ELSEVIER

Journal of Chromatography A, 720 (1996) 409–427

JOURNAL OF CHROMATOGRAPHY A

Development of electrophoretic conditions for the characterization of protein glycoforms by capillary electrophoresis–electrospray mass spectrometry[1]

J.F. Kelly[a,b], S.J. Locke[b], L. Ramaley[a], P. Thibault[b,*]

[a]*Chemistry Department, Dalhousie University, Halifax, Nova Scotia B3H 4J3, Canada*

[b]*Institute for Marine Biosciences, National Research Council, 1411 Oxford Street, Halifax, Nova Scotia B3H 3Z1, Canada*

Abstract

A capillary electrophoresis (CE) method using acidic buffers and capillaries coated with Polybrene, a cationic polymer has been developed for the separation of glycoproteins and glycopeptides. Electrophoretic conditions have been optimized to provide resolution of individual glycoforms observed for different glycoprotein preparations. These conditions were found to be entirely compatible with the operation of electrospray mass spectrometry (ESMS), which facilitated the assignments of possible carbohydrate compositions of glycopeptides arising from digests of glycoproteins. By using operating conditions enhancing the formation of oxonium fragment ions prior to mass spectral analysis, selective identification of glycopeptides was achieved for complex samples such as those from proteolytic digests or chemical cleavages. Examples of applications are presented for ribonuclease B, ovalbumin, horseradish peroxidase, and a lectin from *Erithrina corallodendron* using both CE–ESMS and CE with ultraviolet detection (CE–UV).

1. Introduction

Glycoproteins are a diverse group of complex macromolecules comprising conjugated proteins covalently bonded to one or more heterosaccharide chains [1]. They are ubiquitously distributed in nature, and play important roles in a number of biological recognition processes including hormonal control, defense or decoy functions, cellular adhesion, blood clotting, immunological protection, intercellular interactions, and structural support [2–5].

These macromolecules exhibit extraordinary complexity, and this feature is undoubtedly a required attribute for their unique biological specificity. Their structural diversity is derived not only from the type of linkages between proteins and carbohydrates, but also from the composition and structure of the carbohydrate attached to the polypeptide backbone. Oligosaccharides are usually linked to proteins to the hydroxyl group of a serine or threonine side chain (O-glycosidic bond) or to the amide side-chain of an asparagine residue (N-glycosidic bond). Glycoproteins can also accommodate different oligosaccharide chains at any given attachment site, thus giving rise to a distribution of glycosylation variants referred to as glycoforms. In eukaryotic cells the synthesis of glyco-

* Corresponding author.

[1] #NRCC: 38105.

SSDI 0021-9673(94)01197-4

protein oligosaccharides appears to be reproducible, and highly regulated by exoglycosidase in the endoplasmic reticulum and glycosyltransferase in the Golgi complex [4,6,7]. N-linked oligosaccharides formed in the endoplasmic reticulum are extensively trimmed and undergo elongation reactions in the Golgi complex, while O-linked chains are appended to the glycoprotein [4,8].

Of particular interest to the present investigation are the structures of N-linked glycoproteins. The attachment of carbohydrates to asparagine residues almost always occurs at the tripeptide recognition signal Asn–X–Ser/Thr (where X is any amino acid residue except Pro or Asp [9–11]). In contrast, no O-glycosylation sequon has yet been identified, thus rendering assignment of the location of O-linked carbohydrate more difficult. The series of enzymatic reactions that take place in the Golgi compartments give rise to extensively modified N-linked carbohydrates. The structures of these oligosaccharides have been conveniently classified into three main categories, termed high mannose, complex, and hybrid [9]. Examples of these oligosaccharides are shown in Fig. 1. One of the characteristic features of the N-linked carbohydrates is the fact that all three categories share a common core structure composed of Manα1–3(Manα1–6) Manβ1–4GlcNAcβ1–4 GlcNAc. High mannose oligosaccharides such as ribonuclease B (RNase B), chicken ovalbumin, and the sindbis virus undergo limited transformation in the Golgi compartments, and thus present the lowest level of complexity. High-mannose glycoproteins can contain from 2 to 9 mannose residues appended to the common pentasaccharide core. By contrast, complex glycoproteins such as bovine fetuin, transferrins, and immunoglobulins, show a much higher degree of substitution and have a variable number of GlcNAc, galactose, sialic acid and/or fucose and xylose residues attached to the core. Addition of terminal sialic acids is of special importance since this residue can confer an incremental negative charge to the glycoprotein. Lastly, the hybrid type encompasses features of both high-mannose and complex types, and may also have a GlcNAc linked β1,4 to the β-linked mannose [9].

The glycoform population of glycoproteins confers on them a spectrum of biological ac-

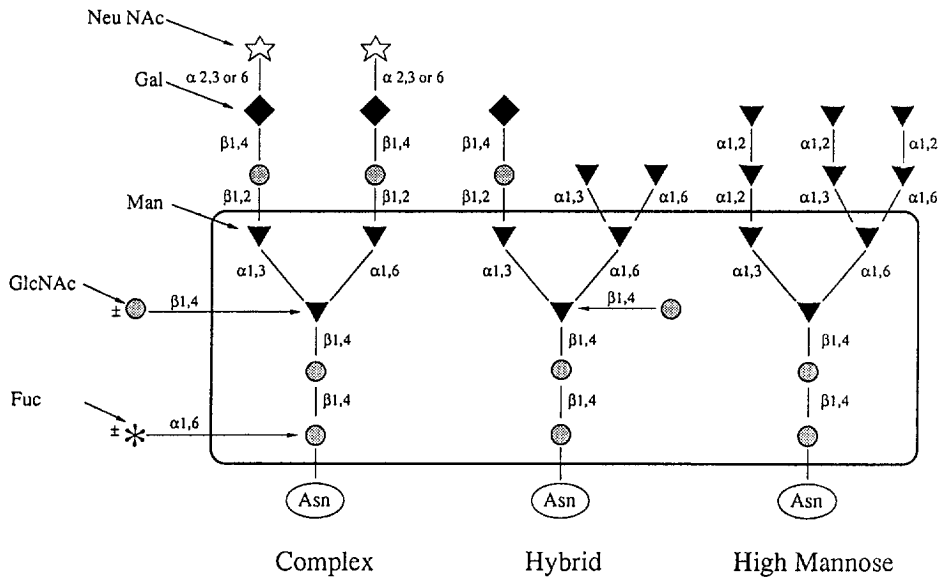

Fig. 1. Structures of the major types of N-linked glycans adapted from Ref. [9]. The pentasaccharide core common to the three types of N-linked oligosaccharides is indicated in the enclosed box.

tivities. Alteration of these types of glycosylation can result in a range of responses from essentially unchanged to the complete loss of particular functions, or even elimination of the protein itself [2]. Advances in the understanding of the biological roles of oligosaccharides and their structure–activity relationships have been gained through the development of different chemical methods providing efficient mapping of glycoform populations. To this end, information relating to the carbohydrate composition can be obtained following the release of the N-linked oligosaccharide from the glycoprotein using mild hydrazinolysis [12] or by selective hydrolysis of the glycosidic bond using endoglycosidases [13].

Separation and preparative purification of the neutral oligosaccharides can be achieved using gel permeation chromatography as demonstrated by Ashford et al. for a series of plant lectins [14]. However analytical scale separation of derivatized carbohydrates is often desirable in view of the limited amount of material available. Previous investigations have reported the use of reagents such as 2-amino pyridine (PA) [15], *p*-aminobenzoic acid ethyl ester (ABEE) [16], and 3-(4-carboxybenzoyl)-2-quinolinecarboxaldehyde (CBQCA) [17], to enhance detectability of the carbohydrate using either UV or fluorescence detection. Such derivatives also confer functionalities upon carbohydrates which have convenient properties for their separation by reversed-phase liquid chromatography [18–20].

Characterization of the released carbohydrates has also been demonstrated by capillary electrophoresis (CE) [21–25], and excellent reviews on this topic have been published recently [26,27]. Numerous applications of zone electrophoresis have used borate buffers at pH conditions favoring the complexation of the vicinal hydroxyl groups of the analyte with the anionic electrolyte [28]. Such conditions have been described for the analysis of oligosaccharides from ovalbumin [29,30], and a series of N-linked asialo-oligosaccharides [31]. Potential structure candidates can also be proposed based on electrophoretic mobility measurements obtained from two-dimensional mapping CE experiments [32]. Alternatively, it is possible to use heptane sulfonic acid and phosphate buffers to suppress the electrosmotic flow and adsorption of the analyte on the silica surface, as demonstrated by Rush et al. for the separation of glycopeptides from erythropoietin [33].

In spite of these advances, structural characterization of oligosaccharides remains a difficult task in view of the complexity of the analyte, and the non-availability of suitable carbohydrate standards. Such experiments do not provide information on glycosylation sites nor do they permit identification of individual glycoforms at a given site. Recently, several mass spectrometric approaches have been used to deduce information on the carbohydrate structure and to localize glycopeptides eluting in a chromatographic run using liquid chromatography–electrospray mass spectrometry (LC–ESMS) [34–37]. The possibility of obtaining molecular mass information for each component as it elutes as a chromatographic peak, coupled with further structural characterization of the carbohydrate moiety using tandem mass spectrometry (MS–MS), provides unique features that make this technique a valuable tool for carbohydrate research.

