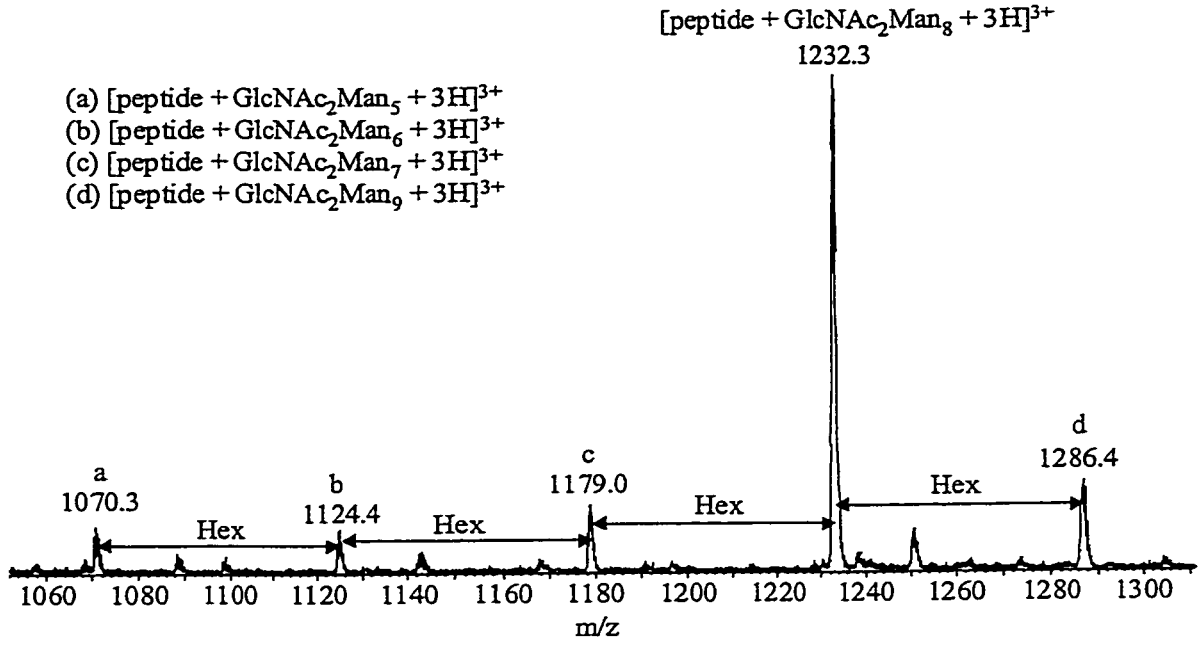


Figure 4.10 Expanded region of the triply charged ion of peak 20 from CBH I (RUT-C30) showing the heterogeneity in high mannose at Asn270 from $Man_{5-9}GlcNAc_2$.

Interestingly, a later eluting peak 13 in Table 4.2, with a molecular mass 2194.8 Da was suspected to contain a single GlcNAc residue at the same glycosylation site. The product ion spectrum of the corresponding $[M+3H]^{3+}$ ion at m/z 732.6 is presented in Figure 4.9b, and shows a series of y-type fragment ions consistent with that expected for tryptic peptide 268-286. An abundant fragment ion at m/z 204 (oxonium ion of GlcNAc) with no detectable signal at m/z 163 (oxonium ion of Man) further supported this proposal. The relative proportion of Asn270 with high mannose to that with single GlcNAc was determined by using the integrated peak area. For CBH I from RUT-C30, the ratio of $Man_8GlcNAc_2$ relative

to single GlcNAc at Asn270 was calculated to be 7:1, while in Iogen-B13, this ratio was found to be 1:2. In contrast, Iogen-M4 only contained a single GlcNAc at Asn270.

The detection of a N-linked glycopeptide bearing a single GlcNAc residue at Asn270 was not totally unexpected. Indeed, both Klarskov et al.[24] and Harrison et al.[25] have reported that the catalytic core domain of CBH I contains three N-linked glycans each consisting of a single GlcNAc. In the present work, CBH I was isolated from *T. reesei* strains of different lineage to that of QM9414 and ALKO2877 (a derivative of QM9414) and variability in site occupancy or glycan heterogeneity could be expected. Alternately, the growth conditions under which CBH I was produced by the various strains could explain some of the glycan heterogeneity at Asn270 as well as the other N-glycosylation sites.

Interestingly, the tryptic peptide for amino acids 355-394 containing a putative N-linked site at Asn384 was not observed in any of these tryptic digests. Rather, a truncated tryptic glycopeptide for amino acids 365-394 bearing a single GlcNAc at Asn384 of mass 3655.2 Da (peak 18 in Table 4.2) was detected. Tandem mass spectrometry experiments on the product ion of m/z 1219.4 ($[M + 3H]^{3+}$) resulted in both the y-type and [y+GlcNAc]-type ions (Figure 4.11a). In a subsequent experiment, the prominent ion at m/z 760.3 corresponding to y_{13} plus a GlcNAc was selected as a precursor and the corresponding product ions showed the GlcNAc attachment commencing at y_{11} (Figure 4.11b), and indicated that the N-linked site is indeed Asn384. In both spectra, the diagnostic oxonium ion at m/z 204.1 was clearly visible.

The tryptic peptide corresponding to amino acids 40-69 was observed as peak 19 (Table 4.2) and its measured molecular mass, 3680.4 Da suggested the incorporation of a single GlcNAc on either Asn45 or Asn64. The first generation of product ion from the corresponding $[M+3H]^{3+}$ at m/z 1227.8 (Figure 4.12a) shows a prominent oxonium ion at

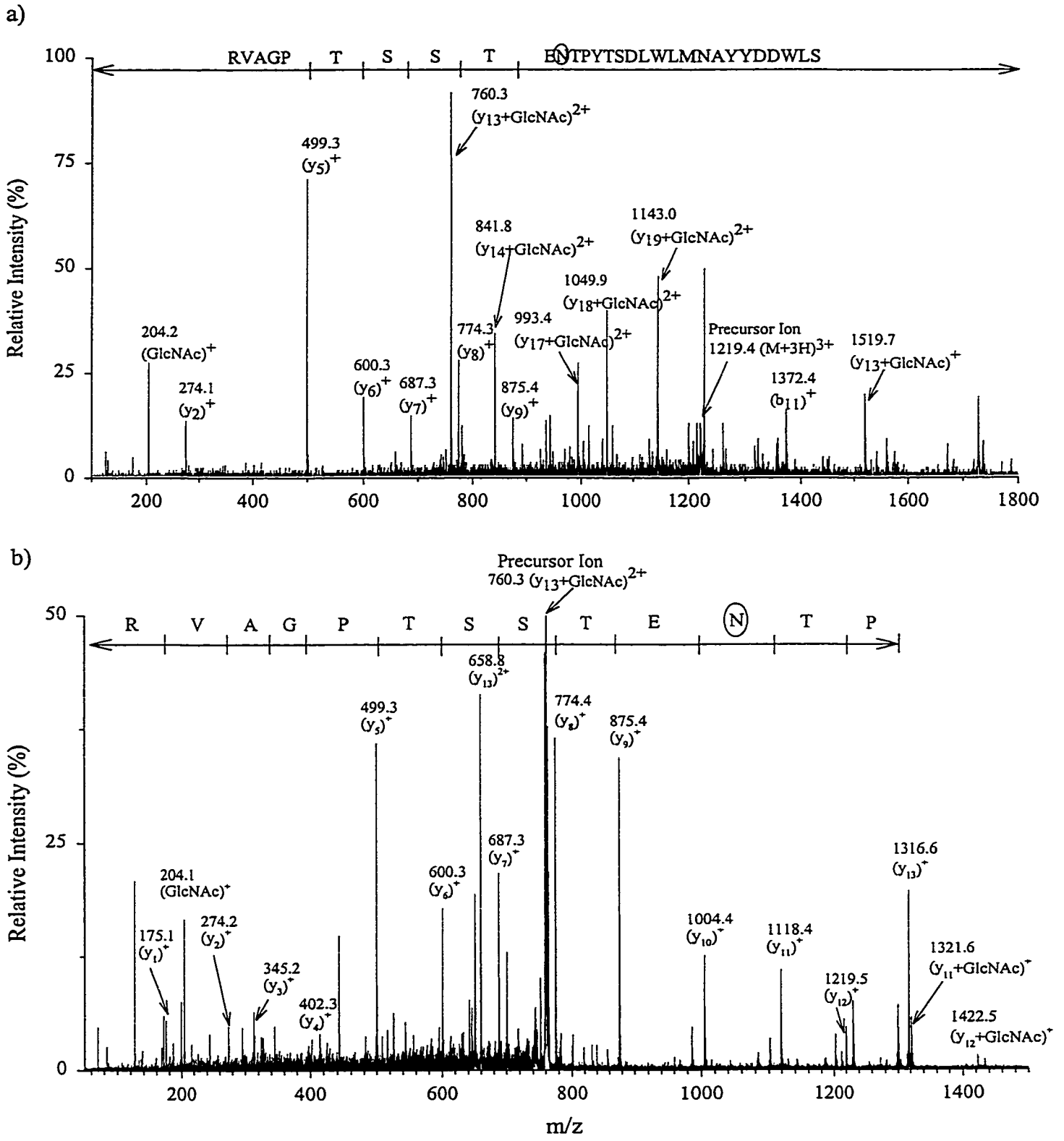


Figure 4.11 MS-MS spectra of CBH I tryptic peak 18 from Iogen-M4 (a) First generation product of m/z 1219. 4 at an orifice voltage of 50V; (b) second generation product ion of m/z 760.3 formed at an orifice voltage of 120 V. Conditions as described in 2.7.3 except collision energy (laboratory frame of reference) of 105 eV for (a) and 70 eV for (b).

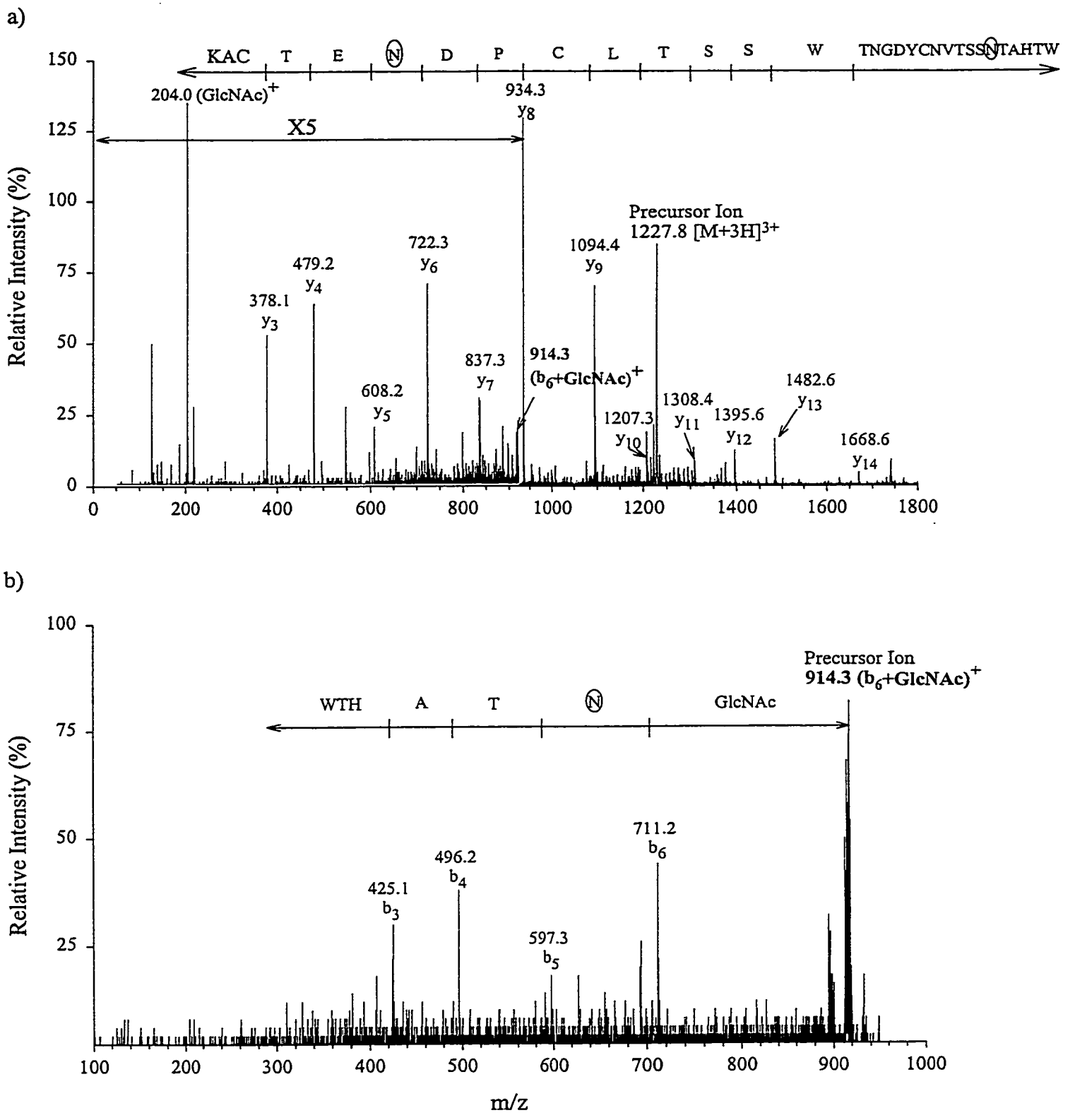


Figure 4.12 MS-MS spectra of CBH I tryptic peak 19 from Iogen-M4 (a) First generation product of m/z 1227.8 at an orifice voltage of 50 V; (b) second generation product ion of m/z 914.3 formed at an orifice voltage of 120 V. Same condition as described in 2.7.3 except collision energy (laboratory frame of reference) of 75 eV for (a) and 50 eV for (b).

m/z 204 together with a number of structurally-related y-type fragment ions. As indicated no single GlcNAc was found on the first y_1-y_{10} fragment ions suggesting that glycosylation occurs on Asn45. This was confirmed by selecting the precursor ion at m/z 914.3 corresponding to the singly-charged fragment ion b_6 plus GlcNAc formed in the orifice/skimmer region of the Q-TOF instrument by raising the orifice voltage to approximately 120 V (Figure 4.12b). The second generation fragment ions thus obtained showed a characteristic fragment at m/z 711.2 resulting from the loss of a single GlcNAc residue from the precursor ion along with a number of consecutive b-type fragment ions from which the glycosylation site could be unambiguously assigned. Based on these tandem mass spectrometric data Asn45 comprised a single GlcNAc residue while Asn64 was not glycosylated consistent with previous reports.[24,25]

The variability in $Man_8GlcNAc_2$:GlcNAc ratios observed for Asn270 in CBH I produced by the three *T. reesei* strains in the present study is unusual and may reflect differences in the activities of trimming glycosidases (such as glucosidase II, α-1,2-mannosidase or endoglycosidase H) produced by the different strains or under different growth conditions. Most notably, Iogen-M4, whose CBH I contained only GlcNAc at this position, was grown at pH 5 while both RUT-C30 and Iogen-B13, whose CBH I showed some $Man_8GlcNAc_2$ at Asn270, were grown at pH 4. The partial deglycosylation at Asn270 may be a direct consequence of steric hindrance at this site whereas glycans appended to residues Asn45 and Asn384 would be more directly exposed to the outer protein core as predicted from the three dimensional crystal structure of CBH I.[13,14] Further experiments to verify this possibility would be warranted.

