MONOLITHIC LIQUID-CHROMATOGRAPHY COLUMNS FOR PROTEIN ANALYSIS

PROTEIN DIGEST SEPARATION AND INTEGRATED SYSTEMS

MONOLITHIC LIQUID-CHROMATOGRAPHY COLUMNS FOR PROTEIN ANALYSIS

PROTEIN DIGEST SEPARATION AND INTEGRATED SYSTEMS

MONOLITHISCHE VLOEISTOFCHROMATOGRAFIE-KOLOMMEN VOOR EIWITANALYSE

SCHEIDING VAN EIWIT-DIGESTEN EN GEÏNTEGREERDE SYSTEMEN

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht,
op gezag van de rector magnificus, prof. dr. J.C. Stoof,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op dinsdag 16 maart 2010 des middags te 2.30 uur

door

Michiel Hendrik Maarten van de Meent
geboren op 9 juli 1978
te Alphen aan den rijn

Promotor : Prof. Dr. G.J. de Jong

'No pressure, no diamonds'
Thomas Carlyle (Scottish Philosepher, 1795-1881)

Aan allen die mij dierbaar zijn

This project was funded by the Netherlands Proteomics Center (NPC):
research project 5.2 – LC systems based on monolithic columns for the separation of peptides and proteins.

The printing of this thesis was financially supported by:
ATAS GL International BV
Dionex Benelux BV

Cover design and lay-out: Roel van de Meent
The picture on the cover represents a scanning electron-micrograph of the cross-section of a poly(styrene-divinylbenzene) capillary monolithic column. Picture courtesy of Dionex Corporation.

Printed by www.proefschriftmaken.nl

ISBN: 978-90-8891-151-4

Chapter 1
GENERAL INTRODUCTION

Ch1

BACKGROUND

The successful sequencing of many genomes, including the human genome, has opened the doors for comprehensive exploration of its functional counterpart, the proteome. While the genome holds the blueprints for all parts of a living organism, the actual work is being done at the protein level. In contrast with the genome, which is constant under all conditions, the proteome is highly dynamic, protein expression levels vary among cell type, tissue and physiological conditions. Post-translational modifications, splice variants and the wide range of protein concentrations encountered also add to the challenge of proteome analysis. In-depth understanding of the proteome is important as it holds great promise for identification of new therapeutic targets and diagnostic and therapeutic biomarkers. Because of the aforementioned properties of the proteome, powerful analytical techniques are needed to create a complete picture of the proteome.

In recent years, several techniques have found their way into proteomics research. The application of powerful separation methods like two-dimensional gel electrophoresis (2D-GE) and soft ionisation techniques for mass spectrometry (MS) has provided the opportunity to identify large numbers of proteins in a single experiment. 2D-GE is now the method of choice for separating complex protein mixtures. Proteins are separated according to their isoelectric point and, subsequently according to size on an SDS-polyacrylamide gel. While 2D-GE offers outstanding resolution for separation of intact proteins, limitations like long separation times, small dynamic range and difficulty in analyzing certain protein classes (very small proteins, very large proteins, proteins with extreme pI, hydrophobic proteins) have fuelled the development of alternative methods. Today, GE-based methods are complemented more and more by liquid chromatography (LC), especially for peptide separation after digestion of protein mixtures. Some strategies combine (1D-) gel electrophoresis for protein separation with (multidimensional) chromatographic analysis of digests obtained from excised gel bands. Other methods completely skip separation at the protein level and apply (multidimensional) LC methods to the separation of complex digests.

Protein identification generally takes place by MS detection followed by data analysis. In top-down proteomics mass spectrometry is performed on intact proteins in order to obtain information on e.g. protein structure or (non-covalent) interactions with ligands or other proteins. Post-translational modifications and other forms of protein processing make identification of intact proteins by mass spectrometry highly challenging. In recent years, the introduction of high resolution mass spectrometry (Orbitrap and Fourier Transform-MS) has improved the quality of intact protein mass data and dissociation techniques for intact proteins have further aided identification. However, mass spectrometric analysis of protein digests is still the major tool for protein

identification as it provides unequivocal confirmation of protein identity as well as information on post-translational modifications. For purified proteins, e.g. proteins spots excised from 2D gels, mapping of peptide masses by MALDI- or ESI-MS is often sufficient for protein identification. For more complex samples, containing peptides from multiple proteins, peptide-sequencing is a better option. Peptides are fragmented in the mass spectrometer and the fragments are used to determine the amino acid sequence of the peptide. This can be done manually or automatically by submitting the MS/MS data files to (online) databases after which the obtained sequences are used to identify the protein(s) in question. One or two matched peptides can be sufficient for positive protein identification. Proteome analysis based on separation and mass spectrometry of digested protein mixtures is called bottom-up proteomics.

As an average protein will already yield some 25-30 fragments after tryptic digestion, even relatively simple protein mixtures will produce rather complex digests. The high resolution and ease of coupling to mass spectrometry means that (capillary) LC is the technology of choice for separation of protein digests. Due to the complexity of these samples, digest analysis requires highly efficient separation methods. In recent years, two new strategies have emerged to meet the analytical challenges posed by the separation of complex protein digests. Attempts to increase the resolution of LC methods have been made by employing relatively long columns or columns packed with very small diameter particles. Both these approaches, however, will lead to increases in column backpressure, creating the need for specialised LC equipment. An inherent disadvantage of the use of packed columns for the separation of peptides and proteins is the low diffusivity of macromolecules, restricting mass transfer in the pores of the stationary phase. In contrast to packed columns, monolithic columns have a bimodal pore structure with large throughpores and shallow mesopores. These unique properties result in high permeability (low column backpressure) and enhanced mass transfer (improved chromatographic efficiency). This makes monolithic columns ideally suited for the analysis of macromolecules. The application of monolithic columns for protein chromatography could provide a viable alternative for gel-electrophoresis based methods. LC-based methods are less time consuming, easier to automate and can be directly coupled to additional analysis steps like on-line proteolysis or detection methods like MS. Additionally, LC-based methods possess great versatility as a result of the possibility to employ and combine different types of columns, providing separation based on different physical-chemical properties of the proteins (charge, size, hydrophobicity, pI, etc.). The specific properties of monolithic columns also offer interesting opportunities for peptide chromatography. The low column backpressure resulting from the large throughpores allows the use of either high flow rates, applicable for fast and efficient separations or long columns at moderate pressure for analyses with high efficiency. Therefore, monolithic columns could offer a low-backpressure alternative to ultra performance LC (UPLC) systems using columns

packed with very small (< 2 μm) particles.

The objectives of the research described in this thesis are to evaluate the applicability of both silica-based and polymeric monolithic columns for protein analysis. The first part describes investigations into the effects of column length and stationary-phase chemistry on the separation of protein digests. The second part is devoted to the development of LC based integrated systems for the separation and identification of protein mixtures.

SCOPE AND OUTLINE OF THE THESIS

LC of protein digests

As current LC methods for peptide separation suffer from lack of resolution (one-dimensional methods) or very long analysis times (multidimensional systems), there is a need for new technologies that allow either increased resolution or higher throughput. The specific properties of monolithic columns allow the use of long columns with high separation efficiency and short columns with high flow rates for fast and efficient separations. Therefore, the separation of protein digests using silica-based and poly(styrene-divinylbenzene) (PS-DVB) monolithic columns has been studied using columns of different length. Some fundamental aspects of the chromatographic behaviour of protein digests have been investigated and the effect of column length and stationary phase on separation performance and MS-based identification were evaluated.

Chapter 2 presents an overview of novel chromatographic materials and columns that are applied for peptide separations in proteomics. Attention is paid to recent developments in reversed-phase column technology, especially UPLC and monolithic columns, as well as other modes of liquid chromatography. The potential of hydrophilic interaction chromatography (HILIC) is discussed as well as its potential in multidimensional separations. Some examples of the application of affinity chromatography in proteomics are given, with special focus on the analysis of phosphorylated proteins and the depletion of high-abundant species.

In **Chapter 3** a 750 x 0.2 mm ID silica monolithic column is presented. The separation of protein digests using this column is studied and the column is compared to a commercially available monolithic column of 150 mm length. A tryptic digest of bovine serum albumin (BSA) is used as a model sample to evaluate the effects of column length and gradient time on peak capacity and protein identification. Due to the application of a UV detector with fiber optics, UV and MS detection can be combined in a single system with minimal band broadening. UV chromatograms are used to determine peak capacity and MS/MS data for protein identification.

In **Chapter 4** 0.1 mm ID silica monoliths of different length are evaluated using a more complex protein digest and longer gradients. The sample is injected directly onto the column (without precolumn), thus focussing exclusively on the performance of the monolithic columns. Peak capacity and protein identification are evaluated using a digest of BSA, α-casein and β-casein. In order to further minimise extra-column effects, the UV detector is also omitted from the experimental set-up and peak capacity is determined from MS extracted ion chromatograms.

Chapter 5 describes the comparison of PS-DVB monolithic columns of 50 mm and 250 mm length for the separation of a 9-protein digest. Column performance was compared using peak capacity, peak production rate en protein identification. Some fundamental aspects of gradient chromatography of peptides are discussed and a prediction is made of the maximum obtainable peak capacities for both columns.

Integrated systems for protein separation and identification

The traditional proteomics approach of gel electrophoresis, in-solution proteolysis and digest analysis suffers from long analysis times, limited reproducibility and lack of automation. A system based on LC separation and on-line digestion could provide a promising alternative to traditional proteomics methods. Therefore an integrated system has been developed for the LC based separation, digestion and identification of protein mixtures. Proteins are separated using a polymeric monolithic column and trapped and (pre-) concentrated on a strong cation exchange (SCX) column. The trapped proteins are desorbed and digested, followed by LC-MS/MS analysis of the digest. First, a system for protein preconcentration, on-line digestion and LC-MS/MS of peptides has been developed. Mobile phases and flow rates are optimised for maximum efficiency and compatibility between different parts of the system. The possibilities of monolithic columns for separation of proteins are studied and the columns are used in the protein analysis system.

Chapter 6 describes the development of an integrated system for the preconcentration and identification of proteins. The system consists of an SCX precolumn for protein trapping and preconcentration, an immobilised trypsin reactor for digestion and an RPLC-MS system for digest analysis and protein identification. Several buffers are evaluated for optimal performance of both desorption of proteins from the SCX column as well as protein digestion by immobilised trypsin. The applicability of the system is demonstrated using some proteins differing in size and isoelectric point (pI).

In **Chapter 7** a system for total analysis of protein mixtures is described. In this system a chromatographic protein separation using a PS-DVB monolith is coupled with a small SCX column for trapping of proteins of interest from the column effluent. This SCX column is then placed in a second system where the trapped proteins are desorbed

using a high-pH buffer and washed to an RP trap via an immobilised trypsin reactor. The digest is analysed by RPLC-MS to identify the trapped protein. The applicability of this system is demonstrated by the identification of β-lactoglobulin A in a mixture of 9 proteins.

Chapter 8 presents some general conclusions and comments on the monolithic columns and systems described in this thesis. Recommendations are made for the applicability of monolithic columns in protein digest analysis and improvements are suggested to expand the possibilities of these columns in proteomics. Developments for the integrated system are also discussed, including possibilities for automation, use of alternative column types and applications.

Chapter 2
NOVEL LIQUID-CHROMATOGRAPHY COLUMNS IN PROTEOMICS

M.H.M. van de Meent and G.J. de Jong

in preparation

ABSTRACT

The enormous interest in proteomics research in recent years has inspired many developments in the field of peptide chromatography. A multitude of strategies has been developed to cope with the vast complexity of proteomics samples, trying to provide a sufficient degree of separation in order to be able to fully exploit the possibilities of mass spectrometric protein identification. As reversed-phase liquid chromatography (RPLC) is still the method of choice for coupling with mass spectrometry (MS), many efforts focus on the development of high efficiency reversed-phase methods like monolithic columns or ultra-high performance liquid chromatography (UPLC). As only RPLC-MS is unlikely to provide sufficient resolution to comprehensively cover highly complex samples, multidimensional methods will remain essential in proteome research. In this area, hydrophilic interaction chromatography (HILIC) seems to be a promising alternative to the traditional strong cation exchange (SCX) - RPLC methods for shotgun proteomics. Affinity based methods have also been successfully employed to provide access to specific sub-proteomes or low abundant plasma proteins. This review describes recent developments in chromatographic methods for proteomics research, focusing on advances in column technology and the application of novel column materials.

1. INTRODUCTION

Proteomics is the area of biomedical research that aims to characterise the full protein complement of a cell, tissue or organism at a certain point in time under given conditions. Any given proteome may comprise hundreds to tens of thousands of different proteins, spanning up to 10 orders of magnitude in concentration [1], thus highlighting the need for efficient analytical technologies to probe the proteome. The general analytical strategy in proteomics consists of separation, identification and interpretation.

Separation strategies for proteomics can generally be divided into two parts: separation on the protein level and separation on the peptide level. The method of choice for separation at the protein level is two-dimensional gel electrophoresis [2] (2D-GE), mostly due to its high resolving power. However, 2D-GE has a limited dynamic range and shows poor performance for very large, very small and very hydrophobic proteins. Additionally, automation, reproducibility and quantification also remain difficult. This has prompted researchers to explore alternatives, like combining 1D SDS-PAGE protein fractionation with enzymatic digestion and a liquid chromatography (LC) based peptide separation [3-5]. LC-based methods coupled to mass spectrometry (MS) are the most widely used methods for peptide analysis because of their high resolution, speed and ease of automation. In recent years the application of several novel reversed-phased columns to proteome analysis has been demonstrated. Furthermore, the availability of various different separation modes offers a high degree of flexibility while combining multiple (orthogonal) LC modes, like strong cation exchange (SCX) and reversed phase (RP), provides a strong increase in separation power [6].

Recently, novel column types like columns packed with relatively small (<2 μm) particles (ultra-high performance liquid chromatography – UPLC), monolithic columns, hydrophilic liquid chromatography (HILIC) and special affinity materials have been introduced in proteomics. In this review we will focus on the use of these materials for the analysis of protein digests. The suitability and potential of these columns in both one- and two-dimensional methods will be discussed and illustrated with relevant applications.