In the present investigation we describe the use of CE combined with electrospray mass spectrometry (CE–ESMS) for the analysis of N-linked glycoproteins and glycopeptides arising from chemical and enzymatic cleavages. Separations performed using strong acidic buffers and coated capillaries were found conducive to the resolution of closely related glycoforms. Analysis of glycoprotein digests by CE–ESMS enables identification of microheterogeneity. In more complex situations where multiple glycosylation sites are present in a given glycoprotein preparation, selective identification of glycopeptides can be made by recording specific oxonium ions arising from collision induced dissociation of oligosaccharides in the ion source region [36]. Examples of applications are demonstrated for the analysis of RNase B, ovalbumin, horse radish peroxidase and a seed lectin from *Erythrina corallodendron*. An abbreviated account of part of this work has been presented previously [38].

2. Experimental

2.1. Proteins and preparation of digests

Bovine pancreatic RNase B, the seed lectin from *Erithrina corallodendron*, and horseradish peroxidase type IV, were obtained from Sigma Chemicals (St Louis, MO, USA). Hen egg ovalbumin was provided by STC Laboratories (Winnipeg, Man., Canada). Prior to tryptic digestion, disulfide bonds of glycoproteins were reduced with dithiothreitol and converted to carboxyamidomethyl cysteine derivatives using iodoacetamide [39] Tryptic digestion was performed in 50 m*M* ammonium bicarbonate pH 7.5 using sequencing grade trypsin (Boehringer Mannheim, Montréal, Qué., Canada) with a substrate:enzyme ratio of 50:1 (w/w). Samples were neutralized with dilute HCl prior to CE analysis in order to avoid bubble formation arising from the injection of bicarbonate or phosphate buffers in a capillary filled with high acidic buffer.

2.2. Capillary electrophoresis

CE–UV analyses were carried out using the P/ACE 2100 system (Beckman Instruments, Fullerton, CA, USA) using a 1.1 m × 50 μm I.D. fused-silica capillary. Cationic coating of the capillary surface was achieved by infusing an aqueous solution of 5% (w/v) hexadimethrine bromide, (Polybrene, Aldrich Chemicals, Milwaukee, WI, USA) and 2% (v/v) of ethylene glycol, with conditioning conditions similar to those described by Wiktorowicz and Colburn [40]. Unless otherwise indicated, all separations were performed using a 2 *M* formic acid buffer and an effective voltage of 30 kV across the capillary. Peaks were detected by absorption at 200 nm.

2.3. Mass spectrometer and interface

CE–ESMS analyses used the same CE system as that described above except that the current monitoring function was disabled to permit normal CE operation once the anodic end of the capillary was connected to the CE–ESMS interface. A hole was drilled at the top left corner of the Beckman CE cartridge to allow the capillary (1.1 m × 50 μm I.D., 180 μm O.D.) to reach the CE–ESMS interface while minimizing the length of the column. In order to prevent breakage of the flexible small O.D. capillary, a small piece of teflon tubing was inserted below the cartridge so that the cathode and the capillary remained together during the normal operation of the CE instrument. A separate power supply (Glassman EH series, Glassman, Whitehouse Station, NJ, USA) was also used to provide a voltage of −25 kV at the injector end of the capillary. A schematic representation of the CE–ESMS interface used in the present study is shown in Fig. 2. Briefly, the interface was constructed from a fully-articulated IonSpray source [41], and is based on a co-axial column arrangement first described by Smith et al. [42]. The interface comprises two tees; the front one is used to introduce the nebulizer gas at a flow rate of 2 l min^{-1}, while the back tee provides a continuous stream of solvent to the interface (25% methanol v/v, 0.2% formic acid v/v) at a flow rate of 5 μl min^{-1}. The electrospray needle traverses the front tee and is inserted in the back tee in such a way that the sheath liquid can be introduced directly into the needle. A Teflon shutter arm, which can be moved back and forth by the action of a contact relay mounted on the side port of the source housing is also shown in Fig. 2. The controllable action of this arm prevents the accumulation of salts and washing solutions on the orifice plate, and provides better performance of the mass spectrometer over a long periods of usage.

Mass spectrometric experiments were conducted using a Perkin Elmer/SCIEX API/III triple quadrupole mass spectrometer (Thornhill, Ont., Canada). Collisional activation in the orifice/skimmer region was achieved by raising the orifice potential to 100 V compared to 50 V in normal operation. Mass spectral acquisition was performed using dwell times of 3 ms per step of 1 Da in full mass scan mode, or 100 ms per channel for selected ion monitoring experiments. Fragment ion spectra obtained from combined

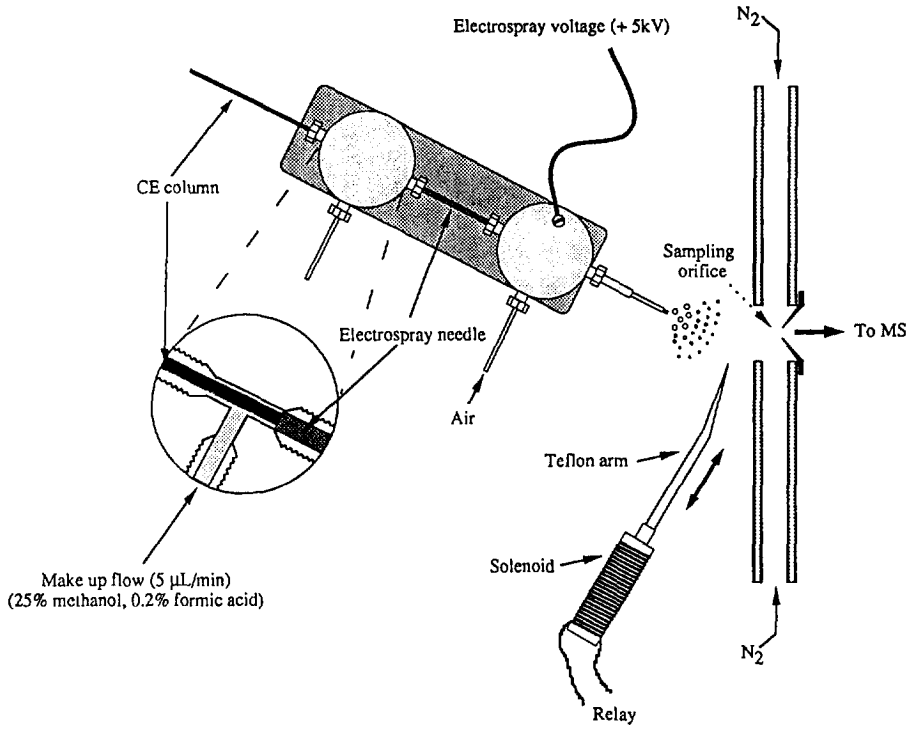

Fig. 2. Schematic representation of the coaxial CE–ESMS interface showing the moveable sidearm which protects the MS orifice. The capillary arrangement inside the back tee is shown as an inset.

CE–ESMS–MS analyses were obtained by selecting the appropriate precursor m/z value in the first quadrupole, and using collision activation with argon target gas in the rf-only quadrupole collision cell. Collision energies were typically 60 eV in the laboratory-frame of reference, and the collision thickness was 3.5×10^{15} atoms cm^{-2}. Tandem mass spectra were acquired using a dwell time of 3 ms per step of 0.5 Da. A Macintosh Quadra 950 computer was used for instrument control, data acquisition and data processing.

3. Results and discussion

3.1. Optimization of electrophoretic conditions under anodal electrosmotic flow conditions with acidic buffers

Previous investigations of the separation of protein glycoforms by CE have typically used borate complexes or ion-pairing agents as buffer additives [21–31]. However, when coupled to the electrospray mass spectrometer, such electrophoretic conditions are not conducive to the detection of either the neutral carbohydrates or the glycoproteins, and can also give rise to high chemical background, suppression of the analyte signal, and non-linear response. The current associated with the CE buffer can be reduced by adding organic solvent or choosing a volatile buffer of low proton affinity (for cationic separation and detection). Acidic buffers composed of ammonium formate or acetate in 25% methanol have been used successfully in a number of CE–ESMS experiments [41–45]. The choice of a suitable buffer system and pH conditions must also take into account the inherent tendency of proteins and basic analytes to strongly interact with the inner wall of the capillary surface, thus reducing the separation performance. In earlier communications we have reported the analysis of peptides and proteins by CE–ESMS using acidic

buffers and capillaries coated with cationic polymers such as Microcoat [46] and Polybrene [41]. These cationic coatings not only reduce the extent of solute/wall interactions but also reverse the direction of the electrosmotic flow (anodal flow), a situation which can magnify the velocity differences of analytes having electrophoretic mobilities in the opposite direction [40].

In preliminary experiments, the effects of Polybrene and ethylene glycol on the separation performance was investigated using a mixture of protein standards containing α-lactalbumin, β-lactoglobulin A, RNase A, Horse heart myoglobin, and lysozyme. The capillary surface was also reconditioned by rinsing the capillary with sodium hydroxide (1.0 *M*), water, and hydrochloric acid (0.1 *M*) between each set of conditions to ensure good reproducibility.

Separations conducted in 0.1 *M* formic acid using a fixed concentration of ethylene glycol of 2% (w/v) indicated that the electrosmotic flow increased slowly from 5.6×10^{-8} to 6.3×10^{-8} $m^2 V^{-1} s^{-1}$ as the concentration of Polybrene was varied from 0 to 10% (w/v) (triplicate injections). For Polybrene concentrations of 0, 1, 2, 5, and 10%, the average theoretical plate numbers for these five proteins as calculated from the peak widths at half height were 340 000, 227 000, 154 000, 202 000 and 146 000, respectively. Although good separation efficiencies were obtained for coating solutions exempt of Polybrene, migration times were generally irreproducible and substantially longer than for solutions containing at least 1% Polybrene. These results alone suggest that there is an interplay between analysis time (as reflected by the electrosmotic flow) and minimization of solute–wall interactions. Although such interactions can be minimized using low Polybrene concentrations, a concentration of 5% was chosen in subsequent experiments for the reasons described above.