In addition, the predominant high mannose-type glycan from RUT-C30 CBH was found to be $Man_8GlcNAc_2$ while Maras et al.[23] previously showed that the predominant N-linked

glycans from RUT-C30 CBH I were $GlcMan_8GlcNAc_2$, $GlcMan_7GlcNAc_2$ and ManP-$GlcMan_7GlcNAc_2$. These differences in glycosylation pattern are likely to be the result of the different growth conditions under which RUT-C30 was grown. In the present study, the enriched medium used to grow RUT-C30 strain may have induced higher glucosidase activity while the minimal medium used in the previous study may have led to higher levels of mannosylphosphotransferase activity. The observation of this type of activity in *Trichoderma* would not be unexpected since mannosyl phosphorylation of N-and O-linked glycans has been observed from a number of *S. cerevisiae* strains[105] and has been shown to occur in response to cellular stress resulting from environmental changes or growth phase.[106]

4.5 MALDI-TOF analysis of CBHI linker domain

Purification of the peptide fragment from papain proteolysis of CBH I using gel filtration chromatographic purification revealed a highly heterogeneous glycopeptide of 8,000-9,500 Da comprising the linker and the cellulose-binding domain. MALDI-TOF analyses of these samples (Figures 4.13a-d) showed a complex profile displaying an incremental number of hexose residues. Degeneracy in the papain digest resulted in peak doublets spaced by 57 Da corresponding to cleavage on either side of Gly439. In addition to this, the glycopeptide obtained from papain digestion of CBH I from RUT-C30 (Figure 4.13a) showed a third set of peaks separated by 80 Da from the major and minor series described above. This mass spacing was also noted in the intact CBH I and suggested the presence of a single sulfate or phosphate group on the linker. Incubation of this glycopeptide with alkaline phosphatase (Figure 4.13b) successfully removed the phosphate group on most linker glycoforms except for the larger oligomer series above m/z 8700. This observation possibly suggests that

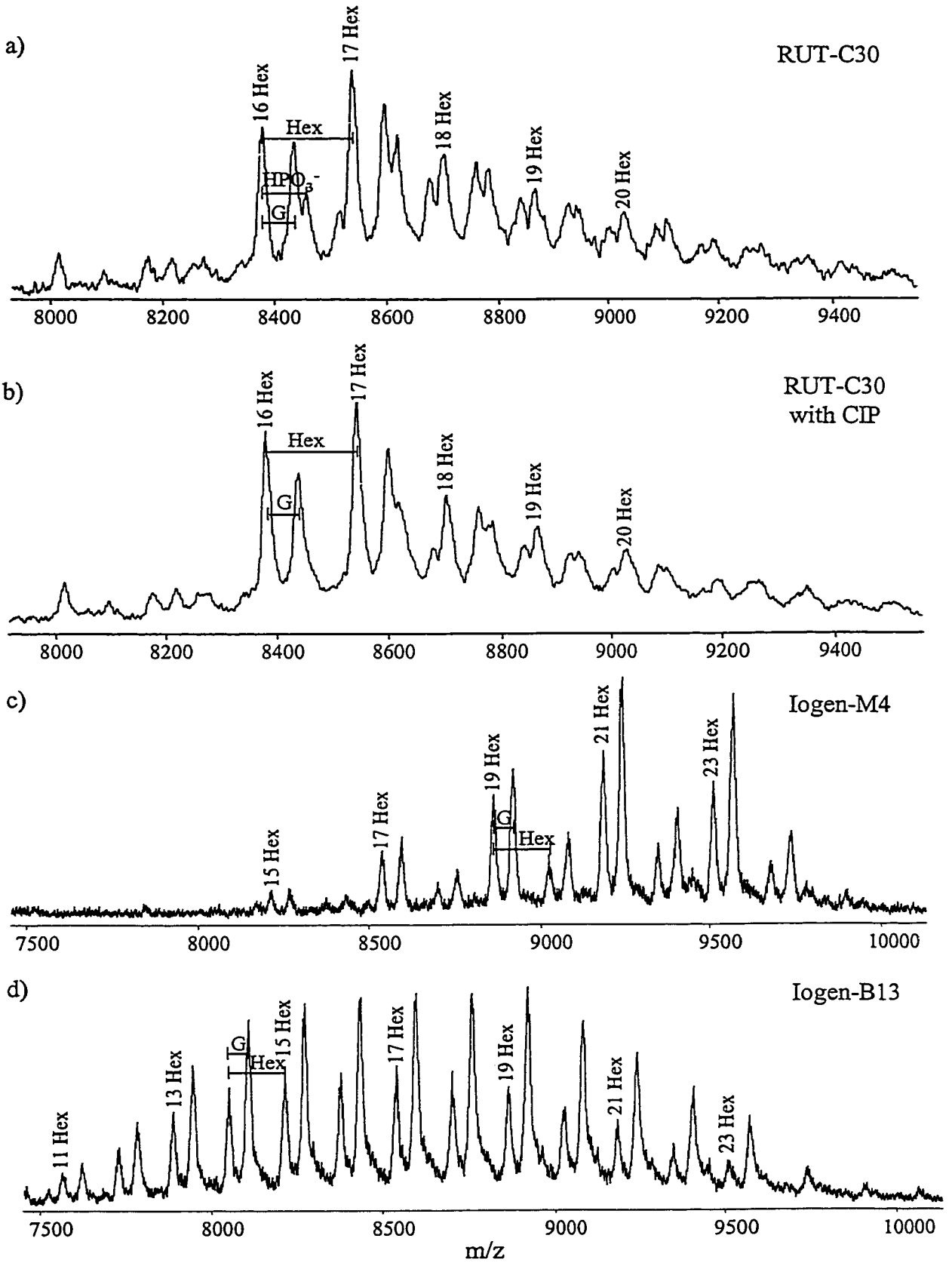


Figure 4.13 MALDI-TOF analysis of the linker/CBD domains following papain digestion of CBH I. (a) RUT-C30; (b) RUT-C30 after alkaline phosphatase incubations; (c) Iogen-M4 and (d) Iogen-B13 (G = glycine).

extensive glycosylation of the linker might impart steric hindrance preventing the catalytic activity of the phosphatase. This result along with purified saccharides from HPAEC-PAD experiments described in the next section indicates that the linker from RUT-C30 is phosphorylated. The presence of a phosphate group on the linker and not a sulfate as reported previously for strain ALKO2877 might reflect the different lineage of the strains from which the CBH I was isolated.[25]

From the mass of the glycopeptide observed in Figure 4.13, it was possible to deduce the number of hexose residues appended to the linker region. In the case of RUT-C30 the major glycoform series corresponded to the mass of the C-terminal peptide (M: 5834.4 Da) containing 16-21 hexose residues. Similar glycoform distributions were observed for the linker of Iogen-M4 (Figure 4.13c) and Iogen-B13 (Figure 4.13d) where the number of hexose varied from 15-24 and 11-24, respectively.

4.6 HPAEC-PAD Analysis of Oligosaccharides Released by Hydrazinolysis

The chromatograms corresponding to oligosaccharides released from the hydrazinolysis of CBH I at 95°C are shown in Figures 4.14a-c for the three different strains. Under the present chromatographic conditions, the small neutral di- and tri-saccharides eluted before 10 min whereas higher oligomer such as $Man_{5-9}GlcNAc_2$ were observed between 15 and 20 min. As indicated, CBH I from Iogen-M4 and Iogen-B13 contained a high proportion of di- and tri-saccharides presumably located on the linker region as reported previously by Harrison et al.[25] CBH I from RUT-C30 contained predominantly disaccharides and a high mannose glycan ($Man_8GlcNAc_2$) on Asn270 as indicated in Figure 4.14a. The same high mannose

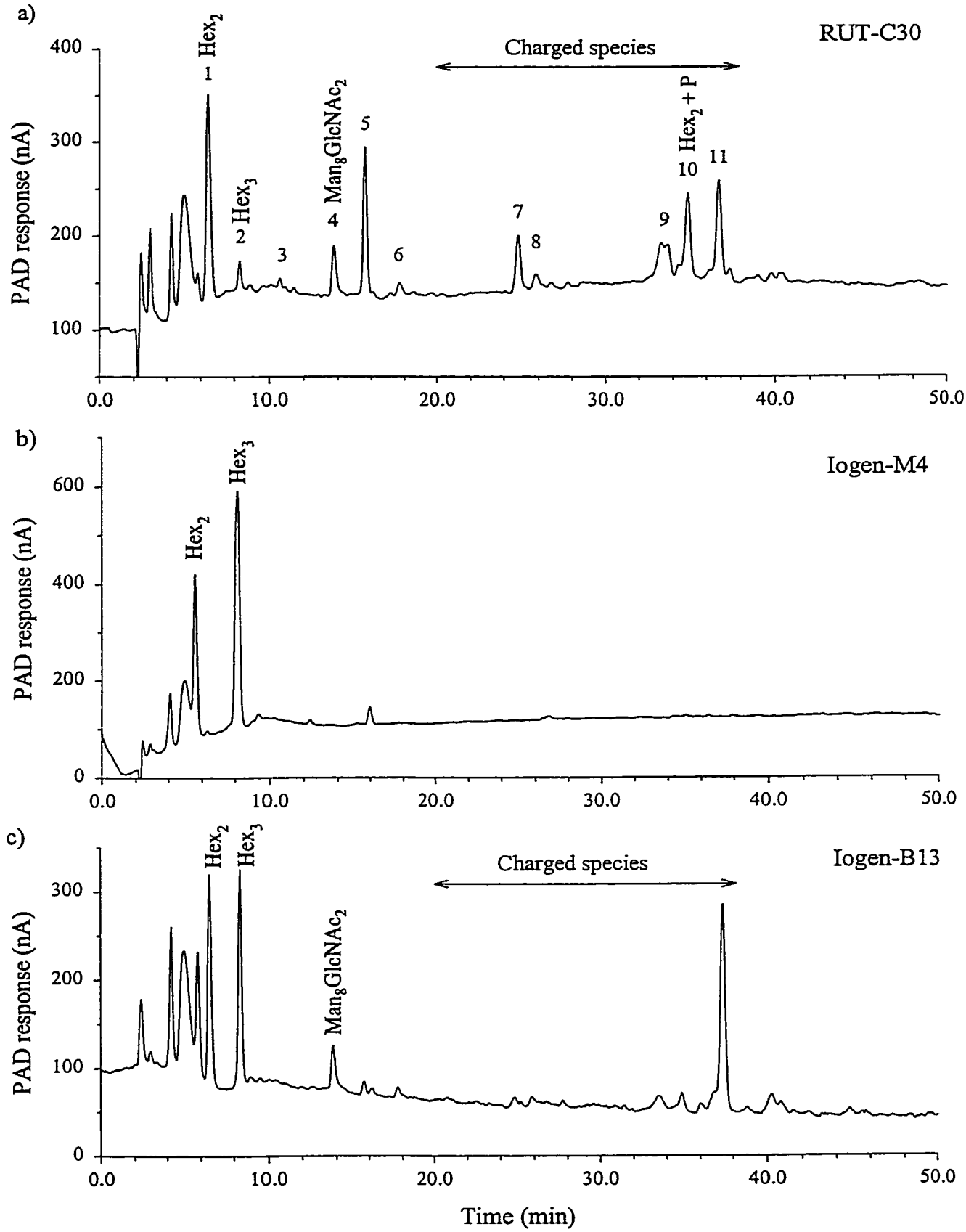


Figure 4.14 HPAEC-PAD analysis of oligosaccharides released from automated hydrazinolysis (N + O linked) of CBH I from the investigated strains. (a) RUT-C30; (b) Iogen-M4 and (c) Iogen-B13.

migrating at a similar position was also observed in CBH I from Iogen-B13 (Figure 4.14c). There was no high mannose or any anionic species detected in CBH I from Iogen-M4 (Figure 4.14b). Interestingly, additional components eluting between 38-44 min were observed mainly in the RUT-C30 strain. These extra peaks were presumed to be strongly retained anionic species. Confirmation of the identity of some of these peaks was achieved by conducting on-line desalting of the separated components followed by ESMS analyses.

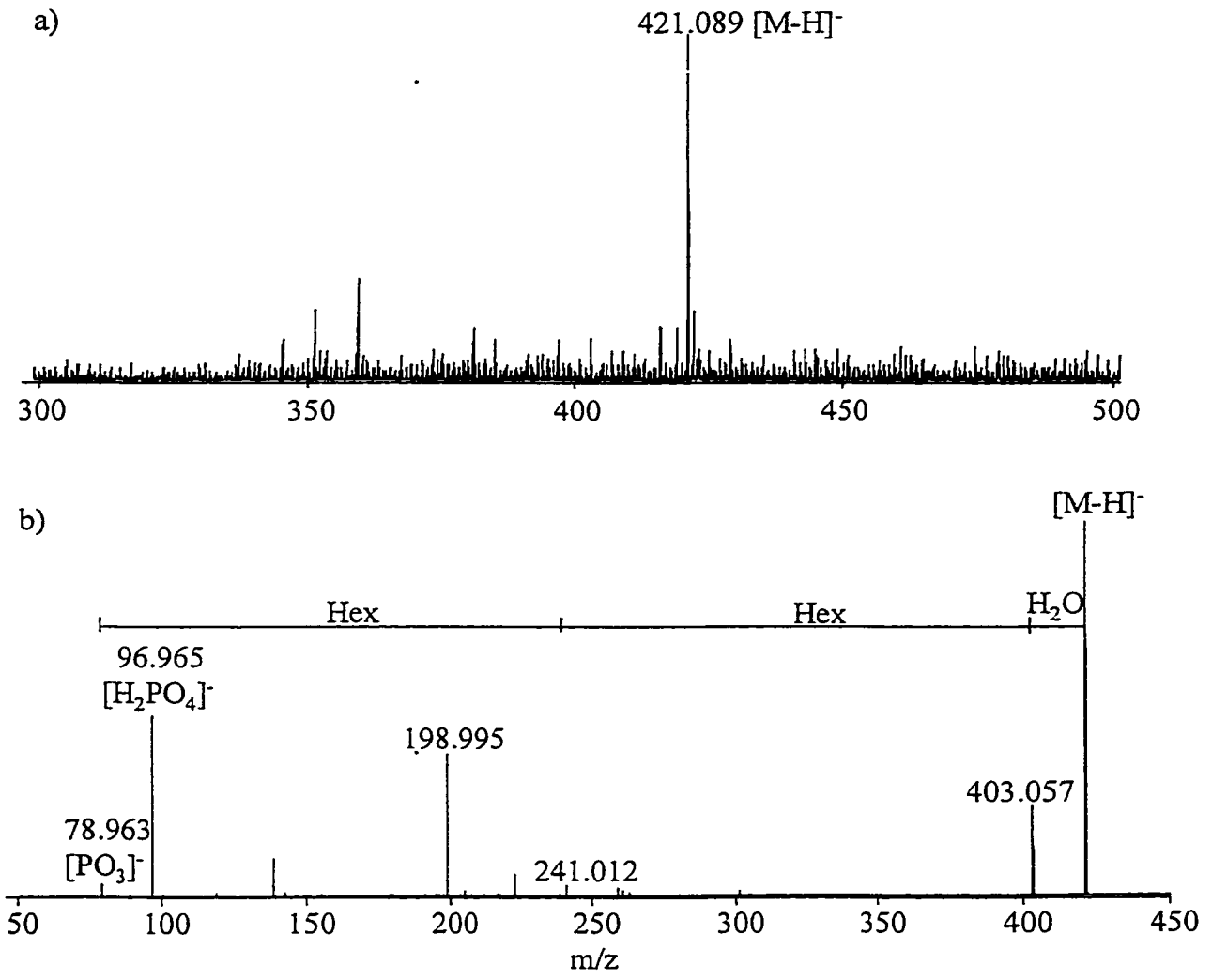


Figure 4.15 ESMS analysis (quadrupole/time-of-flight) of fraction 10 from HPAEC-PAD purification of oligosaccharides derived from CBH I (Figure 4.14). (a) Negative ion ESMS mass spectrum and (b) product ion spectrum of m/z 421 obtained at a collision energy of 40 eV (laboratory frame of reference).

The mass spectral analysis of peak 10 from the HPAEC-PAD separation of glycans from CBH I of RUT-C30 is shown in Figures 4.15a-b. The conventional negative ion ESMS spectrum (Figure 4.15a) of the desalted sample showed a prominent $[M-H]^-$ ion at m/z 421. Accurate mass measurements on this ion (m/z 421.080) were consistent with a phosphorylated di-hexose ($[M-H]^-_{calc}$: 421.075) compared to a sulfate disaccharide ($[M-H]^-_{calc}$: 421.065). The product ion spectrum of the corresponding singly-deprotonated precursor (Figure 4.15b) indicated several fragment ions corresponding to losses of water and cleavage of glycosidic bonds including diagnostic fragments ions at m/z 96.965 and 78.963 supporting the presence of a phosphate group. The mass of this component together with its characteristic fragmentation patterns were consistent with a phosphorylated dihexose glycan. Based on the analysis of the papain digest of CBH I from RUT-C30 (section 4.5), this unusual carbohydrate was assigned as a phosphorylated disaccharide on the linker glycopeptide. An unassigned peak at m/z 198.995 may correspond to a ketene loss ($H_2C = C = O$) from m/z 241.012 due to internal ring cleavage.

Mannosyl phosphorylation on O-linked glycans was also described for mannotriose oligosaccharides in mannosyltransferase mutants (*mnn 1*) strains of *S. cerevisiae* .[107] The occurrence and significance of phosphorylated O-linked glycans in CBH I from *T. reesei* is not yet understood. However, recent report from Jigami and Odani suggests a number of functions for mannosyl transferase in yeast including the supply of GMP for sugar nucleotide transport in the Golgi, cross-linking of mannoproteins to ß-glucan and cellular stress responses.[106] The relationship between biological functions of secreted cellulases and the mannosylphosphate transfer would merit further investigation.

