NATIONAL INSTITUTE OF PUBLIC HEALTH AND THE ENVIRONMENT
BILTHOVEN, THE NETHERLANDS

RIVM Report 518002001

Development of Sensitive Mass Spectrometric Techniques in Protein and Peptide identification

A.P.J.M. de Jong, H.D. Meiring, B. Barosso, G.J. ten Hove and E. van der Heeft

Augustus, 1999

Dit onderzoek werd verricht in opdracht en ten laste van de Directie van het RIVM in het kader van project nr. 518002 (Analytisch Methodiekontwikkeling Volksgezondheid).
This investigation has been performed in order and for the account of the Board of Directors of the RIVM within the framework of project number 518002 (Analytical Methods Public Health).

National Institute of Public Health and the Environment, P.O. Box 1, 3720 BA Bilthoven, The Netherlands
phone: .. 31 30 274 91 11, fax: .. 31 30 274 29 71

MAILING LIST

1	Directie RIVM
2	Depot Nederlandse Publicaties en Nederlandse Bibliografie
3	Dr. A.J.W. van Alphen
4	Dr B. Baumann
5	Dr. C. Beuvery
6	Dr. C.A.C.M. van Els
7	Ing. J. van Gaans-van den Brink
8	Drs. C.A. Herberts
9	Dr. ir. P. Hoogerhout
10	Drs P.M.J.M. Jongen
11	Dr. G. Kersten
12	Dr. A.K.D. Liem
13	Dr. G. de Mik
14	Mw. Drs. E.G. van der Velde
15	Mevr. C van der Velden-de Groot
16	Dr. T. Visser
17	Dr. E.J.H.J. Wiertz
18	Dr. G. Zomer
19	Dr. P. van Zoonen
20	Prof. Dr. B van der Zeijst
21	Hoofd Bureau Voorlichting en Public Relations
22-26	Auteurs
27	Bibliotheek RIVM
28	Bureau Rapportenregistratie
29-50	Bureau Rapportenbeheer

ABSTRACT

This report describes the development of micro and nano-scale liquid chromatographic and capillary zone electrophoretic separation techniques combined with tandem mass spectrometry for the sensitive structure analysis of proteins and peptides. Mass spectrometry is emerging as the method of choice for the characterization and identification of unknown proteins and peptides. Sequence information obtained from derived peptides such as naturally processed antigenic peptides and proteolytic fragments is used to identify the underlying source protein using on-line database searching of protein and DNA databases. Typical sensitivities of current MS methods amounts around the sub-pmol (10^{-12}) of protein on a routine basis. Ongoing refinements in the technology, particularly the development of nano-scale technologies, makes successful applications possible down to the femto (10^{-15}) to attomole (10^{-18}) amounts. Here we report on the results of this laboratory on developments in this field over the past year.

SAMENVATTING

Dit rapport beschrijft de ontwikkeling van micro en nano-scale vloeistofchromatografie en on-line preconcentrering capillaire electrophorese gecombineerd met tandem massaspectrometrie bij gevoelige structuuranalyse analyse van eiwitten en peptiden. Massaspectrometrie is een sterk opkomende techniek voor het verkrijgen van sekwentie-informatie van deze biologisch relevante producten. De verkregen sekwentie-informatie van afgeleide peptiden (proteosomale peptiden en proteolyse producten) worden gebruikt voor het identificeren van het bijbehorende broneiwit met behulp van eiwit en DNA databanken. De gevoeligheid van de huidige MS technologien ligt rond een picomol (10^{-12}) hoeveelheid eiwit. Recente verfijningen, in het bijzonder de ontwikkeling van nano-schaal technologieën, brengen toepassingen op het femto (10^{-15}) tot attomole (10^{-18}) niveau binnen het bereik van de techniek.
Hier rapporteren wij over de resultaten van het R&D programma in het jaar 1998 op het gebied van de karakterisering van eiwitten en peptiden met behulp van massaspectrometrie.

CONTENTS

1 INTRODUCTION

In recent years, mass spectrometry has become a significant tool in the structure analysis of proteins and peptides. Matrix-assisted-laser-desorption/ionization time of flight mass spectrometry (MALDI-ToF MS) has proven to be useful in molecular weight analysis of large proteins [reviewed in Hillenkamp et al] and electrospray ionisation (ESI)-tandem mass spectrometry (MS/MS) coupled to liquid chromatography (LC) has become the method of choice for the sequence analysis of complex peptide mixtures [Hunt et al., 1992]. Main features of ESI/MS are its high sensitivity and the ability to produce multiply charged ions. Since mass spectrometers analyse ions on the basis of their mass-to-charge ratio, higher charge states increase the mass range of the instrument accordingly, thus allowing the determination of moderate to large molecular weight biomolecules by relatively small and low cost instruments. In addition, ESI is well compatible with liquid chromatography and capillary electrophoresis. A most illustrative example of the power of LC/MS techniques is shown in the analysis of major histocompatibility complex (MHC) class I and class II associated peptides [Reviewed in Engelhard et al., 1994].

In recent work of this laboratory we have further refined the on-line LC-ESI/MS technology by the development of micro (refers to micro-liter/min flow rate) and nano (nanoL/min flow rate) column switching systems. Such systems allow the rapid (~5 min) loading of relatively large sample volumes (typically 10-50 µL) onto a micro or nano LC column, thus obtaining two to three orders of magnitude [van der Heeft et al, 1997] increase in sensitivity compared to systems utilising normal bore (4.6 mm ID) LC columns. Other improvements in ESI/MS sensitivity of about one order of magnitude were obtained by the development of a robust sheathless electrospray interface in conjunction with capillary zone electrophoresis and nano-LC. In addition, the applicability of CZE to real samples was improved by the development of an on-line sample pre-concentration system allowing up to 100-1000 times larger sample volumes to be loaded into the CZE column without sacrifying the high separation efficacy of this technique [Barroso and de Jong, 1999]. Collectively, these refinements have contributed to an improved utility of ESI/MS in terms of sensitivity in protein and peptide characterisation at femto-mole (10^{-15}) to attomole (10^{-18}) level on a routine basis. These new technologies have been applied to the identification of MHC class I and II associated peptides [Reviewed in Engelhard et al., 1994; Apella et al., 1995; de Jong, 1998] and to the characterisations of proteins and vaccine products.

Another subject studied in the course of this program concerns the use of on-line protein and DNA database searching. Basically, searches involve the identification of proteins of interest on the basis of the occurrence of matching sequences in derived peptides and the source protein. In cases the protein of interest has not been sequenced before, searching cDNA sequences (expressed sequence tags, ESTs) has proven to be most effective, because EST databases are much larger than protein databases. For instance, about only 5-10% of all human proteins have been currently sequenced whereas about 90% of all the human EST sequences have been determined.

2 EXPERIMENTAL

2.1 Reagents and Standards

All chemicals and solvents were of analytical reagent grade or better. Methanol (HPLC grade), ortho-phosphoric acid, potassium biphosphate, acetic acid, hydrochloric acid, acetic acid (HOAc) and sodium hydroxide were purchased from Merck (Darmstadt, Germany), acetonitrile HPLC grade was from Rathburn (Walkerburn, Scotland) and trifluoroacetic acid (TFA) from Janssen Chimica (Beerse, Belgium). The reversed phases (RP) used in home made micro-HPLC columns were Purosphere RP-C_{18} (Merck, Darmstadt, Germany) Alltima C_8 (Alltech, Deerfield, IL, USA) and Poros 10R1 and 10R2 (PerSeptive Biosystems, Framingham, MA, USA). Model peptides and proteins were obtained both from commercial sources (Sigma, St. Louis, USA) or were in-house synthesised (courtesy: Dr. P Hoogerhout, Laboratory for Vaccine Research).
Standard solutions of peptides were prepared in 0.1% formic acid (v/v) mili-Q water and were stored frozen at -20°C.
HPLC solvents were water containing 0.1mol/L HOAc (solvent A) and acetonitrile containing 0.1mol/L HOAc (solvent B).

2.2 Protein digestion

In all cases, proteins were digested using modified trypsin (Promega, modified sequencing grade, Madison, USA) by two different protocols.