Similarly, the ethylene glycol concentration was varied over the range of 0 to 5% for a fixed concentration of 5% Polybrene. The average plate counts for all five proteins examined reached a maximum value of 202 000 for an ethylene glycol concentration of 2%. For higher or lower ethylene glycol concentrations theoretical plate numbers were generally less than 150 000 for any of the five proteins examined. Optimal separation conditions, with respect to both analysis times and separation efficiencies, were thus obtained using a coating solution containing 5% Polybrene and 2% ethylene glycol.

The separation performance was also affected by the acidity of the electrolyte used in the CE experiment. This effect is shown in Fig. 3 for the separation of peptides resulting from the CNBr cleavage of ovalbumin, and in Fig. 4 for the analysis of RNase B. In both cases the electrolyte was formic acid, and its concentration was varied from 0.1 *M* to 2.0 *M*. For CNBr peptides, analysis conducted using 0.1 *M* formic

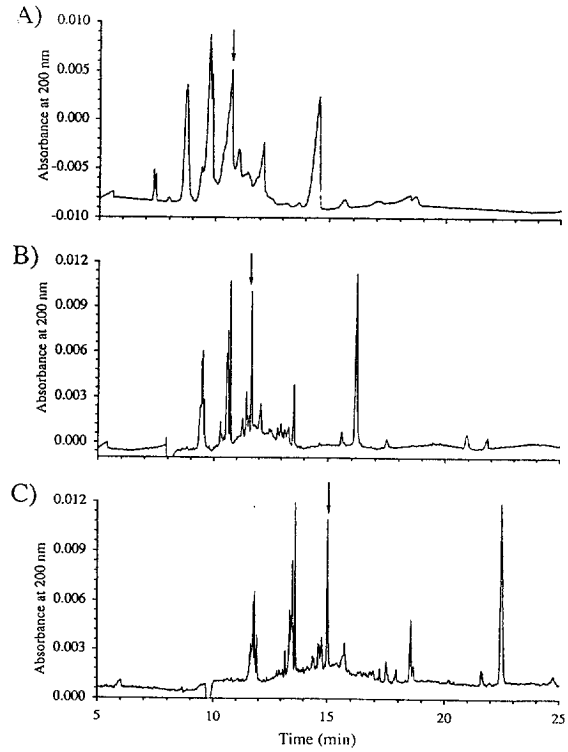

Fig. 3. CE–UV analysis of peptides arising from CNBr cleavage of ovalbumin using formic acid concentrations of 0.1 *M* (A), 0.5 *M* (B) and 2.0 *M* (C). The arrow indicates the position of peptide $CB_{197-210}$. Conditions: Fused-silica capillary 80 cm × 50 μm I.D. coated with a solution of 5% Polybrene and 2% ethylene glycol. The analysis corresponds to the injection of approximately 9 pmoles of ovalbumin.

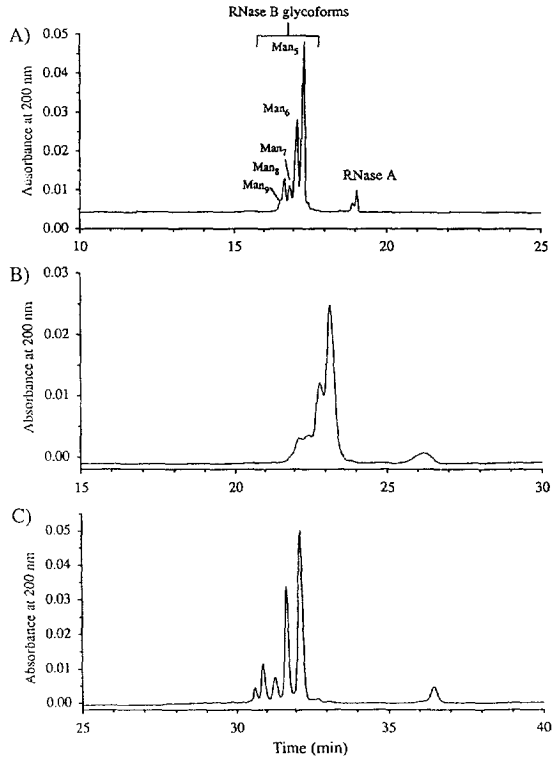

Fig. 4. CE–UV analysis of the glycoprotein RNase B using formic acid concentrations of 0.1 *M* (A), 0.5 *M* (B) and 2.0 *M* (C). Conditions as for Fig. 3 except that the analysis was obtained for the injection of 6 pmol of RNase B.

acid (Fig. 3 A) yielded poor separation performance, with wide peaks and unsatisfactory resolution. This situation was progressively improved as the acid concentration was increased to 2.0 *M* as observed in Figs. 3 A–C. For example, the calculated plate counts for the peptide $CB_{197-210}$ (indicated by an arrow in Fig. 3), were 372 000, 506 000, and 520 000 for formic acid concentrations of 0.5, 1.0 (not shown) and 2.0 *M*, respectively. In this case, the improved performance was mostly attributed to a decrease in electrosmotic flow which varied from 6.7×10^{-8} to 4.3×10^{-8} $m^2 V^{-1} s^{-1}$, in 0.1 to 2.0 *M* formic acid. Since the analyte migrates in the direction opposite to that of the anodal electrosmotic flow, the improvement in resolution comes as a result of the reduced zone velocity.

The resolution of protein glycoforms, and of individual glycopeptides arising from protein digests was also affected by the sample injection size. In a separate series of experiments, the effect of injection volume on peak resolution and zone broadening was investigated for peptides and glycopeptides arising from CNBr cleavage of ovalbumin. When using 2.0 *M* formic acid as electrolyte, ovalbumin glycopeptides were observed as a series of discrete peaks migrating between 11.5 and 12.0 min (Fig. 3 C). The analysis shown in Fig. 3 C corresponds to approximately 40 nl injection (1.5% of the capillary volume) of a 10 mg/ml solution of ovalbumin digest. Under such conditions good resolution of individual glycopeptides was obtained (see following discussion). However as the injection volume was progressively increased up to 110 nl (4.5% of the capillary volume) peak resolution gradually deteriorated due to extensive zone broadening. In order to maintain proper peak definition sample loadings were kept to less than typically 3% of the capillary volume, or approximately 60 nl for a 1 m × 50 μm I.D. column. Such limitations naturally imply that relatively large concentrations of sample be used. The actual amounts injected on the column, for either CE–UV or CE–ESMS separations, were at most 10 pmoles assuming a single protein component. Such sample consumption levels are one to two orders of magnitude lower than those typically used in LC–MS experiments [34–36]. This enhanced sensitivity is related to the high separation efficiencies obtainable under zone electrophoresis conditions.

Improvement in peak resolution with increasing acid concentration was also observed for the analysis of RNase B glycoforms (Fig. 4). This glycoprotein, which has one N-linked glycosylation site at Asn_{38}, is comprised of a mixture of glycoforms containing from two to six mannose units appended to the pentasaccharide core structure (Fig. 1). At a concentration of 0.1 *M* formic acid resolution of all glycoforms was almost achieved. However, when the concentration of formic acid was increased to 0.5 *M* (Fig. 4 B) the resolution was considerably degraded and the peaks became unusually broad, although the electrosmotic flow was significantly lower than that observed in Fig. 4 A. A further

increase of the formic acid concentration to 2.0 *M* reestablished proper resolution of all RNase B glycoforms, though the separation time was much longer than that obtained earlier. The incremental change in molecular mass upon addition of Man residues on the RNase B significantly affects the mobility of the glycoprotein. On average, each extra mannose residue appended on RNase B resulted in a decrease of 2×10^{-10} $m^2 V^{-1} s^{-1}$ or approximately 1% of the electrophoretic mobility. This subtle change in electrophoretic mobility is naturally more perceptible when the electrosmotic flow is reduced, a situation that prevails at a formic acid concentration of 2.0 *M* (Fig. 4 C).

The unusual zone broadening taking place with 0.5 *M* formic acid was attributed to a change in conformational state of RNase B. This was further supported in CE–ESMS experiments (see following discussion) by the observation of a shift of the multiply charged ion envelope toward higher *m/z* ratios as the concentration of formic acid was increased from 0.1 *M* to 2.0 *M*. At a concentration of 0.1 *M* the most abundant charge state of RNase B corresponded to $[MH_{11}]^{11+}$, whereas $[MH_9]^{9+}$ was the most intense ionic species for 2.0 *M* formic acid. This unusual effect is ascribed to a change in the protein conformational state. Such a phenomenon has also been described previously for chymotrypsinogen A [46], and could explain the concurrent changes in electrophoretic mobilities and ionic profiles as the acidity of the solution is varied. An analogous situation was reported recently by Mirza and Chait [47] for cytochrome C analyzed by ESMS whereby the multiply charged ion envelope was progressively shifted toward higher *m/z* values as the pH of the sample solution was lowered. The higher acid concentration used in the present study could possibly lead to a re-coiling of the glycoprotein, thus masking basic residues buried inside the globular structure resulting in a shift of the ionic profile toward lower charge states.