2. REVERSED-PHASE MATERIALS

2.1 Ultra-high performance liquid chromatography

The separation power of gradient LC systems is generally measured in peak capacity (*PC*). As *PC* is proportional to the square root of the column plate number (*N*), factors that improve *N* will also improve *PC*. *N* can be maximised either by increasing the column length (*L*) or by reducing the plate height (*H*), which can be achieved by reducing the particle diameter (d_p). Both approaches, however, cause an increase in column

Table 1: Overview of ultra-high performance LC systems for proteomics applications

Sample	Column	Packing	Analysis time	Remarks	Ref.
One-dimensional systems					
Ovalbumin digest	27 cm x 30 µm ID	1.0 µm nonporous silica	30 min	LIF detection, labelled sample	[8]
Various protein digests	22 cm x 150 µm ID	1.5 µm nonporous silica	30 min		[9]
Human Serum	10 cm x 2.1 mm ID	1.7 µm Acquity BEH	5 min		[10]
β-casein digest	5 cm x 2.1 mm ID	1.7 µm Acquity BEH	7.5 min		[11]
Protein kinase α digest	15 cm x 75 µm ID	1.7 µm Acquity BEH	30 min		[12]
5-protein digest	15-50 cm x 75 µm ID	1.7 and 3.0 µm C18	24-432 min		[13]
Shewanella oneidensis lysate	40 cm x 50 µm ID	1.4 µm porous silica	120-720 min		[15]
S. oneidensis lysate	20 cm x 50 µm	0.8 µm porous silica	8-50 min	Fast separations	[16]
Yeast lysate	87 cm x various ID	3 µm porous silica	180 min		[17]
Deinococcus radiodurans lysate	80 cm x 150 µm ID	3 µm porous silica			
3 µm nonporous silica	180 min		[18]		
Yeast lysate	150 cm x 50 µm ID	3 µm porous silica	200 min	Dual column system	[20]
Yeast lysate	80 cm x 150 µm ID	1.5, 2, 3 µm porous silica	180 min	Multi-column system	[21]
Yeast lysate	100 cm x 75 µm ID	3 µm porous silica	120 min	Dual column, SPE preconcentration	[22]
Yeast lysate, *D. radiodurans lysate, S. oneidensis lysate*	86 cm x 15-150 µm ID	3 µm porous silica	180-300 min	SPE preconcentration	[24]
Multidimensional systems					
Yeast lysate	5 cm x 100 µm ID	5 µm Partisphere SCX	7-27 h	Ultra-high-pressure MudPIT	[23]
	50 cm x 50 µm ID	3 µm Aqua C18			
Human plasma lysate	80 cm x 320 µm ID	3 µm polysulfoethyl aspartamide	About 80 h	Off-line system; 200 min SCX gradient, 15 fractions, 300 min RP gradient	[25]
	85 cm x 30 µm ID	3 µm porous silica			

backpressure which limits their application. This limitation can be overcome by the use of LC equipment designed to cope with these increased pressures. An overview of the use of ultra-high performance LC systems in proteomics is given in Table 1.

The use of capillary reversed phase columns packed with small C18-modified silica particles, operated at very high pressures, was pioneered by the group of Jorgenson in the late 1990s [7,8]. With a system operating at pressures in excess of 2500 bar, they obtained plate numbers of over 200,000 for small organic molecules using 30 µm id columns packed with 1.5 µm non-porous particles. The use of small internal diameter columns facilitates the efficient dissipation of the heat that is generated at high pressures [7]. The technique, termed ultra-high performance liquid chromatography (U(H)PLC), was also expanded to gradient chromatography. A peak capacity of 300 was obtained for a 30 min LC-UV analysis of an ovalbumin digest (fig.1) [8] and a UPLC-MS/MS system was used to identify an unknown protein from a 2D gel separation of rat liver lysate [9].

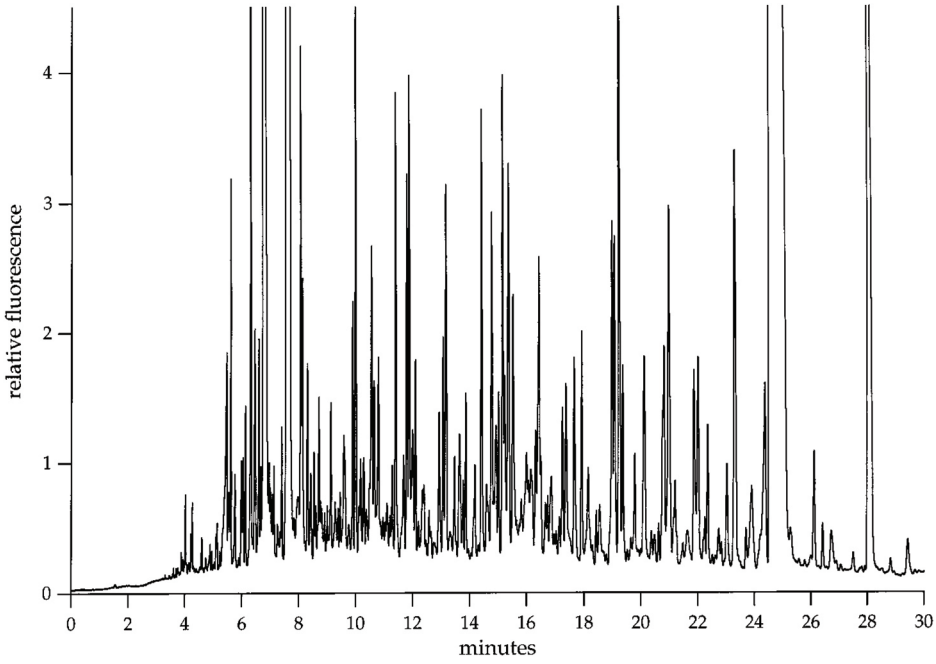

Figure 1

Separation of a TRITC-labelled ovalbumin digest using a 27 cm x 33 µm column packed with 1.0 µm C18-modified nonporous silica particles. The column was operated at an inlet pressure of 2550 bar (reprinted with permission from [8]).

The use of porous particles with small diameters has increased the loadability of UPLC columns and UPLC is gradually finding wider application in proteome analysis [10-12]. The applications show that these high-efficiency systems can also be used for high-speed analysis. The gain in peak capacity that can be achieved for peptide separations by a reduction particle size from 3 µm to sub-2 µm was shown to be about 1.5-fold for both capillary [13] and 4.6 mm ID [14] columns. Small-diameter particles have also been used to pack relatively long capillary columns in order to achieve high peak capacities. The group of Smith used a 40 cm x 50 µm capillary column packed with 1.4 µm porous particles for the analysis of a *Shewanella oneidensis* protein tryptic digest [15]. The column was operated at 1400 bar and a peak capacity of 1000 was obtained with a 400 min gradient and more than 12000 peptides and about 2000 proteins were identified using a 12 h gradient and a linear ion trap mass spectrometer. A 20 cm column packed with 0.8 µm particles was used for the fast analysis of the same digest, identifying about 250 proteins in less than 10 min [16]. The peak widths of 3-4s obtained in these experiments are indicated by the authors to be the limit of being identifiable by MS/MS.

As demonstrated by the same group, high backpressure systems can also be used to operate long capillary columns packed with more standard 3-µm particles. 87 cm long columns of different internal diameters, packed with C18-bonded porous silica particles, were used at 690 bar for the analysis of bacterial and yeast cell extract proteolytic digests [17,18]. Using this approach, Smith *et al.* demonstrated the effect of reducing column diameter and flow rate on nanoESI-MS signal intensity. The number of peptides, detected in 100 ng of a yeast protein tryptic digest, increased 200-fold upon reduction in column diameter from 74.5 µm to 14.9 µm (fig.2) while increasing the sample amount from 25 to 1000 ng on a 29.7 µm column only achieved a 14-fold increase. The reduction in column diameter resulted in a flow rate of just 20 nl/min at 690 bar, which is well suited for highly sensitive ESI-MS detection as ion suppression is suggested to be absent at these flow rates [19]. Peak capacities of over 1000 were obtained for 3 h gradients using columns with different internal diameters [18]. The group of Lee reached a peak capacity of about 1500 for a 200 min gradient using a 1.5 m x 50 µm ID column packed with 3 µm particles [20]. When combining UPLC with high-resolution Fourier transform ion cyclotron resonance (FTICR-) MS, over 100,000 peptides could be identified from a single LC-MS experiment [21]. A dual-column system using preconcentration by solid phase extraction (SPE) was developed for high-throughput analysis, providing high sensitivity of protein identification [22]. Almost 500 proteins were identified from 10 µg of a yeast soluble protein fraction using a system with six valves and two 100 cm x 75 µm columns packed with 3-µm particles.

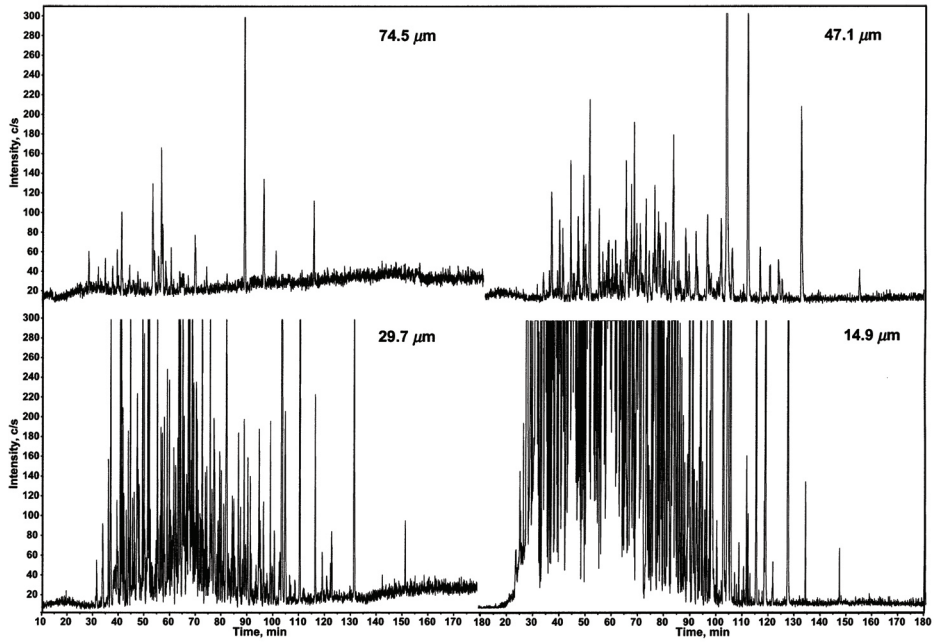

Figure 2

Chromatograms illustrating the effect of downscaling column diameter on ESI-MS response. All chromatograms were produced by injecting 100 ng of a yeast soluble protein tryptic digest on 87 cm long columns with different inner diameters. Capillaries were packed with 3.6 μm C18-bonded porous particles and were operated at 690 bar (reprinted with permission from [17]).

Recently, UPLC has also been incorporated in multidimensional LC systems. The MudPIT technology, pioneered by Yates' group [6], was recently equipped with a 50 cm x 50 μm ID column operated at 1400 bar [23]. The number of salt-gradient steps in the MudPIT experiment was reduced from 15 to 4 to utilise the full potential of the relatively long RP column. This resulted in a 1.3-fold increase in the number of identified proteins from 20 μg of a yeast digest and a reduction in total analysis time from 27.5 to 24.3 hours.

2.2 Monolithic columns

An alternative to packed beds is the use of continuous beds or monolithic materials. Monolithic columns consist of a single piece of separation media which displays high permeability and low resistance to mass transfer, making monolithic materials especially suited for the separation of analytes with low diffusion constants, like proteins, peptides, nucleic acids and synthetic polymers.

One of the first continuous bed materials was an ion exchange column prepared by strongly compressing a gel plug consisting of acrylic acid/N,N'-methylenebisacrylamide

copolymer in order to increase its pressure resistance [26]. A mixture of 5 proteins was successfully separated using a sodium chloride gradient at various flow rates. The first rigid monoliths were synthesised in the early nineties, first as methacrylate-based membranes [27], later also as methacrylate- and polystyrene-based rods [28,29]. In 1996 a silica-based monolith was synthesised for use in HPLC [30]. Silica based monoliths are reported to be more suited to the separation of small molecules while organic monoliths are more useful for the analysis of macromolecules [31]. However, both types of column have been applied to peptide separations [32,33].

Short (60 mm) poly(styrene-co-divinylbenzene) (PS-DVB) capillary monolithic columns posses high separation efficiency for the separation of both peptides and proteins [33]. A comparison of 100 and 200 µm ID PS-DVB monolithic columns with a 75 µm ID packed capillary column for mass spectrometric identification of proteins demonstrated the superior performance of the monoliths [34]. Average peak widths measured on both monolithic columns were around 19 s, which is about 17% smaller than the 23 s obtained for the packed column (fig.3). The 100 µm monolith achieved the best results for protein identification, showing the highest number of peptide identifications, mascot score and sequence coverage. The 200 µm monolith outperformed the packed column for large sample amounts but the relatively high flow rate resulted in less sensitivity at lower concentrations. The high efficiency of PS-DVB monolithic columns for separation of peptides was also illustrated in an earlier study of the same group. Peak widths at half height of less than three seconds were reported for the analysis of a standard peptide mixture at 70°C [35].

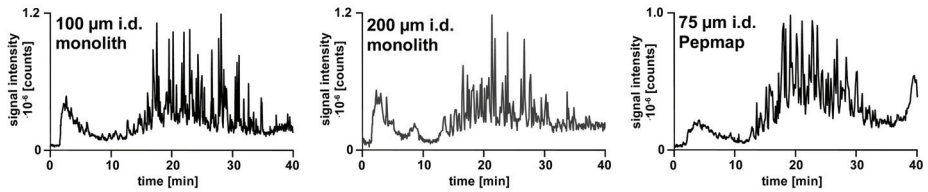

Figure 3

Comparison of the separation of a reference 10-protein digest using monolithic and packed columns. From left to right; 60 x 0.1 mm ID poly(styrene-divinylbenzene) (PS-DVB) monolith, 60 x 0.2 mm ID PS-DVB monolith, 75 x 0.075 mm ID Pepmap packed column. The columns were operated at 500, 2000 and 200 nl/min, respectively with gradients of 0-40% ACN for the monoliths and 5-60% for the Pepmap column, all in 40 minutes (adapted from [34]).

In-situ polymerisation techniques make monolithic columns highly amenable to preparation of very narrow columns. A 10 cm x 20 µm ID PS-DVB monolithic capillary column was shown to possess an efficiency of 100,000 plates/m for the LC-MS analysis of peptides [36]. The same column was used for highly sensitive analysis of protein digests, 36 proteins were identified from a sample that was equivalent to only about 1000

cells of breast carcinoma tissue. A 25 cm x 10 µm ID silica monolithic capillary column was used to identify 1332 proteins from 100 ng of an *S. Oneidensis* tryptic digest in a 3 h analysis [37]. This compares very favourably to the about 2000 proteins identified by the same group from a 12 h UPLC analysis of 10 µg of a similar sample [15].

Silica based monolithic columns generally display an even lower backpressure than PS-DVB monolithic columns. Because of this, several groups have studied long columns, either by connecting multiple short columns in series [38-40] or by using single-piece long (capillary) columns [41-43]. Columns of up to 1 m length have been constructed by column coupling, reaching efficiencies of up to 80,000 plates for a mixture of alkylbenzenes [40]. Even longer columns have been made by producing monoliths inside fused silica capillaries. Plate numbers of up to 500,000 have been reported for a 100 µm ID column of 440 cm long and 1 million plates have been obtained by connecting 3 of these columns in series [44]. The drawback of using these columns is a long analysis time.

Monolithic silica capillary columns of up to 100 cm have been used for proteomics and metabolomics applications. Tolstikov *et al*. identified several hundred plant metabolites using a 90 cm x 200 µm column [43], while Luo *et al*. reached a peak capacity of ca. 420 for a bacterial tryptic digest using a 70 cm x 20 µm column with an exponential gradient [41]. Rieux *et al*. investigated plate height, loadability and repeatability for the analysis of a simple protein digest using a 56 cm x 50 µm column [42]. Peak widths at half height were around 10 s with relative standard deviations (RSDs) of retention times under 0.5%. Loadability was about 10 pmol and could be increased 10-fold by using a trapping column. Comparison of 15 and 75 cm long silica monolithic columns showed a gain in chromatographic performance that was slightly higher than what is predicted by chromatographic theory [45,46]. As peak capacity is proportional to $\sqrt{N}$, the increase in peak capacity (*PC*) should be proportional to the square root of the column length ratio [47]. Peak capacity ratios varied from 2.7 to 4.0, which is slightly higher than $\sqrt{5}$ [46].