Protocol 1 (agarose gel concentration) [Meiring et al, 1998 and references therein): The protein of interest was isolated on a primary polyacrylamide gel by one-dimensional or two-dimensional electrophoresis. Subsequent agarose-based gel concentration is performed as described before. The gel slabs were washed thoroughly with water and the protein of interest, visualized by a conventional staining procedure (Coomassie Blue or Silver staining) was excised from the gel and cut into 1 x 1 mm slices. The gel pieces were washed with two successive portions of 5% trichloroacetic acid (TCA) for 15 min. The TCA was then removed by three consecutive 10 min washes with distilled water. Next, the gel pieces were equilibrated for at least 1 h in 0.1% SDS, 10% glycerol, 50 mM dithiothreitol (DTT), 12 mM Tris/HCl pH 6.8 (100-200 µl) and 0.01% bromophenol blue. Secondary gels consisting of 1% agarose in 125 mM Tris/HCl pH 6.8 were prepared in Pasteur pipettes. After solidification, the primary gel pieces including the sample buffer were transferred onto the agarose gel. The secondary gels were mounted into a cylindrical electrophoresis tank containing electrophoresis buffer (90 mM Tris, 90 mM boric acid and 0.1% SDS) in both reservoirs. The electrophoresis was carried out at a constant potential of 250 V. During electrophoresis the proteins elute from the primary gel pieces and reconcentrate at the moving boundary of the sample and electrophoresis buffer. Bromophenol blue was used to visualize the course of protein stacking. The concentration/elution procedure was stopped when the dye approaches approximately 2 cm from the lower end of the Pasteur pipette.
The narrow part of the agarose gel was isolated and stained for 30 min in a calconcarboxylic staining solution (0.05% calconcarboxylic acid in 45% methanol and 9% acetic acid). The gel was destained for 1-2 h in 5% methanol and 7.5% acetic acid and washed in distilled water for at least 1 h. The protein spot was isolated and the gel

piece was equilibrated and subsequently melted in digestion buffer (50 mM ammonium bicarbonate and 5 mM calcium chloride). After cooling down trypsin was added for overnight digestion at 37 °C.

Protocol 2 (In-gel digestion). A second digestion method was adapted from Mann et al, referred to as in-gel-digestion [Mann and Wilm, 1994]. Briefly, the protein isolated on a polyacrylamide gel was excised and consecutively washed with water, twice with water/acetonitrile (1/1, v/v), with neat acetonitrile and finally with 0.1 M ammonium carbonate/acetonitril (1/1, v/v) and dried in vacuum. Next, the protein is incubated with ~ 0.5 µg trypsin in 50 mM ammonium bicarbonate (20 µL) and 5 mM sodium chloride (20 µL) for one hour. The exess of digestion buffer is replaced by 20 µL of the same buffer but without trypsin and left overnight (16 h) at 37°C. The reaction is stopped by the addition of 10% TFA (1 µl). Tryptic peptides occurring in the aqueous supernatant were collected and the remaining peptides on the gel were extracted twice in water/acetonitril (1/1, v/v) containing 5% formic acid. Finally, 10 mM DTT (9 µl) was added and the pooled solvents lyophilised and stored until analysis at -20°C.

2.3 Liquid chromatography-electrospray ionisation mass spectrometry

All experiments were performed on a quadrupole ion trap instrument (LCQ, Finnigan MAT, San Jose, CA, USA) equipped with both a standard and a nano-ESI source [Wilm & Mann, 1996].
The standard ESI source was used in conjunction with the standard in-house developed LC system [Van der Heeft, 1998]. The system comprises of a 100 µm ID LC column operated at a typical flow of 0.5 µL/min, coupled to a pre-column in a switching system. Sheath liquid (60:40:0.1, methanol, water, acetic acid, v/v/v) was added to the column outlet flow in a ratio of 2:1 (v/v) by means of a syringe pump.

Sheathless LC/ESI/MS was used in conjunction with both CZE and nano-LC experiments. For this purpose, both the CZE and the LC column comprises a tapered tip (emitter) at the outlet end (see Fig.1) which was mounted into the nanoESI source assembly, with the emitter positioned in line with the orifice of ESI sampling capillary at a distance of approx. 5 mm.

One piece capillary-emitter

Emitters at the CZE column termini were prepared according to Bateman et al.[Bateman et al., 1997]. Briefly, fused silica (fs) capillaries (50 µm ID x 350 µm OD x 85 cm length) were tapered at the back end by heating the capillary 2-3 cm from the end in a flame with an appropriate temperature of a microwelding torch meanwhile gently pulling manually until the capillary enlarged and separates. The resulting long tapered tip was than inserted into a 50 or 75 µm ID capillary and bent until the tapered end snapped. Next, tips were etched in 40% hydrofluoric acid for 15 min to reduce the outer diameter. During etching, distilled water was pumped through the capillary using a syringe pump at a flow rate of 1 ul/min to prevent the hydrofluoric acid entering the tip where it will etch the interior surface. After etching the capillary was rinsed with water, purged with helium and gold coated.

Disposable emitters

Disposable emitters were manufactored by pulling short pieces (L ~ 15 cm, 50 µm ID x 360 µm OD) of fused silica tubing (fs) in an appropriate pulling device (courtesy: Protana, Odense, Denmark) at appropriate temperature, pulling force and speed untill it separates. This results in two emitters with an orifice diameter of approximately 7 µm ID (~50 µm OD). After metalisation (see below), emitters were zero-dead-volume connected to the outlet end of the 50 µm ID LC or the CZE column by means of a 300 µm ID Teflon sleeve.

Metal coating procedure

Gold coating of emitters was performed in two consecutive steps:

A. Sputter coating. Tapered tips were placed in a sputtering chamber (SC7610 Sputter Coater, VG Microtech, Manchester UK), at an angle of approx. 45° off the vertical defined by the sputter target and the counter electrode. The chamber was sealed, pumped down to 20-30 Pa, and purged with argon. Gold vapour depositing was performed at 1.8 kV at a current of 18 mA for 75 s resulting in a layer of approx. 250 Å.

B. Electroplating. The gold layer was enlarged by means of electroplating. The capillary tip was electroplated by placing approx. 1 cm of the vapour deposited tip into a gold III solution in nitric acid (500 mg/L Au(III) in 10% nitric acid) at a voltage of 1.5 V applied to the tip from an house-hold battery for 5 min. A platinum wire was used as counter-electrode. To prevent blockage during plating, the capillary was continuously rinsed with de-ionised water (syringe pump, 1 µL/min).

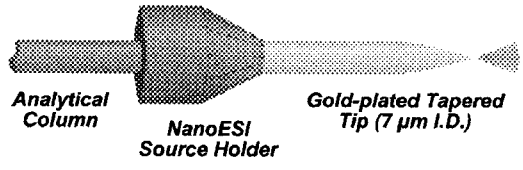

Fig. 1. Outline of the nanoLC sheatless ESI interface

NanoLC column switching system

The nano-LC switching system was very similar to the µ-column switching system published previously [Van der Heeft et al, 1998], but with a slightly modified switching unit (see fig. 2). The 6-port switching valve opens and closes the waste side of a low dead volume (29 nL) T. During loading, the open waste side of the T allows the relatively high flow delivered by the isocratic pump (3 µL/min; water + 0.075% TFA) to pass the loop and the pre-column (PC) to waste. Next, typically after 6 min, the valve is switched resulting in the closure of the waste-line and the connection of the PC in line with gradient pump and analytical column, resp. The gradient elutes the sorbed peptides from the precolumn into the analytical column and are subsequently gradient eluted into the mass spectrometer.

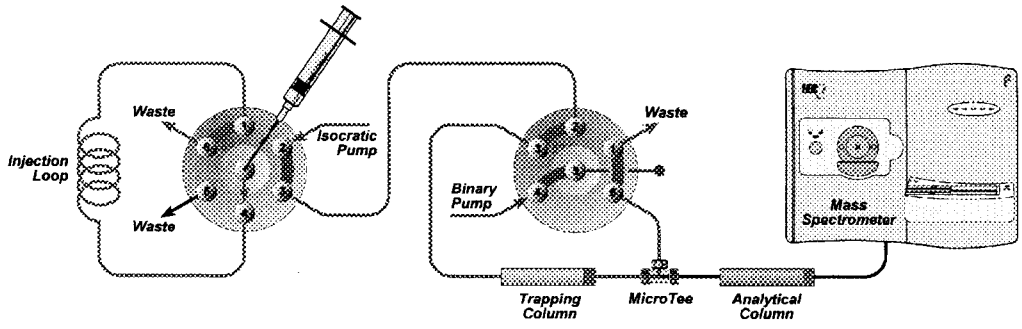

Fig. 2 Nano-LC column switching system comprising of a pre-column (PC) or trapping column and an analytical column (AC). The AC is connected to the PC by a low dead volume T (29 nL). During loading of the sample from the injection loop (10 uL) the flow of the isocratic pump transports the contents of the loop over the PC where the peptides are retained by the HPLC particles and passes by the open T to waste. Switching the 6 port valve results in the closure of the T which directs the flow of the gradient pump accross the PC into the AC.