4.7 Conclusion

The present report described the application of electrospray mass spectrometry and a number of enzymatic digestions for the analysis of cellobiohydrolase I purified from *T. reesei* strain RUT-C30 and two of its derivative strains. The profile of N- and O-linked glycans appended to the protein backbone of CBH I was obtained by flow injection mass spectrometry on a hybrid quadrupole time-of-flight instrument and highlighted the presence of unexpected patterns including the variability in N-linked glycans and phosphorylated residues on the O-linked glycans in the linker regions of the CBH I from RUT-C30.

The precise identification of N-linked glycans on four putative sites was achieved using cLC-ESMS and tandem mass spectrometry. A high/low orifice stepping function was incorporated as part of the mass spectral acquisition enabled the identification of glycopeptides from the complex tryptic digests of CBH I from the three different *T. reesei* strains. The suspected glycopeptides were further subjected to MS-MS to confirm the nature of glycans and their attachment sites. A single GlcNAc was found at Asn45 and Asn384 in CBH I from all strains while Asn270 predominantly contained a high mannose oligosaccharide, $Man_8GlcNAc_2$, in CBH I from RUT-C30, a mixture of $Man_8GlcNAc_2$ and GlcNAc in CBH I from Iogen-B13, and a single GlcNAc in CBH I from Iogen-M4. No N-linked glycans were observed at Asn65 in any of the CBH I samples. The presence of single GlcNAc residues at Asn45 and Asn384 and the heterogeneity in the glycan population at Asn270 may reflect some endoglycosidase H activity in *T. reesei* whose accessibility to the N-linked glycan at Asn270 may be limited by steric hindrance.

The linker region of all three *T. reesei* strains examined showed an heterogenous population of di- and tri-mannosyl saccharides as indicated from the HPAEC-PAD profiles

of the hydrazinolysis products of CBH I. The molecular mass profiles of CBH I from RUT-C30 showed additional heterogeneity in the form of a phosphorylated glycan. Alkaline phosphatase treatment of the linker-CBD peptide together with mass spectral analyses of HPAEC-PAD fractions of acidic oligosaccharides supported the presence of a phosphorylated disaccharide on the linker region. Further investigations will be required to address the significance and functional relationship of the phosphate transfer to CBH I.

You know my methods. Apply them.

Sir Arthur Conan Doyle (1859-1930)

The Sign of Four (1890), Ch. 6

Characterization of Posttranslational Modification in CBH II, EG I and EG II from *T. reesei* Strain RUT-C30

5.1 Introduction

Studies of CBH I from strain RUT-C30 have led to the assignment of a single phosphorylated di-saccharide in the linker domain. This is a first time report in the literature of phosphorylated O-linked glycans in CBH I from *T. reesei*, although previous investigation on mutant strains of *S. cerevisiae* have described a mannosylphosphorylated mannotriose. [107] Such a finding prompted us to investigate the occurrence of mannosylphosphorylation of O-linked oligosaccharide in other cellulases secreted from *T. reesei*. These results also raise the question as to whether CBH I is a specific substrate for this mannosylphosphotransferase or if protein phosphorylation is dependent on the cell cycle or environmental changes. Furthermore, the observation of both single GlcNAc and high mannose at Asn270 of CBH I from RUT-C30 has suggested the expression of endogenous Endo H enzymatic activity. To address these questions and their influences on other cellulases, we undertook the structural characterization of N-linked glycans in other cellulase components.

The current experiments focus on the cellulase analyses from the Iogen strain RUT-C30. This allows for further study of the substrate specificity of the mannosylphosphotransferase and the suspected endogenous Endo H enzymes. Following similar strategy displayed in Figure 2.1, three cellulolytic enzymes from *T. reesei*, namely CBH II, EG I and EG II, were purified from Iogen strain RUT-C30 using both ion-exchange chromatography and chromatofocusing described in section 2.4. Their glycosylation profiles were analyzed by a combination of enzymatic digestions and mass spectrometric techniques.

5.2 Mass Spectral Analysis of Intact CBH II, EG I and EG II from RUT-C30

Individual purified components from RUT-C30 were initially analyzed by MALDI-TOF at a wide mass range (m/z 5,000-100,000, Figures 5.1a-c). In all three spectra, the singly charged ions indicated form the base peaks. The doubly charged ions are also clearly visualized. As expected, CBH II shows the highest mass with average mass of m/z 57649 among the investigated cellulases, followed by EG I and EG II with m/z 53014 and 48196, respectively. These masses are summarized in Table 5.1, together with the masses derived from the protein sequence and the % glycosylation. Interestingly, an additional peak was observed in the high mass range of EG II (Figure 5.1c), which is attributed to the dimer formation of EG II. The observation of dimer formation is not unusual in MALDI-TOF analysis, especially in samples with high concentration.

MALDI-TOF analysis has provided a rapid and convenient way to identify the cellulase, but it failed to resolve the glycoform distribution. As a result, these cellulases were further subjected to flow injection analysis by ESMS to profile the microheterogeneity in glycosylation (Figures 5.2a-d). Contrary to the electrospray mass spectum of CBH I analyzed by the Micromass Q-Tof instrument, CBH II, EG I and EG II from RUT-C30 were analyzed by the PE/Sciex Q-Star. In all three latter cases, the majority of the multiply-charged ion envelope was displayed in the range m/z 2000-3000 (Figure 5.2a). The reconstructed molecular mass profiles of CBH II, EG I and EG II from RUT-C30 are compared in Figures 5.2b-d and their mass distributions agree with the results from MALDI-TOF analyses (Figures 5.1a-c). The CBH II from RUT-C30 showed a minor glycoform distribution from 55,500 – 56,500 Da, which was later proven to be the series with single GlcNAc at Asn310 possibly resulting from expression of endogenous Endo H activity. In all three reconstructed

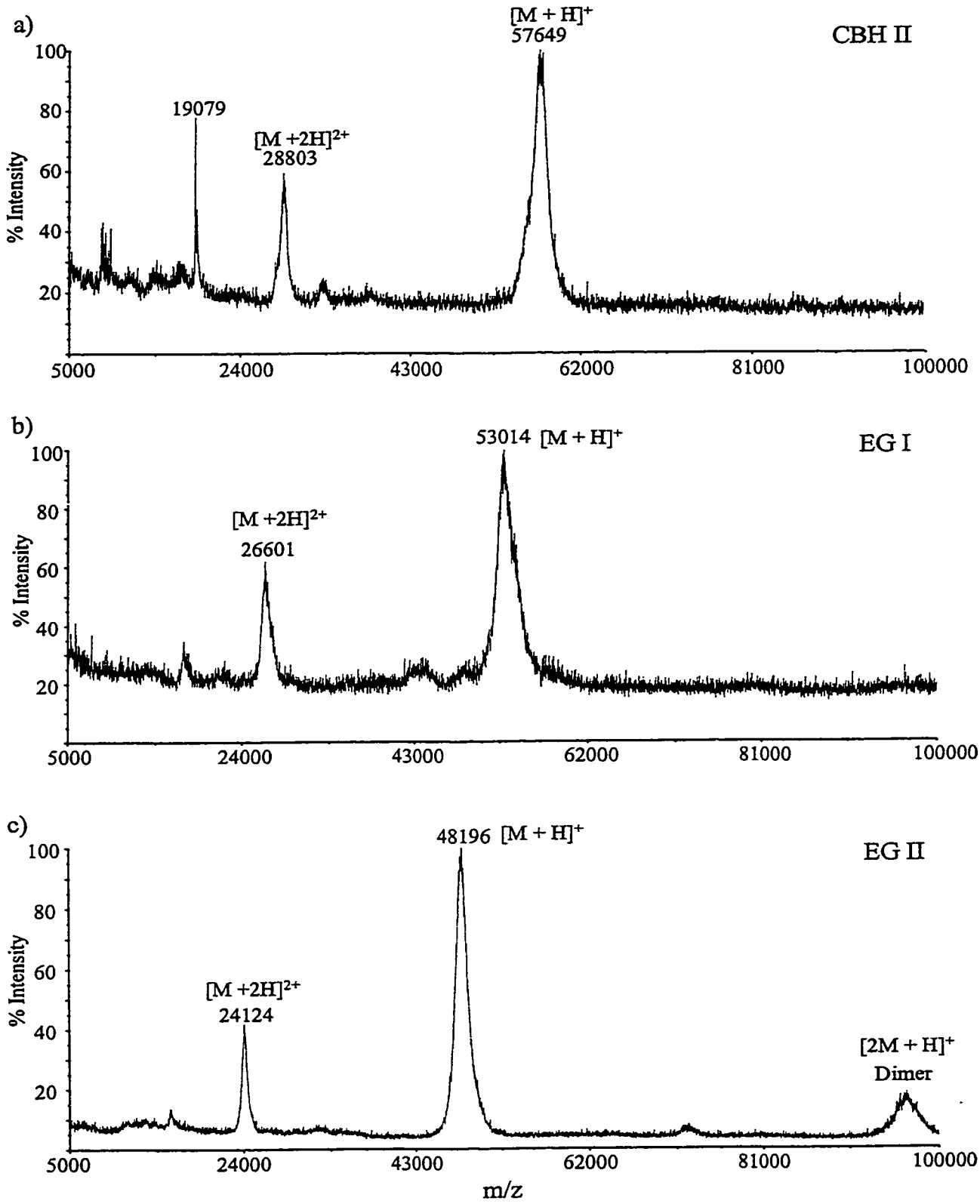


Figure 5.1 MALDI-TOF mass spectra of purified cellulase enzymes from Iogen strain RUT-C30. (a) CBH II; (b) EG I and (c) EG II.

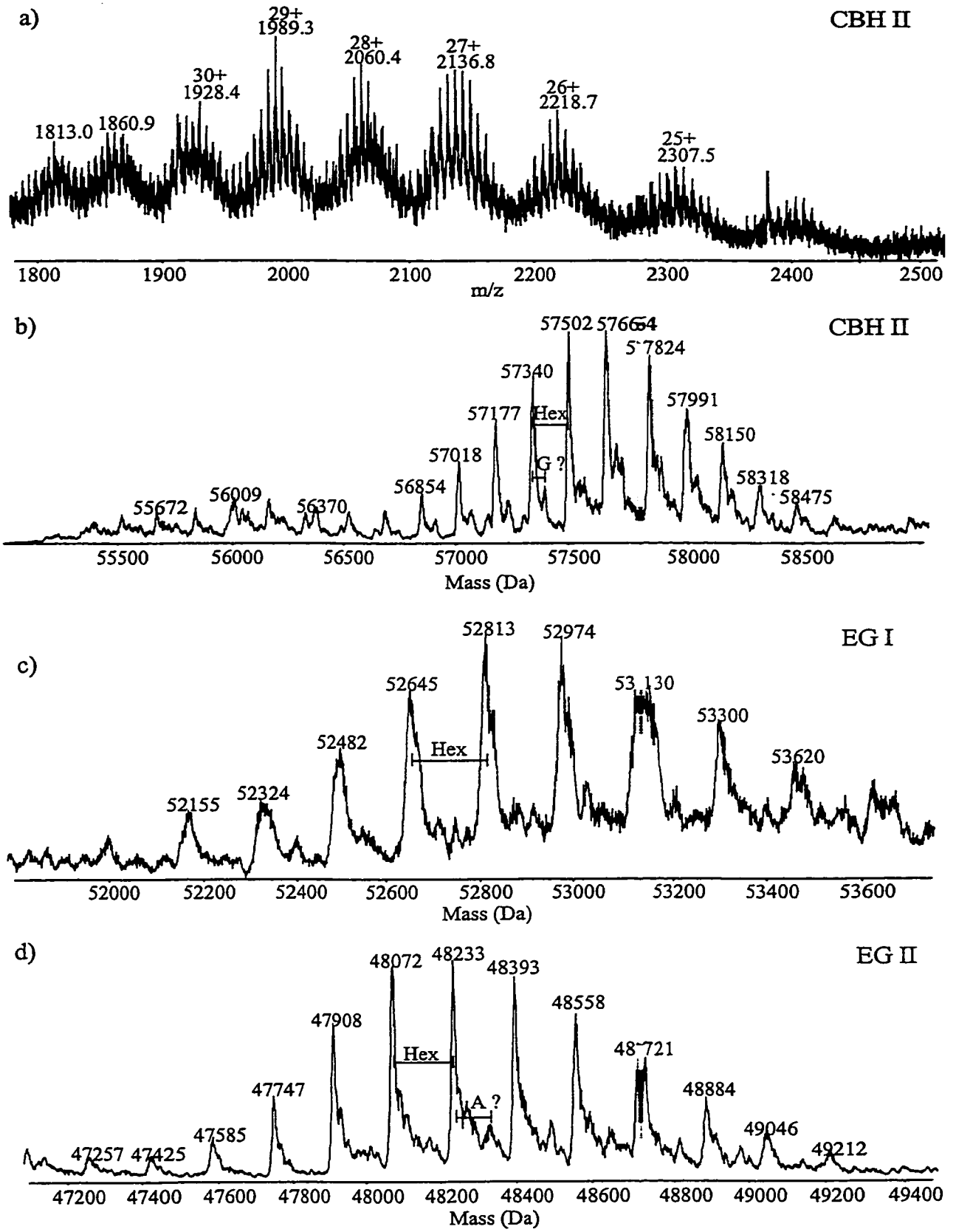


Figure 5.2 ESMS mass spectra of (a) CBH II and reconstructed molecular mass profiles of (b) CBH II; (c) EG I and (d) EG II. All cellulases were purified from Iogen RUT-C30 (G=glycine; A=alanine).

mass spectra, the glycoforms shown are spaced by hexose unit (162 Da) and the corresponding % glycosylation are summarized in Table 5.1. The relative proportion of glycans in CBH II, EG I and EG II with respect to the protein mass was calculated to be 17.7–24.0 %, 13.0-16.5 % and 11.9-16.7 %, respectively. Unlike the CBH I from RUT-C30, there are no 80 Da series of peaks embedded between the glycoforms in these three cellulases, suggesting that no occurrence of mannosylphosphorylation in the linker domains. Interestingly, an additional series of peaks with average mass corresponding to a glycine residue was observed in CBH II (Figure 5.2b), while an extra set of peaks corresponding to an alanine residue was found embedded between the hexose spacings in EG II (Figure 5.2d). These unusual modifications are yet to be confirmed by sequencing the *cbh 2* and *eg 2* genes from *T. reesei* strain RUT-C30.

Table 5.1 Comparison of molecular weight and % glycosylation of major cellulase enzymes from RUT-C30 observed by MALDI-TOF and ESMS

		MALDI-TOF		ESMS	
Cellulase	$Mass_{avg.}$ (Da)	$Mass_{avg.}$ (Da)	% glycosylation	$Mass_{avg.}$ (Da)	% glycosylation
CBH I	52205.8	56829	8.9	56500-58000	8.2-11.1
CBH II	47162.4	57648	22.2	55500-58500	17.7-24.0
EG I	46011.3	53013	15.2	52000-53600	13.0-16.5
EG II	42164.5	48195	14.3	47200-49200	11.9-16.7

5.3 Papain Proteolysis

Papain is a thiol protease extracted from crude papaya latex.[108] It is a non-specific protease generally targeted at the regions of disordered protein structure. Papain proteolysis was reported to release the core protein from cellobiohydrolases.[109,110,111] Purified cellulases from RUT-C30 were digested with papain and subsequently analyzed by MALDI-TOF MS (Figure 5.3a-c). In each case, the released linker/CBD glycopeptide was enlarged in the inset and was shown to be highly heterogeneous. Consistent with the ESMS mass spectra of the intact cellulases, there were no additional 80 Da peak series from phosphate modifications appeared between the hexose spacings.

Interestingly, the released CBH II core from RUT-C30 resulted in two peaks spaced by 1564 Da (Figure 5.4a). From previous MALDI-TOF investigations on CBH I core, this spacing was attributed to variability in N-linked site occupancy. Incubation with Endo H enzymes effectively removed the peak with high mannose, resulting in a single CBH II core protein mass (Figure 5.4b). This experiment also enabled the identification of a single high mannose glycan from three putative N-linked sites. This high mannose oligosaccharide was later identified to be at Asn310 by LC-MS and tandem MS analyses (sections 5.4, 5.6). The intact CBH II in Figures 5.4a-b was also detected due to the limited proteolysis performed for only three hours and its observed mass also decreased after the endoglycosidase H incubation.