2.3 Porous-layer open tubular (PLOT) columns

A different approach to the production of small ID columns is the use of porous-layer open tubular (PLOT) columns. PLOT columns have a very high permeability, making it possible to employ the very long columns with high separation efficiency. A disadvantage of most open tubular columns is the limited loadability resulting from the low phase ratio, but the use of a porous layer increases the loading capacity. A 4.2 m x 10 µm ID PLOT column was shown to provide a peak capacity of about 400 (fig.4), facilitating the identification of 238 proteins from only 4 ng of a tryptic digest of a bacterial SDS-PAGE fraction (>70 kDa) [48]. Incorporation of a 3.2 m long PLOT column into a two-dimensional chromatography system led to the identification of 536 proteins from the 15-40 kDa gel fraction of a cervical cancer (*SiHa*) cell line [49].

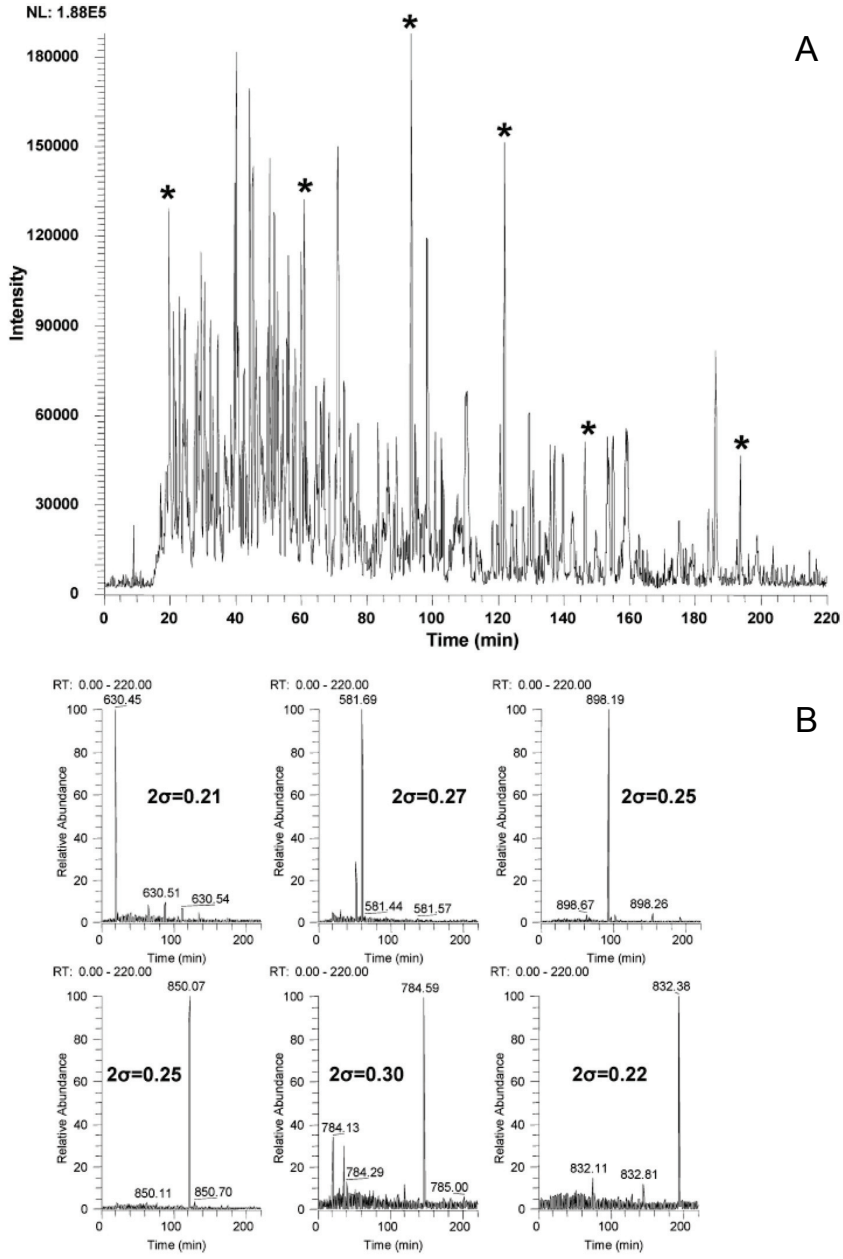

Figure 4

Peak capacity calculation for the analysis of a tryptic digest of a cut from an SDS-PAGE analysis of *Methanosacrina acetivorans* protein extract. A – base peak chromatogram of a 260 min gradient (0-36% ACN) on a 4.2 m x 10 µm ID PS-DVB PLOT column. B – Extracted ion chromatograms of six peaks used to calculate a peak capacity of ~400 (reprinted with permission from [48]).

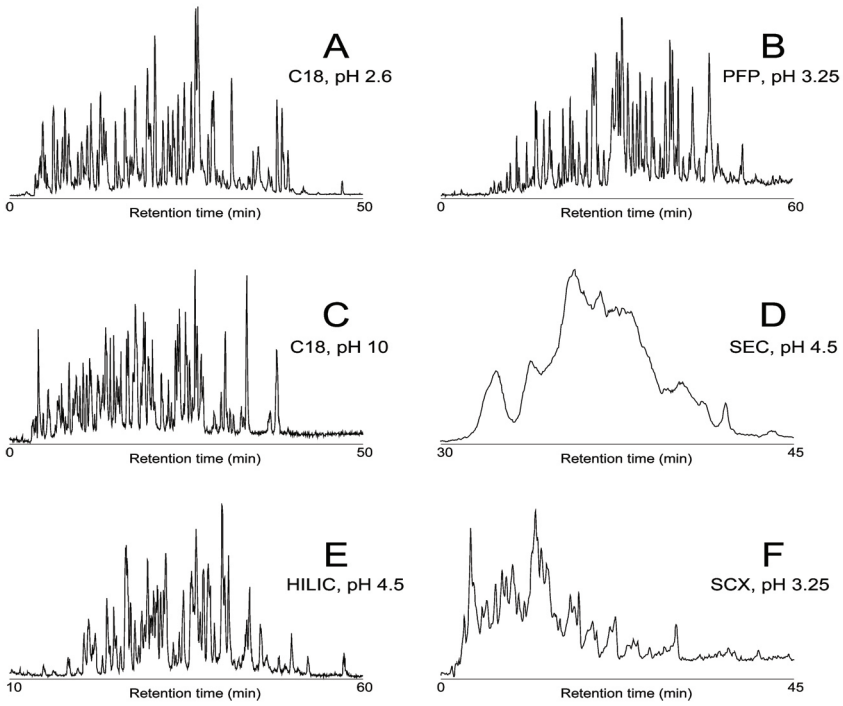

Figure 5

Comparison of the separation of phosphorylase b tryptic digest using different LC modes. A – C18-bonded, 3 µm (150 x 2.1 mm); B – Pentafluorophenyl, 5 µm (150 x 2.1 mm); C- XTerra C18, 3.5 µm (150 x 2.1 mm); D – Diol 60 Å, 5 µm (250 x 4.6 mm); E – Bare silica (150 x 2.1 mm); F – Polysulfoethyl A, 5 µm (150 x 2.1 mm) (reprinted with permission from [51]).

3. HYDROPHILIC-INTERACTION CHROMATOGRAPHY

Hydrophilic-interaction chromatography (HILIC) is a mode of chromatography that is related to normal-phase (NP) chromatography as it uses a hydrophilic stationary phase and an organic mobile phase. Contrary to NP, however, HILIC generally employs water-miscible solvents and elution is achieved by a gradient of increasing water content. Due to its high organic content (typically acetonitrile), the HILIC mobile phase is highly compatible with mass spectrometry, especially when buffered with volatile salts like ammonium formate or acetate. Comparison of LC-MS analysis of peptide mixtures for two different HILIC columns and an RP column showed that, under optimised conditions, the performance of HILIC-MS systems is comparable to that of RPLC-MS and that differences in selectivity may yield complementary results for identification of tryptic peptides [50]. In another recent study [51], the performance of bare silica sorbent in HILIC mode for the separation of peptides was found to be superior to strong cation exchange (SCX) and size exclusion (SEC) and inferior only to RPLC (fig.5). HILIC was

also shown to display a higher degree of orthogonality to RPLC compared to SCX, which suffers from unwanted clustering of similarly charged peptides. Thus, HILIC-RPLC appears to be a promising alternative to SCX-RPLC, the current workhorse of shotgun proteomics. In a recent paper, Boersema *et al.* [52] describe the development of a multidimensional system for peptide separation. The system consists of a zwitterionic (ZIC-)HILIC phase in the first dimension, connected off-line to a capillary RPLC column in the second dimension. When the HILIC separation was performed at pH 6.8, 1284 proteins could be identified from 10 μg of a digest of nuclear proteins.

The relative polarity of the stationary phase makes HILIC columns highly suited for the investigation of protein glycosylation. Glycosylation is a very important post-translational modification (PTM), which is involved in molecular recognition and control of biological activity. Additionally, glycosylation is a highly heterogeneous PTM, making global profiling of protein glycosylation very difficult. Glycopeptide analysis is further complicated by MS difficulties, originating from the large molecular weights of glycosylated peptides and the complex fragmentation patterns caused by a combination of peptide and glycan cleavages [53].

Fragmentation problems can be circumvented by enzymatically deglycosylating HILIC-enriched glycopeptide fractions prior to MS analysis [53-55], but in this way detailed information about glycoside structure is lost. A HILIC-MS method for the analysis of intact glycopeptides was developed by Wuhrer *et al.* [56]. Briefly, a protein of interest is digested with pronase, a mixture of bacterial proteases, yielding a digest consisting of relatively small peptides (2-8 aminoacids). The whole digest is fractionated using a HILIC column with the glycosylated peptides eluting last. The glycopeptides are subjected to a multistage MS cycle, giving information about glycoside structure as well as the amino acid sequence of the peptide.

4. AFFINITY-BASED STATIONARY PHASES

Affinity chromatography, in the broadest sense of the term, comprises all those chromatographic modes that rely on a specific interaction between functional groups on the stationary phase and the target molecules. Affinity chromatography has already been used for decades to selectively target specific molecules, e.g. an antigen using an antibody column or enzymes using immobilised substrates or vice versa. In proteomics, affinity chromatography is more commonly used to enrich a specific subset of proteins or peptides from a complex mixture.

Phosphorylated peptides are probably the most important target for affinity methods in proteomics as they play an important role in the regulation of many intracellular processes, like signal transduction and cell division. Affinity chromatographic methods

for the analysis of phosphopeptides can be broadly divided into two categories [57]: (1) immunoaffinity chromatography using immobilised antibodies against phosphorylated amino acids or (2) methods based on the direct interaction between phosphate groups and metal ions. The latter category includes metal loaded phases (IMAC and metal-loaded affinity chromatography) and titanium dioxide (TiO_2) particles. Titanium dioxide was recently introduced as a stationary phase material for peptide chromatography [58,59] and was quickly accepted as a robust and efficient method for the enrichment of phosphopeptides in proteomics. This is illustrated in Figure 6, which illustrates the efficiency of TiO_2 columns for phosphopeptide enrichment. The top chromatogram shows the RP-MS analysis of an SCX fraction from the 2D-LC analysis of the digest from a D. melanogaster cell lysate and the middle and bottom chromatograms show the TiO_2 flow-through and eluate fractions, respectively. The use of a TiO_2 column increased the number of identified phosphopeptides 8-fold in this SCX fraction and from 549 to 2152 identified phosphopeptides for all SCX fractions combined [60]. The speed and ease of operation of the TiO_2 material has led to application for phosphoproteome analysis in various organisms, like yeast [61], *Drosophila* [62] and human kidney cells [63]. TiO_2 has also been incorporated in chip-LC-MS methods [64].

As about 50% of all proteins are estimated to be glycosylated [65], glycoproteomics is another important area for selective enrichment. A highly effective method for glycoprotein and glycopeptides enrichment is lectin-based affinity chromatography. Lectins are proteins with strong specific affinity for carbohydrates, which ranges from affinity for specific sugars to groups of carbohydrate chains. This makes it possible to target very specific (combinations of) glycoproteins by tailoring both the specificity and the number of different lectins. Both single lectin [66] and multi-lectin [65,67] purification methods have been described.

A very specific affinity method is high-abundant protein depletion. Depletion is a very common prefractionation technique, applied especially in plasma proteomics. High abundant protein depletion in plasma proteomics is necessary since the 10 most abundant proteins comprise approximately 90% of the protein mass in plasma and the top 22 represents about 99% [68]. Most systems for high-abundant protein depletion are based on immunoaffinity chromatography and systems removing as much as 58 proteins have been described in literature [69] and systems depleting the 14 most abundant proteins are commercially available [70]. The group of El Rassi developed a system consisting of several monolithic microcolumns with immobilised affinity ligands placed in series [71]. This creates a flexible depletion system which can be readily coupled to subsequent columns for e.g. separation or tryptic digestion. High-abundant protein depletion systems have been shown to remove more proteins from the sample than intended [72-74]. Removal of unintended proteins is a drawback of these systems that will affect both sensitivity and reproducibility and can thus limit their applicability.

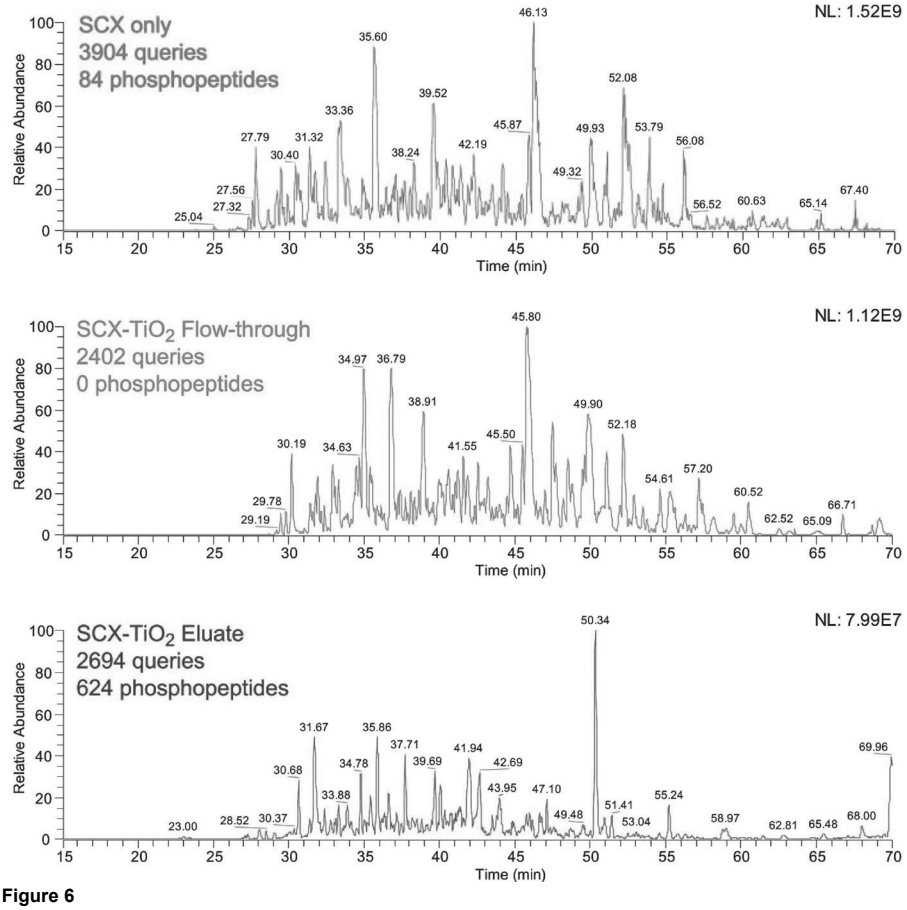

Figure 6

Selective enrichment of phosphopeptides from an SCX fraction in the 2D-LC analysis of a *drosophila melanogaster* cell lysate tryptic digest. Top chromatogram – direct RPLC-MS analysis of the SCX fraction. Middle chromatogram – RPLC-MS analysis of the TiO_2 flow through fraction. Bottom chromatogram – RPLC-MS analysis of the TiO_2 column eluate (reprinted with permission from [58]).