2.4 Preconcentration -capillary electrophoresis-mass spectrometry

Preconcentration capillaries were manufactured by placing 1 mm length RP-C18 bed packed into a 50 µm x 365 µm x 25 cm (ID x OD x length) fused silica tubing containing a retaining frit at the outlet end (Fig. 3). The C18 particles in 2-propanol were propelled into the capillary from a slurry placed in a pressurised ($5x10^5$ Pa) vessel. After the desired packing length (1 mm) was obtained, a terminating piece of SBD membrane was inserted into the capillary and pushed up to the bed with water at a pressure of 10^6 Pa. Before use, the pre-column was successively rinsed with methanol (5 min), water (2 min) and running buffer (10 min). The PC was connected with the CE column by means of a 300 ID x 1 cm Teflon sleeve.

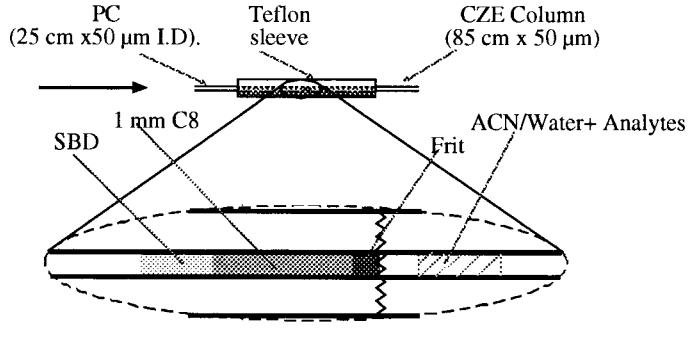

Fig. 3 Lay-out of pre-concentration/capillary zone electrophoresis interface

Samples were injected using the following sequence of events, all performed with pressure: (1) injection of sample at a flow rate of 1 μL/min at p = $2x10^5$ Pa; (2) rinsing of the column with the background buffer at p = $2x10^5$ Pa for 180 s; (3) introduction of 20 nL elution mixture, acetonitrile/water/acetic acid (50/49.9/0.1) at p =$2.5x10^4$ Pa for 10 s; (4) pushing-up of elution mixture with background buffer so that it passed precisely the trap area (≥26 cm displacement in sampling capillary corresponding to ≥ 510 nL (50 μm ID) or ≥ 1115 nL (75 μm ID) at p = 10^5Pa for approx. 65 s and 135 s, resp.; (5) followed by switching on the CE voltage (30 kV).
In case of transient-isotachophoresis (tITP), step (3) is preceded by the introduction of 5 nL of LSB at p=10^4 Pa for 6 s.
The running buffer was a mixture of 2 mM ammonium acetate the pH adjusted to 2.9 with acetic acid (1 vol%).

2.5 Protein and polynucleotide database searching
Protein databases were searched by two protocols through the Internet using:
(i) peptide sequences (Edman -type data: AA-AA-AA-AA): http://www.man-heidelberg/services/peptidesearch/FR_sequenceonlyform.html
(ii) sequence tags [(mz1) AA-AA-AA (mz2)]: http://www./man-heidelberg/services/peptidesearch/FR_peptidepatternform.html

Expressed Sequence Tags (EST) were searched using a locally running program PeptideSearch [Mann & Wilm, 1994]. The EST database was downloaded from the Internet (National Centre for Biotechnology Information)
Results of EST searches, i.e. those amino acid sequences that correspond to certain pieces of cDNA sequences in a particular reading frame, were used in homology searches using the search algorithm BLAST. The corresponding proteins were identified by the search program ENTREZ (URLs:www.ncbi.nlm.nih.gov/BLAST/ and www.ncbi.nlm.nih.gov/ENTREZ/)

2.6 Mass dependent scanning
was developed using a specific feature in the instrument acquisition software. This software selects the most abundant peptide ion from each primary MS scan (MS-1 domain). Next, this ion is selectively isolated by the ion trap and subjected to collision induced dissociation (CAD) conditions followed by mass analysis (MS/MS) of the resulting product ions. A threshold value was set above which peptide ion selection takes place (5 x 10^5 counts). This strategy allows the user to perform unattended MS/MS analysis of peptides eluting from a LC or a CZE column. In addition, the summed MS spectrum of all primary MS scans during the entire LC elution time course provides information on the molecular weight of the different peptides (peptide mapping) present in the mixture analysed. The complexity of the mixture determines which fraction of all peptides present will be sequence analysed, the more complex the mixture the lower this fraction will be (due to more simultaneous eluting peptides).

3. RESULTS AND DISCUSSION

Proteins and peptides are key elements in the action and regulation of biological systems. They are the effectorates of many biological processes and play central roles in the structural and metabolic organization of living cells. The proteins are the most polymorphic macromolecules and represent quantitatively the main biochemical material in tissues where they account for more than two-third of the organic substance.
The primary structure of a protein or peptide includes the amino acid sequence plus any post-translational modifications (*i.e.* alterations to specific amino acid side chains, addition of carbohydrates, formation of disulfide bonds or other cross-links). The distinction between peptides and proteins is based on size: peptides generally are below 10 kDa in mass, proteins are larger.

Many of the recent advantages in microbiology and biotechnology can be attributed to the recognition and identification of their primary structure (*i.e.* determination of the amino acid sequences of peptides and proteins) of these key elements. Hence, protein and peptide analysis of biological materials still is one of the most important assays in clinical and biomedical research, not only to elucidate its biological activity, but also as a part of molecular pathology.

3.1 Protein Identification following digestion and MS analysis.

Although molecular weight analysis of large proteins can be done by MALDI-ToF on the intact molecule, identification of proteins, i.e., to establish their amino acid sequences, requires shorter pieces of the protein. Common protocols therefore include enzymatic or chemical digestion (see table 1) of the protein followed by the analysis of the resulting fragments.
A most convenient way of characterisation/identification of unknown proteins is to analyse the molecular masses of all digestion peptides, referred to as peptide mapping, followed by a search of the matching protein. For this purpose, programs are available to search sequences in large protein databases (o.a SwissProt). This approach has been adapted by many protein chemistry laboratories, mainly using the most sensitive technique MS technology known to date, MALDI-ToF [Wilkins et al, 1997]. This technology, however, is not available in our laboratories. In stead we use a 'second best' technology, being electrospray ionisation mass spectrometry, using a relatively small and cheap but otherwise powerful ion trap instrument. Over the past year, our R&D program aimed to improve the sensitivity of protein and peptide identification analysis to the sub-picomolar level including
(i) optimization of protein digestion; (ii) development of nanoscale electrospray techniques coupled to nanoscale liquid chromatography; (iii) coupling of ESI/MS to preconcentration capillary zone electrophoresis; and (iv) implementation of protein and DNA data base searching.

Table 1: Specific cleavage agents for polypeptides and their cleavage sites.