The electrophoretic conditions described above were also found compatible with the operation of electrospray ionization. The conditioning of the capillary surface with Polybrene and ethylene glycol did not lead to any unusual extraneous background signal or leaching of the coating reagents during the normal CE–ESMS operation. The CE–ESMS analysis of RNase B using 2.0 *M* formic acid is shown in Fig. 5. The total ion current (TIC) for the full mass scan acquisition (*m/z* 1300–2000) is presented in Fig. 5 A, along with extracted mass spectra in Figs. 5 B–D for peaks observed at 35.6, 35.2 and 40.8 min, respectively. The analyte ions were easily distinguished from the background signal, and a series of ions corresponding to multiply protonated forms of the molecule (attachment of 8 to 11 protons) were observed in each mass spectrum. For each electrophoretic peak, the *m/z* values were used to calculate the molecular mass

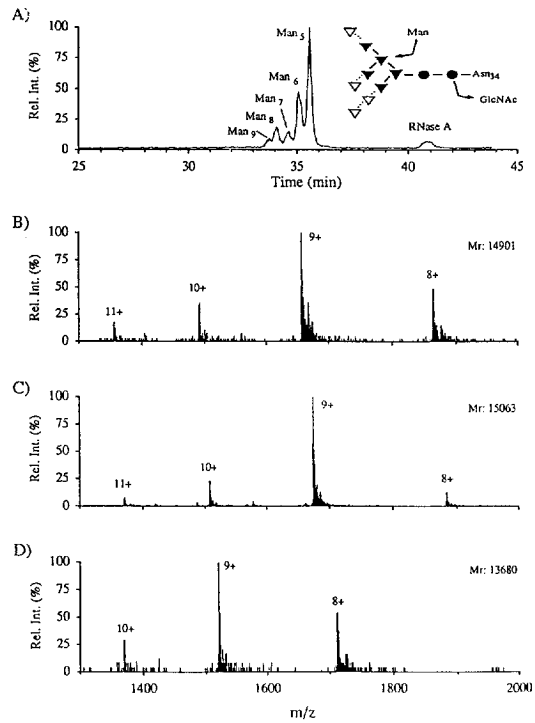

Fig. 5. CE–ESMS analysis of RNase B. (A) Total ion electropherogram for the full mass scan acquisition (*m/z* 1300–2000). Extracted mass spectra for peaks migrating at 35.2 (B), 35.6 (C), and 40.8 min (D). The calculated molecular mass (M_r) is shown on the right corner of each spectrum. Conditions: Fused-silica capillary 110 cm × 50 μm I.D. coated with a solution of 5% Polybrene and 2% ethylene glycol, 2.0 *M* formic acid, injection of 6 pmol of RNase B. Sheath flow of 5 μl/min of an aqueous solution of 0.2% formic acid and 25% methanol.

(isotope-averaged, M_r) of the proteins. Mass assignments under these scanning conditions were typically within 0.01–0.02% of the calculated values. The molecular mass of the major RNase B glycoform observed at 35.2 min (Fig. 5 B) was 14 901, in good agreement with the expected value for a Man_5–$GlcNAc_2$ moiety covalently bonded to the non-glycosylated RNase A. It is interesting to note that in addition to these RNase B glycoforms, a small peak of RNase A (calc. M_r 13 682), assumed to be a minor contaminant of this preparation, was also observed at 40.8 min in Fig. 5 A.

It is noteworthy that the RNase B glycoform distribution obtained using CE–ESMS (Fig. 5 C) is virtually indistinguishable from that obtained previously using CE–UV (Fig. 4 A). For example, CE–UV analyses, conducted using a 2 *M* formic acid buffer, indicated that the relative proportions of individual glycoforms as calculated from their respective peak areas were 3, 10, 7, 29 and 51%, for Man_9, Man_8, Man_7, Man_6, and Man_5 glycoforms, respectively. In comparison, CE–ESMS analyses performed under similar electrophoretic conditions yielded relative proportions of 3, 8, 7, 31 and 51%, respectively, for the same glycoproteins. These results are within 2–5% of those reported previously by Rudd et al. [25] on a different batch of RNase B. These workers used a combination of chemical and enzymatic digestion followed by analyses with CE–UV, ion exchange, and gel permeation chromatography. Although the intent of the present investigation was not to establish detection limits of individual glycoforms, data presented here and by other research groups [25,28–32] indicate that glycoproteins (or glycopeptides), present at levels as low as 1–3%, can be detected using CE with either UV or mass spectrometric detection.

The application of CE–ESMS to the analysis of protein glycoforms was also investigated for more complex samples in which the microheterogeneity of the glycoprotein arose from both glycosylation at different Asn residues and variability in the carbohydrate composition at a given site. An example of this is shown in Fig. 6 for the analysis of horseradish peroxidase

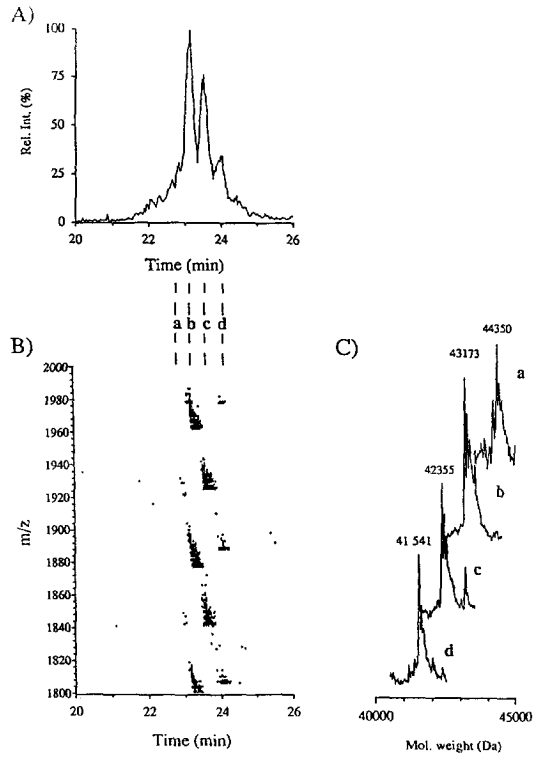

Fig. 6. Analysis of horseradish peroxidase glycoforms using CE–ESMS. (A) Total ion electropherogram for a narrow scan acquisition from *m/z* 1800–2000. (B) Contour profile displaying the distribution of *m/z* values vs. time. (C) Reconstructed molecular mass profile calculated from the multiply charged ions observed for each peak annotated a–d. Conditions as for Fig. 5 except that 5 pmol of the glycoproteins were injected on the column.

(HRP). This glycoprotein has a total of 9 glycosylation sites. The molecular mass of the carbohydrate-free protein, calculated from the published sequence is 33 893 [48]. In this particular example, a narrow mass range was scanned rapidly to provide adequate temporal and mass spectral resolution of the different glycoforms. Under such scanning conditions charge states corresponding to the attachment of 22–24 protons on the molecule were detected. The TIC (*m/z* 1800–2000) is shown in Fig. 6 A, along with the two dimensional depiction of ion intensity as a function of both *m/z* and time for the same analysis (Fig. 6 B). The reconstructed molecular mass profiles obtained for selected

groups of HRP glycoforms, annotated a–d in Fig. 6 are also presented in Fig. 6 C. Higher molecular mass glycoforms with lower electrophoretic mobilities migrate first, followed by components with lower m/z values. The concurrent changes in both electrophoretic mobilities and M_r values are evidenced in Fig. 6 B by the observation of diagonal lines of negative slope. As will be mentioned in the following section, this characteristic feature, which is unique to glycoproteins and glycopeptides, has been used in the present study to identify microheterogeneity at a given glycosylation site.

The complexity of the glycoform distribution of HRP is easily visualized from the contour profile of Fig. 6 B by the observation of a series of closely spaced m/z values for each group of peaks (a–d). The observed M_r values ranged from 41 000 to 44 000, suggesting that the carbohydrate content of HRP could account for as much as 25–30% of the protein molecular mass. Earlier studies by Harthill and Ashford [49] have indicated that N-glycans released from HRP include a series of complex carbohydrates having Xyl, Fuc and possibly GlcNAc residues attached to the pentasaccharide core. The most abundant glycoform identified in Fig. 6 has a molecular mass of 43 176 (peak b), a value which is in good agreement with that obtained by Green and Oliver [50]. This particular glycoform has been assigned as the HRP protein devoid of a C-terminus Ser, and to which is attached 8 carbohydrate chains having the common structure $GlcNAc_2$, Man_3, Fuc_1, Xyl_1 (calc. M_r 43 175) [50]. Interestingly, a co-migrating glycoprotein, with a molecular mass 87 Da higher than that of the major component was observed in each peak series. This observation reflects the microheterogeneity of the C-terminal Ser of HRP [47]. Under the present electrophoretic conditions, the higher molecular mass group of glycoforms was found in peak a. The major component (M_r 44 350) of this group has a total of 9 possible N-glycans bonded to the protein. A summary of the mass measurements obtained for the individual glycoforms of HRP is presented in Table 1 together with tentative carbohydrate compositions. It is evident from this example that assignment of glycoforms still remains a complex task even when the molecular mass can be determined with relatively good accuracy and precision. More elaborate studies on the location of N-linked glycans within the glycoprotein are thus required to facilitate the interpretation of carbohydrate composition and the identification of sites of heterogeneity.

Table 1
Tentative assignment of carbohydrate chains in glycoforms of horseradish peroxidase based on mass measurements taken from data of Fig. 6

Peak assignment	M_r (obs.)	M_r (calc.)	Carbohydrate chains[a]
a	44 431	44 434	9 $Carb_1$
	44 355	44 346	(−Ser) + 9 $Carb_1$
b	43 345	43 337	8 $Carb_1$ + Hex
	43 264	43 263	8 $Carb_1$
	43 176	43 175	(−Ser) + 8 $Carb_1$
c	42 520	42 516	(−Ser) + 7$Carb_1$ + 1 $Carb_2$ + Hex
	42 443	42 441	7 $Carb_1$ + 1 $Carb_2$
	42 355	42 354	(−Ser) + 7 $Carb_1$ + 1 $Carb_2$
d	41 627	41 635	6 $Carb_1$ + $Carb_2$ + $Carb_3$
	41 541	41 548	(−Ser) + 6 $Carb_1$ + $Carb_2$ + $Carb_3$

[a] Assignment based on proposed structures of Green and Oliver [50]. −Ser indicates the protein sequence [47] devoid of C-terminus serine. $Carb_1$: $GlcNAc_2$, Man_3, Fuc_1, Xyl_1. $Carb_2$: $GlcNAc_1$, Fuc_1. $Carb_3$: $GlcNAc_1$, Hex_1.