5. CONCLUDING REMARKS

Recent years have shown a tremendous increase in chromatographic methods available to proteomics researchers, some of which are improvements of pre-existing technologies and others providing completely new analysis tools. The development of high-efficiency and low flow rate RPLC methods has allowed efficient coupling of separation methods to high resolution MS, yielding up to 2000 identified proteins from one dimensional (1D) LC-MS experiments. It is unlikely however that 1D chromatographic methods will be able to provide comprehensive protein coverage for any proteome. It is therefore important to see how separation methods like HILIC-RPLC are now applied to peptide chromatography, providing a promising alternative to standard SCX-RPLC based multidimensional methods. It is probably safe to say that no one method will be able to ever probe the entire proteome. Affinity-based methods are mainly used for selective isolation and preconcentration in the analysis of specific sub-proteomes like the phosphoproteome, the glycoproteome and others.

Future developments will likely include the introduction of different separation modes like HILIC and ion-exchange in monolithic column technology. Furthermore, both UPLC and monolithic columns will be developed further, achieving higher efficiencies for peptide separation and providing new possibilities in multidimensional chromatography. Moreover, polymeric monolithic materials also seem promising for the separation of intact proteins. Miniaturisation of chromatographic systems, including chip-based LC, will also continue and this may also help to fully exploit the advantages of higher ionisation efficiency in LC-ESI-MS at very low flow rates.

REFERENCES

[1] N.L. Anderson, N.G. Anderson, *Mol Cell Proteom* **1** (2002) *845*.

[2] P.H. O'Farrell, *J Biol Chem* **250** (1975) *4007*.

[3] S. Gauci, L.M. Veenhoff, A.J.R. Heck, J. Krijgsveld, *J Proteome Res* **8** (2009) *3451*.

[4] J. Reinders, A. Sickmann, *Proteomics* **5** (2005) *4052*.

[5] J. Krijgsveld, S. Gauci, W. Dormeyer, A.J.R. Heck, *J Proteome Res* **5** (2006) *1721*.

[6] M.P. Washburn, D. Wolters, J.R. Yates III, *Nat Biotechnol* **19** (2001) *242*.

[7] J.E. MacNair, K.C. Lewis, J.W. Jorgenson, *Anal Chem* **69** (1997) *983*.

[8] J.E. MacNair, K.D. Patel, J.W. Jorgenson, *Anal Chem* **71** (1999) *700*.

[9] L. Tolley, J.W. Jorgenson, M.A. Moseley, *Anal Chem* **73** (2001) *2985*.

[10] R.G. Kay, B. Gregory, P.B. Grace, S. Pleasance, *Rapid Commun Mass Spectrom* **21** (2007) *2585*.

[11] M. Lima, C. Moloney, J.M. Ames, *Amino Acids* **36** (2009) *475*.

[12] J. Seidler, M. Adal, D. Kübler, D. Bossemeyer, W.D. Lehmann, *Anal Bioanal Chem* (2009) *1*.

[13] H. Liu, J.W. Finch, M.J. Lavallee, R.A. Collamati, C.C. Benevides, J.C. Gebler, *J Chromatogr A* **1147** (2007) *30*.

[14] M. Gilar, A.E. Daly, M. Kele, U.D. Neue, J.C. Gebler, *J Chromatogr A* **1061** (2004) *183*.

[15] Y. Shen, R. Zhang, R.J. Moore, J. Kim, T.O. Metz, K.K. Hixson, R. Zhao, E.A. Livesay, H.R. Udseth, R.D. Smith, *Anal Chem* **77** (2005) *3090*.

[16] Y. Shen, R.D. Smith, K.K. Unger, D. Kumar, D. Lubda, *Anal Chem* **77** (2005) *6692*.

[17] Y. Shen, R. Zhao, S.J. Berger, G.A. Anderson, N. Rodriguez, R.D. Smith, *Anal Chem* **74** (2002) *4235*.

[18] Y. Shen, R. Zhao, M.E. Belov, T.P. Conrads, G.A. Anderson, K. Tang, L. Pasa-Tolic, T.D. Veenstra, M.S. Lipton, H.R. Udseth, R.D. Smith, *Anal Chem* **73** (2001) *1766*.

[19] K. Sandra, M. Moshir, F. D'hondt, K. Verleysen, K. Kas, P. Sandra, *J Chromatogr B* **866** (2008) *48*.

[20] M.S. Kim, W.S. Choie, Y.S. Shin, M.H. Yu, S.W. Lee, *B Kor Chem Soc* **25** (2004) *1833*.

[21] Y. Shen, N. Tolic, R. Zhao, L. Pasa-Tolic, L. Li, S.J. Berger, R. Harkewicz, G.A. Anderson, M.E. Belov, R.D. Smith, *Anal Chem* **73** (2001) *3011*.

[22] H.K. Min, S.W. Hyung, J.W. Shin, H.S. Nam, S.H. Ahn, H.J. Jung, S.W. Lee, *Electrophoresis* **28** (2007) *1012*.

[23] A. Motoyama, J.D. Venable, C.I. Ruse, J.R. Yates III, *Anal Chem* **78** (2006) *5109*.

[24] Y. Shen, R.J. Moore, R. Zhao, J. Blonder, D.L. Auberry, C. Masselon, L. Pasa-Tolic, K.K. Hixson, K.J. Auberry, R.D. Smith, *Anal Chem* **75** (2003) *3596*.

[25] Y. Shen, J.M. Jacobs, D.G. Camp, R. Fang, R.J. Moore, R.D. Smith, W. Xiao, R.W. Davis, R.G. Tompkins, *Anal Chem* **76** (2004) *1134*.

[26] S. Hjertén, J.-L. Liao, R. Zhang, *J Chromatogr A* **473** (1989) *273*.

[27] T.B. Tennikova, M. Bleha, F. Svec, T.V. Almazova, B.G. Belenkii, *J Chromatogr A* **555** (1991) *97*.

[28] F. Svec, J.M.J. Frechet, *Anal Chem* **64** (1992) *820*.

[29] Q.C. Wang, F. Svec, J.M.J. Frechet, *Anal Chem* **65** (1993) *2243*.

[30] H. Minakuchi, K. Nakanishi, N. Soga, N. Ishizuka, N. Tanaka, *Anal Chem* **68** (1996) *3498*.

[31] F. Svec, C.G. Huber, *Anal Chem* **78** (2006) *2100*.

[32] B. Barroso, D. Lubda, R. Bischoff, *J Proteome Res* **2** (2003) *633*.

[33] A. Premstaller, H. Oberacher, W. Walcher, A.M. Timperio, L. Zolla, J. Chervet, N. Cavusoglu, A.v. Dorsselaer, C.G. Huber, *Anal Chem* **73** (2001) *2390*.

[34] H. Toll, R. Wintringer, U. Schweiger-Hufnagel, C.G. Huber, *J Sep Sci* **28** (2005) *1666*.

[35] W. Walcher, H. Toll, A. Ingendoh, C.G. Huber, *J Chromatogr A* **1053** (2004) *107*.
[36] A.R. Ivanov, L. Zang, B.L. Karger, *Anal Chem* **75** (2003) *5306*.
[37] Q. Luo, K. Tang, F. Yang, A. Elias, Y. Shen, R.J. Moore, R. Zhao, K.K. Hixson, S.S. Rossie, R.D. Smith, *J Proteome Res* **5** (2006) *1091*.
[38] T. Bamba, E. Fukusaki, Y. Nakazawa, A. Kobayashi, *J Sep Sci* **27** (2004) *293*.
[39] J. Bones, C. Duffy, M. Macka, B. Paull, *Analyst* **133** (2008) *180*.
[40] T. Ikegami, E. Dicks, H. Kobayashi, H. Morisaka, D. Tokuda, K. Cabrera, K. Hosoya, N. Tanaka, *J Sep Sci* **27** (2004) *1292*.
[41] Q. Luo, Y. Shen, K.K. Hixson, R. Zhao, F. Yang, R.J. Moore, H.M. Mottaz, R.D. Smith, *Anal Chem* **77** (2005) *5028*.
[42] L. Rieux, D. Lubda, H.A.G. Niederlander, E. Verpoorte, R. Bischoff, *J Chromatogr A* **1120** (2006) *165*.
[43] V.V. Tolstikov, A. Lommen, K. Nakanishi, N. Tanaka, O. Fiehn, *Anal Chem* **75** (2003) *6737*.
[44] K. Miyamoto, T. Hara, H. Kobayashi, H. Morisaka, D. Tokuda, K. Horie, K. Koduki, S. Makino, O. Nuñez, C. Yang, T. Kawabe, T. Ikegami, H. Takubo, Y. Ishihama, N. Tanaka, *Anal Chem* **80** (2008) *8741*.
[45] M.H.M. Van De Meent, G.J. De Jong, *Anal Bioanal Chem* **388** (2007) *195*.
[46] M.H.M. Van De Meent, G.J. de Jong, *J Sep Sci* **32** (2009) *487*.
[47] X. Wang, W.E. Barber, P.W. Carr, *J Chromatogr A* **1107** (2006) *139*.
[48] G. Yue, Q. Luo, J. Zhang, S.-L. Wu, B.L. Karger, *Anal Chem* **79** (2007) *938*.
[49] Q. Luo, Y. Gu, S. Wu, T. Rejtar, B.L. Karger, *Electrophoresis* **29** (2008) *1604*.
[50] Y. Yang, R.I. Boysen, M.T.W. Hearn, *J Chromatogr A* **1216** (2009) *5518*.
[51] M. Gilar, P. Olivova, A.E. Daly, J.C. Gebler, *Anal Chem* **77** (2005) *6426*.
[52] P.J. Boersema, N. Divecha, A.J.R. Heck, S. Mohammed, *J Proteome Res* **6** (2007) *937*.
[53] P. Hagglund, R. Matthiesen, F. Elortza, P. Hojrup, P. Roepstorff, O.N. Jensen, J. Bunkenborg, *J Proteome Res* **6** (2007) *3021*.
[54] P. Hagglund, J. Bunkenborg, F. Elortza, O.N. Jensen, P. Roepstorff, *J Proteome Res* **3** (2004) *556*.
[55] H. Kaji, Y. Yamauchi, N. Takahashi, T. Isobe, *Nat Protocols* **1** (2007) *3019*.
[56] M. Wuhrer, C.A.M. Koeleman, C.H. Hokke, A.M. Deelder, *Anal Chem* **77** (2004) *886*.
[57] T.E. Thingholm, O.N. Jensen, M.R. Larsen, *Proteomics* **9** (2009) *1451*.
[58] M.W.H. Pinkse, P.M. Uitto, M.J. Hilhorst, B. Ooms, A.J.R. Heck, *Anal Chem* **76** (2004) *3935*.
[59] A. Sano, H. Nakamura, *Anal Sci* **20** (2004) *565*.
[60] M.W.H. Pinkse, S. Mohammed, J.W. Gouw, B. van Breukelen, H.R. Vos, A.J.R. Heck, *J Proteome Res* **7** (2007) *687*.
[61] J.T. Wilson-Grady, J. Villén, S.P. Gygi, *J Proteome Res* **7** (2008) *1088*.
[62] B. Bodenmiller, L.N. Mueller, M. Mueller, B. Domon, R. Aebersold, *Nat Meth* **4** (2007) *231*.
[63] H. Molina, D.M. Horn, N. Tang, S. Mathivanan, A. Pandey, *P Natl Acad Sci USA* **104** (2007) *2199*.
[64] S. Mohammed, K. Kraiczek, M.W.H. Pinkse, S. Lemeer, J.J. Benschop, A.J.R. Heck, *J Proteome Res* **7** (2008) *1565*.
[65] R. Qiu, F.E. Regnier, *Anal Chem* **77** (2005) *7225*.
[66] J. Zhao, D.M. Simeone, D. Heidt, M.A. Anderson, D.M. Lubman, *J Proteome Res* **5** (2006) *1792*.
[67] Z. Yang, W.S. Hancock, *J Chromatogr A* **1053** (2004) *79*.
[68] R.S. Tirumalai, K.C. Chan, D.A. Prieto, H.J. Issaq, T.P. Conrads, T.D. Veenstra, *Mol Cell Proteomics* **2** (2003) *1096*.
[69] M. Gao, C. Deng, W. Yu, Y. Zhang, P. Yang, X. Zhang, *Proteomics* **8** (2008) *939*.
[70] N. Zolotarjova, P. Mrozinski, H. Chen, J. Martosella, *J Chromatogr A* **1189** (2008) *332*.

[71] Y. Jmeian, Z. El Rassi, *Electrophoresis* **29** (2008) *2801*.

[72] J. Granger, J. Siddiqui, S. Copeland, D. Remick, *Proteomics* **5** (2005) *4713*.

[73] T. Liu, W.-J. Qian, H.M. Mottaz, M.A. Gritsenko, A.D. Norbeck, R.J. Moore, S.O. Purvine, D.G. Camp, II, R.D. Smith, *Mol Cell Proteomics* **5** (2006) *2167*.

[74] N. Zolotarjova, J. Martosella, G. Nicol, J. Bailey, B.E. Boyes, W.C. Barrett, *Proteomics* **5** (2005) *3304*.

EFFECT OF COLUMN LENGTH ON PROTEIN DIGEST SEPARATION

Chapter 3

IMPROVEMENT OF THE LIQUID CHROMATOGRAPHIC ANALYSIS OF PROTEIN TRYPTIC DIGESTS BY THE USE OF LONG CAPILLARY MONOLITHIC COLUMNS WITH UV AND MS DETECTION

M.H.M. van de Meent and G.J. de Jong

Analytical and Bioanalytical Chemistry **388** (2007) *195*

ABSTRACT

Optimisation of peak capacity is an important strategy in gradient LC. This can be achieved by using either long columns or columns packed with small particles. Monolithic columns allow the use of long columns at relatively low back-pressure. The gain in peak capacity using long columns was evaluated by the separation of a tryptic bovine serum albumin digest with an LC-UV-MS system and monolithic columns of different length (150 and 750 mm). Peak capacities were determined from UV chromatograms and MS/MS data were used for Mascot database searching. Analyses with similar gradient slope for the two columns produced ratios of the peak capacities that were close to the expected value of the square root of the column length ratio. Peak capacities of the short column were 12.6 and 25.0 with 3 and 15 min gradients, respectively, and 29.7 and 41.0 for the long column with 15 and 75 min gradients, respectively. Protein identification scores were also higher for the long column, 641 and 750 for the 3 and 15 min gradients with the short column and 1376 and 993 for the 15 and 75 min gradients with the long column. Thus, the use of long monolithic columns provides improved peptide separation and increased reliability of protein identification.