Reagent	*Cleavage site*
Chemical cleavage	
Cyanogen bromide	Carboxyl side of methionine residues
Phenyl isothiocyanate (Edman)	Uncharged terminal amino group of the peptide
Hydroxylamine	Asparagine-glycine bonds
2-Nitro-5-thiocyanobenzoate	Amino side of cysteine residues
Enzymatic cleavage	
Trypsin	Carboxyl side of lysine and arginine residues
Chymotrypsin	Carboxyl side of tyrosine, phenylalanine and tryptophan
Endoproteinase Asp-N	Amino side of aspartic acid and cysteic acid
Endoproteinase Lys-C	Amide, ester and peptide bonds at the carboxylic side of lysine

Digestion, either enzymatically (e.g., trypsin) or chemically (i.e., cyanobromide) of the protein is a first step in the identification/characterisation of proteins. In general, proteins are too big to break them down in bits by means of collisions with relatively small molecules (air, helium, argon or heavier gaseous molecules in the collision cell) inside the mass spectrometer. The energy that can be added to multiple charged molecules by means of collisions ranges typically between some tens to hundreds eV. Larger large molecules are capable of distributing this energy over numerous chemical bonds so lowering the average energy increment per individual bonding to a significant extent. Experiences indicate an upper MW limit for dissociation of polypeptides to about 3-5,000 Da, but smaller peptides may be difficult to break down, depending on their amino acid composition. Particularly, the presence of one or a few arginyl (Arg, R) residues in combination with glutaric acid (Glu, E) or aspartic acid (Asp, D) residues in the peptide increases the required energy significantly [Summerfield et al., 1997].
Digestion rather than the analysis form the critical step in the identification process. Sensitivities of advanced MS technologies are now well below the fmol range quantities, but the digestion of proteins and/or the recovery of the peptide fragments formed remains difficult to perform below the pmol (nanogram-microgram) range. In addition, contaminating proteins, e.g., occurring from the digestion itself (autolysis) or from sample manipulation (keratin, from dust, human skin), may readily mask the tryptic peptides if really small protein amounts have been processed.
Analysis of the tryptic peptides was performed by μLC/MS using mass dependent collision induced dissociation (CID) tandem mass spectrometry. The data obtained comprise the so-called peptide map (total ion current LC-profile, fig 4 A), m/z values of individual peptide ions in the digestion mixture (summed primary MS spectra, fig 4B) and a number of CAD MS/MS spectra. In this example, approx. 40 CAD spectra were recorded from a single LC run using mass dependent scanning.

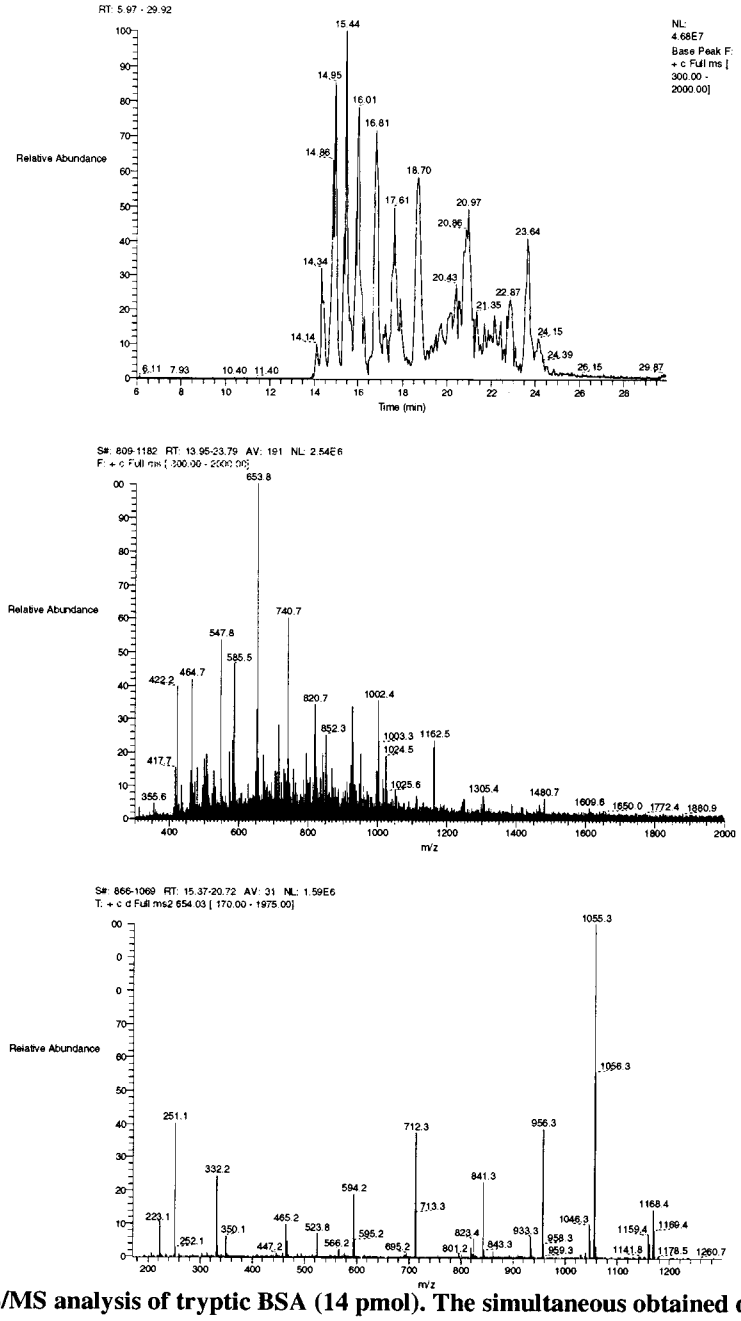

Fig. 4. LC/MS/MS analysis of tryptic BSA (14 pmol). The simultaneous obtained data by mass dependent scanning comprise the so-called peptide map (LC trace, top), masses of individual tryptic peptides (occurring as singly and multiply charged ions, middle) and product ion or MS/MS spectra of collision activation dissociated (CAD) peptides (bottom). The CAD spectrum of m/z 653.8 (MW 1305.7) contains the sequence tag (712.3) EDV[IL] (1168.5) corresponding with the partial sequence HLVDEPQNLIK of BSA 402-412 eluting at 15.44 min

An example of the characterisation of a vaccine product was carried out for *N. Meningitis* class I outer membrane protein. The overall spectrum, recorded from direct nanoESI analysis of the tryptic peptides is shown in Fig. 5. Fig A represents the peptide ions present in the digestion mixture (singly, multiply charged peptides) and Fig. B a CAD product ion spectrum of the 21-mer peptide $[M+2H]^{2+}$ at m/z 1075. The y-ion series, i.e. fragments resulting from cleavages across the amino acid backbone

incorporating the C-terminus, allows the deduction of the partial sequence [KQ][IL]TEA[KQ]AADGGASG. [KQ] and [LI] denotes the isobaric K or Q and the isomeric L or I at this position, resp. A closer view (fig. 5C) into the spectrum reveals the presence of an alteration in the F91 Hidden 74 protein. The standard P 1.7.16 protein contains a N (residue mass 114 Da) whereas the present variant possess a D (residue mass 115 Da, de-amination) at position 48.

3.2 Nano-scale LC/ESI/MS

Following the in-house developed micro-LC system (Van der Heeft et al., 1998), efforts were made to reduce the interior diameter of the LC column by a factor two to 50 μm. Benefits of such small interior diameter columns in conjunction with a large volume injection system include (i) a further enhancement of the sensitivity – theoretically by a factor of four– and (ii) the use of sheathless ESI. Collectively, these refinements would give a gain in sensitivity of more than one order of magnitude, providing μL-size injection volumes can be made. The column is interfaced to the nanoESI source assembly using the same emitter as was used with CZE/MS, a gold coated tapered tip with 7-10 μm orifice (see fig. 1). Technical details are given in experimental section and fig.2.

Fig 6 shows the sensitivity of the system as well as performance characteristics of the system. Injection of 1 fmol of standard peptides gave a signal to noise ratio for the protonated molecular ions of > 100. The width at half height of eluting peaks were less than 7 s. Compared to the micro-LC system, the sensitivity increase amounts a factor of 25 to 50. This advantage was converted in the use of smaller sample volumes by a factor of 4 (injection volume10 in stead of 40 uL), which still resulted in a better sensitivity by a factor of 5 – 10. This feature allows the identification of proteins at the low fmol level or low nanogram amounts on gel. Fig. 7 shows the analysis of 1 fmol of tryptic beta-lactoglobulin.