Tomme et al. has reported the papain cleavage sites of CBH II from strain QM 9414 by partial sequencing at the N terminus of CBH II core.[109] This position corresponds to the amide linkage between Gly82 and Ser83. Assuming papain has the same cleavage position in

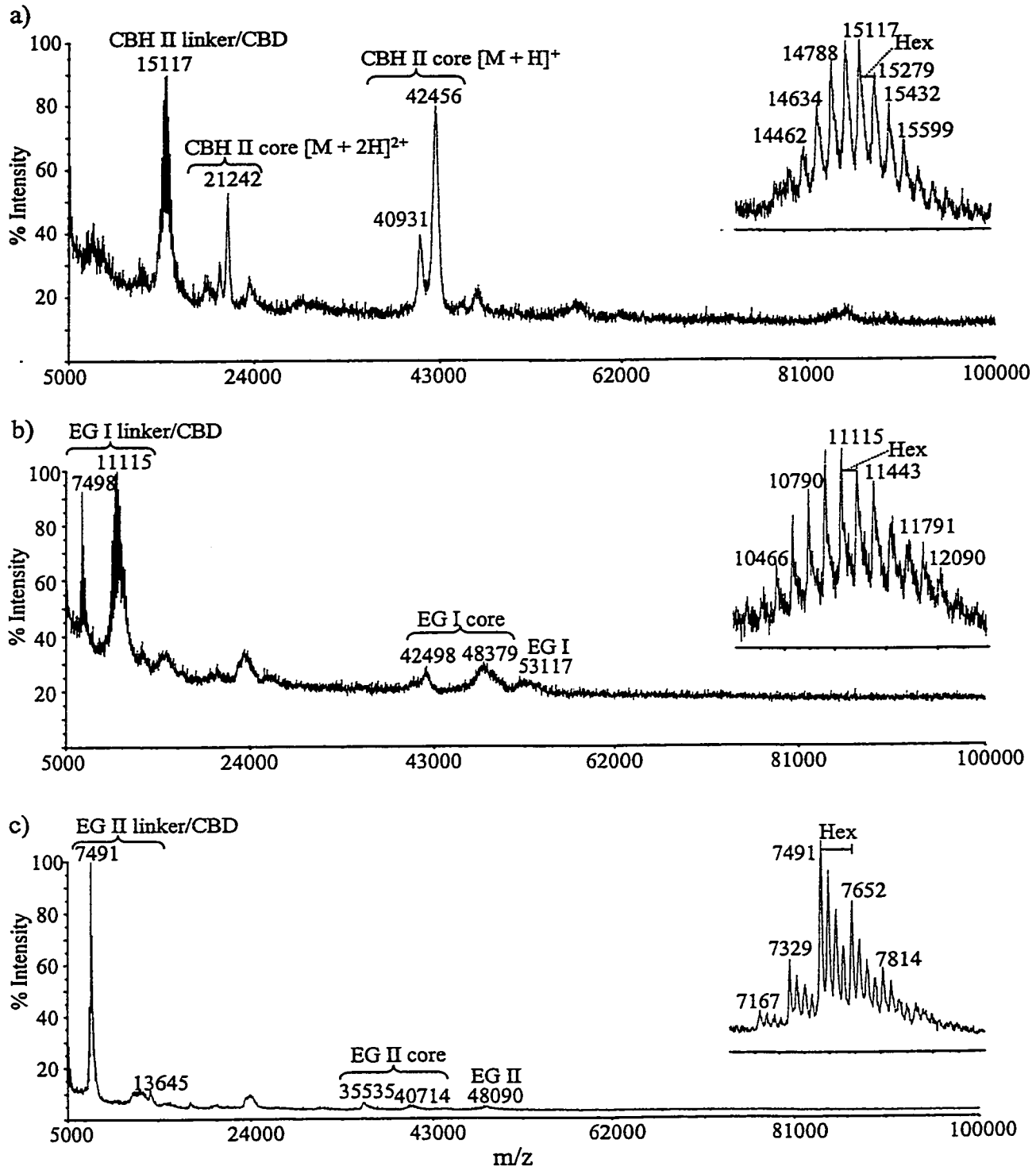


Figure 5.3 MALDI-TOF analysis of papain proteolysis of purified cellulases from RUT-C30. (a) CBH II; (b) EG I and (c) EG II. The released linker and CBD domains were enlarged and shown in the insets.

CBH II from RUT-C30, the heavily heterogenous 15,000 Da linker/CBD shown in the inset of Figure 5.3a would correspond to the O-linked glycans and possibly a N-linked glycan

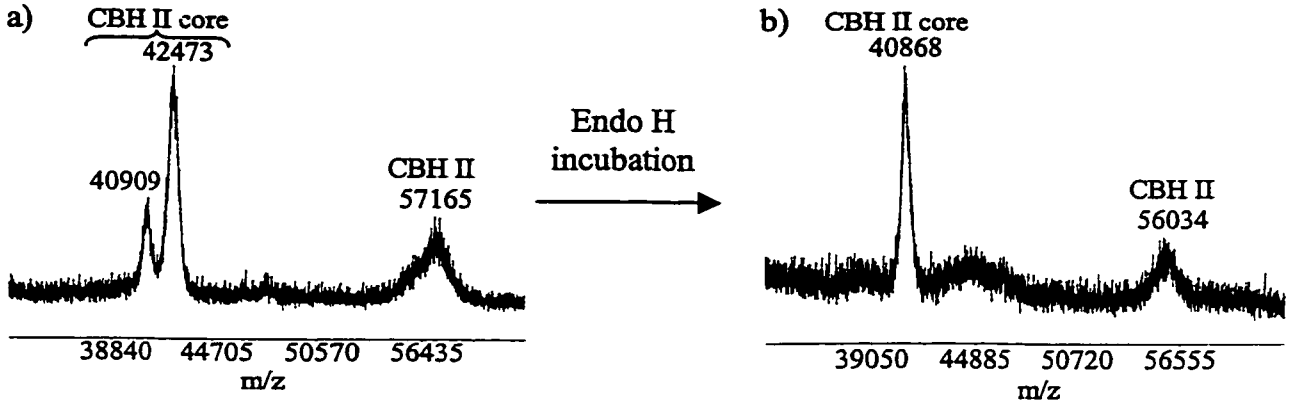


Figure 5.4 MALDI-TOF analysis of CBH II from RUT-C30 incubated with a) papain and b) papain followed by Endo H. Conditions for digestions in experimental section 2.7.2 were followed, except that partial papain digest was carried out in 3 hrs.

at Asn14 appended to the native peptide with a predicted mass of 8150.8 Da. Results from chymotryptic digest has shown that the putative N-linked position at Asn14 is not glycosylated (section 5.5). It was thus concluded that there are 39 to 46 O-linked hexoses attached to the CBH II linker peptide. It is not surprising that CBH II has roughly twice the number of O-linked hexoses compared to CBH I because of its longer linker domain (Figures 3.1a,b).

In contrast to the successful coring of cellobiohydrolases, papain proteolysis of endoglucanases was not effective and have shown inconclusive results (Figures 5.3b-c). MALDI-TOF analyses of EG I and EG II have revealed that there are more than one cleavage site. Even though a high enzyme dosage was used (enzyme:substrate 5:1 for EG I and 10:1 for EG II), a relatively large amount of intact cellulases compare to the cores were still visible after 24 hours of digestion, indicating the poor coring ability of papain on the endoglucanases. The large EG I linker/CBD glycopeptide about 11,000 Da displayed an

incremental number of O-linked hexose residues (inset of Figure 5.3b). The smaller of the two EG II linker/CBD glycopeptides of about 7,500 Da showed a relatively small O-linked heterogeneity with extra three series of potassium adduct ions (39 Da) embebbed between the hexose spacings (inset of Figure 5.3c). The large EG II linker/CBD glycopeptide of about 13,000 Da was poorly resolved. Due to the low coring rate of endoglucanases by papain and the lack of knowledge of cleavage site in the literature, the number of O-linked glycans attached to EG I and EG II was not successfully assigned.

5.4 Tryptic Digestion

Similar to CBH I, tryptic digestion was used to characterize the attachment and the nature of N-linked oligosaccharides in CBH II, EG I and EG II from RUT-C30. The digests were preliminary analyzed by MALDI-TOF using 2,5 DHB as a matrix for small peptides (m/z 500-6,000) and sinapinic acid for relatively large peptide (m/z 4,000-20,000). The observed tryptic peptides from MALDI-TOF analyses of CBH II, EG I and EG 2 are listed in Tables 5.2, 5.3 and 5.4, respectively. Isotopic resolution was not typically achieved above m/z 4,000 and the average masses are reported. These masses had larger error (± 5 Da) from the calculated values due to the wide range used with external calibration.

Subsequent analysis by LC-ESMS provided a more accurate approach for peptide matching (Figures 5.5a-c, Tables 5.2-5.4). The mass accuracy for below m/z 3,000 was generally ±0.1 Da. In contrast to the capillary C18 reversed-phase column used in CBH I analysis, a larger bore C8 column (2.1 x 150 mm) was used to separate tryptic digests from CBH II, EG I and EG II. This allows for the observation of large hydrophobic peptides above 4,000 Da and also enabled the collection of fractions for subsequent MS-MS analyses.

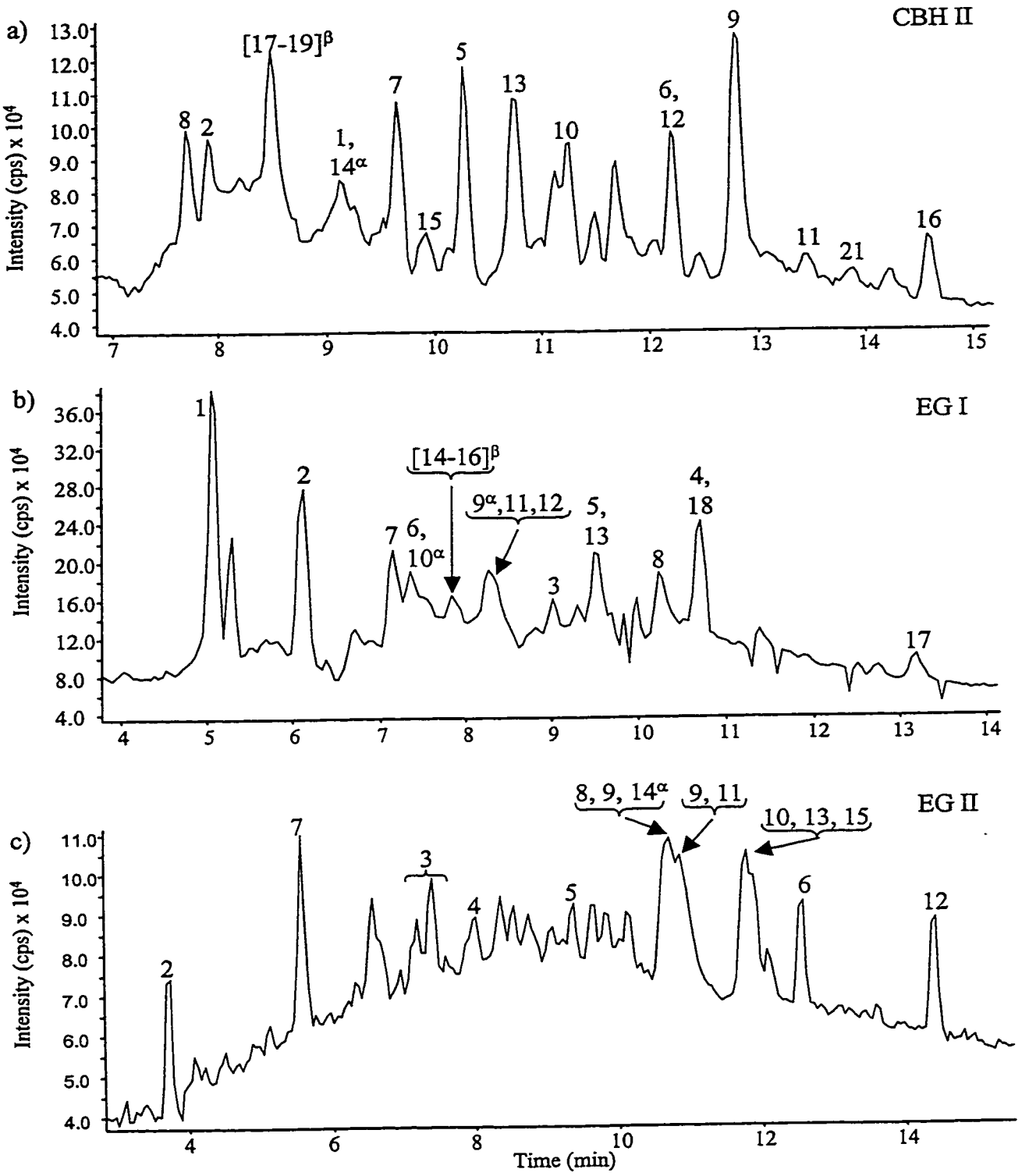


Figure 5.5 LC-ESMS analyses of tryptic digests of purified cellulase from RUT-C30. (a) CBH II; (b) EG I and (c) EG II. The number and corresponding peptide assignment for (a), (b) and (c) are shown in Table 5.2, 5.3 and 5.4, respectively (numbers above with α superscript represent peptides with single GlcNAc while β are those with high mannose).

Table 5.2 Assignment of Tryptic masses of CBH II from RUT-C30

Peak #	Residues	Assignment	$Mass_{mono}$(Da)[a]	Observed Mass (Da)	
				LC-MS	MADLI-TOF
1*	135-141	WLDTLDK	889.5	889.4	889.5
2	197-203	NYIDTIR	893.5	893.4	893.7
3	195-203	YKNYIDTIR	1184.6	-	1184.8
4	204-213	QIVVEYSDIR	1220.6	-	1220.7
5	142-153	TPLMEQTLADIR	1386.7	1386.6	1386.7
6	130-141	VPSFMWLDTLDK	1450.7	1450.7	1451.0
7	158-174	NGGNYAGQFVVYDLPDR	1883.9	1883.8	1883.7
8	175-194	DCAALASNGEYSIADGGVAK	1967.9	1967.8	-
9	214-234	TLLVIEPDSLANLVTNLGTPK	2207.2	2207.2	2206.8
10*	357-378	QPTGQQQWGDWCNVIGTGFGIR	2504.2	2504.1	2504.6
11	130-153	VPSFMWLDTLDKTPLMEQTLADIR	2819.4	2819.4	2819.7
12	328-353	LYIHAIGPLLANHGWSNAFFITDQGR	2910.5	2910.3	2910.7
13*	379-410	PSANTGDSLLDSFVWVKPGGECDGTSDSSAPR	3308.5	3308.4	3309.2
14*	298-327	GLATNVANYNGWNITSPPSYTQGNAVYNEK + GlcNAc	3445.6	3446.4	3446.3
15	158-194	NGGNYAGQFVVYDLPDRDCAALASNGEYSIADGGVAK	3833.8	3834.6	3834.1
16	411-447	FDSHCALPDALQPAPQAGAWFQAYFVQLLTNANPSFL	4105.0	4104.9	4105.2
17	298-327	GLATNVANYNGWNITSPPSYTQGNAVYNEK + $Man_7GlcNAc_2$	4783.1	4783.9	4780.4
18*	298-327	GLATNVANYNGWNITSPPSYTQGNAVYNEK + $Man_8GlcNAc_2$	4945.1	4946.1	4944.4[b]
19	298-327	GLATNVANYNGWNITSPPSYTQGNAVYNEK + $Man_9GlcNAc_2$	5108.2	5108.0	5104.1[b]
20	235-288	CANAQSAYLECINYAVTQLNLPNVAMYLDAGHAGWLGW-PANQDPAAQLFANVYK	5981.8	5982.8	5978.1[b]