1. INTRODUCTION

Identification of unknown proteins is a key step in proteome analysis. The standard method of protein identification consists of enzymatic digestion of the protein(s), usually by using trypsin, followed by mass spectrometric analysis of the resulting digest. When analyzing a single protein, e.g. from an excised 2D-gel spot, a peptide separation is usually unnecessary. When analyzing a digest of a mixture of proteins, the resolution of only a mass spectrometer is usually not sufficient. In those cases, a peptide separation is applied before MS detection. In shotgun proteomics, where a whole proteome is digested without prior protein fractionation, 2D-LC is the method of choice due to the high complexity of the peptide mixtures. A drawback of multidimensional methods is the time-frame, as typical analysis times for 2D-LC analysis of peptide mixtures range from several hours to more than a day [1-4]. One-dimensional LC methods can be used for separation of less complex samples (digests of <100 proteins), but this requires very efficient separations in combination with the additional separation power of mass spectrometric detection. Such high-efficiency separations have been reported for analyses with long, packed columns, using (ultra-) high pressure (~70 MPa) systems [5,6].

Peak capacity is the primary parameter for evaluation of efficiency in gradient chromatography [7]. Peak capacity was first defined by Giddings [8] as the maximum number of peaks that can be separated by a phase system. This theory was later adapted for gradient chromatography by Horváth and Lipsky [9]. There are generally two approaches for optimisation of the peak capacity. The first is increase of the gradient length, gradients of up to 10 hours have been reported on single columns [10]. However, according to theory, peak capacity increases to a maximum and then decreases as the gradients are longer [11,12]. The second approach is the use of longer columns, as the peak capacity increases linearly with the square root of the plate number and thus with the square root of the column length. Wang *et al.* illustrated this by connecting several columns packed with a pellicular stationary phase [13]. They showed that the ratio of the peak capacity and the square root of the column length were constant for columns ranging in length from 7.5 to 60 cm. The main limitation to simply increasing column length is the higher back-pressure. A possible solution for this problem is the use of monolithic columns. Monolithic columns have a higher permeability compared to packed columns, facilitating fast separations or, alternatively, the use of long columns [10,14-16]. Wang *et al.* reported a back-pressure of 28.5 MPa for a 60 cm x 2.1 mm packed column at a linear flow of 0.94 mm/s [13]. In contrast, other groups have obtained 3 times higher linear flow rates for monolithic columns of similar length [10,17]. The higher permeability of monolithic columns makes it possible to operate long columns at higher flow rates while using conventional LC equipment. Luo *et al.* separated a bacterial protein digest at a linear flow rate of 2.4 mm/s with a 70 cm monolithic column

at 34.5 MPa (5000 psi) [10]. Tolstikov *et al.* analysed plant metabolomic extracts with 2.6 mm/s for a 60 cm column [17].

The quality of the analysis of a protein digest can be expressed in different ways. One way is to describe a chromatographic efficiency, in terms of parameters like peak width, peak capacity and resolution. Another way is to use the reliability of protein identification, like SEQUEST [18] or Mascot (www.matrixscience.com) [19] scores. Such an approach is also useful, as the mass spectrometer adds a second dimension to the separation which is overlooked when only chromatographic efficiency is measured.

In this paper the evaluation of capillary monolithic silica columns of different lengths for the LC-UV-MS analysis of a BSA tryptic digest is described. Columns of 150 and 750 mm length were investigated using gradient times varying from 3 to 75 min. Chromatographic peak capacities, based on UV detection, and protein identification, based on Mascot scoring data of the MS detection, were determined as a measure for the efficiency of peptide separation.

2. THEORETICAL ASPECTS

The peak capacity is defined as the maximum number of bands that will fit within a chromatogram with a resolution of R_s = 1.0 [20]. In gradient-LC, where peak width is about constant throughout the separation, the theoretical peak capacity is given by equation 1:

$$PC = 1 + \frac{t_G}{w_{av}} \tag{1}$$

where t_G is the gradient time and w_{av} is the average baseline peak width. For large values of *PC*, this approaches t_G/w_{av}. Because sample peaks often do not occupy the entire length of the gradient, the sample peak capacity can then be defined as,

$$PC^{**} = \frac{(t_z - t_a)}{w_{av}} = \frac{\Delta t}{w_{av}} \tag{2}$$

where t_a and t_z are the retention times of the first and last eluting peaks, respectively.
Wang *et al.* [13] investigated the effect of column length on peak capacity in packed columns and found that the peak capacity is proportional to $\sqrt{L}$ if the gradient slope is proportional to *L*. Gradient slope can be defined as:

$$\frac{\Delta\phi t_0}{t_G} \tag{3}$$

where $\Delta\phi$ is the change in organic modifier fraction during the gradient ($0 \le \phi \le 1$) and t_0 is the column dead time. If the linear flow rate is constant, t_0 is only dependent on

column length. Therefore, t_G should be varied proportionally to column length in order to keep gradient slope constant.

3. EXPERIMENTAL

3.1 Materials and reagents

Bovine serum albumin (BSA), trypsin (porcine, type IX-S, EC 3.4.21.4) and 1,4-dithiothreitol (DTT) were purchased from Sigma (St.Louis, MO, USA), iodoacetamide (IAA) and ammonium hydrogencarbonate from Fluka (Buchs, Switzerland). Acetonitrile was HPLC gradient-grade (Biosolve, Valkenswaard, the Netherlands) and spectroscopy grade trifluoroacetic acid (TFA) was obtained from Merck KGaA (Darmstadt, Germany). All solutions were prepared using water from a Milli-Q water purifying system (Millipore, Bedford, MA, USA).

All reagents for the digestion of BSA were prepared in 200 mM NH_4HCO_3 buffer, pH 8. The tryptic digest was prepared as follows; 100 µl of BSA stock solution (3.5 µg/µl in water) was set to pH 8 by adding 25 µl of a 1 M NH_4HCO_3 buffer (pH 8). After addition of 25 µl of a 10 mM DTT solution the sample was incubated at 50°C for 30 minutes to reduce disulfide bonds, after cooling to room temperature 25 µl of a 30 mM IAA solution was added and the sample was incubated in the dark for 60 min to alkylate the free thiols. Trypsin was dissolved in 10 µl of buffer to reach a trypsin/protein mass ratio of 1:50 in the final solution, the trypsin solution was added to the sample, which was incubated at 37°C overnight for 15 hours. The digestion was stopped by addition of 15 µl of 10% TFA. The sample was diluted to 200 ng/µl (3 µM) with LC mobile phase A (water + 0.05% TFA) and injected without further purification.

3.2 Apparatus and LC columns

All analyses were performed with an Agilent 1100 nanoLC system (Agilent technologies, Waldbronn, Germany), consisting of a vacuum degasser, a binary Nano-Pump, a µ-well plate sampler and a column switching module with a trapping column in the 1-4 position of the 6-port column-switching valve. The trapping pump was a Gynkotek Model 480 (Gynkotek GmbH, Germering, Germany). Detection was performed by UV and MS detection, connected in series. The UV detector was an MU 701 UV-VIS detector (ATAS GL International, Veldhoven, The Netherlands), equipped with an external optical-fiber flow cell (6 nl, 3mm light path); peptides were detected at 215 nm. The Mass Spectrometer was an Agilent LC/MSD Trap XCT (Agilent Technologies, Waldbronn, Germany) ion trap mass spectrometer, equipped with an orthogonal ESI interface. The external flow cell of the UV detector allows minimal time delay and band broadening between UV and MS detection.

The monolithic columns were provided by GL Sciences Inc. (Tokyo, Japan). The columns

were a 150 x 0.1 mm MonoCap for nano-flow C18-silica monolithic column and a 750 x 0.2 mm MonoCap High Resolution C18-silica monolithic column. For trapping of the digest a 5 x 0.3 mm column packed with 5 µm Zorbax 300 SB-C18 (Agilent Technologies, Waldbronn, Germany) was used.

3.3 Method and data analysis

LC solvent A was water + 0.05% TFA (v/v), solvent B was acetonitrile + 0.04% TFA (v/v). The trapping solvent was a mixture of 5% (v/v) solvent B in solvent A. After injection of the digest (0.25 µl for the 0.1 mm i.d. column and 1.0 µl for the 0.2 mm i.d. column), the sample was trapped on the trapping column at a flow rate of 5 µl/min. After 5 min, the trapping column was switched on-line with the separation column and the gradient was started. All separations were performed at room temperature using a gradient of 5-50% solvent B with gradient times varying from 3-75 min. MS spectra were acquired in the positive ion mode over the 400-2000 m/z range, after which the two most intense ions (with a preference for doubly charged ions) were selected for fragmentation. MS/MS-fragmentation spectra were acquired over the 100-2200 m/z range. An ESI spray voltage of -3 kV was used for all experiments.

The effect of separation efficiency on protein identification was evaluated using the Mascot search engine (www.matrixscience.com) [19]. LC-MS/MS data were converted to the mascot generic format (.mgf file) using the data analysis software, the .mgf files were searched against the MSDB database using Mascot's MS/MS ion search module. The database was searched for tryptic peptides from all entries in the database, allowing 1 missed cleavage per peptide and containing carbamidomethyl cysteine as a variable modification. Mass tolerances were set to default values: Peptide mass tolerance ± 2.0 Da, MS/MS tolerance ± 0.8 Da.

4. RESULTS AND DISCUSSION

4.1 LC-UV

Because of the difference in diameter the 150 x 0.1 mm and the 750 x 0.2 mm columns were used with different flow rates. For the 150 mm and the 750 mm columns, the flow rates were 0.5 µl/min and 2.0 µl/min, respectively, resulting in a linear flow rate of 1.06 mm/s. Injection volumes were also proportional to the square of the column diameter, 0.25 µl of the digest for the 0.1 mm column and 1.0 µl onto the 0.2 mm column. During the gradient, the maximum back-pressure of the 750 mm column was below 20 Mpa, which is well below the manufacturers limit of 30 Mpa.

Figure 1 shows the LC-UV chromatograms of 3 and 15 min gradients run on the 150 mm column and 15 and 75 min gradients run on the 750 mm column. When the chromatograms of the analyses with similar gradient slope are compared (Fig. 1A/B

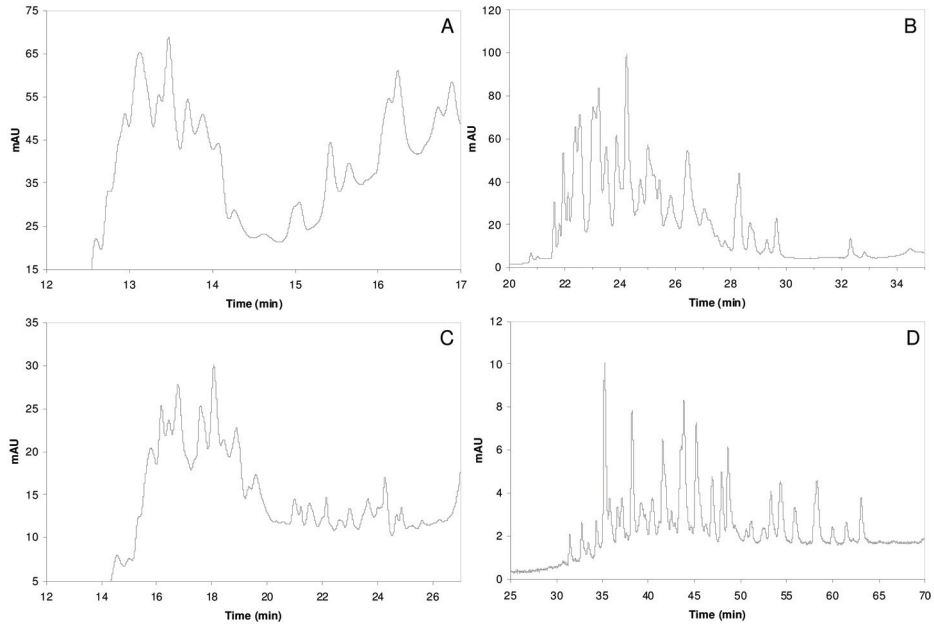

Figure 1

LC-UV chromatograms of a tryptic BSA digest, separated by monolithic silica capillary columns of 150 x 0.1 mm (A, C) and 750 x 0.2 mm (B, D) using a gradient of 5-50% acetonitrile (0.04% (v/v) TFA) in water (0.05% (v/v) TFA). A: 3 min gradient (15% min^{-1}), B and C: 15 min gradient (3% min^{-1}), D: 75 min gradient (0.6% min^{-1}). The run times include 5 min of trapping and 2.5 and 12.5 min of gradient delay time for the 150 mm and 750 mm column respectively.

and 1C/D), it is clear that an increase in column length improves the peptide separation. In order to quantify the efficiency of the separation the sample peak capacity was calculated for all analyses. Because of the incomplete resolution of the digest the peak capacity was estimated by using the average peak width of a selected number of peaks that appeared to contain only a single peptide. Using this method, peak capacities were calculated for all analyses and the results are summarised in Table 1. The peak capacities found for the short column are comparable to those found in literature for similar columns [21,22]. As expected, the peak capacities of the long column are higher than those of the short column, but they are relatively low compared to the values reported in ref. 10. However, when gradient time is taken into consideration, the difference is significantly less: PC^{**}/t_G is 0.55 peaks/min with the present system (75 min gradient on the 750 mm column) and 1.62 peaks/min for the 260 min gradient reported in ref. 10. A possible negative effect on the resolution of our system is the use of a trapping column, filled with a different stationary phase, which probably has a selectivity that differs from that of the analytical column.

If chromatograms with the same gradient slope are compared (3 and 15 min gradients for the 150 mm column and 15 and 75 min gradients for the 750 mm column, respectively) the ratio of the peak capacities should be close to the square root of the column length ratio (= 2.24). For the 3/15 min gradient pair this ratio is 29.7/12.6 = 2.37, for the 15/75 pair the ratio is 41.0/25.0 = 1.64. For the short column, no increase in peak capacity is observed with a gradient of more than 15 min but for the long column the peak capacity increases up to a 75 min gradient. A possible explanation can be found in the study of Stadalius *et al*. [11,12], who have demostrated that peak capacity will increase with gradient time, until it reaches a maximum after which it will even decrease. The gradient time at which this maximum is obtained is higher for longer columns.

4.2 LC-MS

To assess the influence of peptide separation on protein identification, MS/MS data were investigated by database searching using the MS/MS ions search module from the Mascot search engine. Results are expressed as a Mascot score, the number of unique identified peptides and the percentage of the BSA amino acid sequence covered by these peptides (Table 2). These results were compared to the scores obtained for direct infusion of the BSA digest at the same flow rates as the LC separations. The protein identification parameters for the infusion experiments were about similar for both flow rates. All database searches gave bovine albumin as the top protein match and the only other significant matches were albumins from other species.