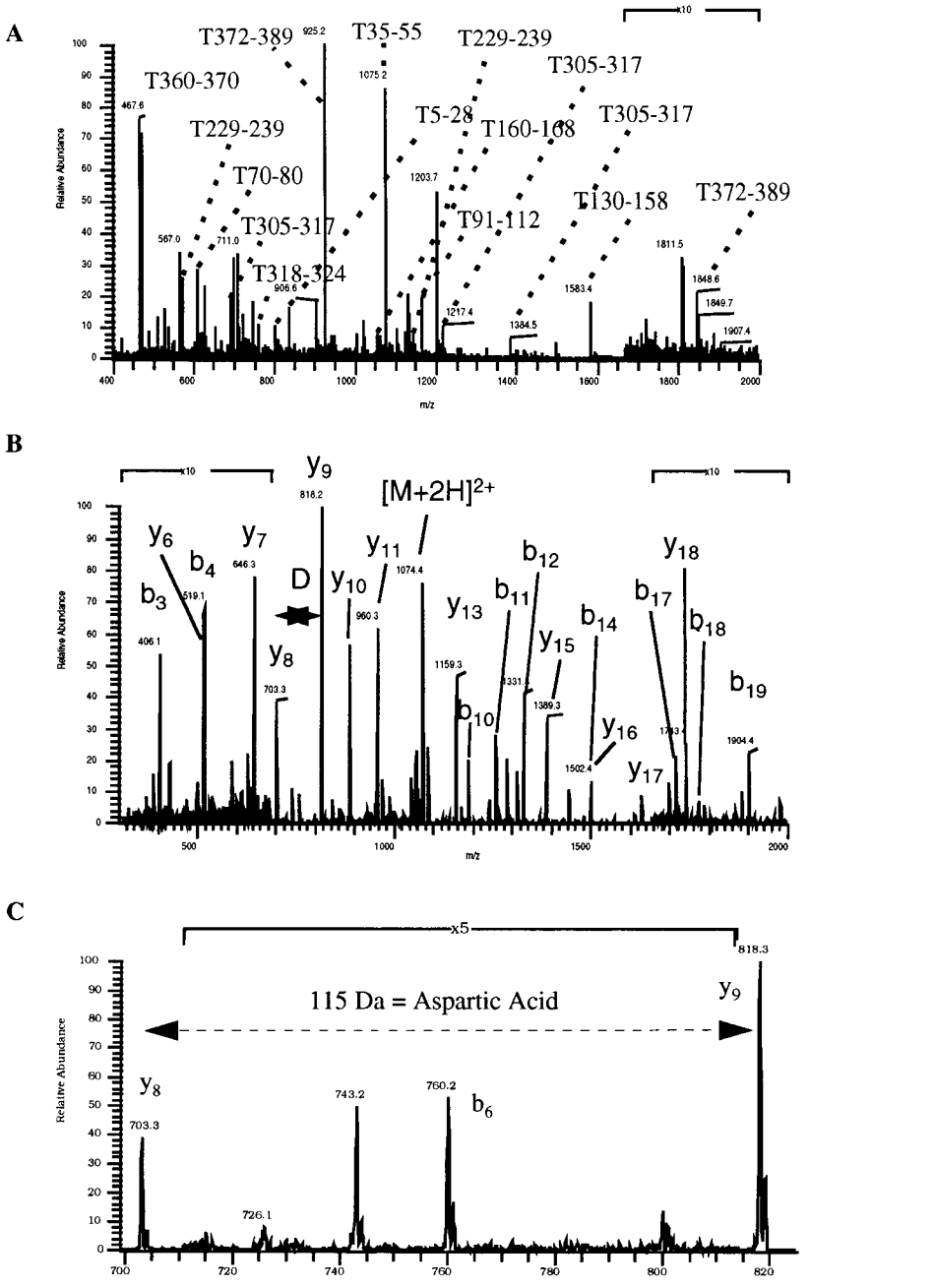

Fig. 5. NanoESI analysis of the tryptic peptides of N. Meningitis class I outer membrane protein. (A): Averaged MS spectrum, showing singly and multiply protonated peptide ions. (B): CAD product ion spectrum (MS/MS) of $[M+2H]^{2+}$ 1075.2 (MW 2148 Da), corresponding with the sequence 35-56 of M. Meningitis por A protein (NYQLQLTEAQAADGGASGQVK), in which a N-D conversion was observed compared to the subtype P1.7.16 gene. (C): exploded view of the MS/MS spectrum B, showing an Asp (D) residue at position 48 in the outer membrane loop 1 part of the protein.

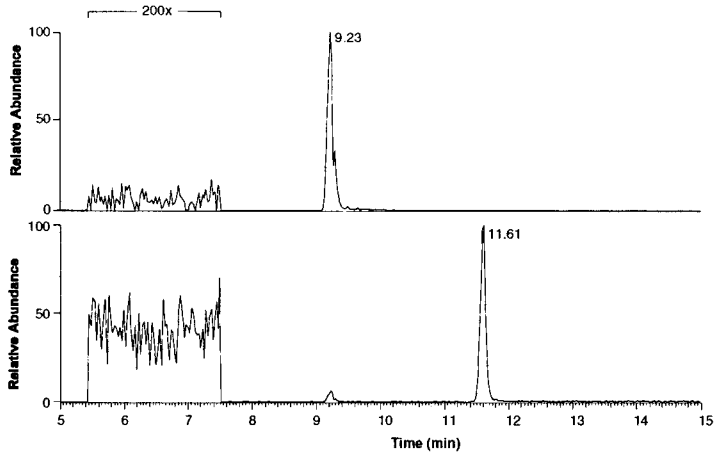

Fig. 6. NanoLC-MS analysis of 1 fmol each of standard peptides angiotensin III (upper) and oxytocin (lower). Signal-to-noise ratios for the peptide ions are> 200.

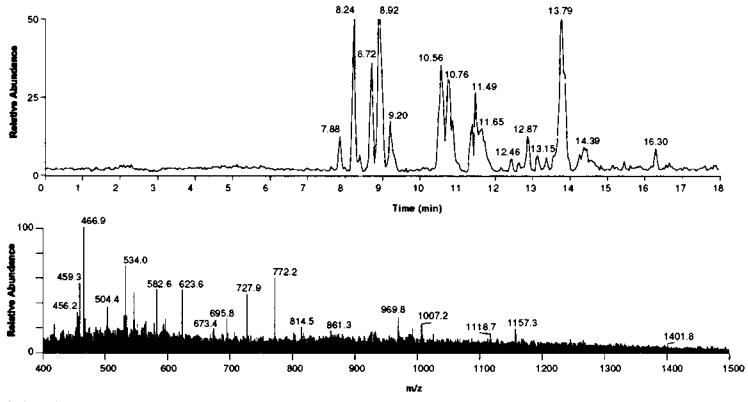

Fig. 7. Nano-LC/MS analysis of 1 fmol tryptic beta-lactoglobulin. Top, LC/MS base peak trace, bottom summed MS spectra over elution range 7-17 min showing singly and multiply charged tryptic peptide ions.

3.3 Preconcentration-CZE-ESI coupled to mass spectrometry

Capillary zone electrophoresis (CZE) is emerging an attractive alternative to liquid chromatography (LC) for the analysis of biological derived samples with mass spectrometric detection [Landers, J.P]. CZE offers better resolution, relatively short analysis times and complementary separation mechanism to LC. However, unlike for micro and nano-scale LC, coupling of CZE to electrospray ionisation mass spectrometry (ESI/MS) has been a serious obstacle for many years. Much of this was due to impairment of the CZE performance by the MS interface, like zone broadening and or dilution of the capillary effluent [Wahl, J.H].

In the recent past, CZE has been coupled to continuos flow fast-atom bombardment (CF-FAB) [Reinhoud et al., 1989, Minard et al, 1989] and ion spray (ISP) MS [Smith

et al, 1988] utilising liquid junction and co-axial type interfaces, respectively. The purpose of these interfaces was to create a means to apply the CZE terminating potential and to add the FAB matrix liquid or make-up solvent for the high flow ISP source. Recent refinements in CZE-MS interfacing started with the development and use of low-flow or nano-scale ESI sources [Wilm and Mann 1996]. Unlike standard ESI sources, this source does not require a make-up or sheath liquid for a stable operation. Olivares et al. [1987] have designed for the first time so-called sheathless CZE emitter utilising a blunt-end metal coated capillary. Later Chowdhury and Chait [Chowdhury and Chait, 1991] refined this technology by the construction of a sharply tapered tip at the CZE capillary terminus allowing to electrospray aqueous solutions at lower potentials, thus avoiding the onset of corona discharges.