[a] Refers to ^{12}C monoisotopic component

[b] Average masses observed in MALDI-TOF analyses using sinapinic acid as matrix

* Peptides subjected to MS/MS analyses

Cysteine is carbamidomethylated

Table 5.3 Assignment of Tryptic masses of EG I from RUT-C30

Peak #	Residues	Assignment	$Mass_{mono}$(Da)[a]	Observed Mass (Da)	
				LC-MS	MALDI-TOF
1	1-13	pEQPGTSTPEVHPK	1387.7	1387.7	1387.5
2	235-247	SYYGPGDTVDTSK	1388.6	1388.6	-
3*[c]	165-177	IPDNLAPGNYVLR	1440.8	1440.8	1440.8
4	109-122	LYLLDSDGEYVMLK	1657.8	1657.8	1657.7
5	22-39	SGGCVAQDTSVVLDWNYR	2025.9	2025.9	2026.2
6	419-437	TCTSGTTCQYSNDYYSQCL	2307.9	2307.8	2308.0
7*	213-234	SCTATACDSAGCGFNPYGSGYK	2329.9	2329.9	2330.0
8*	248-271	TFTIITQFNTDNGSPSGNLVSITR	2582.3	2582.3	2582.6
9	182-205	NGTLNTSHQGFCCNEMDILEGNSR + GlcNAc	2956.2	2956.2	-
10*	40-68	WMHDANYNSCTVNGGVNTTLCPDEATCGK + GlcNAc	3474.4	3475.3	3475.8
11	273-307	YQQNGVDIPSAQPGGDTISSCPSASAYGGLATMGK	3484.6	3484.5	-
12	272-307	KYQQNGVDIPSAQPGGDTISSCPSASAYGGLATMGK	3612.7	3612.6	3614.0[b]
13	69-108	NCFIEGVDYAASGVTTSGSSLTMNQYMPSSSGGYSSVSPR	4150.8	4150.8	4153.2[b]
14	182-205	NGTLNTSHQGFCCNEMDILEGNSR + $Man_7GlcNAc_2$	4293.7	4293.6	4297.2[b]
15*	182-205	NGTLNTSHQGFCCNEMDILEGNSR + $Man_8GlcNAc_2$	4455.7	4455.6	4458.4[b]
16	182-205	NGTLNTSHQGFCCNEMDILEGNSR + $Man_9GlcNAc_2$	4617.8	4617.7	4621.2[b]
17	308-363	LSSGMVLVFSIWNDNSQYMNWLDSGNAGPCSSTEGNPSNI-LANNPNTHVVFSNIR	6095.8	6095.8	-
18*	123-181	VMLKLNGQELSFDVDLSALPCGENGSLYLSQMDENGGAN-QYNTAGANYGSGYCDAQCPVQTWR	6461.8	6463.8	6467.7[b]

[a] Refers to ^{12}C monoisotopic component

[b] Average masses observed in MALDI-TOF analyses using sinapinic acid as matrix

[c] Unexpected tryptic peptides from EG IV of *T. reesei*

* Peptides subjected to MS/MS analyses

Table 5.4 Assignment of Tryptic masses of EG II from RUT-C30

Peak #	Residues	Assignment	$Mass_{mono}$(Da)[a]	Observed Mass (Da)	
				LC-MS	MALDI-TOF
1	320-323	QNNR	530.3	-	530.1
2	205-210	YASQSR	710.3	710.3	710.2
3	97-102	VYPPLK	715.4	715.4	715.3
4*	169-178	CIVDIHNYAR	1259.6	1259.6	1259.5
5*[c]	165-177	IPDNLAPGNYVLR	1440.8	1440.8	1440.7
6*	306-319	NIDGAFSPLATWLR	1559.8	1559.8	1559.8
7*	290-305	YLDSDNSGTHAECTTN	1783.7	1783.6	1783.7
8	74-96	FAGVNIAGFDFGCTTDGTCVTSK	2424.1	2424.0	2424.1
9	131-154	LPVGWQYLVNNNLGGNLDSTSISK	2588.3	2588.3	2587.7
10	179-204	WNGGIIGQGGPTNAQFTSLWSQLASK	2717.4	2717.3	2717.4
11	155-178	YDQLVQGCLSLGAYCIVDIHNYAR	2828.3	2828.3	2828.5
12	211-239	VWFGIMNEPHDVNINTWAATVQEVVTAIR	3309.7	3309.6	3309.7
13	290-319	YLDSDNSGTHAECTTNNIDGAFSPLATWLR	3325.5	3325.5	3326.6
14*	103-130	NFTGSNNYPDGIGQMQHFVNDDGMTIFR + GlcNAc	3377.5	3377.4	3377.5
15	240-289	NAGATSQFISLPGNDWQSAGAFISDGSAAA-LSQVTNPDGSTTNLIFDVHK	5077.4	5077.4	5079.9[b]

[a] Refers to ^{12}C monoisotopic component
[b] Average mass observed in MALDI-TOF analysis
[c] Unexpected tryptic peptides from EG IV of *T. reesei*
* Peptides subjected to MS/MS analyses
Cysteine is carbamidomethylated

The current LC-ESMS analyses were carried out on a PE/Sciex Q-Star hybrid quadrupole/time-of-flight instrument. The suspected tryptic glycopeptides with additional mass modification of hexose and N-acetyl-hexosamine units were tentatively assigned and later confirmed by tandem MS. The assignment of glycopeptides with high mannose was also elucidated by incubating the tryptic digests with N-glycosidase F. MALDI analysis has shown that peaks 14 to 16 from EG I (Table 5.3) were removed after subjecting to PNGase F incubations while an additional peak corresponding to the native sequence 182-205 at m/z 2755.1 appeared (Figure 5.6).

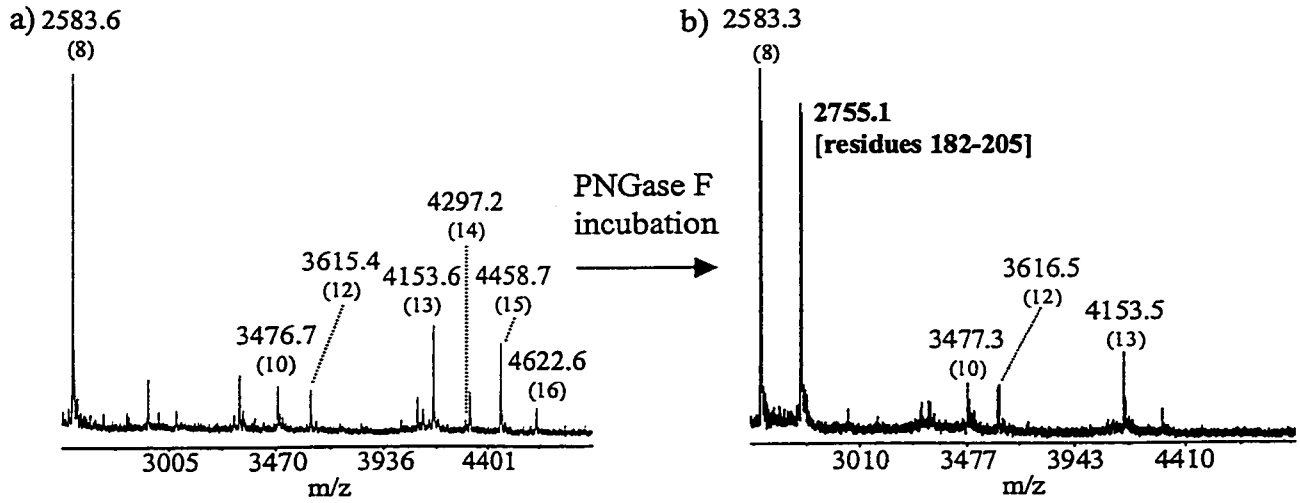


Figure 5.6 MALDI-TOF analysis of EG I tryptic digest (a) before and (b) after PNGase F incubations. See experimental section 2.7.1 for digestion protocol. Numbers in parentheses underneath the indicated masses correspond to those in Table 5.3.

The sequence coverage of tryptic digests by combined results from both MALDI-TOF and LC-ESMS are listed in Table 5.8 in section 5.5. The % coverage for CBH II, EG I and EG II including the large heterogeneous linker peptides observed in MALDI-TOF was calculated to be 84%, 98.3% and 81.9%, respectively. Among the three putative N-glycan sites in CBH II (Asn14, 289, 310), only Asn310 showed a high mannose predominantly

$Man_8GlcNAc_2$ and single GlcNAc in the ratio of 4:1 (peaks 14, 17-19 of Table 5.2) as indicated from the corresponding LC-ESMS experiment (Figure 5.5 a). By using integrated peak height from extracted mass spectra, the ratio of Man_7:Man_8:Man_9 at Asn310 was calculated to be about 2:5:4. No information was obtained for the other two sites.

There are overall six putative N-glycan sites in EG I core (Asn56, 142, 182, 186, 259, 372). A high mannose, predominantly $Man_8GlcNAc_2$ was found at either Asn182 or Asn186, with minor occupancy of GlcNAc in the ratio of 13:1 (peaks 9, 14-16 of Table 5.3). The occurrence of these two sites in a single tryptic peptide makes it difficult to assign the specific site for this high mannose oligosaccharide. This N-glycan is also heterogeneous and the ratio of Man_7:Man_8:Man_9 was found to be 2:20:5. A single GlcNAc was identified at Asn56 (peak 10 of Table 5.3). There was no N-linked glycan found at Asn142 and Asn259 (peaks 8, 18 of Table 5.3). No information for glycosylation on Asn372 was successfully obtained. In contrast, only one putative N-glycan site is expected in EG II core (Asn103) and this site was found to be occupied with a single GlcNAc (peak 14 of Table 5.4).

Analysis by MS/MS partial sequence have resulted in a number of unexpected peptides. Peaks 10 and 13 from CBH II (Table 5.2) originated from the same tryptic peptide (sequence 357-410), but was unexpectedly cleaved at the arginine adjacent to a proline residue. Similarly, peaks 6 and 7 from EG II (Table 5.4) corresponding to the same tryptic peptide (sequence 290-306) resulted from cleavage between two asparagine residues. Interestingly, the sequence read "IPDNLAPGNYLR" listed in peak 3 of EG I and peak 5 of EG II (Tables 5.3-5.4) was identified as a EG IV tryptic peptide by MS-MS sequencing. These unexpected results will be discussed further in section 5.7.

5.5 Chymotryptic Digestion

Chymotrypsin is a pancreatic enzyme that is specifically used to hydrolyze the amide linkages at the C terminus of bulky hydrophobic residues such as phenylalanine, tryptophan and tyrosine (except when the adjacent residue is proline) while asparagine, histidine, methionine and leucine are cleaved at a slower rate.[112] The objective of performing chymotryptic digestion was to complement the information on the N-linked glycans that were not successfully assigned previously. In contrast to trypsin, chymotrypsin digestion results in numerous small peptides that complicated the subsequent mass assignments. The mass of peak 8 of CBH II chymotryptic peptides has two matches, and no definite assignment could be made (Table 5.5). On the other hand, a strong signal was observed at m/z 906.3 (peak 9 of Table 5.5). This corresponds to only one sequence "GQCGGQNW" with no modified Asn14

Table 5.5 Assignment of chymotryptic masses of CBH II (RUT-C30) observed by MALDI-TOF (cysteine is carbamidomethylated)

Peak #	Residues	Assignment	$Mass_{mono}$ (Da)	$Mass_{obs.}$ (Da)
1	432-435	QAYF	527.2	527.2
2	343-347	SNAFF	584.3	584.3
3	94-99	VGVTPW	657.4	657.4
4	325-329	NEKLY	665.3	665.4
5	337-342	LANHGW	696.3	696.4
6	330-336	IHAIGPL	719.4	719.4
7	243-248	LECINY	810.4	810.4
8	263-270 262-269	DAGHAGWL LDAGHAGW	825.4 825.4	825.4
9	8-15	GQCGGQNW	905.4	905.3
10	186-195	SIADGGVAKY	979.5	979.5
11	297-306	RGLATNVANY	1077.6	1077.6
12	170-179	DLPDRDCAAL	1144.5	1144.5
13	199-209	IDTIRQIVVEY	1347.7	1347.8
14	348-364	ITDQGRSGKQPTGQQQW	1913.9	1913.9
15	347-364	FITDQGRSGKQPTGQQQW	2061.0	2060.9
16	348-367	ITDQGRSGKQPTGQQQWGDW	2272.1	2271.9

and thus confirming that no glycosylation occurred at this position.

Analysis of EG I chymotryptic peptides failed to provide additional information on N-glycosylation. A strong signal at m/z 560.2 had three potential matches, with a putative N-glycan site Asn182 appeared in the sequence "RNGTL" (peak 1 of Table 5.6). Previous results from tryptic peptides suggest this site is either unglycosylated or occupied by a high mannose glycan. As a result, no conclusion could be reached about the state of occupancy at this position. The only glycosylation site in EG II at Asn103 was successfully assigned by tryptic analysis and revealed a single GlcNAc attached to this residue. The corresponding chymotrypsin analysis is also summarized in Table 5.7 for comparison purpose.

Table 5.6 Assignment of chymotryptic masses of EG I (RUT-C30) observed by MALDI-TOF (cysteine is carbamidomethylated)

Peak #	Residues	Assignment	$Mass_{mono}$ (Da)	$Mass_{obs.}$ (Da)
1	124-128	NGQEL	559.3	559.2
	234-237	KSYY	559.3	
	181-185	RNGTL	559.3	
2	35-38	DWNY	596.2	596.2
3	42-46	HDANY	618.2	618.2
4	90-94	TMNQY	655.3	655.2
5	429-433	SNDYY	660.2	660.2
6	360-364	SNIRW	674.4	674.3
7	160-166	NTAGANY	709.3	709.2
8	41-46	MHDANY	749.3	749.2
	306-313	GKALSSGM	749.4	
9	95-102	MPSSSGGY	784.3	784.3
10	267-273	VSITRKY	865.5	865.5
11	103-110	SSVSPRLY	907.5	907.3
12	350-359	ANNPNTHVVF	1111.5	1111.5
13	171-180	CDAQCPVQTW	1263.5	1263.4
14	237-249	YGPGDTVDTSKTF	1386.6	1386.5
15	256-273	NTDNGSPSGNLVSITRKY	1922.0	1922.6
16	18-36	KCTKSGGCVAQDTSVVLDW	2110.0	2109.7
17	274-299	QQNGVDIPSAQPGGDTISSCPSASAY	2606.2	2605.7

Table 5.7 Assignment of chymotryptic masses of EG II (RUT-C30) observed by MALDI-TOF (cysteine is carbamidomethylated)

Peak #	Residues	Assignment	$Mass_{mono}$ (Da)	$Mass_{obs.}$ (Da)
1	195-198	TSLW	505.3	505.2
2	130-135	RLPVGW	726.4	726.4
3	206-212	ASQSRVW	832.4	832.4
4	248-255	ISLPGNDW	900.4	900.3
5	284-290	IFDVHKY	920.4	920.4
6	169-176	CIVDIHNY	1032.5	1032.4
7	319-327	RQNNRQAIL	1111.6	1111.5
8	318-327	LRQNNRQAIL	1224.7	1224.6
9	180-194	NGGIIGQGGPTNAQF	1429.7	1429.6
10	248-261	ISLPGNDWQSAGAF	1461.7	1461.4
11	214-227	GIMNEPHDVNINTW	1638.8	1638.4
12	105-120	TGSNNYPDGIGQMQHF	1764.8	1764.4
13	228-247	AATVQEVVTAIRNAGATSQF	2033.1	2032.8

The sequence coverage for chymotryptic peptides of CBH II, EG I and EG II are listed in Table 5.8, together with the tryptic results. It is clear from this comparison that tryptic analysis is by far superior to chymotrypsin in terms of site specificity and % coverage. The numerous small peptides resulting from chymotrypsin digestion may have suppressed the ionization of analytes including glycopeptides and thus limiting carbohydrate assignment.

Table 5.8 Comparison of percent sequence coverage observed in tryptic and chymotryptic analyses of purified cellulases (number in parentheses under the trypsin column represents % coverage discounting the large heterogenous linker peptides)

	% sequence coverage	
Cellulases (RUT-C30)	Trypsin	Chymotrypsin
CBH II	84.8 (67.6)	23.9
EG I	98.3 (85.8)	32.5
EG II	81.9 (63.0)	30.7

5.6 Analysis of Selected Tryptic Peptides by MS/MS

All tentative glycopeptides and unidentified peaks were further subjected to tandem MS for partial peptide sequencing. The observed tryptic peptides in Figure 5.5 were purified by fraction collection and selected masses were ionized by nanoelectrospray to facilitate MS/MS analyses. The variability in site occupancy at Asn310 of CBH II was confirmed by conducting product ion scan on peaks 14 and 18 indicated in Table 5.2 (Figures 5.7a, b). By using low collision energy (30 eV), the triply charged series (m/z 1595.8, 1541.9, 1487.9, 1433.8, etc.) of peaks corresponding to the loss of hexose and N-acetylhexosamine units were clearly displayed (Figure 5.7a). This characteristic fragmentation pattern together with the diagnostic oxonium ions at m/z 163.1, 204.1 and 366.1 strongly support the proposal of a high mannose oligosaccharide attached to Asn310. The mass of parent ion excluding the high mannose $Man_8GlcNAc_2$ corresponds to the peptide 298-327 containing a putative N-glycan site at Asn310 (consensus sequence NIT). This high mannose site is heterogeneous and contains glycans ranging from Man_7 to Man_9. The occurrence of $Man_8GlcNAc_2$ as the major glycoform is consistent with the earlier results from CBH I.