3.2. Identification of microheterogeneity at a single glycosylation site

The characterization of individual carbohydrate chains attached to any given glycosylation site presents a considerable challenge to the analyst. Patterns of microheterogeneity are sometimes difficult to recognize in CE–UV analyses of proteolytic digests, as glycopeptides might be obscured by co-migrating peptides. Furthermore, single parameter detection with assignment based on peak position offers limited information on the carbohydrate structure. The analysis of glycoprotein digests by CE–ESMS can therefore provide an additional level of specificity. More particularly, contour intensity plots as a function of m/z and time, such as that presented in Fig. 6, can be used to identify patterns of glycosylation at a single attachment site. Glycopeptides that vary by the extension of hexose residues display an incremental change in both molecular mass and electrophoretic mobility. This feature is usually reflected by the appearance of diagonal lines in the corresponding contour profile, as mentioned above.

The analysis of a tryptic digest of a sample of RNase B, previously reduced and alkylated with iodoacetamide, is shown in Fig. 7. As indicated in the preceding section, this glycoprotein has only one N-glycosylation site at Asn_{34}. The pattern of heterogeneity can easily be identified in Fig. 7 B by the observation of a single diagonal line between 15 and 17 min. These ions correspond to the doubly-protonated glycopeptide molecules (T_{34-37}), and have m/z values extending from 750 to 1200. Under the present electrophoretic conditions, baseline resolution of each glycopeptide was obtained as evidenced from the total ion electropherogram (m/z 600–1500) shown in Fig. 7 A. The glycopeptide profile obtained for the tryptic digest of RNase B is similar to that observed previously for the native glycoprotein (Fig. 4 C). It is interesting to note that an additional glycopeptide having only 4 mannose residues was observed in Fig. 7. The abundance of the corresponding glycoform was much lower in the original RNase B sample

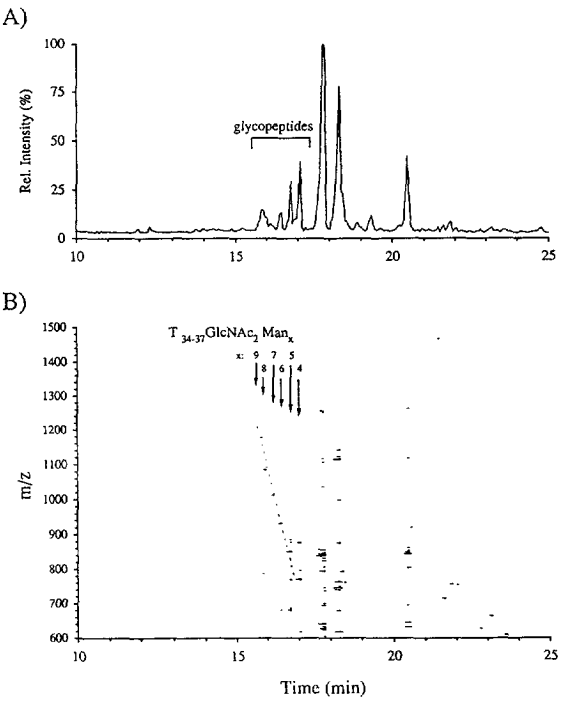

Fig. 7. CE–ESMS analysis of tryptic peptides from RNase B. (A) Total ion electropherogram for m/z 600–1500. (B) Contour profile showing characteristic glycoform pattern as indicated by the dotted diagonal line. Each glycopeptide is spaced by m/z 81 ($M_{r,Man}/2$) from its adjacent peak. Conditions as for Fig. 5.

(peak at 32.8 min in Fig. 4 C), suggesting that the formation of this glycopeptide might be due to hydrolysis of the glycosydic bond during sample preparation.

The potential of CE for analysis of closely related glycoforms was also investigated for more complex samples such as ovalbumin. This glycoprotein has a molecular mass of 42 950.2, and has two potential sites of phosphorylation on serine 68 and 344 and an N-terminal acetyl group [30,51]. The electrospray mass spectrum of this glycoprotein (Fig. 8) shows a complex series of peaks extending from M_r 44 000 to 45 000. One of the more abundant glycoform is observed at M_r 44 330, and possibly corresponds to the high mannose carbohydrate $GlcNAc_2Man_6$ attached

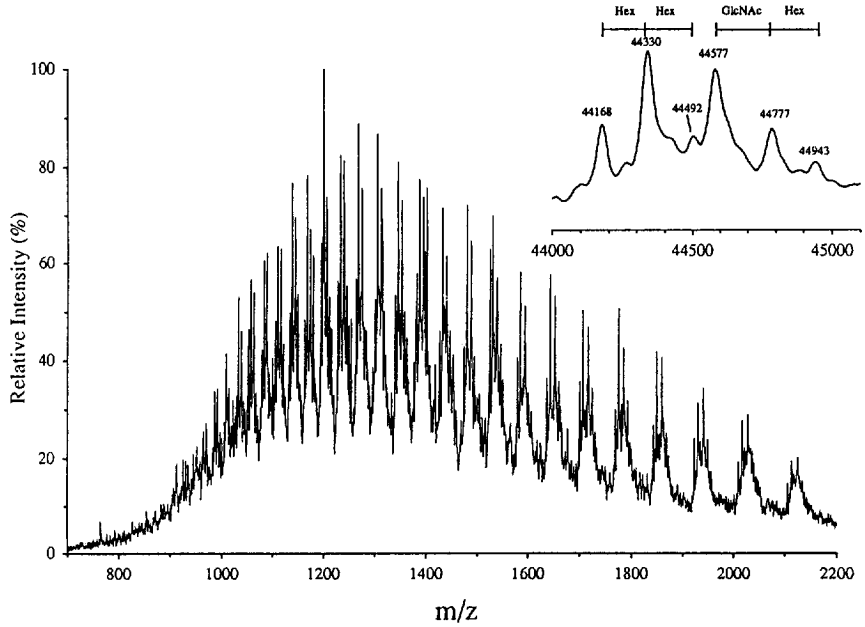

Fig. 8. Electrospray mass spectrum of ovalbumin. Flow injection analysis corresponding to the injection of approximately 500 pmol of the glycoprotein. Conditions: 15 μl/min flow rate, 0.1% aqueous trifluoroacetic acid in 50% acetonitrile.

to the diphosphate protein (calc. M_r 44 329.3). In addition to this major component, other high mannose glycoproteins showing simple addition and removal of one hexose residue from this major glycoform are also observed in Fig. 8 at 44 492 and 44 168, respectively. However the glycoform population of ovalbumin appears much more complex as evidenced by the broadening of the peaks corresponding to higher molecular mass members. Although ovalbumin has only one glycosylation site, considerable heterogeneity has been noted in the types of carbohydrate bonded to the Asn_{292} residue. CE–UV analyses of PA derivatives of oligosaccharides from ovalbumin, using borate buffers, have indicated that this glycoprotein is comprised of both high mannose and hybrid-type carbohydrates [29].

The CE analysis of CNBr peptides from ovalbumin, using both UV and mass spectrometric detection, is shown in Figs. 9 A and 9 B, respectively. Glycopeptides corresponding to the protein segment 288–298 are observed as a series of closely spaced peaks migrating between 16 and 18 min. The CE–UV profile of the glycopeptides is qualitatively similar to that observed for the analysis of the native glycoprotein [38], and confirmed that ovalbumin heterogeneity is mostly associated with its glycoform population.

The mass spectra of individual glycopeptides were characterized by abundant signals for doubly-, and triply-protonated molecules, $[MH_2]^{2+}$ and $[MH_3]^{3+}$. The molecular mass of each glycopeptide was calculated using the observed m/z values. Correlation of these mass measurements with those predicted from available information on possible carbohydrate structures of ovalbumin [29] facilitated the identification of individual glycopeptides annotated in Fig. 9 A. Examples of extracted mass spectra are shown in Figs. 9 C and 9 D for glycopeptides migrating at 16.8 and 17.2 min in Fig. 9 B. The observed M_r value for the glycopeptide of Fig. 9 C is 2656 Da, in good agreement with a high mannose type carbohydrate having the structure $GlcNAc_2$, Man_6 (calc. M_r 2656.7). Doubly-charged fragment ions corresponding to sequential losses of Man residues (spacing of m/z 81) from the

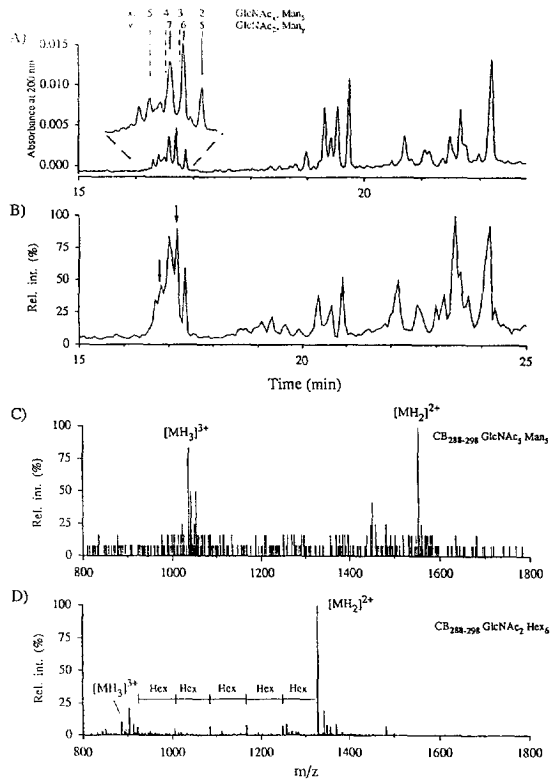

Fig. 9. Analysis of peptides arising from CNBr cleavage of ovalbumin. (A) CE–UV analysis; (B) Total ion electropherogram obtained for the CE–ESMS analysis (m/z 800–1800). Extracted mass spectra taken at 16.8 and 17.2 min in B are shown in C and D, respectively. The glycan assignments shown in A are based on mass measurement obtained from the full mass spectral analysis. Conditions for both UV and ESMS: fused-silica capillary 110 cm × 50 μm I.D. coated with a solution of 5% Polybrene and 2% ethylene glycol, electrolyte: 2.0 *M* formic acid. The analysis corresponds to the injection of approximately 2 pmoles of ovalbumin digest.

parent molecule are observed in the m/z region between the $[MH_2]^{2+}$ and $[MH_3]^{3+}$ ions in Fig. 9 C. These fragment ions arise from collisional activation in the orifice/skimmer region prior to mass analysis, and provide further confirmation of the proposed structure.