Alternative designs include the 'sheath-less pinhole' approaches by Wahl et al.and Cao et al., utilizing a microhole near the capillary outlet, where the CZE terminating voltage was applied through a conductive gold epoxy composite or a tiny platinum wire, respectively. Others have made use of a short piece of polysulphone micro-dialysis tubing [Severs et al., 1996] or a porous glass joint [Settlage et al., 1998]. All the above mentioned designs were aimed at both minimising extra-column dead volumes and eliminating solvent supplements to the CZE outlet to preserve separation efficiency and analyte band concentrations of CZE separated analytes, respectively.

Another limitation of CZE/MS in its application to dilute biological samples is the limited concentration sensitivity of the technique. Different to LC, the volumetric loading capacity of the CZE column is limited to about 5% of the total capillary volume, i.e. nL ranges for common CZE column dimensions ($L \leq 1$ m x 50-75 µm I.D.). Stacking techniques may enlarge the injection volume up to 50-90% of the column volume (typically 2 µL for 1 m x 50 µm I.D. column), but even this volume may represent only a small proportion of the total sample volume. For this purpose, Tomlinson [Tomlinson et al., 1997] and others [Guzman et al., 1991 Strausbauch et al., 1995, Beattie et al., 1995] have developed an on-line pre-concentration technique. Pre-concentration was achieved by means of a sorption medium (hydrophobic polymeric membrane or reversed phase (RP) LC particles) placed in front of the CZE column where analytes with appropriate affinity to the sorption medium were retained from relatively large aqueous sample volumes. It was claimed that the PC interface would not affect CZE performance when CZE was preceded by transient-isotachophoresis (tITP) stacking of the eluted analytes from the PC into the CZE column.

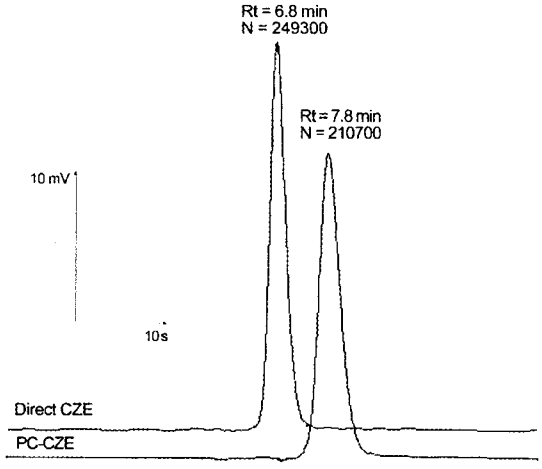

Fig. 8. Comparison of column separation efficiency obtained by direct CZE and preconcentration CZE.

Fig 8 shows the comparison of direct CZE and PC-CZE on the model peptide oxytocin. It follows, that the separation efficiency obtained by direct CZE (N = 249300) and by PC-CZE (N= 210700, 85%) were in good agreement. However, such relatively high plate numbers were attainable only if the column has been conditioned between each run. The observed small loss of separation efficiency in PC-CZE (-15%) has been attributed to the somewhat diminished EOF (-13%) in the PC-CZE system. The reason for a lower EOF in conjunction with the PC remains unclear yet. Strausbach et al. [Strausbauch] have ascribed this phenomenon to a reversal of the EOF at the C18 particles (locally, here over a length of <2 mm).

CZE-ESI Interface.

The emitter at the capillary terminus is one of the most critical parts in interfacing CZE to mass spectrometry. It should maintain the electrical contact across the CZE buffer and produce a fine aerosol mist from aqueous CZE buffers at a moderate elelctrospray voltage, well below the corona onset voltage. In addition, it should be rugged and durable and not introduce any dead volume, because this will dramatically detiorade the CZE separation efficiency. Lack of durability is frequently cited in the literature as a major drawback in many designs [Valaskovic et al, 1995].

Silver coatings appear to suffer from poor chemical stability due to electrochemical oxidation of the silver [Smith]. In addition, silver/water clusters are often observed in the mass spectra [McGee, 1982].

Silanisation of the glass surface is known to improve the adhesion of metals to various surfaces including glass [McGee, 1982; Wasserman et al., 1989; Kriger et al 1995; Goss et al ,1991], but emitters prepared by such procedures remained susceptible to damage even by minor electrical discharges.

Significantly better results were obtained using emitters prepared according to Kelly et al. [1997]. The sharply tapered, gold metalised tips produced stable, durable ESI sprays at moderate voltage of typically 1.4 - 2.2 kV. Contrary to previous reports, we observed that butt-connected (disposable) emitters caused slight tailing on all peaks

and caused bubble formation during the CZE run. Therefore, the emitter was constructed directly onto the CZE column terminus.

The size of the emitter orifice (range: 5-30 μm) was studied in terms of the operating electro-osmotic flow (EOF), the quality and stability of the spray, the sensitivity and the ease of manufacturing and use. The EOF is determined by the capillary diameter, the CZE field strength (V/m) and the buffer pH and composition. The 5 μm orifice produced a good quality spray at relatively low ESI voltage (typically 700-1100 V) but could handle only small EOFs of typically <50 nL/min. If higher EOFs were used, the excess flow caused a droplet at the emitter, which increases with time until it discharged off. In addition, blocking may readily occur, particularly when analysing unfiltered real life samples. The high range orifice diameter (30 μm) required a larger flow than is normally produced by 50 μm capillaries operated at 30 kV/m and a buffer pH of 2.9. Optimal results were obtained for the mid-range emitters measuring around 15 μm ID orifice, which produced stable and low background electrospray for common EOFs (100-200 nL/min) at moderate ESI potential (1.4-2.2 kV). The critical balance between emitter orifice size and EOF requires maintaining between-run variations in the EOF within relatively small ranges. Variation of the EOF may occur with use due to adsorption of analytes to the capillary inner surface, so reducing the density of the free silanol groups with time. Frequent, e.g. daily, reconditioning by flushing the capillary with NaOH is recommended, rather than to apply pressure at the column head, in cases the EOF has been diminished below a minimum level required for a steady ESI cone (e.g. 150 nL/min). Pressure driven flow has a parabolic flow profile which causes severe peak zone broadening resulting in deterioration of the electrophoretic performance.

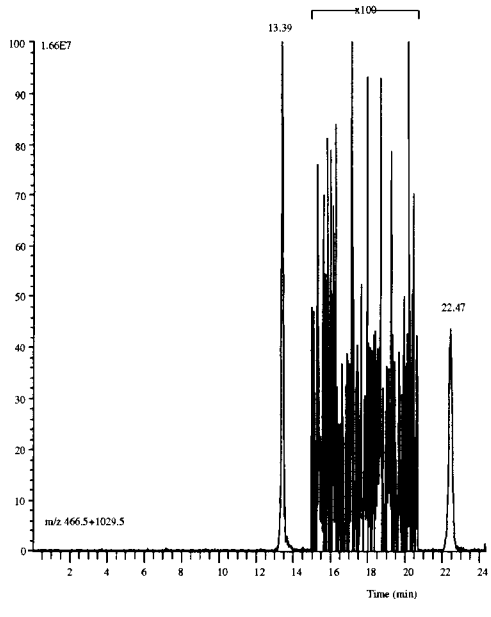

Fig. 8. PC-CZE analysis of 1 fmol each of standard peptides angiotensin and oxytocin.

The sensitivity of the system was determined from the analysis of 1 fmol each of two standard peptides (Fig. 8). Both peptides were detected at a signal-to-noise ratio of

≥100:1. These data indicate a lower limit of detection (LOD) for peptides of about 25 attomole.

Application to real samples

The PC-CZE system has been applied to the analysis of tryptic peptides and MHC class I bound peptides. Five µL or 500 fmol tryptic bovine serum albumin (BSA) digest were analysed (Fig. 9). The total ion electropherogram (Fig 9A) shows approx. 40 resolved peaks above a threshold level of 0.5% relative abundance. The sensitivity of the sequence analysis of individual tryptic peptides is demonstrated in Fig. 9C. The CAD MS/MS spectrum was obtained from a minor tryptic peptide at m/z 572 (Fig. 9B, retention time 17.88 min, relative abundance 6%). The product ion spectrum contained prominent fragment ions corresponding with the partial sequence KQTALVELLK (BSA 236-245).