A later eluting peak with a molecular mass corresponding to peptide 298-327 with a single GlcNAc was listed as peak 14 in Table 5.2. The product ion spectrum of the corresponding triply charged ion at m/z 1150.5 shows a series of y-type ions for the peptide 298-327 and a strong diagnostic oxonium ion of GlcNAc at m/z 204.1 (Figure 5.7b). The attachment site was further supported by the appearance of y-type ions with GlcNAc starting from y_{18}, which corresponded to Asn310 at the N-terminal residue. Unlike CBH I, the extent of glycosylation in CBH II has not been extensively studied. Rouvinen and co-workers have

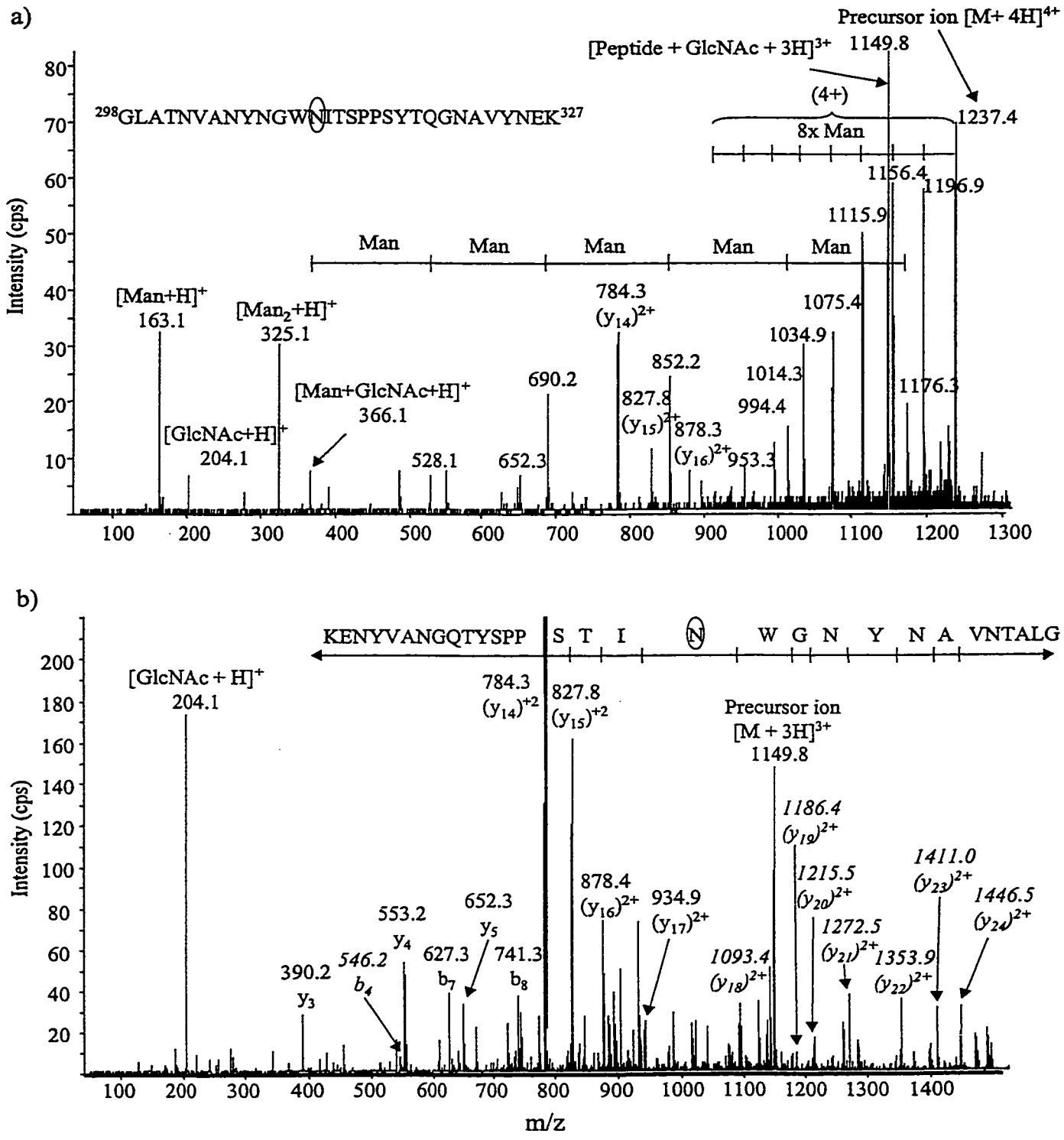


Figure 5.7 MS-MS spectra of N-linked glycopeptides of CBH II tryptic digest with oligosaccharide attached at Asn310. (a) Product ion scan of m/z 1238.0 (peak 18 of Table 5.2) ; (b) Production ion scan of m/z 1150.05 (peak 14 of Table 5.2). See experimental section 2.7.3 for conditions except collision energy(laboratory frame of reference) of 90 eV for (a) and 120 eV for (b).

reported the glycosylation at Asn289 and Asn310 in CBH II without any assignment of their respective N-linked oligosaccharide.[18]

The assignment of high mannose in EG I is more challenging due to the presence of two N-glycan sites (Asn182, 186) in the same tryptic peptide. The product ion spectrum of the suspected glycopeptide with $[M + 4H]^{4+}$ at m/z 1115.4 was dissociated at relatively low energy (35 eV) and a triply charged series consistent with a single GlcNAc at m/z 986.4 and a $Man_7GlcNAc_2$ at m/z 1432.8 could easily be observed (Figure 5.8a). Subsequent selection of the peptide with GlcNAc as precursor ion for second generation MS-MS analysis failed to provide conclusive evidence for the attachment site. The putative site at Asn56 was suspected to contain a single GlcNAc and was confirmed by selecting the corresponding $[M+3H]^{3+}$ at m/z 1159.8 as parent ion to undergo CID (Figure 5.8b). The first y_1-y_{12} ions was matched and GlcNAc attachment was observed starting from y_{13} onwards. This confirms the attachment site is at Asn56. The appearance of HexNAc diagnostic ion with no Hex and Hex-HexNac ions further supports this proposal. Kleywegt et al. have reported the occupancy of single GlcNAc at position Asn56 and Asn182 by electron-density mapping.[19] This is consistent with the current result at Asn56, but differs at Asn182. Nevertheless, the presence of both high mannose and single GlcNAc in EG I serve as an other indication of Endo H activities in *T. reesei*.

The only putative N-glycan site at Asn103 in EG II was suspected to have a single GlcNAc attachment (Figure 5.9a). Interestingly, the mass of the precursor ion is 14 Da less than the calculated molecular mass of peptide 103-130 (peak 14 of Table 5.4). The product ion spectrum of the $[M+3H]^{3+}$ at m/z 1238.0 showed a strong HexNAc oxonium ion at m/z 204.1 and a series of y-type ions from y_1-y_7. Starting from y_8 ion onwards, the observed

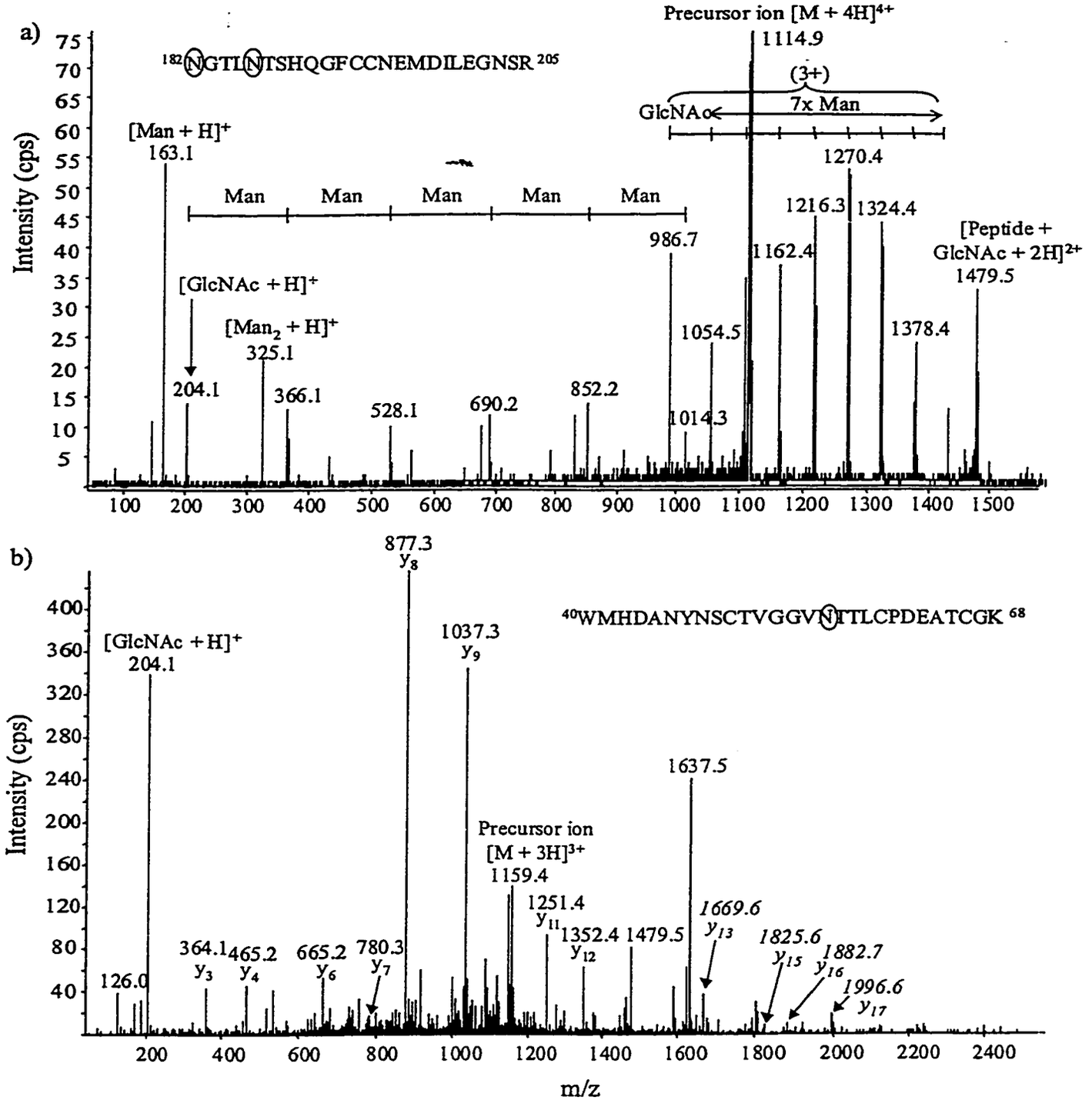


Figure 5.8 MS-MS spectra of N-linked glycopeptides of EG I tryptic peptide. (a) Product ion scan of m/z 1114.9 (peak 15 of Table 5.3) to confirm the presence of high mannose at either Asn182 or Asn186; (b) Production ion scan of m/z 1159.4 (peak 10 of Table 5.3) to confirm single GlcNAc at Asn56. The y-type ions in italics indicate addition of a GlcNAc. See experimental section 2.7.3 for conditions except collision energy (laboratory frame of reference) of 140 eV for (a) and 150 eV for (b).

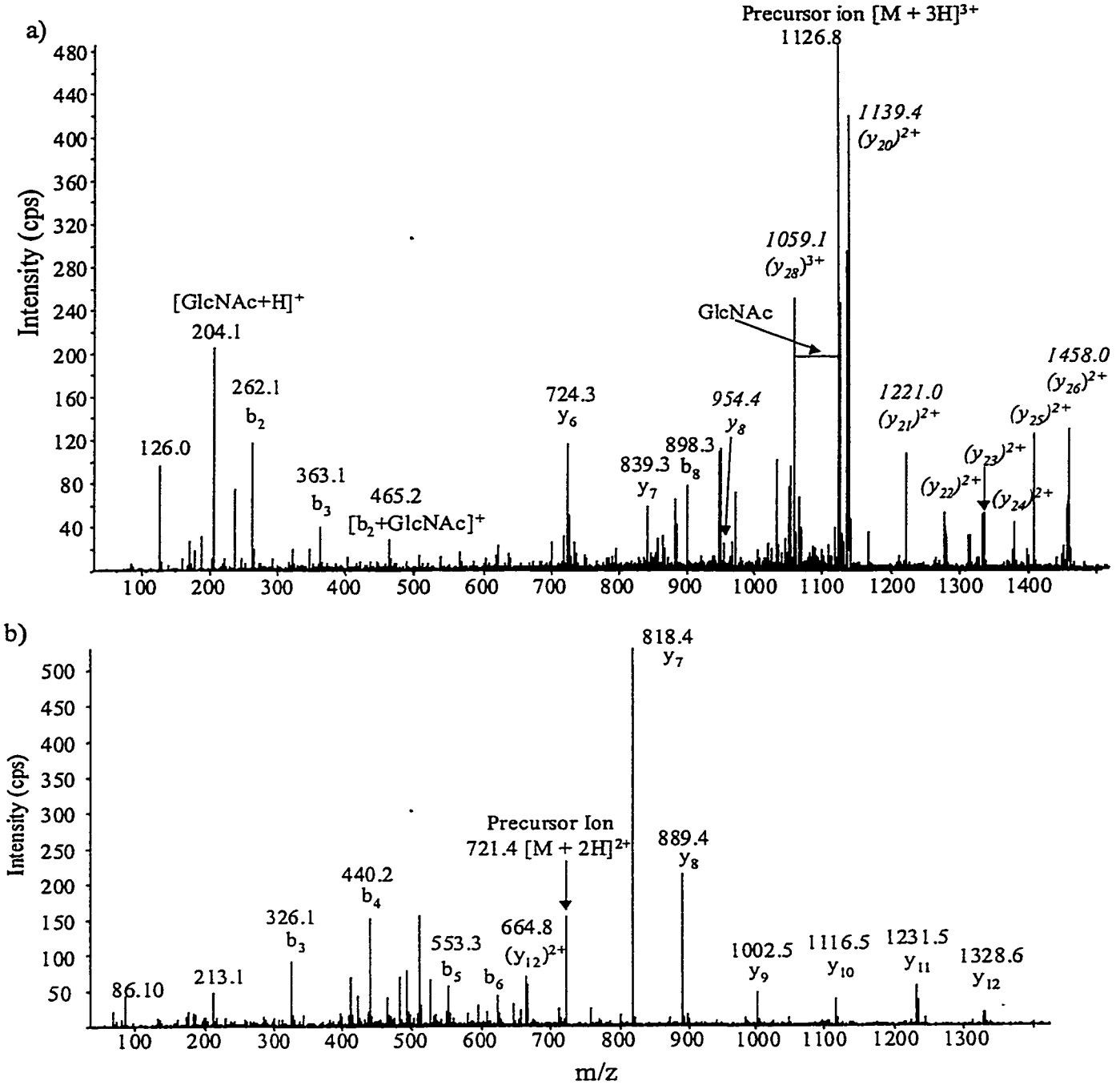


Figure 5.9 MS-MS spectra of N-linked glycopeptides of EG II tryptic digest. (a) Product ion scan of m/z 1126.8 (peak 14 of Table 5.4) confirming the single GlcNAc at Asn103. The y-type ions in italics indicate a decrease in 14 Da from Glu144 to Asp144; (b) Product ion scan of m/z 721.4 (peak 3 of Table 5.3 and peak 5 of Table 5.4). See experimental section 2.7.3 for conditions except collision energy (laboratory frame of reference) of 150 eV for (a) and 80 for (b).

masses were 14 Da less than the expected one. The y_8 ion corresponds to a glutamic acid (129 Da) residue at the N-terminus and the 14 Da shift strongly suggests a substitution of the glutamic acid by an aspartic acid (115 Da). Therefore, MS-MS of m/z 1238.0 from EG II confirmed both the single GlcNAc at Asn103 and the single amino acid substitution of Glu144 with Asp144.