Figure 2 shows base peak chromatograms of separations with the same gradient slope, a 15 min gradient for the 150 mm column and a 75 min gradient for the 750 mm column. The Mascot scores for the 750 mm column are higher than those for the 150 mm column. The average score per identified peptide is between 52 and 60 for the long column and between 42 and 51 for the short column. Combined with the larger number of identified peptides on the long column this adds up to a higher Mascot score (Table 2).

Table 1: Chromatographic parameters from LC-UV

	150 x 0.1 mm			*750 x 0.2 mm*		
t_G [a]	w_{av} [b]	Δt [c]	*PC*** [d]	w_{av}	Δt	*PC***
3 min	0.24	3.05	12.6	0.28	3.16	11.2
15 min	0.44	11.0	25.0	0.39	11.6	29.7
75 min	1.27	37.6	29.6	0.77	31.7	41.0

[a] *Gradient time*

[b] *Average width of 7 to 12 peaks which were found to contain only a single peptide (based on MS data).*

[c] *Elution window between the first and last eluting peptides*

[d] *Sample peak capacity ($\Delta t/w_{av}$)*

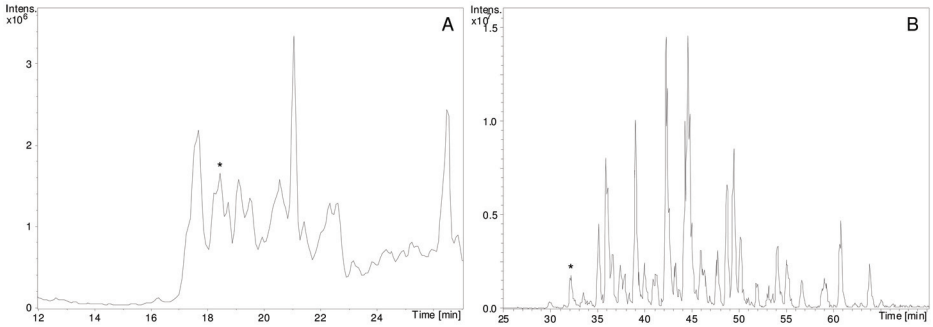

Figure 2

LC-MS Base Peak Chromatograms and Mass Spectra of tryptic BSA digest, separated by monolithic silica columns of 150 x 0.1 mm (A) and 750 x 0.2 mm (B), using a gradient of 5-50% acetonitrile (0.04% (v/v) TFA) in water (0.05% (v/v) TFA). A: 15 min gradient (3% min^{-1}); B: 75 min gradient (0.6% min^{-1}). The run times include 5 min of trapping and 2.5 and 12.5 min of gradient delay time for the 150 mm and 750 mm column respectively.

Despite providing only a low resolution, even the shortest gradients cause a considerable increase in the reliability of protein identification. Increase in gradient length leads to cleaner mass spectra (Fig. 3) and a higher number of identified peptides and sequence coverage, compared to direct infusion. These numbers, however, decrease beyond a certain gradient time. This could possibly be attributed to the lower peak heights for longer gradient times. This result indicates that there is a gradient slope where an increase in chromatographic separation no longer improves protein identification.

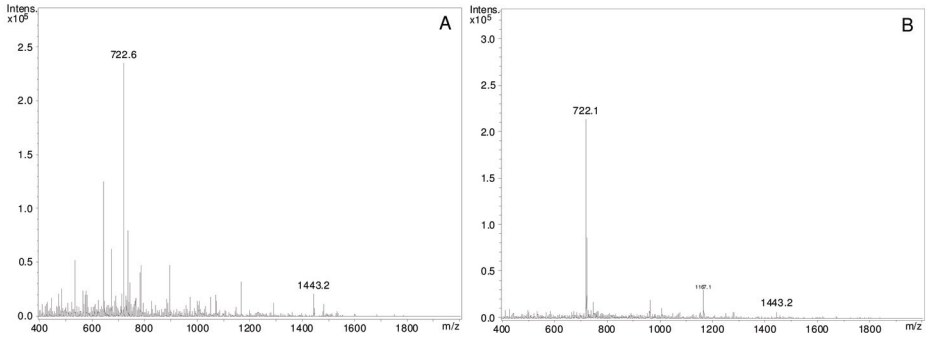

Figure 3

Averaged Mass Spectra of peptide YICDNQDTISSK (m/z 722.32, $M2H^{2+}$) as identified from Extracted Ion Chromatograms in the LC-MS analysis of a tryptic BSA digest. A; 150 x 0.1 mm silica monolithic column, 15 min gradient of 5-50% acetonitrile (0.04% (v/v) TFA) in water (0.05% (v/v) TFA). B; 750 x 0.2 mm column, 75 min gradient.

5. CONCLUSIONS

The use of long silica-based capillary monolithic columns provides a clear advantage over shorter columns, i.e. an increase of chromatographic efficiency and reliability of protein identification. As expected from chromatographic theory, a factor 5 longer column gives a 1.6 – 2.4 times increase in peak capacity for separations with similar gradient slope. The use of longer gradients also leads to an initial improvement of the protein identification score but the score seems to have a maximum at longer gradient times.

While the use of longer columns for the separation of peptides has a clear advantage because of the gain in chromatographic efficiency, this also gives a longer analysis time. As maximum protein identification scores for rather simple digests are reached at relatively short gradient times, it is important to find a compromise between chromatographic efficiency and analysis time. However, if the sample is more complex, the use of longer columns is more attractive as longer gradients are necessary to achieve sufficient separation.

In the near future, short and long columns of the same diameter (0.1 mm i.d.) will be compared. Further improvement of the separation might be obtained by optimisation of the combination of the trap column and the analytical column. Moreover, the potential of longer monolithic capillary columns will be demonstrated by the analysis of more complex and real samples.

ACKNOWLEDGEMENTS

The authors would like to thank Mr. M. Kurano of ATAS GL International for generously providing the monolithic capillary columns.

REFERENCES

[1] C. Delahunty, J.R. Yates III, *Methods* **35** (2005) *248*.

[2] J. Dai, C.H. Shieh, Q. Sheng, H. Zhou, R. Zeng, *Anal. Chem.* **77** (2005) *5793*.

[3] M.P. Washburn, D. Wolters, J.R. Yates III, *Nat. Biotechnol.* **19** (2001) *242*.

[4] G.J. Opiteck, J.W. Jorgenson, R.J. Anderegg, *Anal. Chem.* **69** (1997) *2283*.

[5] Y. Shen, R.D. Smith, *Electrophoresis* **23** (2002) *3106*.

[6] Y. Shen, R. Zhao, M.E. Belov, T.P. Conrads, G.A. Anderson, K. Tang, L. Pasa-Tolic, T.D. Veenstra, M.S. Lipton, H.R. Udseth, R.D. Smith, *Anal. Chem.* **73** (2001) *1766*.

[7] U.D. Neue, *J. Chromatogr. A* **1079** (2005) *153*.

[8] J.C. Giddings, *Anal. Chem.* **39** (1967) *1027*.

[9] C.G. Horváth, S.R. Lipski, *Anal. Chem.* **39** (1967) *1893*.

[10] Q. Luo, Y. Shen, K.K. Hixson, R. Zhao, F. Yang, R.J. Moore, H.M. Mottaz, R.D. Smith, *Anal. Chem.* **77** (2005) *5028*.

[11] M.A. Stadalius, M.A. Quarry, L.R. Snyder, *J. Chromatogr. A* **327** (1985) *93*.

[12] L.R. Snyder, M.A. Stadalius, in C. Horvath (Editor), *High-Performance Liquid Chromatography, Advances and Perspectives, Academic Press, New York* (1986) *195*.

[13] X. Wang, W.E. Barber, P.W. Carr, *J. Chromatogr. A* **1107** (2006) 139.

[14] L. Xiong, R. Zhang, F. Regnier, *J. Chromatogr. A* **1030** (2004) *187*.

[15] T.P. Hennessy, R.I. Boysen, M.I. Huber, K.K. Unger, M.T.W. Hearn, *J. Chromatogr. A* **1009** (2003) *15*.

[16] F.C. Leinweber, D.G. Schmid, D. Lubda, K. Wiesmüller, G. Jung, U. Tallarek, *Rapid Commun. Mass Spectrom.* **17** (2003) *1180*.

[17] V.V. Tolstikov, A. Lommen, K. Nakanishi, N. Tanaka, O. Fiehn, *Anal. Chem.* **75** (2003) *6737*.

[18] J.K. Eng, A.L. McCormack, I. Yates, John R., *J. Am. Soc. Mass Spectrom.* **5** (1994) *976*.

[19] D.N. Perkins, D.J.C. Pappin, D.M. Creasy, J.S. Cottrell, *Electrophoresis* **20** (1999) *3551*.

[20] J.W. Dolan, L.R. Snyder, N.M. Djordjevic, D.W. Hill, T.J. Waeghe, *J. Chromatogr. A* **857** (1999) *1*.

[21] B. Barroso, D. Lubda, R. Bischoff, *J. Prot. Res.* **2** (2003) *633*.

[22] J. Rozenbrand, W.P. van Bennekom, K.K. Unger, G.J. de Jong, *Anal. Bioanal. Chem.* **385** (2006) *1055*.

Chapter 4

POTENTIAL OF LONG CAPILLARY MONOLITHIC COLUMNS FOR THE ANALYSIS OF PROTEIN DIGESTS

M.H.M. van de Meent and G.J. de Jong

Journal of Seperation Science **32** (2009) *487*

Ch4

ABSTRACT

The gain in separation efficiency for protein digests using long monolithic columns has been evaluated for a liquid chromatography-mass spectrometry (LC-MS) system with capillary monolithic columns of different lengths (150 and 750 mm). A mixture of bovine serum albumin, α-casein and β-casein tryptic digests was used as a test sample. Peak capacity and productivity (peak capacity per unit time) were determined from base peak chromatograms and MS/MS data were used for protein identification by Mascot database searching. Peak capacity and protein identification scores were higher for the long column. Analyses with similar gradient slope for the two columns produced ratios of the peak capacities that were slightly higher than the expected value of the square root of the column length ratio. Peak capacity ratios varied from 2.7-4.0 for four different gradient slopes, while protein identification scores were 2-4 times higher for the long column. Similar values were obtained for the productivity of both columns and the highest productivity was obtained at gradient times of 45 min and 75 min for the short and long column, respectively. The use of long monolithic columns improves peptide separation and increases reliability of protein identification for complex digests, especially if longer gradients are chosen.

1. INTRODUCTION

Modern-day proteome analysis relies heavily on the power of mass spectrometry (MS) to identify unknown proteins. Essential steps in this process are the digestion of intact proteins into peptides (usually by trypsin) and separation of the peptide mixtures. Due to its high resolution and ease of coupling to (nano-)electrospray ionisation-mass spectrometry (ESI-MS), reversed-phase liquid chromatography (RPLC) is the most widely used separation tool for analysis of peptide mixtures. Commercially available columns typically have a length of 150-250 mm and are mostly suited for analysis of relatively simple peptide mixtures. When sample complexity increases, e.g. digests of complex protein mixtures, the performance of these standard columns is often insufficient. One possible approach is the use of multidimensional LC methods, which combine multiple (orthogonal) separation techniques like ion-exchange, size-exclusion or hydrophilic interaction chromatography with RPLC as a final separation step. These methods, however, require complex instrumental set-ups and often lead to relatively long run-times [1-5]. Another possibility is to increase the efficiency of the reversed-phase column. This can be achieved either by reducing the particle size of the column packing material [6-8] or by increasing the length of the column [9,10]. Both strategies, however, lead to an increase in column backpressure which means that special equipment is often required. Due to their high porosity, monolithic materials only cause a relatively small pressure drop per unit of column length and this makes it possible to use much longer columns with conventional LC apparatus.

Some research groups have created monolithic columns of up to one meter in length by connecting several 100 mm long monolithic columns in series, reaching efficiencies of up to 90,000 plates [11-13]. Commercially available capillary monolithic columns generally have a maximum length of 150 mm, but some research groups have published results obtained with home-made monolithic columns of up to 1.4 m long. Tolstikov *et al*. [14] identified several hundred plant metabolites using a 900 x 0.2 mm column, while Luo *et al*. reached a peak capacity of ca. 420 for a bacterial tryptic digest using a 700 mm x 20 µm column using an exponential gradient [15]. Rieux *et al.* [16] investigated plate height, loadability and repeatability for the analysis of a simple protein digest using a 560 mm x 50 µm column, while Ikegami *et al*. [13] reported over 160,000 plates for an isocratic separation of alkylbenzenes with a 1.4 m x 200 µm column at 190 bar.

The primary parameter to evaluate the performance of a gradient separation is the peak capacity (*PC*). Peak capacity was first defined by Giddings [17] and adapted for gradient chromatography by Horváth and Lipsky [18]. Peak capacity is defined as the maximum number of peaks that will fit within a chromatogram with a resolution of 1.0. Since peak width is about constant throughout a gradient separation, the theoretical peak capacity can be calculated using equation 1:

$$PC = 1 + \frac{t_G}{w_{av}} \quad (1)$$

where t_G is the gradient time and w_{av} is the average peak width. For large values of *PC*, this approaches t_G/w_{av}. A disadvantage of using *PC* as a measure for separation performance is that it is dependent on the length of the gradient, rather than on the elution window of the analytes. As a result, empty spaces in the chromatogram at the beginning or the end of the gradient can artificially increase *PC*. An alternative could be to calculate a sample peak capacity (Equation 2). Here the time difference between the first and last eluting peaks (Δt) is used rather than the gradient time.

$$PC^{**} = \frac{(t_z - t_a)}{w_{av}} = \frac{\Delta t}{w_{av}} \quad (2)$$

A disadvantage of *PC*** is the dependence of Δt on sample composition and detection method. The use of different detection methods or variation in sample complexity may give changes in Δt and thus *PC***. Since *PC* is only dependent on gradient programming, it is not sensitive to changes in instrumental set-up and sample composition and therefore more suited for comparing different systems.

Another way to assess separation performance and, at the same time, circumvent the disadvantages of both the *PC* and *PC*** methods is to use the productivity, or number of separated components per unit of time, as introduced by Gilar *et al.* [19]. The productivity can be calculated by dividing the peak capacity by the analysis time. The productivity provides different information than peak capacity, which focuses on the maximum number of analytes that can be separated by a given system, while the productivity describes the maximum number of compounds that can be separated in a given time interval.

A second way to describe the quality of the analysis of a protein digest is the identification of the proteins from mass spectrometry data using database search engines like SEQUEST [20] or Mascot (http://www.matrixscience.com) [21]. The number of identified peptides and proteins and the degree of sequence coverage are indicative of the quality of the separation. Co-elution of analytes can lead to mutual ion suppression [22] and the potential loss of sequence information because some peptides might not be fragmented since the mass spectrometer can only fragment a limited number of ions from a single spectrum without an undesirable increase of the duty-cycle. If the duty-cycle becomes too long, precursor ion concentrations will already decrease before the mass spectrometer starts the fragmentation. A better separation reduces the number of necessary MS/MS scans during each duty-cycle, allowing more full-MS scans and more effective precursor selection.