The glycopeptide shown in Fig. 9 D has a molecular mass of 3104, and corresponds to a hybrid-type carbohydrate having a $GlcNAc_5$, Man_5 structure (calc. M_r 3104.2). It is interesting to note that the $[MH_3]^{3+}$ ions are more abundant for the complex type than for the high mannose carbohydrates suggesting that extra GlcNAc residues confer additional sites of protonation. The CE–ESMS analysis of CNBr peptides of ovalbumin also indicated that the highest member of the series of hybrid carbohydrates observed at 16.7 min in Fig. 9 B has an extra hexose residue in addition to that of the glycopeptide shown in Fig. 9 D. In fact, the family of hybrid carbohydrates appears to have a number of GlcNAc units extending from 2 to 5 residues while the high mannose series, which possess only 2 GlcNAc, has from 5 to 7 mannose residues. As the total number of carbohydrate residues of either type of N-linked carbohydrates approaches similar values, baseline resolution of all glycoforms becomes increasingly difficult to achieve. This situation was observed in Fig. 9 B for the broader peak at 17.0 min, which contains glycopeptides having the following compositions: $GlcNAc_5$ Man_4 (M_r 2942.1), $GlcNAc_4$ Man_5 (M_r 2900.9) and $GlcNAc_2$ Man_7 (M_r 2818.8). Although these glycopeptides have similar electrophoretic mobilities they can be distinguished in the corresponding mass spectra based on their different molecular masses. The combined use of CE separation and mass spectral detection thus provides additional selectivity and facilitates the identification of closely related protein glycoforms.

3.3. Selective identification of carbohydrates in proteolytic digests of glycoproteins

In situations where no pattern of carbohydrate heterogeneity can be visualized from the contour profile of m/z vs. time, more elaborate mass spectral strategies are required to identify glycopeptides in complex digests. This task can be facilitated by using operating conditions that promote dissociation of the protonated molecules and formation of fragment ions specific to the oligosaccharides prior to their mass detection [34–36]. By increasing the orifice voltage of the mass spectrometer to typically 100 V, N-linked carbohydrates which possess GlcNAc and Man residues will dissociate to give rise to oxonium fragment ions at m/z 204 and 163. Although other oxonium ions, such as m/z 133 and 147,

can be formed for xylose and fucose residues, fragment ions at m/z 163 and 204 were generally more abundant for a number of glycopeptides examined in the present study.

The CE–ESMS analysis of a tryptic digest from the seed lectin of *E. corallodendron* is shown in Fig. 10. This analysis used both selected ion monitoring for m/z 163 and 204 with a high orifice voltage (Fig. 10 A), and a separate full mass scan acquisition from m/z 500–1500 acquired with an orifice voltage setting of 50 V(Fig. 10 B). Earlier investigations in this laboratory of a group of legume lectins [52] have indicated that *E. corallodendron* lectin has two

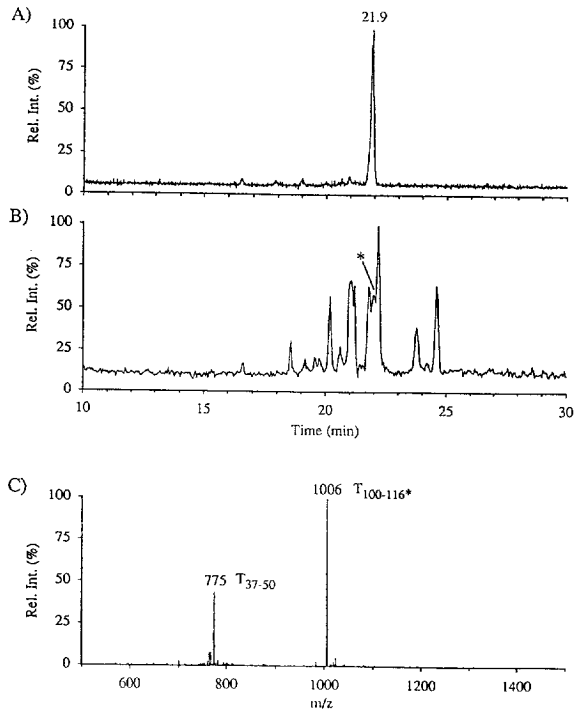

Fig. 10. Analysis of tryptic digest of the seed lectin *E. corallodendron*. (A) CE–ESMS analysis performed under SIR mode (combined profile for m/z 163 and 204) using an orifice voltage of 100 V. (B) Total ion electropherogram for a narrow scan acquisition from m/z 500–1500 (orifice voltage: 50 V). (C) Extracted mass spectrum taken at 21.9 min in B, asterisk indicates the glycopeptide. Conditions as for Fig. 5 except that 1 pmol of the tryptic digest of *E. corallodendron* was injected on the column.

possible sites of glycosylation with no apparent heterogeneity. The CE–ESMS analysis presented in Fig. 10 A shows only a single glycopeptide migrating at 21.9 min. This unexpected result was also confirmed by LC–ESMS analysis of the same tryptic digest sample, suggesting that the second glycopeptide might have been lost through the sample purification process or that its m/z value lies outside the selected m/z range. In contrast to the relatively simple electropherogram of Fig. 10 A, the CE–ESMS analysis performed under full mass scan acquisition (Fig. 10 B) gave a number of peaks some of which were only partly resolved. The comparison of these two analyses thus provides a selective means of identifying glycopeptides in relatively complex digests.

The mass spectrum of the glycopeptide peak indicated by an arrow at 21.9 min in Fig. 10 B is presented in Fig. 10 C. Two co-migrating components at m/z 775 and 1006 were observed in this mass spectrum. The peptide of lower intensity was presumed to be the sequence segment 37–50 based on the available cDNA sequence information [53]. However, no molecular mass could be matched for the more abundant component at m/z 1006. More selective analysis using combined tandem mass spectrometry (Fig. 11) revealed that fragment ions originating from dissociation of m/z 1006 yield characteristic features associated with the carbohydrate structure of the glycopeptide. In particular, MS–MS spectra of N-linked glycopeptides typically exhibit a series of fragment ions corresponding to cleavage of glycosidic bonds with almost no dissociation of the peptide backbone structure. In the example shown in Fig. 11, an intense fragment ion is observed at m/z 1024 corresponding to cleavage of the glycosydic bond between the two GlcNAc residues leaving a single GlcNAc attached to the doubly-charged peptide. This information alone suggested that the glycopeptide has a M_r value of 3015, and could correspond to the tryptic peptide segment 100–116 to which is attached a complex carbohydrate $GlcNAc_2$, Man_3, Fuc_1, Xyl_1. Further support of this proposal was obtained through

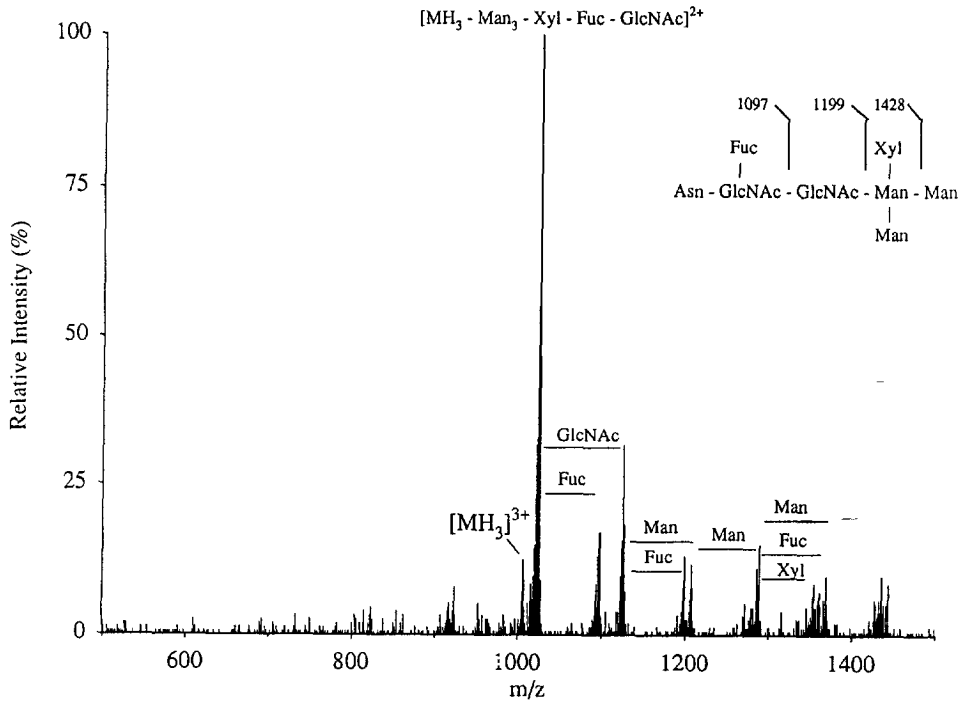

Fig. 11. Tandem mass spectrum of m/z 1006 corresponding to the $[MH_3]^{3+}$ ion of the glycopeptide $T_{100-116}$ from the tryptic digest of *E. corallodendron*. Conditions: Argon target gas, collision gas thickness of 3.5×10^{15} atoms cm^{-2}.

rationalization of fragment ions observed in Fig. 11, which confirmed the carbohydrate sequence of this glycopeptide.