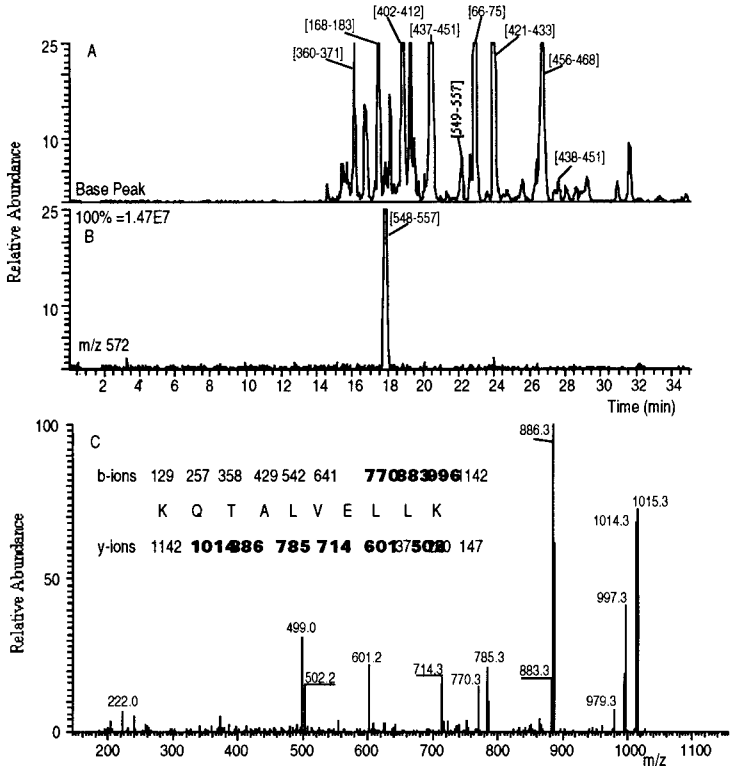

Fig. 9. PC-CZE/MS analysis of tryptic BSA peptides. The sample amount was 500 fmol on column. The Y-axis in A and B was enlarged 4 times.

A second example of the application of the PC-CZE/ESI/MS system to the analysis of biologically relevant peptides is shown in Fig. 10. MHC molecules are known to bind an extremely large array of self and foreign (antigenic) peptides resulting in extremely complex elution mixtures known to contain more than 5,000 constituents [Hunt et al, 1992]. Currently, microscale LC coupled to mass spectrometry is considered as the method of choice for analysis of antigenic peptides [Engelhard, 1994, Van der Heeft et al, 1998]. Additional separation efficiency would be very useful in those analysis, particularly in cases where peptides of interest, i.e., disease specific epitopes, co-elute with several other peptides in the mixture [Den Haan et al., 1995].

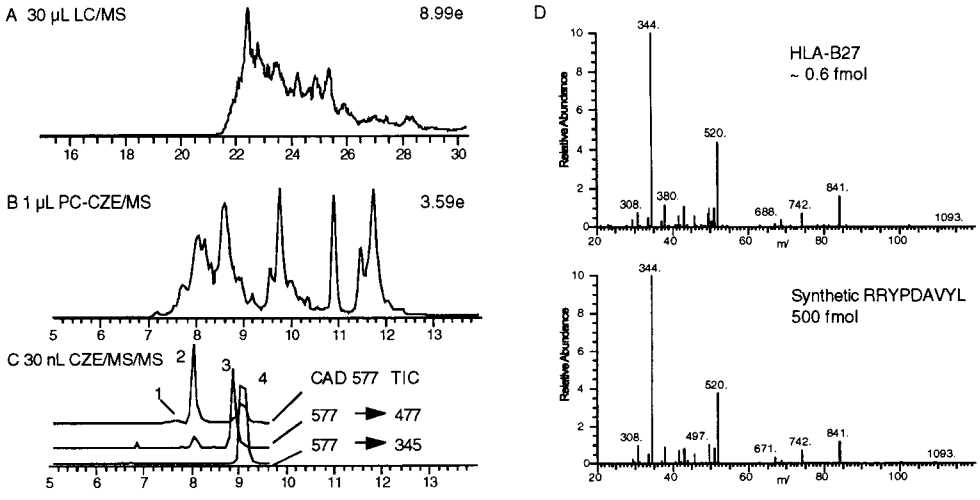

Fig. 10. Comparison of uLC and CZE analysis of peptides associated with HLA-B2705 molecules. For details see text.

The application represents the analysis of a measles virus-specific antigen, RRYPDAVYL, associated with HLA-B*2705 (human lymphocyte antigen type B2705) molecules. From prior LC/MS analysis, it was found that at least 3 other peptides were present in the mixture with the same molecular mass (MW = 1152), which eluted in 2 peaks from the µLC column. Consequently, production spectra contained significant interferences. With CZE, these four peptides were baseline separated (Fig. 10C) and the CAD MS/MS spectrum of the last eluting MW 1152 peptide compared well with that of the standard peptide (fig 10D). The sensitivity in this analysis was derived from the estimated complex number of 2425 per individual cell. The peptide amount consumed in the analysis in Fig 10C and D is estimated on 0.6 fmol ($156x10^3$ x 2425 = $0.6x10^{-15}$ mole). Comparison of the data obtained by µLC/MS and CZE/MS indicate a enhancement of the sensitivity in CZE/MS by about two orders of magnitude, or 10-100 attomoles. The sensitivity of the µLC/MS system generally is between 1 and 10 fmol consumed [Van der Heeft et al, 1998].

3.4 Protein and DNA database searching

MS and MS/MS data of proteolytic fragments have proven to be very useful in the identification of unknown proteins. The first domain MS data provides information on the molecular masses (MW's) of the peptides formed during enzymatic hydrolysis . In addition, the second domain or MS/MS data provide information on (short) amino acid sequences in the various parts of the protein.

Protein data bases can be searched with different features:

(1) on the basis of MWs of the tryptic peptides observed in the digest (peptide map);

(2) on the basis of a matching partial sequence in one or more of the tryptic peptides (Edman-type data);

(3) on the basis of matching sequence tags (see Fig. 11).

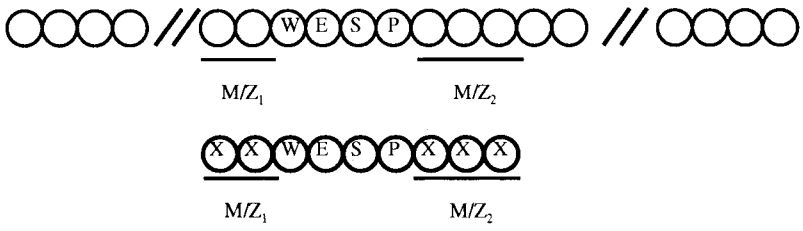

Fig. 11. Example of protein data base searching using the so-called sequence tag. All proteins in the database are searched for the following characteristics:
(i) the possibility to form a tryptic peptide of mass M from the protein under search, that (ii) contains the sequence WESP and (iii) has N and C terminal parts with combined masses of m/z1 and m/z2, respectively. If all three requirements have been fulfilled, the protein is selected as a candidate. Note: similarity in mass of the N and C terminal parts does not necessarily mean that those parts contain the same amino acids and sequence, but this criterion excludes proteins that does not posses this similarity.

In addition, the same searches can be performed in expressed sequence tags (EST's), databases. ESTs are established smaller or larger nucleotide sequences (cDNA pieces) that encode for certain proteins in a particular reading frame.
Protocol (1) represents the simplest search of all that does not require any amino acid sequence information. The disadvantage, however, is that it is less specific compared to searches using amino acid sequences in the derived peptides. In addition, the use of sequence tags (combined mass and sequence information) is much more specific than partial sequence information alone. The requirements to fulfil, i.e. the MW of the peptide; the combined mass of the otherwise unidentified N and C terminal residues and the correct sequence somewhere in the peptide, strongly reduces the number of proteins that will comply. This is illustrated in Table 2.

Table 2. Comparison of results of protein database searches using sequence only information (Edman-type data) and sequence tags. Searches were carried out for a partial sequence of the peptide KLWESPQEI matching sequence 84-92 in *measles virus non-structural protein* (SwissProt P35977). Note: this protein occurs with 8 entries in the non-redundant database.