A number of unexpected tryptic peptides were identified by MS-MS sequencing. Interestingly, a particular sequence which was later found to be a EG IV tryptic peptides by database searching via the European Molecular Biology Laboratory (EMBL) peptide search program (http://www.mann.embl-heidelberg.de/Services/PeptideSearch/) was observed in both EG I and EG II (Figure 5.9b). It is not clear why we only observe one EG IV tryptic peptide, but traces amount of unstable EG IV could be co-purified with the other endoglucanases. Peaks 10 and 13 in Table 5.2 were originated from the same tryptic peptide (Gln357-Arg410) and had an expected cleavage at Arg378 prior to Pro379 (Figures 5.10a, b). Cleavage of Arg-Pro bond is very unusual in tryptic digest, although examples of situations have been reported in the literature. Nyman et al. had reported such cleavage in the sequence -Trp-Arg-Pro-Ala- while Grand et al. had observed this activity at the -Ala-Arg-Pro-Ala-.[113,114] The current sequence is –Ile-Arg-Pro-Ser-Ala- and does not display any consensus domain. Another unexpected cleavage occurred between the Asn305 and Asn306 of EG II tryptic peptide Tyr290-Arg319. The corresponding peptides are numbered as peaks 6 and 7 in Table 5.4 and their product ion spectra are shown in Figures 5.11a and b. The majority of ions indicated in Figure 5.11a belongs to the b-type ions as opposed to the y-type ions due to the absence of a basic residue at the C-terminal end. This cleavage may be attributed to the

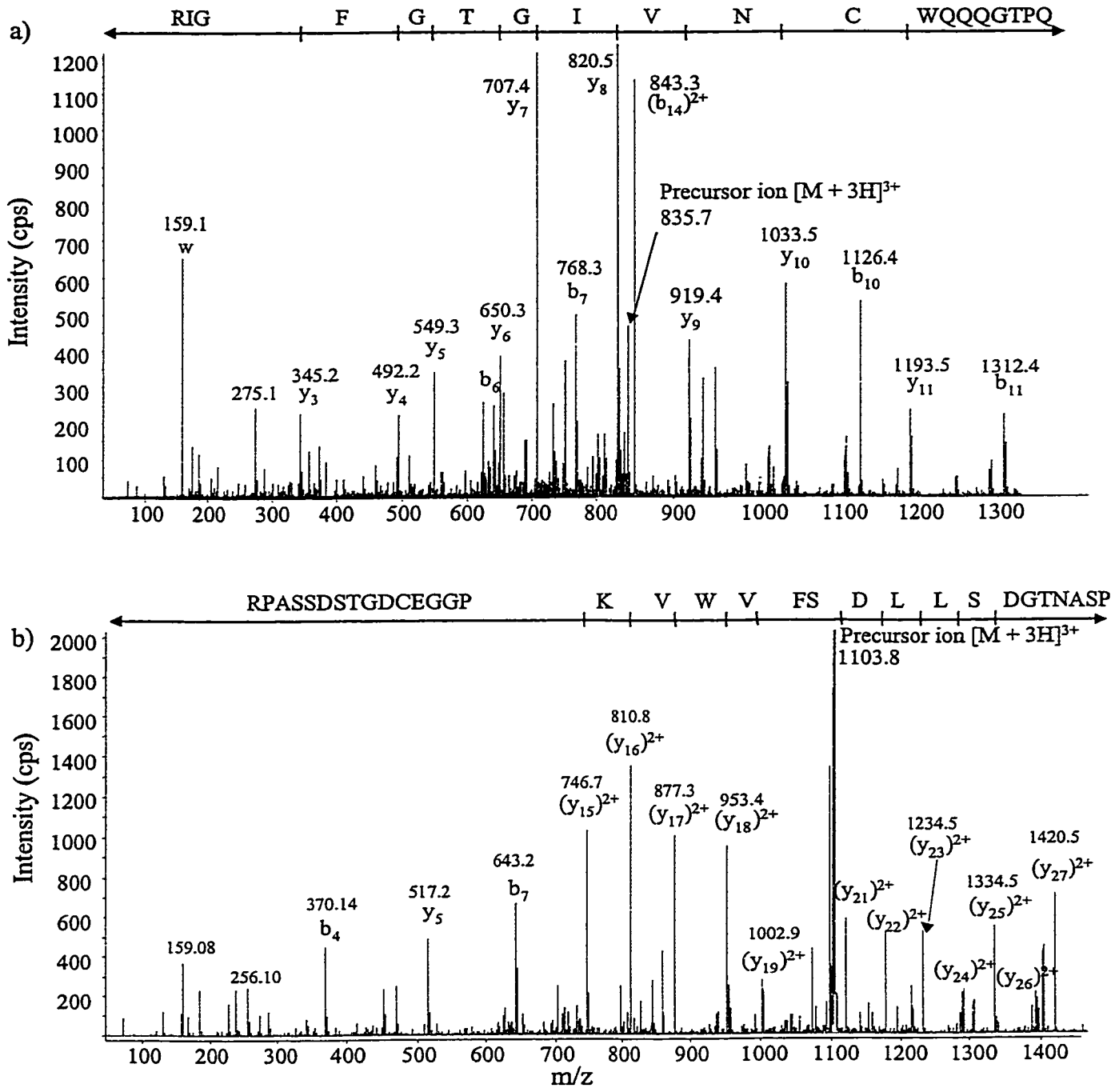


Figure 5.10 MS-MS spectra of two unexpected tryptic peptides from CBH II. (a) Product ion scan of m/z 835.7 (peak 10 of Table 5.2); (b) Product ion scan of m/z 1103.8 (peak 13 of Table 5.2). Experimental conditions as described in section 2.7.3 except collision energy (laboratory frame of reference) of 120 eV for (a) and 150 eV for (b).

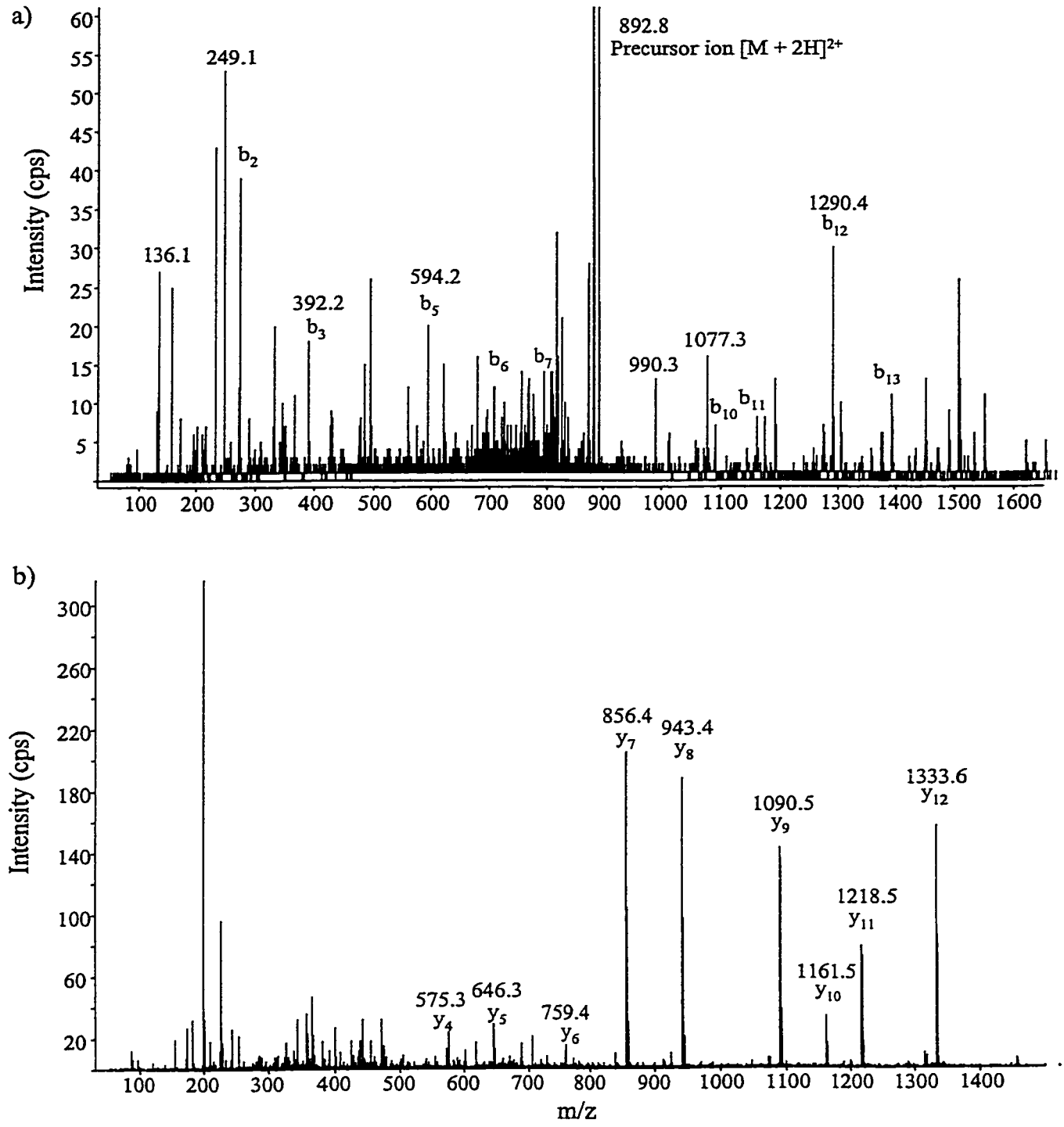


Figure 5.11 MS-MS spectra of two tryptic peptides from EG II with unexpected cleavage site. (a) Product ion scan of m/z 892.8 (peak 7 of Table 5.4); (b) product ion scan of m/z 780.9 (peak 6 of Table 5.4). See section 2.7.3 for experimental conditions except collision energy (laboratory frame of reference) is 100 for (a) and 80 for (b).

enzymatic activities of small amount of asparagine protease present in the trypsin or in the original crude enzymes.

5.7 Conclusion

This is a first time report on the assignment of N-linked glycosylation in CBH II, EG I and EG II purified from strain RUT-C30. Electrospray mass spectrometry was initially used to probe the complexity of glycosylation. A number of enzymatic digestions including papain, trypsin, chymotrypsin, Endo H and PNGase F were used to elucidate the structure of N-linked oligosaccharide. These provided an exhaustive assignment of the majority of putative N-glycan sites. A single GlcNAc was found at Asn56 of EG 1 and Asn103 of EG II. A high mannose predominantly $Man_8GlcNAc_2$ (Man_7:Man_8:Man_9 2:5:4) was found at Asn310 in CBH II, and this site also contains small proportion of single GlcNAc (~20 %). A high mannose predominantly $Man_8GlcNAc_2$ (Man_7:Man_8:Man_9 2:20:5) was also found in EG I with attachment site at either Asn182 or Asn186. This site also has a minor glycoform comprising a single GlcNAc (~7 %). No glycosylation was observed at Asn14 in the CBD domain of CBH II, nor was any glycosylation attachment found at Asn142 or Asn259 of EG I. The N-linked information on Asn289 of CBH II and Asn372 of EG I still remained to be identified.

With all these results, the question as to whether cellulases from RUT-C30 can serve as the specific substrates for the mannosylphosphotransferase can be undoubtedly answered. From the electrospray mass spectra and papain analyses, there were no additional series of 80 Da peaks which indicate the absence of phosphate modifications in CBH II, EG I and EG II in the purified fractions. It is not known at present why CBH I is the only cellulase displaying

a phosphorylated residue and further investigations would be warranted. The appearance of high mannose with predominantly $Man_8GlcNAc_2$ was observed in three of the four major cellulases (CBH I, CBH II and EG I). The presence of single GlcNAc in all investigated cellulases strongly supports the proposal of an endogenous Endo H secreted by *T. reesei*. Purification and sequencing of this endoglycosidase enzyme will further lead to a better understanding of the overall enzymatic systems exhibited in this filamentous fungus.

References

[1] Nevalainen, H.; Suominen, P.; Taimisto, K. *J. Biotechnol.* **1994**, *37*, 193.

[2] Carter, G. L.; Allison, D.; Rey, M. W.; Dunn-Coleman, N. S. *Bio/Technology* **1992**, *6*, 2167.

[3] Harkki, A.; Mantyla, A.; Penttilä, M.; Muttilainen, S.; Buhler, R.; Suominen, P.; Knowles, J.; Nevalainen, H. *Enzyme. Microb. Technol.* **1991**, *13*, 227.

[4] Kubicek, C. P. *Advances in Biochemical Engineering/Biotechnology* **1992**, *45*, 1.

[5] Nidetzky, B.; Steiner, W.; Hayn, M.; Claeyssens, M. *Biochem. J.* **1994**, *298*, 705.

[6] Henrissat, B.; Driguez, H.; Viet, C.; Schulein, M. *Bio/Technology* **1985**, *3*, 722.

[7] Srisodsuk, M.; Reinikainen, T.; Penttilä, M.; Teeri, T. T. *J. Bio. Chem.* **1993**, *268*, 20756.

[8] Clarke, A. J. *Biodegradation of Cellulose ENZYMOLOGY AND BIOTECHNOLOGY*, Ch. 2, Technomic Publishing Co., Inc., Pennsylvania, USA (1997).

[9] Kleywegt, G. J.; Zou, J. Y.; Divne, C., Davies, G. J.; Sinning, I.; Ståhlberg, J; Reinikainen, T.; Srisodsuk, M.; Teeri, T. T.; Jones, T. A. *J. Mol. Biol.* **1997**, *272*, 383.

[10] Teeri, T. T.; Lehtovaara, P.; Kauppinen, S.; Salovuori, I.; Knowles, J. *Gene* **1987**, *51*, 43.

[11] Lee, H. J.; Brown, R. M. Jr. *J. Biotechnol.*, **1997**, *57*, 127.

[12] Enari, T. M. Microbial cellulases: *Microbial Enzymes and Biotechnology*, W.M. Fogarty Eds, **1983**, pp.183, Applied Science Publishers, London and New York.

[13] Divine, C., Ståhlberg, J., Reinikaninen, T., Rouhonen, L., Petterson, G., Knowles, K. J. C., Teeri, T. T., Jones, T. A. *Science*, **1994**, *265*, 524.

[14] Divine, C., Ståhlberg, J., Teeri, T. T., Jones, T. A. *J. Mol. Biol.* **1998**, *275*, 309.

[15] Reinikainen, T. The cellulose-binding domain of cellobiohydolase I from *Trichoderma reesei* Interaction with cellulose and application in protein immobilization **1994**, VTT publications 206, Technical research centre of Finland.

[16] Sinnott, M. L. *Chem. Rev.* **1990**, *90*, 1171.

[17] Koivula, A.; Reinikainen, T.; Ruohonen, L.; Valkeajärvi, A.; Claeyssens, M.; Teleman, O.; Kleywegt, G. J.; Szardenings, M.; Rouvinen, J.; Jones, T. A.; Teeri, T. T. *Protein Eng.* **1996**, *9*, 691.

[18] Rouvinen, J.; Bergfors, T.; Teeri, T.; Knowles, J. K. C. *Science* **1990**, *249*, 380.

[19] Kleywegt, G. J.; Zou, J. Y.; Divne, C., Davies, G. J.; Sinning, I.; Ståhlberg, J; Reinikainen, T.; Srisodsuk, M.; Teeri, T. T.; Jones, T. A. *J. Mol. Biol.* **1997**, *272*, 383.

[20] Macarron, R.; van Beeumen, J.; Henrissat, B.; de la Mata, I.; Claeyssens, M. *FEBS Lett.* **1993**, *316*, 137.

[21] Saloheimo, M.; Lehtovaara, P.; Penttila, M. Teeri, T. T.; Ståhlberg, J.; Johansson, G.; Petterson, G.; Claeyssens, M.; Tomme, P.; Knowles, J. K. C. Gene **1988**, *63*, 11.

[22] Kubicek C. P. *Advances in Biochemical Engineering/Biotechnology* **1992**, *45*, 1.

[23] Maras, M.; De Bruyn, A.; Schraml, J.; Herdewijn, P.; Claeyssens, M.; Fiers, W.; Contreras, R. *Eur. J. Biochem.* **1997**, *245*, 617.

[24] Klarskov, K.; Piens, K.; Ståhlberg, J.; Høj, P. B.; Van Beeumen, J.; Claeyssens, M. *Carbohyd. Res.* **1997**, *304*, 143.

[25] Harrison, M. J.; Nouwens, A. S.; Jardine, D. R.; Zachara, N. E.; Gooley, A. A., Nevalainen, H.; Packer, N. H. *Eur. J. Biochem.* **1998**, *256*, 119.

[26] Kubicek, C. P.; Panda, T.; Schreferal-Kunar, G.; Gruber, F.; Messner, R. *Can. J. Microbiol.* **1987**, *33*, 698.

[27] Carter, G. L.; Allison, D.; Rey, M. W.; Dunn-Coleman, N. S. *Mol. Microb.* **1992**, *6*, 2167.

[28] Mandels, M.; Reese, E. T. *J. Bacteriol.* **1957**, *73*, 269.

[29] Nevalainen, H.; Palva, E. T.; Bailey, M. J. *Enzyme Microb. Technol.* **1980**, *59*, 1980.

[30] Montenecourt, B. S.; Eveleigh, D. E. *Appl. Environ. Microbiol.* **1977**, *34*, 777.

[31] Monnig, C. A.; Kennedy, R. T. *Anal. Chem.* **1994**, *66*, 280R.

[32] Kuhr, W. G. *Anal. Chem.* **1990**, *62*, 403R.

[33] Baker, D. R. *Capillary Electrophoresis*; John Wiley & Sons, Inc.: New York, 1995; pp. 2.