In this paper capillary monolithic silica columns of two different lengths are compared for the LC-MS analysis of a mixture of tryptic peptides. Contrary to our previous publication [23], both columns now have the same internal diameter and a trapping column was omitted from the LC system in order to focus on the performance of the separation columns. Moreover, the bovine serum albumin (BSA) tryptic digest used previously was replaced by a mixture of tryptic digests of BSA, α-casein and β-casein. This mixture contains about twice the number of peptides over 500 Da, including several peptides larger than 2500 Da and a number of phosphopeptides and is therefore more similar to a real sample. Columns of 150 x 0.1 and 750 x 0.1 mm were investigated using gradients from 3 to 225 min. Peak capacity, productivity and Mascot protein identification scores were compared for the two columns.

2. EXPERIMENTAL

2.1 Materials and reagents

BSA, α-Casein, β-Casein, trypsin (porcine, type IX-S, EC 3.4.21.4) and 1,4-dithiothreitol (DTT) were purchased from Sigma (St.Louis, MO, USA), MS grade formic acid, iodoacetamide (IAA) and ammonium hydrogencarbonate from Fluka (Buchs, Switzerland). Acetonitrile was HPLC gradient-grade (Fisher Scientific, Loughborough, UK) and biochemistry grade trifluoroacetic acid (TFA) was obtained from Acros Organics (Geel, Belgium). All solutions were prepared using water from a SynergyUV water purifying system (Millipore, Bedford, MA, USA).

All reagents for the digestion of the proteins were prepared in 200 mM NH4HCO3 buffer (pH 8) unless stated otherwise. The tryptic digests were prepared as follows; 500 µl of protein stock solutions (4.0 µg/µl in water) was set to pH 8 by adding 125 µl of a 1 M NH_4HCO_3 buffer (pH 8). After addition of 125 µl of a 10 mM DTT solution the sample was incubated at 50°C for 30 minutes to reduce disulfide bonds, after cooling to room temperature 125 µl of a 30 mM IAA solution was added and the sample was incubated in the dark for 60 min to alkylate the free thiols. Trypsin was dissolved in 50 µl of buffer to reach a trypsin to protein mass ratio of 1:50 in the final solution, the trypsin solution was added to the sample, which was incubated overnight (15 hours) at 37°C. The digestion was stopped by addition of 75 µl of 10% TFA. The digests were diluted to 1.0 µg/µl using LC mobile phase A (water + 0.1% FA / 0.01% TFA) and stored at -20°C until further use. Before injection, the separate digests were mixed in a 1:1:1 ratio and diluted to 200 ng/µl per protein using mobile phase A.

2.2 Apparatus and LC columns

All analyses were performed with an Agilent 1100 nanoLC system (Agilent technologies, Waldbronn, Germany), consisting of a vacuum degasser, a binary Nano-Pump and a µ-well plate sampler. Detection was performed using an Agilent LC/MSD Trap XCT

(Agilent Technologies, Waldbronn, Germany) ion trap mass spectrometer, equipped with an orthogonal nanoESI interface.

The 150 mm long column was a 150 x 0.1 mm i.d. Chromolith CapRod Monolithic Capillary (Merck KGaA, Darmstadt, Germany). The 750 x 0.1 mm MonoCap High Resolution C18-silica monolithic column was provided by GL Sciences Inc. (Tokyo, Japan).

2.3 Method and data analysis

LC solvent A was 2% acetonitrile + 0.1% FA / 0.01% TFA (v/v/v), solvent B was 90% acetonitrile + 0.08% FA / 0.01% TFA (v/v/v). After injection of 0.5 µl the digest, the sample was trapped on the top of the column. After 5 min, the gradient was started. All separations were performed in duplicate using a gradient of 5-55% solvent B with gradient times varying from 3 to 225 min at a flow rate of 0.5 µl/min, resulting in a maximum backpressure of approximately 35 bar on the 150 mm column and about 180 bar on the 750 mm column. The total scan time was below 1s for a full cycle, consisting of one precursor scan followed by up to three fragment ion scans.

A nano-electrospray voltage of -2.75 kV was used for all experiments, MS spectra were acquired at an m/z range of 400-2000 and MS/MS-fragmentation spectra were acquired over a 100-2200 m/z range using the peptide fragmentation settings of the mass spectrometer. The effect of separation efficiency on protein identification was evaluated using the Mascot search engine (www.matrixscience.com). LC-MS/MS data were converted to mascot generic format (.mgf) using the XCT's data analysis software, the peak lists were searched against the Swiss-Prot database using Mascot's MS/MS ion search module. The database was searched for tryptic peptides from all Mammalia in the database, allowing 2 missed cleavages per peptide and containing carbamidomethyl cysteine as a fixed modification and M-oxidization and S/T/Y-phosphorylation as variable modifications. Peptide mass tolerance was set to ± 1.2 Da, MS/MS tolerance to ± 0.9 Da.

3. RESULTS AND DISCUSSION

3.1 Peak capacity

The choice of the LC gradient conditions is important for optimisation of the peak capacity. The gradient slope is proportional to the ratio t_0 / t_G, therefore the ratio of the gradient lengths should correspond with the ratio of the column lengths in order to keep the slope constant. The use of a longer column results in a proportional increase in t_0, therefore t_G should also be increased proportionally to column length in order to keep the gradient slope constant. When comparing analyses with equal gradient slope, the peak capacity ratio for two columns should be close to the square root of the column length ratio [24].

Figure 1 shows the base peak chromatograms of analyses using two different gradient slopes for both the 150 and 750 mm columns. When comparing the chromatograms with equal gradient slope, the improvement in separation by use of the long column is clear. Peptide separation was evaluated by the peak capacities, which were calculated according to equation 1. The average peak width was calculated from the peak widths in extracted ion chromatograms of eight peptides, covering all the proteins in the sample. The peptides used are shown in table 1, the numbers correspond to those in figure 1. The average peak widths, peak capacity and productivity are shown in table 2. Minimum peak widths are about 15s, while the MS scan rate is higher than 1 scan/s, thus the sampling rate is sufficient for accurate determination of peak widths. Peak capacity ratios were between 2.7 and 4.0, which is slightly higher than the theoretical prediction of $\sqrt{5}$. The peak capacities calculated for the short column are similar to literature values [23,25,26], although the performance of the present system is better for long gradients with respect to the previously published data [23]. This is also true for the long column when peak capacities are compared to those obtained with the 750 x 0.2 mm column. It should be noted that the peak capacities in ref. [23] were calculated from

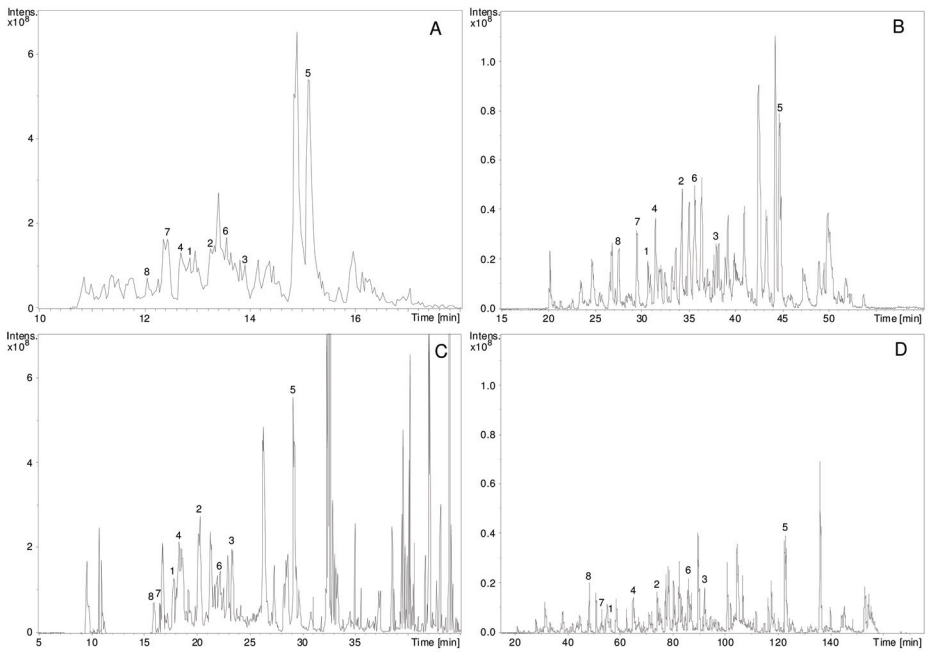

Figure 1

Base peak chromatograms of a mixture of BSA, α-Casein and β-Casein tryptic digests, separated on capillary silica monolithic columns of 150 mm (A,C) and 750 mm (B,D) x 0.1 mm ID, using a gradient of 5-55% B (90% ACN + 0.08% FA, 0.01% TFA) in solvent A (2% ACN + 0.1% FA, 0.01% TFA). A: 9-min gradient, B: 45-min gradient, C: 45-min gradient, D: 225-min gradient. Peak numbers correspond to peptides listed in table 1.

Table 1: Peptides used for peak capacity calculations

	Peptide	*m/z*	*Protein*
1	LVVSTQTALA	1002.58 (1+)	BSA
2	LVNELTEFAK	582.32 (2+)	BSA
3	LGEYGFQNALIVR	740.40 (2+)	BSA
4	KVPQVSTPTLVEVSR	820.47 (2+)	BSA
5	FFVAPFPEVFGK	692.87 (2+)	α-S1-Casein
6	HQGLPQEVLNENLLR	880.48 (2+)	α-S1-Casein
7	NAVPITPTLNR	598.34 (2+)	α-S2-Casein
8	FQpSEEQQQTEDELQDK	1031.41 (2+)	β-Casein

Table 2: Average peak width (w_{av}), peak capacity (PC) and productivity (PC/t_{anal}[a]) for two columns and various gradient times (t_G)

	150 x 0.1 mm			*750 x 0.1 mm*		
t_G	w_{av} *(min)*	*PC*	PC/t_{anal} *(peaks/min)*	w_{av} *(min)*	*PC*	PC/t_{anal} *(peaks/min)*
3 min	0.28	10.7	0.60	-	-	-
9 min	0.34	26.6	1.11	0.33	27.7	0.81
15 min	0.37	40.5	1.35	0.35	43.1	1.08
45 min	0.52	86.7	1.45	0.50	90.3	1.29
75 min	0.73	102.7	1.14	0.56	134.7	1.35
225 min	-	-	-	0.95	237.5	0.95

[a] *t_{anal}: t_G + 15 for the 150 mm column, t_G + 25 for the 750 mm column*

UV rather than MS data and that the experiments were performed using an LC system containing a trapping column.

The productivity of a separation can be calculated by dividing the peak capacity by the total analysis time (t_{anal}). The total analysis time consists of the gradient length and the time needed for sample trapping and column re-equilibration. In the present systems t_{anal} can be obtained by adding 15 min to the gradient time for the short column. Due to the longer equilibration time, 25 min need to be added to the gradient time for the long column. As is shown in table 2 the productivity is similar for both columns. Productivity initially increases with gradient length for both the short and the long column but subsequently decreases as gradients are still longer. For the short column the maximum productivity is obtained at a shorter gradient time, while the long column provides higher productivity for relatively long gradient times. This is the consequence of the higher peak capacity for the long column and the smaller relative contribution of

the re-equilibration time.

3.2 Mass Spectrometry

Figure 2 shows the mass spectra of β-casein phosphopeptide FQpSEEQQQTEDELQDK (*m/z* 1031.41 (2+), 687.94 (3+)), peak 8 from the chromatograms in figure 1. It is clear that the use of a longer column improves peptide separation, resulting in cleaner mass spectra with less interfering ions. The influence of peptide separation on protein identification was assessed by Mascot database searching. Data files from the duplicate MS/MS experiments were submitted to MS/MS ion search using the Mascot search engine. Search results from duplicate runs were combined and the number of identified peptides and total sequence coverage were calculated. The results are shown in table 3. Mascot scores are generally 2-4 times higher for the long column, when comparing analyses with equal gradient slope. For the shortest gradients (3-min/15-min gradient pair) however, there is more variation in the gain factor. This is probably the result of the limited separation efficiency for the 3-min gradient on the short column. The higher

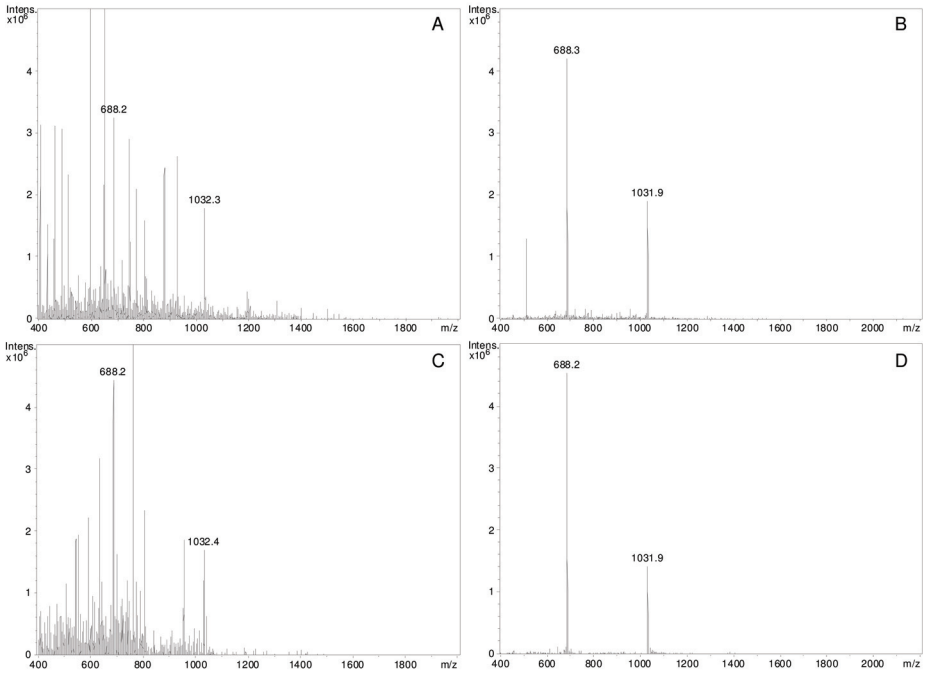

Figure 2

Averaged Mass Spectra of β-Casein phosphopeptide FQpSEEQQQTEDELQDK (m/z 1031.41 (2+), 687.94 (3+)) as identified from Extracted Ion Chromatograms in the LC-MS analysis of a tryptic digest mixture. Separation on capillary silica monolithic columns of 150 mm (A,C) and 750 mm (B,D) x 0.1 mm ID, using a gradient of 5-55% B (90% ACN + 0.08% FA, 0.01% TFA) in solvent A (2% ACN + 0.1% FA, 0.01% TFA). A: 9-min gradient, B: 45-min gradient, C: 45-min gradient, D: 225-min gradient.