The selective identification of glycopeptides via detection of characteristic oxonium ions was found particularly valuable in the assignment of glycans from horseradish peroxidase. This glycoprotein probably represents the most challenging situation examined in this study. As mentioned previously, this peroxidase exhibits both heterogeneity in the type of glycans attached to a particular site and a total of 9 possible sites of glycosylation. The more abundant glycoform group, annotated b in Fig. 6, has a molecular mass distribution from 43 176 to 43 345 (Table 1) and comprises eight N-linked glycans. From the contour profile of Fig. 6, variability in the type of oligosaccharides is also clearly evident, but the exact site of substitution cannot be established from these data alone. In order to address key questions regarding the location and the type of glycans attached to a particular Asn residue it was necessary to conduct CE–ESMS experiments on a proteolytic digest of this glycoprotein. In this particular case the analysis of tryptic peptides was suitable since arginine and lysine residues are evenly distributed through the entire protein structure, thus producing glycopeptides of convenient size.

The CE–ESMS analyses of the tryptic digest of horseradish peroxidase (reduced and alkylated with iodoacetamide), performed under high orifice voltage conditions, is shown in Fig. 12 A together with full mass spectral acquisition in Fig. 12 B. As demonstrated in the previous example, selective identification of glycopeptides was facilitated through the observation of oxonium ions at m/z 163 and 204, and enabled rejection of peaks not containing any glycan. From Fig. 12 A, it is possible to identify at least

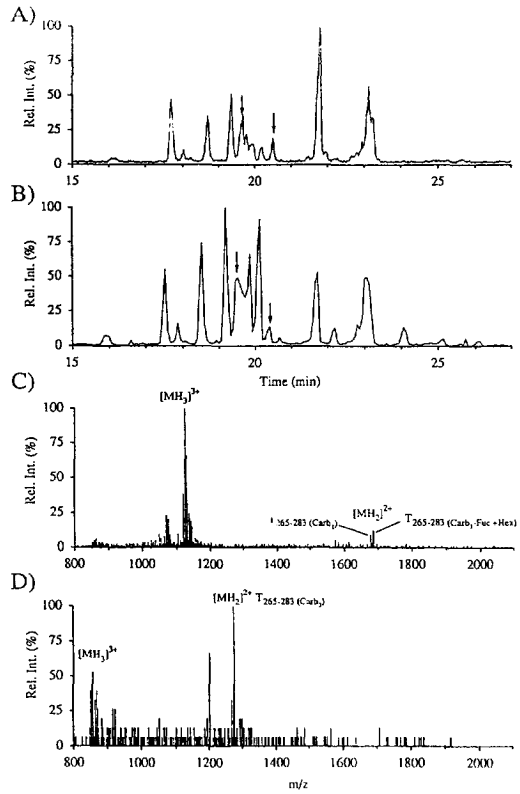

Fig. 12. Analysis of tryptic digest of horseradish peroxidase. (A) CE–ESMS analysis conducted under SIR mode (combined profile for m/z 163 and 204) using an orifice voltage of 100 V. (B) Total ion electropherogram for a narrow scan acquisition from m/z 800–2100 (orifice voltage: 50 V). Extracted mass spectra for peaks identified by arrows in B at 19.5 min (C) and 20.3 min (D). Conditions as for Fig. 5 except that 5 pmol of the tryptic digest of horse radish peroxidase were injected on the column.

8 potential glycopeptide peaks, some of which show heterogeneous patterns. Assignment of possible glycan structures for individual glycopeptides could be made using the observed M_r values and matching these with the tryptic peptide molecular masses predicted from the known protein sequence [48].

The information obtained from these analyses is conveniently summarized in Table 2 for peaks observed in Fig. 12 B. As expected, most of the glycopeptides bear a complex N-linked glycan having a xylose and a fucose appended to the pentasaccharide core. This glycan, labelled $Carb_1$ in Table 2, has been identified for 8 glycosylation sites. It is interesting to note that no ion was detected under the present conditions for tryptic peptide $T_{284-298}$ modified by this complex carbohydrate attached at Asn_{286}. Rather, a signal at m/z 1587, corresponding to the tryptic peptide free of any oligosaccharide was detected at 22.2 min in Fig. 12 B. This result suggests that a large proportion of Asn_{286} residues are devoid of an oligosaccharide, although a glycoprotein group of low abundance bearing 9 N-linked glycans was observed in Fig. 6 (peak a).

From these CE–ESMS analyses it was also possible to verify whether any of the other Asn residues showed variability in the carbohydrate distribution or in the type of substituent attached at any particular position. Except for Asn_{286}, all of the other Asn residues presenting the characteristic tripeptide signal for N-glycosylation (positions 13, 57, 158, 186, 198, 214, 255 and 268) appeared to bear a N-linked carbohydrate. Heterogeneity in the glycan composition was observed for three Asn residues. In some cases the type of substitution led to a significant change in electrophoretic mobilities, such that two different N-linked glycans could be resolved from each other.

In other situations only substitution of a single carbohydrate residue was involved, and no observable change in migration time could be observed. An example of this is shown in Fig. 12 C for the tryptic peptide $T_{265-283}$, for which two closely related carbohydrate chains were observed. This mass spectrum, which was extracted from the peak at 19.5 min in Fig. 12 B, shows doubly- and triply-protonated molecules for a glycopeptide consisting of a $Carb_1$ structure (M_r 3355) together with another glycan corresponding to a simple substitution of fucose for a hexose residue (M_r 3371). It is noteworthy that another closely related glycopeptide was observed at 20.3 min in Fig. 12 B. The extracted mass spectrum of this latter glycopeptide is shown in Fig. 12 D, and indicates that this smaller component has a molecular mass of 2549. The only match that could be obtained for this glycopeptide was for a shorter chain carbohydrate having a GlcNAc–Hex attached to the

Table 2
Identification of peptides and glycopeptides arising from the tryptic digest of horseradish peroxidase shown in Fig. 12

Migration time (min)	M_r (obs.)	M_r (calc.)	Protein segment	Carbohydrate chains*
17.5	3381	3379.6	1–19	$Carb_1$
17.9	2558	2557.8	1–19	$Carb_2$
18.5	4987	4986.1	184–206	2 $Carb_1$
19.2	3675	3674.0	242–264	$Carb_1$
19.5	3355, 3371	3354.5, 3370.5	265–283	$Carb_1$, $Carb_1$–Fuc + Hex
19.6	3147, 3162	3145.5, 3161.5	207–224	$Carb_1$, $Carb_1$–Fuc + Hex
19.9	2746	2745.1	125–149	–
20.1	2604	2602.9	94–118	–
	2642	2643.0	184–206	–
20.3	2548	2548.7	265–283	$Carb_3$
21.7	1844	1842.8	154–159	$Carb_1$
22.2	1586	1586.8	284–298	–
22.9	4174, 4210	4173.5, 4203.5	39–62	–
23.1	4012, 4044	4011.4, 4041.4	39–62	$Carb_1$ + Hex, $Carb_1$–Xyl + 2Hex
24.1	1022	1022.5	66–75	$Carb_1$, $Carb_1$–Xyl + Hex
	1186	1185.3	233–241	–
25.1	959	960.1	20–27	–
26.1	935	935.1	225–232	–
27.7	1475	1475.5	160–174	–

[a] $Carb_1$: $GlcNAc_2$, Man_3, Fuc_1, Xyl_1. $Carb_2$: $GlcNAc_1$, Fuc_1. $Carb_3$: $GlcNAc_1$, Hex_1.

same peptide segment. A similar situation was also noticed for the glycopeptide T_{1-19} for which two closely migrating peaks were observed at 17.9 min. Based on the observed molecular mass, this smaller component is presumed to contain a glycan composed of GlcNAc–Fuc.

In view of this latter finding it appears that the variability in the carbohydrate composition first noticed in the native horseradish peroxidase analysis (Fig. 6) arises mostly from two particular sites Asn_{13} and Asn_{268}. For example, the series of glycoforms with molecular masses in the range 41 500–41 700 (peak d in Fig. 6) could presumably have both sites occupied by the small N-linked glycan. In the case of the glycoform labelled as peak c in Fig. 6, the $Carb_3$ glycan attached to Asn_{268} could be substituted for the higher molecular mass $Carb_1$, thus accounting for the shift in the observed molecular mass of the glycoprotein. Naturally the above proposal is only tentative for the moment, and further confirmation of these assignments will require both glycan analyses and gas phase sequencing. However the simple CE–ESMS experiments outlined above do provide valuable structural information that can be applied when sample availability is limited.