	Partial sequence	Sequence tag	Number of matching proteins
1	–WES———	-	>500
2	–WESP——	-	203 (195+8)
3	–WESPQ——	-	31 (23+8)
4	–WESPQE—	-	9 (1+8)
5	-	(242.2)WES(644.3); MW 1128.6	8

An unknown HLA-A*0201 bound self-peptide with MW 944.5±0.2 Da was partially sequenced as xxx-xxx-xxx-xxx-xxx-Asp-Val-Val-Ser-xx. The MS/MS spectrum

indicates 5 unknown N-terminal residues with a combined mass of 428.2 Da. This information is translated into the sequence tag

(428.2) DVVS (828.4).

The C terminal residue can be assigned to a Val from the mass difference (117 Da) between the peptide MH+ at m/z 945.5 and the combined mass of the first 8 residues (m/z 828.4). The dbEST search of this sequence tag yielded 5 DNA sequences with accession numbers (gi) 1068794, 1283230, 1289039, 1317696, 2272965 and 2916698.

A homology search (BLAST) using the nucleotide sequence gi/1068794 comprising 328 bases, yielded 35 proteins with different number of aligning DNA sequences. The largest alignment (208 bases) was obtained for the human gene coding for the glucocorticoid-induced leucine zipper protein. Much lower numbers of aligning bases (<40) were obtained for genes encoding for o.a. heat shock proteins in the various species. The leucine-zipper protein represents the following amino acid sequence:

MNTEMYQTPMEVAVYQLHNFSTSFFS**SLLGGDVVSV**KLDNSA
SGASVVALDNKIEQAMDLVKNHLMYAVREEVEVLKEQIRELLE
KNSQLERENTLLKTLASPEQLEKFQSRLSPEEPAPEAPETPETPEA
PGGSAV

The partial sequence 26-36 (indicated in bolt face), complies with the sequence tag used: (1) the molecular weight (MW) of the peptide amounts 944.5 Da; (2) the 5 N-terminal residues (SLLGG) have a combined mass of 427 Da and, (3) it contains the sequence DVVS in the right position and (4) possesses a C-terminal V residue.

Another example of the use of EST searching concerns the identification of proteins isolated from S. pneumonia. Several derived proteins were digested and sequenced. However in none of the cases any positive result was obtained in both protein and EST searches. Therefore, the pneumonia genome (www.tigr.org/data/s_pneunoniae) was downloaded and the so-called ORFs (open reading frames) were translated into amino acid sequences using all six reading frames. Using this dedicated database, it was possible to identify several S. pneumonia proteins or homologous proteins in related species (Lactobacillus paracasei, Bactillus subtilis).

4 CONCLUSIONS

Nano-scale separation techniques coupled to mass spectrometry provides sub fmol sensitivity in the sequencing of peptides. Biochemical methods rather than mass spectrometric analysis forms the sensitivity-limiting step in structure analysis of unknown peptides and proteins.

REFERENCES

1 Apella, E.; Padlan, E.A.; Hunt, D.F. , EXS 1995, 73, 105.
2 Barroso, M.B.; de Jong A.P.; J. Cap. Electrophoresis 1999, to appear.
3 Bateman, K.P.; White, R.L.; Thibault, P. Rapid Commun. Mass Spectrom. 1997, 11, 307.
4 Beattie, J.H.; Self, R.; Richards, M.P. Electrophoresis 1995, 16, 322.
5 Boguski,M.S.; Lowe, T.M.; Tolstoshev, C.M. Nat Genet 1993, 4, 332.
6 Cao, P.; Moini, M. J. Am. Soc. Mass Spectrom. 1997, 561.
7 Chowdhury, S.K.; Chait, B.T. Anal Chem 1991, 63, 1660.
8 Engelhard, V.H. Ann Rev Immunol 1994, 12, 181.
9 Goss, C.A.; Charych, D.H.; Majda, M. Anal Chem 1991, 63, 85.
10 Guzman, N.A.; Trebilcock, M.A.; Advis, J.P. J Liq Chromatogr 1991, 14, 997.
11 den Haan, J.M.; Sherman, N.E.; Blokland, E.; Huczko, E.; Koning, F.; Drijfhout, J.W.; Skipper, J.; Shabanowitz, J.; Hunt, D.F.; Engelhard, V.H. et al. Science 1995, 268, 1476.
12 Hillenkamp, F.; Karas, M.; Chait, B. Anal Chem 1991, 63, 1193A De Jong, A. Mass Spectrom Rev 1998, 17, 311.
13 van der Heeft, E.; ten Hove, G.J.; Herberts, C.A.; Meiring, H.D.; van Els, C.A.C.M.; de Jong, A.P.J.M. Anal. Chem. 1998, 70, 3751.
14 Hunt, D.F.; Henderson, R.A.; Shabanowitz, J.; Sakaguchi, K.; Michel, H.; Cox, A.L.; Apella, E.; Engelhard, V. Science 1992, 255, 1261.
15 de Jong, A. Mass Spectrom Rev, 1998, 17, 311.
16 Kelly, J.F.; Ramaley, L.; Thibault, P. Anal Chem 1997, 69, 51.
17 Kriger, M.S.; Cook, K.D.; Ramsey, R.S. Anal Chem 1995, 67, 385.
18 Landers, J.P. Handbook of Capillary Electrophoresis, CRC Press Inc., Boca Raton, Florida, 1997.
19 Mann, M.; Wilm, M. Anal Chem 1994, 66, 4390.
20 McGee, J.B. United States Patent. 1982,
21 Meiring, H.D.; ten Hove, G.J.; de Jong, A.P.J.M. 1997, RIVM Report 502501059.
22 Minard, R.D.; Luckenbill, D.; Curry, P.; Ewing, A.G. Adv Mass Spectrom, 1989, 11, 253.
23 Moseley, M.A.; Deterding, L.J.; Tooner K.B.; Jorgenson, W. Anal Chem 1991, 63, 1467.
24 Olivares, J.A.; Nguyen, N.T.; Yonker, C.R.; Smith, R.D. Anal Chem 1987, 59, 1230.
25 Reinhoud, N.J.; Niessen, W.M.A.; Tjaden, U.R.; Gramberg, L.G.; Verheij, E.R.; van der Greef, J. Rapid Commun Mass Spectrom, 1989, 3, 348.
26 Severs, J.L.; Harms, A.C.; Smith, R.D. Rapid Commun Mass Spectrom, 1996, 10, 1175.
27 Settlage, R.E.; Russo, P.S.; Shabanowitz, J.; Hunt, D.F. J Micro Separat 1998, 10, 281.
28 Smith, R.D.; Barinaga, C.J.; Udseth, H.R. Anal Chem, 1988, 60, 1948.
29 Smith, R.D.; Wahl, J.H.; Goodlett, D.R.; Hofstadler, S.A. Anal Chem 1993, 65, 547A.
30 Strausbauch, M.A.; Xu, S.Z.; Fergurson, J.E.; Nunez, M.; Machacek, D.; Lawson, G.M.; Wettstein, P.J.; Landers, J.P. J Chromatogr A 1995, 717, 279.

30 Strausbauch, M.A.; Landers, J.P.; Wettstein, P.W. Anal Chem 1996, 68, 306.
31 Summerfield, S.G.; Cox, A.K.; Gaskell, S.J. J Am Soc Mass Spectrom, 1977, 8, 25.
32 Tomlinson, A.J.; Guzman, N.A.; Naylor. S. J Cap Elec 1995, 2, 247.
33 Tomlinson, A.J.; Benson, L.M.; Jameson, S.; Johnson, D.H.; Naylor, S. J Am Soc Mass Spectrom 1997, 8, 15.
34 Valaskovic, G.A.; Kellerher, N.L.; Little, D.P.; Aaserud, D.J.; McLafferty, F.W. Anal Chem 1995, 67, 3802.
35 Valaskovic, G.A.; McLafferty F.W. J Am Soc Mass Spectrom 1996, 7, 1270.
36 Wahl, J.H.; Gale, D.C.; Smith R.D. J Chromatogr A 1994, 659, 217.
37 Wahl, J.H.; Smith, R.D. J Cap Electrophoresis 1994, 1, 62.
38 Wasserman, S.R.; Biebuyck, H.; Whitesides, G.M. J Mater Res 1989, 4, 886.
39 Wilkins, M.R.; Williams, K.L.; Appel, R.D.; Hochstrasser, D.F. (editors) *Proteome research: New frontiers in functional genomics.* Springer Verlag Berlin, 1997.
40 Wilm, M.; Mann, M. Anal Chem 1996, 68, 1.