[34] Tadey, T. *Effects of Buffer Modifiers and Polymeric Surface Coatings in Capillary Electrophoresis*, Ph.D. Thesis , Department of Chemistry, McGill University, 1995, pp. 7.

[35] Baker, D. R. et. al, pp. 23.

[36] St. Claire, III R. L. *Anal. Chem.* **1996**, *68*, 569R.

[37] Hjertén, S.; Zhu M. *J. Chromatogr.* **1985**, *346*, 265.

[38] R-Diaz, R.; Zhu, M.; Wehr, T. *J. Chromatogr. A* **1997**, *772*, 145.

[39] Hjertén, S. *Chromatogr. Rev.* **1967**, *9*, 122.

[40] Hjertén, S. *J. Chromatogr.* **1985**, *347*, 191.

[41] Gamble, T. N.; Ramachandran, C.; Bateman, K. P. *Anal. Chem.* **1999**, *71*, 3469.

[42] Clarke, N. J.; Tomlinson, A. J.; Schomburg, G.; Naylor, S. *Anal. Chem.* **1997**, *69*, 2786-92.

[43] Kilár, F.; Végvári, Á.; Mód, A. *J. Chromatogr. A.* **1998**, *813*, 349.

[44] Mazzeo, J. R.; Krull, I. S. *Anal. Chem.* **1991**, *63*, 2852.

[45] Hjertén, S.; Liao, J. L.; Yao, K. *J. Chromatogr.* **1987**, *387*, 127.

[46] Wei, J.; Yang, L.; Harrata, A.K.; Lee, C.S. *Electrophoresis* **1998**, *19*, 2356.

[47] Cifuentes, A.; Moreno-Arribas, M.V.; de Frutos, M.; Díez-Masa, J.C. *J. Chromatogr. A*, **1999**, *830*, 453.

[48] Pantazaki, A.; Taverna, M.; Vidal-Madjar, C. *Analytica Chimica Acta* **1999**, *383*, 137.

[49] Conti. M.; Gelfi, C.; Bosisio, A.B.; Righetti, P. G. *Electrophoresis* **1996**, *17*, 1590.

[50] Mairo, N.; Baudin, B.; Giboudeau, J. *J. Chromatogr. B*, **1998**, 706, 123.

[51] Medve, J.; Lee, D.; Tjerneld, F. *J. Chromatogr. A*, **1998**, *808*, 153.

[52] Burlingame, A. L.; Boyd, R. K.; Gaskell, S. J. *Anal. Chem.* **1994**, *66*, 634R.

[53] Ehring, H.; Karas, M.; Hillenkamp, F. *Org. Mass Spectrom.* **1992**, *27*, 472.

[54] Gross, J.; Strupat, K. T. *Trends Anal. Chem.* **1998**, *17*, 1998.

[55] Yamashita, M.; Fenn, J. B. *J. Phys. Chem.* **1984**, *88*, 4451.

[56] Gaskell, S. J. *J. Mass Spectrom.* **1997**, *32*, 677.

[57] Kebarle, P.; Tang, L. *Anal. Chem.* **1993**, *65*, 972A.

[58] Iribarne, J. V.; Thomson, B. A. *J. Chem. Phys.* **1976**, *64*, 2287.

[59] Thomson, B. A.; Iribarne, J. V. *J. Chem. Phys.* **1979**, *71*, 4451.

[60] Wilm, M.; Mann, M. *Anal. Chem.* **1996**, *68*, 1.

[61] Guilhaus, M. *J. Mass Spectrom.* **1995**, *30*, 1519.

[62] Kaufmann, R. *J. Biotechnol.* **1995**, *41*, 155.

[63] Gross, J.; Strupat, K. T. *Trends Anal. Chem.* **1998**, *17*, 1998.

[64] Wiley, W. C.; McLaren, I. H. *Rev. Sci. Instrum.* **1953**, *26*, 1150.

[65] Vestal, M. L.; Juhasz, P.; Martin, S. A. *Rapid Commun. Mass Spectrom.* **1995**, *9*, 1044.

[66] Rose, M. E.; Johnstone, R. A. *Mass Spectrometry for Chemists and Biochemists*, 1st Ed., Cambridge University Press, NY, **1982**, Ch. 2.

[67] Skoog, D. A.; Leary, J. J. *Principles of Instrumental Analysis*, 4th Ed., Saunders College Publishing, USA, **1992**, Ch. 18.

[68] Watson, J. T. *Introduction to Mass Spectrometry*, Raven Press, NY, **1985**, Ch. 2.

[69] Jennings, K. R.; Dolnikowski G. G. *Methods in Enzymology Mass Spectrometry*, edited by McCloskey J. A., Academic Press, NY, **1990**, *193*, 37.

[70] Busch, K. L.; Glish, G. L.; McLuckey, S. A. Mass Spectrometry/Mass Spectrometry Techniques and Applications of Tandem Mass Spectrometry, VCH Publishers, Inc., NY, 1988, pp. 3.

[71] Gaskell, S. J. *Triple Quadrupoles and Hybrid Instruments for Tandem Mass Spectrometry, Biological Mass Spectrometry: Present and Future*, edited by Matsuo, T; Caprioli, R. M.; Gross, M. L.; Seyama, Y. John Wiley & Sons, Ltd., 1994, pp. 165.

[72] Fenselau, C. *Annu. Rev. Pharmacol. Toxicol.* **1992**, *32*, 555.

[73] Cole, M. J.; Enke, C. G. *Anal. Chem.* **1991**, *63*, 1032.

[74] Volmer, D. A.; Hui, J. P. M. *Rapid Commun. Mass Spectrom.* **1998**, *12*, 123.

[75] Morris, H. R.; Paxton, T.; Dell, A.; Langhorne, J.; Berg, M.; Bordoli, R. S.; Hoyes, J.; Bateman, R. H. *Rapid Commun. Mass Spectrom.* **1996**, *10*, 889.

[76] Morris, H. R.; Paxton, T.; Panico, M,; McDowell, R.; Dell, A. *J. Protein Chem.* **1997**, *16*, 469.

[77] Dawson, J. H. J.; Guilhaus, M. 1989, *Rapid Commun. Mass Spectrom.* **1989**, *3*, 155.

[78] Biemann, K. *Annu. Rev. Biochem* **1992**, *61*, 977.

[79] Shevchenko, A.; Chernushevich, I.; Ens, W.; Standing, K. G.; Thomson, B.; Wilm, M.; Mann, M. *Rapid Commun. Mass Spectrom.* **1997**, *11*, 1015.

[80] Wilm, M.; Shevchenko, A.; Houthaeve, T.; Breit, S.; Schweigerer, L.; Fotsis T.; Mann, M. *Nature*, **1996**, *379*, 466.

[81] Mørtz, E.; Vorm, O.; Mann, M.; Roepstorff, P. *Biol. Mass Spectrom.* **1994**, *23*, 249.

[82] http://www.ncbi.nlm.nih.gov/BLAST/blast_overview.html

[83] Mandels, M. and Reese T. *J. Bacteriol.* **1957**, *73*, 269-278.

[84] Bhikhabhai, R.; Johansson, G.; Pettersson, G. *J. Appl. Biochem.* **1984**, *6*, 336-345.

[85] Ståhlberg, J.; Divne, C.; Koivula, A.; Piens, K.; Claeyssens, M.; Teeri, T. T.; Jones, T. A. *J. Mol. Biol.* **1996**, *264*, 337.

[86] Woodward, J.; Brown, J. P.; Evans, B. R.; Affholter, K. A.; *Biotechnol. Appl. Biochem.* **1994**, *19*, 141.

[87] Kelly, J. F.; Locke, S. J.; Ramaley, L.; Thibault, P. *J. Chromatogr. A* **1996**, *720*, 409.

[88] Wiktorowicz, J. E.; Colburn, J. C.; *Electrophoresis*, **1990**, *11*, 769.

[89] Mairo, N.; Baudin, B.; Giboudeau, J. *J. Chromatogr. B*, **1998**, *706*, 123.

[90] Tarentino, A. L.; Plummer, T. H. Jr. *Guide to Techniques in Glycobiology, in Methods in Enzymology*, pp. 44. Academic press, New York (1994).

[91] Carr, S. A.; Huddleston, M. J.; Bean, M. F. *Protein Sci.* **1993**, *2*, 183.

[92] Bateman, K. P.; White, R. L.; Yaguchi, M.; Thibault, P. *J. Chromatogr. A*, **1998**, *794*, 327.

[93] Tang, Q.; Harrata, K.; Lee, C. S. *Anal. Chem.* **1996**, *68*, 2482.

[94] Tang, Q.; Harrata, K.; Lee, C. S. *J. Mass Spectrom.* **1996**, *31*, 1284.

[95] Tang, Q.; Harrata, K.; Lee, C. S. *Anal. Chem.* **1997**, *69*, 3177.

[96] Gavel, Y.; von Heijne, G. *Protein Eng.* **1990**, *3*, 433.

[97] Ward, M., Wu, S., Dauberman, J., Weiss, G., Larenas, E., Bower, B., Rey, M., Clarkson, K., Bott, R. *Proceedings of the second TRICEL symposium in TRICHODERMA REESEI CELLULASES AND OTHER HYDROLASES*, Finland, ed. by Suominen, P. and Reinikainen, T., *Foundation for Biotechnical and Industrial Fermentation Research* **1993**, *8*, 153.

[98] Landers, J. P.; Oda, R. P.; Madden, B. J.; Spelsberg, T. C. *Anal. Biohem.* **1992**, *205*, 115.

[99] Harris, D. C. *Quantitative Chemical Analysis*, 3rd Ed., W. H. Freeman and Company, New York, 1982, pp.662.

[100] Tang, S.; Nesta, D. P.; Maneri, L. R.; Anumula, K. R. *J. Pharm. Biomed. Anal.* **1999**, *19*, 569.

[101] Salovouri, I.; Makarow, M.; Rauvala, H.; Knowles, J.; Kääriänen, L. *Bio/technology* **1987**, *5*, 152.

[102] El-Shinawy, N. A.; Basta, A. H.; Yacoub, S. F.; Mohamed, S. H. *Polymer-Plastics Technology and Engineering* **1998**, *37*, 141.

[103] Nair, B. *Int. J. Toxicol.* **1998**, *17*, 67.

[104] Shoemaker, S.; Schweickart, V.; Ladner, M.; Gelfand, D.; Kwok, S.; Myambo, K.; Innis M. *Biot.* **1983**, *1*, 691.

[105] Ballou, C. E. *Methods Enzymol.* **1990**, *185*, 440.

[106] Jigami, Y., Odani, T. *Biochim. Biophys. Acta* **1999**, *1426*, 335.

[107] Nakayama, K.; Feng, Y.; Tanaka, A.; Jigami, Y. *Biochim. Biophys. Acta* **1998**, *1425*, 255.

[108] Kamphuis, I. G.; Kalk, K. H.; Swarte, M. B. A.; Drenth, J. *J. Mol. Biol.* **1984**, *179*, 233.

[109] Tomme, P.; Van Tilbeurgh, H.; Pettersson, G.; Van Damme, J.; Vandekerckhove, J.; Knowles, J.; Teeri, T.; Claeyssens M. *Eur. J. Biochem.* **1988**, *170*, 575.

[110] Divne, C.; Sinning, I.; Ståhlberg, J.; Pettersson, G.; Bailey, M.; Siika-aho M.; Margolles-Clark E.; Teeri T.; Jones T. A. *J. Mol. Biol.* **1993**, *234*, 905.

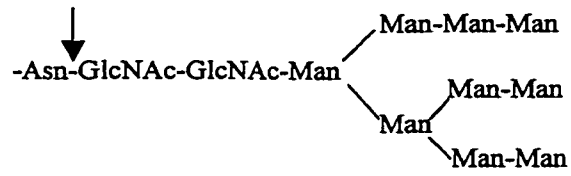
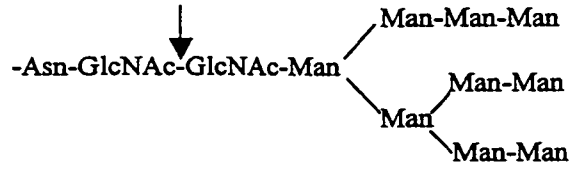
[111] Woodward, J.; Brown, J. P.; Evans, B. R.; Affholter, K. A. *Biotechnol. Appl. Biochem.* **1994**, *19*, 141.

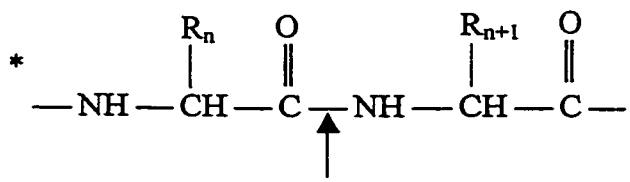
[112] Voet, D.; Voet, J. *Biochemistry*, 2nd Ed., John Wiley and Sons, Inc., USA, 1995, pp.112.

[113] Nyman, P. O.; Strid, L.; Westermark, G., *Biochim. Biophys. Acta.* **1966**, *122*, 554.

[114] Grand, R. J. A.; Wilkinson, J. M.; Mole, L.E. *Biochem. J.* **1976**, *159*, 633.

APPENDIX - Substrate Specificity of Relevant Proteolytic Enzymes

Enzyme	Specificity	Comments
(1) Trypsin	* R_n = Arg, Lys; $R_{n+1} \neq$ Pro	Hydrolyses the peptide bond at the C-terminals of basic residues such as Arg and Lys; highly specific.
(2) Chymotrypsin	* R_n = Phe, Trp, Tyr; $R_{n+1} \neq$ Pro	Hydrolyses the peptide bond at the C-terminals of aromatic residues such as Phe, Trp, Tyr; cleavage proceeds more slowly when R = Ala, Asp, Glu, Met, Leu.
(3) Alkaline Phosphatase	$HPO_3 - X \longrightarrow HO - X$ H_2O PO_3^- where X is commonly the 5' end of DNA.	Can also be used to remove phosphate from phosphorylated tyrosine, serine and threonine residues in proteins.
(4) PNGase F (N-glycosidase F)		Cleaves between the Asn and GlcNAc residues of all N-linked high mannose, hybrid and complex oligosaccharides.
(5) Endo H (Endoglycosidase H)		Cleaves the chitobiose core of N-linked high mannose and some hybrid oligosaccharides.



Claims to Original Research

(1) development of capillary isoelectric focusing for the quantitation of cellulases in crude fermentation extracts from *T. reesei*. This technique allows for the detection of variations in composition of cellulase enzyme in mutant strains.

(2) first time report on the use of electrospray mass spectrometry to characterize the glycoform heterogeneity in the purified cellulase enzymes

(3) assignment of N- and O-linked glycosylation in CBH I purified from *T. reesei* RUT-C30 and two mutant strains. The presence of single GlcNAc oligosaccharide and heterogeneity in the glycan population characterized by both high mannose predominantly $Man_8GlcNAc_2$ and single GlcNAc residue in putative N-linked glycan sites were observed. The majority of O-linked glycan in the linker domain is found to be di- and tri-saccharides.

(4) first time report on the occurrence of phosphorylation in CBH I glycans from *T. reesei* in the literature. This mannosylphosphorylation was located on a O-linked di-saccharide in the linker region.

(5) assignment of N-linked glycosylation in CBH II, EG I and EG II from RUT-C30 strain. Consistent with CBH I, both the single GlcNAc oligosaccharide and heterogeneity in the glycan population characterized by high mannose (predominantly $Man_8GlcNAc_2$) and single GlcNAc residue in putative N-linked glycan sites were observed. There was no occurrence of phosphorylated O-linked di-saccharide in these purified glycoproteins.

List of Publications and Presentations

(1) Hui, J. P. M.; Lanthier, P.; White, T. C.; McHugh, S. G.; Yaguchi, M.; Roy, R.; Thibault, P. Characterization of cellobiohydrolase I (Cel7A) glycoforms from extracts of *Trichoderma reesei* using capillary isoelectric focusing and electrospray mass spectrometry, *J. Chromatogr. B*, in press.

(2) Hui, J. P. M.; White, T. C.; Yaguchi, M.; Roy, R.; Thibault, P. Characterization of protein glycoforms in cellobiohydrolase and endoglucanase from *Trichoderma reesei*, Proc. of the 48th ASMS Conference on Mass Spectrometry and Allied Topics, Long Beach, CA, 2000.

(3) Hui, J. P. M.; Giroux, G.; Lanthier, P.; White, T. C.; Yaguchi, M.; Roy, R.; Thibault, P. Monitoring protein glycoforms in cellulase extracts from *Trichoderma reesei*, Proc. of the 47th ASMS Conference on Mass Spectrometry and Allied Topics, Dallas, TX, 1999.