4. Conclusion

Capillary electrophoresis offers a valuable tool for the analysis of glycoconjugates and provides a means of determining glycoform populations in different types of glycoprotein. Separations conducted under anodal electrosmotic flow with Polybrene coated capillaries and acidic buffers permitted resolution of glycopeptides and glycoproteins according to the number of carbohydrate residues appended on the peptide backbone. In the case of RNase B, a glycoprotein of approximately 15 000 having from three to seven mannose residues attached to the pentasaccharide core structure, the extension of each extra mannose is reflected by a corresponding decrease of 2×10^{-10} $m^2 V^{-1} s^{-1}$ in electrophoretic mobility. However, in order to achieve resolution of closely related glycoforms it was necessary to reduce the extent and magnitude of the opposing electrosmotic flow. Under these

electrophoretic conditions the electrosmotic flow was reduced from 6.7×10^{-8} to 4.3×10^{-8} $m^2 V^{-1} s^{-1}$ when the concentration of formic acid was increased from 0.1 to 2.0 *M*. For high mannose glycoproteins, which exhibit the simplest form of carbohydrate heterogeneity, baseline resolution of each individual glycoform was achieved using 2.0 *M* formic acid as electrolyte.

Such capillary conditioning and buffer conditions were also found entirely compatible with the use of electrospray mass spectrometry. Although detection limits were not investigated in the present study, it was possible to obtain full mass spectral acquisition of glycoproteins and their digests using low pmole injection sizes. Mass spectral detection of CE peaks was found particularly useful in obtaining information on the structure of the analyte, and in locating and identifying the carbohydrate attached to the protein. Glycoform patterns arising from a single glycosylation site were easily identified using contour intensity profiles of m/z vs. time by the observation of a diagonal line corresponding to a concurrent change of electrophoretic mobilities and molecular mass. The use of ESMS was also found necessary for the analysis of glycoproteins such as ovalbumin which consist of two families of closely related N-linked carbohydrates attached to a single glycosylation site. CE–ESMS analyses of CNBr peptides from ovalbumin enabled identification of a high mannose series extending from five to seven residues together with a family of hybrid type glycoforms having from two to five GlcNAc for a fixed number of five mannose residues.

The application of CE–ESMS was also investigated for the analysis of glycoforms of horse radish peroxidase and a lectin from *E. corallodendron*, both containing complex-type carbohydrates. Selective identification of glycopeptides present in digests of the glycoprotein was achieved using mass spectrometric conditions that promote the formation of characteristic oxonium fragment ions. Using this approach it was possible to identify not only the major carbohydrates attached to eight different glycosylation sites of horseradish peroxidase but also to characterize which site exhibit more extensive heterogeneity. Although the present work emphasizes the utility of CE–ESMS for analysing N-linked glycoproteins, a similar approach can be used to probe the structures of O-linked glycoproteins.

Acknowledgments

The authors would like to thank Dr. Martin Young from the Institute for Biological Sciences for valuable discussions, and for providing some of the samples of ovalbumin and *E. corallodendron* used in the present study. The technical assistance of Mrs. Pearl Blay from IMB is also gratefully acknowledged.

References

[1] A. Gottschalk, in A. Gottschalk (Editor), Glycoproteins, Elsevier, New York, 1966, Ch. 1, p. 1.
[2] A. Varki, Glycobiology, 3 (1993) 97.
[3] H. Lis and N. Sharon, Eur. J. Biochem., 218 (1993) 1.
[4] C.F. Goochee, M.J. Gramer, D.C. Andersen, J.B. Bahr and J.R. Rasmussen, in P. Todd, K. Sikdar and M. Bier (Editors), Frontiers in Bioprocessing II, American Chemical Society, Washington, D.C., 1992, p. 199.
[5] K. Olden, T. Yeo and K. Yeo, in H.J. Allen and E.C. Kisailus (Editors), Glycoconjugates, Structure and Functions, Marcel Dekker, New York, 1992, p. 403.
[6] J. Roth, Biochim. Biophys. Acta, 906 (1988) 405.
[7] C.B. Hirschberg and M.D. Snider, Annu. Rev. Biochem., 56 (1987) 63.
[8] J. Roth, J. Electron Microsc. Tech., 17 (1991) 121.
[9] R. Kornfeld and S. Kornfeld, Ann. Rev. Biochem., 54 (1985) 631.
[10] R.D. Marshall, Annu. Rev. Biochem., 41 (1972). 673.
[11] R.D. Marshall, Biochem. Soc. Symp., 40 (1974) 17.
[12] S. Takasaki, T. Mizuochi and A. Kobata, in V. Ginsburg (Editor), Methods in Enzymology, Academic Press, San Diego, CA, 1982, Vol. 83, p. 263.
[13] F. Maley, R.B. Trimble, A.L. Tarentino and T.H. Plummer, Anal. Biochem., 180 (1988) 195.
[14] D.A. Ashford, R.A. Dwek, T.W. Rademacher, H. Lis and N. Sharon, Carbohydr. Res., 213 (1991) 215.
[15] S. Hase, S. Hara and Y. Matsushima, J. Biochem. (Tokyo), 85 (1979) 217.
[16] W.T. Wang, N.C. LeDonne Jr., B. Ackerman and C.C. Sweeley, Anal. Biochem., 180 (1984) 366.
[17] J. Lieu, O. Shirota, D. Wiesler and M. Novotny, Proc. Natl. Acad. Sci. USA, 88 (1991) 2302.

[18] S. Honda, E. Akao, S. Suzuki, M. Okuda, K. Kakehi and J. Nakamura, Anal. Biochem., 180 (1989) 351.
[19] J. Suzuki-Sawada, Y. Umeda, A. Kondo and I. Kato, Anal. Biochem., 207 (1992) 203.
[20] A. Kondo, M. Kiso, A. Hasegawa and I. Kato, Anal. Biochem., 219 (1994) 21.
[21] W. Nashabeh and Z. El Rassi, J. Chromatogr., 514 (1990) 57.
[22] W. Nashabeh and Z. El Rassi, J. Chromatogr., 600 (1990) 279.
[23] M. Taverna, A. Baillet, D. Biou, D. Schlüter, R. Werner and D. Ferrier, Electrophoresis, 13 (1992) 359.
[24] S. Honda, A. Makino, S. Suzuki and K. Kakehi, Anal. Biochem., 176 (1989) 72.
[25] P.M. Rudd, I.G. Scragg, E. Coghill and R.A. Dwek, Glycoconjugate J., 9 (1992) 86.
[26] Z. El Rassi in P.R. Brown and E. Grushka (Editors), Advances in Chromatography, Marcel Dekker, New York, 1994, Vol. 34, p. 177.
[27] J.D. Olechno and K.J. Ulfelder, in J.P. Landers (Editor), CRC Handbook of Capillary Electrophoresis: A Practical Approach, CRC Press, Boca Raton, FL, 1994, p. 255.
[28] J.P. Landers, R.P. Oda and M.D. Schuchard, Anal. Chem., 64 (1992) 2846.
[29] S. Honda, A. Makino, S. Suzuki and K. Kakehi, Anal. Biochem., 191 (1990) 228.
[30] J.P. Landers, R.P. Oda, B.J. Madden and T.C. Spelsberg, Anal. Biochem., 205 (1992) 115.
[31] S. Suzuki, K. Kakehi and S. Honda, Anal. Biochem., 205 (1992) 227.
[32] S. Honda, S. Suzuki, A. Nitta, S. Iwase and K. Kakehi, A Companion to Methods in Enzymology, 4 (1992) 233.
[33] R.S. Rush, P.L. Derby, T.W. Strickland and M.F. Rohde, Anal. Chem., 65 (1993) 1834.
[34] S.A. Carr, M.J. Huddleston and M.F. Bean, Protein Sci., 2 (1993) 183.
[35] J.J. Conboy and J.D. Henion, Proceedings of the 39 Conference on Mass Spectrometry and Allied Topics, May 19–24, 1994, Nashville, TN, p. 1418.
[36] M.J. Huddleston, M.F. Bean and S.A. Carr, Anal Chem., 65 (1993) 877.
[37] A.W. Guzzetta, L.J. Basa, W.S. Hancock, B.A. Keyt and W.F. Bennett, Anal. Chem., 65 (1993) 2953.
[38] J.F. Kelly, S.J. Locke and P. Thibault, Discovery Newsletter, Beckman, Fullerton, 1993, Vol. 2, Issue 2.
[39] G. Allen, in R.H. Burdon and P.H. van Knippenberg (Editors), Laboratory techniques in biochemistry and molecular biology, Elsevier, New York, 1989, Vol. 9, p. 58.
[40] J.E. Wiktorowicz and J.C. Colburn, Electrophoresis, 11 (1990) 769.
[41] S. Pleasance and P. Thibault in P. Camillieri (Editor), Capillary Electrophoresis, Theory and Practice, CRC Press, Boca Raton, FL, 1993, p. 311.
[42] R.D. Smith, C.J. Barinaga and H.R. Udseth, Anal. Chem., 60 (1988) 1948.
[43] R.D. Smith, J.H. Wahl, D.R. Goodlett and S.A. Hofstadler, Anal. Chem., 65 (1993) 574A.
[44] I.M. Johansson, E.C. Huang, J.D. Henion and J. Zweigenbaum, J. Chromatogr., 554 (1991) 311.
[45] J.H. Wahl and R.D. Smith, J. Capillary Electrophor., 1 (1994) 62.
[46] P. Thibault, C. Paris and S. Pleasance, Rapid Commun. Mass Spectrom., 5 (1991) 484.
[47] U.A. Mirza and B.T. Chait, Anal. Chem., 66 (1994) 2898.
[48] K.G. Welinder, Eur. J. Biochem., 96 (1979) 483.
[49] J.E. Harthill and D.A. Ashford, Biochem. Soc. Trans., 20 (1992) 113S.
[50] B.N. Green and R.W.A. Oliver, Biochem. Soc. Trans., 19 (1991) 929.
[51] P.E. Perlmann, Adv. Protein Chem., 10 (1955) 1.
[52] P. Thibault and N.M. Young, unpublished results.
[53] R. Arango, S. Rozenblatt and N. Sharon, FEBS Lett., 264 (1990) 109.