Table 3: Peptides used for peak capacity calculations

t_G	Protein	150 x 0.1 mm			750 x 0.1 mm		
		Score[a]	Peptides[b]	Coverage[c]	Score[a]	Peptides[b]	Coverage[c]
3	BSA	111	6	11%	-	-	-
	α-S1-Casein	174	6	27%			
	α-S2-Casein	not identified					
	β-Casein	77	2	10%			
9	BSA	300	11	21%	800	17	26%
	α-S1-Casein	328	7	34%	518	10	43%
	α-S2-Casein	64	2	10%	183	8	27%
	β-Casein	108	2	10%	192	5	20%
15	BSA	247	9	19%	769	22	35%
	α-S1-Casein	399	9	37%	483	11	46%
	α-S2-Casein	68	2	10%	144	5	22%
	β-Casein	131	3	13%	83	5	13%
45	BSA	399	13	27%	986	19	32%
	α-S1-Casein	279	8	37%	570	9	42%
	α-S2-Casein	not identified			203	5	22%
	β-Casein	79	2	10%	268	6	18%
75	BSA	574	17	34%	1010	22	33%
	α-S1-Casein	352	8	36%	715	17	56%
	α-S2-Casein	46	1	5%	230	6	26%
	β-Casein	103	2	10%	313	7	22%
225	BSA	-	-	-	918	17	26%
	α-S1-Casein				546	9	37%
	α-S2-Casein				134	4	18%
	β-Casein				194	4	16%

[a] *Average Mascot score (duplicate runs)*

[b] *Total number of unique peptides identified (duplicate runs)*

[c] *% of amino acid sequence covered by identified peptides*

scores for the analyses on the long column are the result of a higher number of identified peptides and a higher average score per peptide (43.6 ± 8.9 for the short column and 52.3 ± 8.8 for the long column). The higher number of identified peptides also leads to considerably higher sequence coverage for the analyses on the long column.

The four proteins in the sample were identified in all (duplicate) analyses on the long column, while α-S2-casein was not identified in several analyses for the short column. In two cases (3-min and 45-min gradients) α-S2-casein was not identified in both duplicate runs. This is probably a result of the fact that this protein is a minor component (~20%)

in α-casein and is therefore present in relatively low amounts compared to the other proteins.

The peak capacity increases for longer gradients on both columns and this is mostly accompanied by higher Mascot scores and more identified peptides. However, the Mascot scores and sequence coverage clearly decrease for the longest gradient on the long column. This can be explained by the limited complexity of the sample. The digest mixture used in these experiments contains about 100 peptides while the 75-min gradient on the long column already results in a peak capacity of 135. The observed increase in peak capacity for the 225-min gradient does not yield any further significant improvement in protein identification for this sample. A longer gradient time gives broader peaks, lower peak concentrations and fewer identified peptides. Similar behaviour has also been observed when analysing less complex samples. A simple BSA tryptic digest (about 55 peptides) gives the highest Mascot score and sequence coverage using a 15 min gradient on a 750 mm column and this decreases for longer gradient times [23]. The data presented here show that the optimum gradient length for protein identification increases for more complex samples.

4. CONCLUSIONS

The choice of the LC system is very important for the analysis of protein digests. The quality of protein identification by mass spectrometry is highly dependent on the LC separation. As is demonstrated in this paper, relatively long columns are superior to normal (150-250 mm) columns for the analysis of digests of protein mixtures. The longer column gives less band broadening at longer gradient times and a relatively high increase in peak capacity for analyses with equal gradient steepness. The enhanced separation efficiency yields less co-eluting peptides and cleaner mass spectra, which increases protein identification scores and the number of identified peptides. Productivity is similar for both columns, with a maximum at different gradient lengths for the two columns. For short columns a higher productivity was obtained at relatively short gradient times. Long columns provide higher productivity at the longer gradient times, which are preferred for the analysis of complex digest mixtures. Generally, both analysis time and separation power should be considered for the selection of the optimum column.

Monolithic materials are an interesting alternative for improvement of peptide separation with respect to columns packed with smaller particles, which require special equipment. The advantages of monolithic materials are the high permeability and the resulting low backpressure. This allows the use of relatively long columns for complex mixtures or the use of high flow rates for fast analysis, as has also been demonstrated for protein digests [26, 27].

ACKNOWLEDGEMENTS

This work was supported by The Netherlands Proteomics Center (http://www.netherlandsproteomicscenter.nl). The authors would like to thank Mr. M. Kurano of ATAS GL International for generously providing the 750 mm monolithic capillary column.

REFERENCES

[1] J. Dai, C.H. Shieh, Q. Sheng, H. Zhou, R. Zeng, *Anal. Chem.* **77** (2005) *5793*.

[2] C. Delahunty, J.R. Yates III, *Methods* **35** (2005) *248*.

[3] M.P. Washburn, D. Wolters, J.R. Yates III, *Nat. Biotechnol.* **19** (2001) *242*.

[4] G.J. Opiteck, J.W. Jorgenson, R.J. Anderegg, *Anal. Chem.* **69** (1997) *2283*.

[5] P.J. Boersema, N. Divecha, A.J.R. Heck, S. Mohammed, *J Proteome Res.* **6** (2007) *937*.

[6] J.E. MacNair, K.D. Patel, J.W. Jorgenson, *Anal. Chem.* **71** (1999) *700*.

[7] Y. Shen, R.D. Smith, K.K. Unger, D. Kumar, D. Lubda, *Anal. Chem.* **77** (2005) *6692*.

[8] L. Tolley, J.W. Jorgenson, M.A. Moseley, *Anal. Chem.* **73** (2001) *2985*.

[9] Y. Shen, R. Zhao, M.E. Belov, T.P. Conrads, G.A. Anderson, K. Tang, L. Pasa-Tolic, T.D. Veenstra, M.S. Lipton, H.R. Udseth, R.D. Smith, *Anal. Chem.* **73** (2001) *1766*.

[10] Y. Shen, R. Zhao, S.J. Berger, G.A. Anderson, N. Rodriguez, R.D. Smith, *Anal. Chem.* **74** (2002) *4235*.

[11] T. Bamba, E. Fukusaki, Y. Nakazawa, A. Kobayashi, *J Sep. Sci.* **27** (2004) *293*.

[12] J. Bones, C. Duffy, M. Macka, B. Paull, *Analyst* **133** (2008) *180*.

[13] T. Ikegami, E. Dicks, H. Kobayashi, H. Morisaka, D. Tokuda, K. Cabrera, K. Hosoya, N. Tanaka, *J Sep. Sci.* **27** (2004) *1292*.

[14] V.V. Tolstikov, A. Lommen, K. Nakanishi, N. Tanaka, O. Fiehn, *Anal. Chem.* **75** (2003) *6737*.

[15] Q. Luo, Y. Shen, K.K. Hixson, R. Zhao, F. Yang, R.J. Moore, H.M. Mottaz, R.D. Smith, *Anal. Chem.* **77** (2005) *5028*.

[16] L. Rieux, D. Lubda, H.A.G. Niederlander, E. Verpoorte, R. Bischoff, *J. Chromatogr. A* **1120** (2006) *165*.

[17] J.C. Giddings, *Anal. Chem.* **39** (1967) *1027*.

[18] C.G. Horvath, S.R. Lipski, *Anal. Chem.* **39** (1967) *1893*.

[19] M. Gilar, A.E. Daly, M. Kele, U.D. Neue, J.C. Gebler, *J. Chromatogr. A* **1061** (2004) *183*.

[20] J.K. Eng, A.L. McCormack, I. Yates, John R., *J. Am. Soc. Mass. Spectrom.* **5** (1994) *976*.

[21] D.N. Perkins, D.J.C. Pappin, D.M. Creasy, J.S. Cottrell, *Electrophoresis* **20** (1999) *3551*.

[22] A. Schmidt, M. Karas, T. Dulcks, *J. Am. Soc. Mass. Spectrom.* **14** (2003) *492*.

[23] M.H.M. Van De Meent, G.J. De Jong, *Anal. Bioanal. Chem.* **388** (2007) *195*.

[24] X. Wang, W.E. Barber, P.W. Carr, *J. Chromatogr. A* **1107** (2006) *139*.

[25] B. Barroso, D. Lubda, R. Bischoff, *J Proteome Res* **2** (2003) *633*.

[26] J. Rozenbrand, W.P.v. Bennekom, K.K. Unger, G.J. De Jong, *Anal. Bioanal. Chem.* **385** (2006) *1055*.

[27] V. Guryča, S. Kieffer-Jaquinod, J. Garin, C.D. Masselon, *Anal. Bioanal. Chem.* (2008) *in press (DOI: 10.1007/s00216-008-2433-x)*

Chapter 5

POTENTIAL OF POLY(STYRENE-DIVINYLBENZENE) MONOLITHIC COLUMNS FOR THE ANALYSIS OF PROTEIN DIGESTS

M.H.M. van de Meent, S. Eeltink and G.J. de Jong

submitted (Journal of Seperation Science)

ABSTRACT

Two polystyrene-based capillary monolithic columns of different length (50 and 250 mm) were used to evaluate the effects of column length on gradient separation of protein digests. A tryptic digest of a 9-protein mixture was used as a test sample. Peak capacities were determined from selected extracted ion chromatograms and MS/MS data were used for database matching using the MASCOT search engine. Peak capacities and protein identification scores were higher for the long column with all gradients. Peak capacities appear to approach a plateau for longer gradient times, maximum peak capacity was estimated to be 294 for the short column and 370 for the long column. Analyses with similar gradient slope produced a ratio of the peak capacities of 3.36 for the long and the short column, which is slightly higher than the expected value of the square root of the column length ratio. The use of longer PS-DVB monolith improves peptide separation, as reflected by higher peak capacity, and also increases protein identification, as observed from higher identification scores and a larger number of identified peptides.

1. INTRODUCTION

Identification of proteins by mass spectrometry of proteolytic digests is an important tool in proteomics research. In order to obtain a maximum amount of data from a digest sample, a separation of the digest prior to introduction into the mass spectrometer is necessary. Due to its high resolution and ease of coupling with (nano)electrospray ionisation-mass spectrometry, reversed-phase LC is the method of choice for the separation of peptide mixtures. Columns packed with C18-bonded silica particles are most widely used for peptide separation but organic polymer based materials, like styrene/divinylbenzene copolymers (PS-DVB), are also employed.

Attempts to increase the throughput of protein identification from LC-MS experiments can be broadly divided into two categories. The first strategy is improving the separation efficiency of the LC column by either increasing column length [1,2] or reducing the particle size of the column packing material [3-5]. These approaches offer high efficiency separations but suffer from high backpressure, which necessitates either the use of low flow rates, leading to long run times, or the use of special equipment compatible with the high pressure. The second approach for increase of the throughput is using high-porosity materials which allow very fast analyses. Monoliths are a highly promising type of column materials, possessing a bimodal pore structure with large throughpores and shallow mesopores. These unique properties result in high permeability (low column backpressure) and enhanced mass transfer (improved chromatographic efficiency). This allows the use of either high flow rates for fast separations or relatively long columns combined with conventional LC apparatus. Moreover, the high mass transfer makes monolithic columns well suited for the analysis of macromolecules.

Monolithic columns are available in several stationary phases. The first (polymeric) monoliths were synthesised in the early nineties, first as methacrylate-based membranes [6], later also as methacrylate- and polystyrene-based rods [7,8]. In 1996 a silica-based monolith was synthesised for use in HPLC [9]. Due to the absence of stagnant mesopores in the polymer backbone, these stationary phases typically offer high-efficiency peptide separations (no inter-particle mass transfer). The efficiency can be optimized further by tuning the morphology, while covalent attachment of the monolith against the fused-silica capillary wall ensures high robustness and eliminates channeling effects. While the advantages of using longer columns have been proven for silica-based monoliths [10-14], the use of polymer-based monoliths has been limited to the application of short (5-6 cm) PS-DVB based columns for fast separations of relatively simple mixtures [15-17].

Evaluation of the performance of gradient separations is usually carried out by comparison of peak capacities (*PC*). Peak capacity was first defined by Giddings [18]

as the maximum number of peaks that will fit within the applied elution window with a resolution of 1.0. The concept was adapted for gradient chromatography by Horváth and Lipsky [19]. Since peak width is about constant throughout a gradient separation, the experimental peak capacity can be calculated using Equation 1:

$$PC = 1 + \frac{t_G}{w_{av}} \quad (1)$$

where t_G is the gradient time and is the average peak width (4σ). For large values of *PC*, this approaches t_G / w_{av}. When comparing gradient separations using columns of different length, one should also compare different gradients. A theoretical description of peak capacity is given in Equation 2 [20]:

$$PC = 1 + \frac{\sqrt{N}}{4} \frac{S\Delta\varphi}{S\Delta\varphi(t_0/t_G) + 1} \quad (2)$$

where *N* is the column plate number, *S* is the slope of the plot of the natural logarithm of the retention factor versus solvent composition and Δφ is the change in volume fraction organic modifier during the gradient. When stationary phase and gradient composition are constant, *PC* is only dependent on t_0 / t_G and $\sqrt{N}$ (column length). Thus the effect of column length can be evaluated when the ratio t_0 / t_G is kept constant, i.e. by scaling t_G to the column length. When comparing columns in this way, the peak capacity ratio should be close to the square root of the column length ratio [20].

An alternative means of evaluating the efficiency of a digest separation is by the identification of the proteins using database search engines like SEQUEST [21] or MASCOT (www.matrixscience.com) [22]. The number of identified peptides and the degree of protein sequence coverage are an indication of separation performance. Co-elution of peptides can lead to mutual ion suppression and the loss of sequence information when too many peptides co-elute [23].

In this paper, capillary poly(styrene-divinylbenzene) monolithic columns of two different lengths are compared for the LC-MS analysis of a tryptic digest of a 9-protein mixture. The digest contains approximately 175 peptides with a mass higher than 500 Da, including several peptides larger than 2500 Da and a number of potential phosphorylation sites. PS-DVB monoliths of 50 x 0.2 mm and 250 x 0.2 mm were investigated using gradients between 5 and 300 min. Column performance was evaluated with regard to peak capacity, peak production rate (PPR, peaks/min) and MASCOT identification scores.

2 EXPERIMENTAL

2.1 Materials and reagents

Cytochrome c (bovine), serum albumin (bov.), β-lactoglobulin A (bov.), carbonic anhydrase (bov.), lysozyme (chicken), myoglobin (horse), ribonuclease A (bov.), α-lactalbumin (bov.), trypsin (porcine, type IX-S, EC 3.4.21.4) and 1,4-dithiotreitol (DTT) were purchased from Sigma (St. Louis, MO, USA). Catalase (bov.), MS grade formic acid (FA) and acetonitrile (ACN), iodoacetamide (IAA) and ammonium hydrogen carbonate were obtained from Fluka (Buchs, Switzerland). All solutions were prepared using water from a Synergy UV water purifying system (Millipore, Bedford, MA, USA).

All reagents for the digestion of the protein mixture were prepared in 200 mM NH_4HCO_3 buffer (pH 8) unless stated otherwise. Protein stocks were prepared in water at 5.0 mg/ml for BSA and 2.0 mg/ml for all other proteins, resulting in concentrations as listed in Table 1. A protein mixture was prepared by mixing equal amounts of the stock solutions and adding an extra aliquot of water to achieve a 10-fold dilution. The mixture was digested as follows: 100 µl of the protein solution was set to pH 8 by addition of 25 µl of a 1 M NH_4HCO_3 buffer (pH 8). After addition of 25 µl of a 10 mM DTT solution the sample was incubated at 50 °C for 30 min to reduce disulfide bonds. After cooling to room temperature 25 µl of a 30 mM IAA solution was added and the sample was incubated in the dark for 60 min to alkylate the free thiols. Trypsin was dissolved in 10 µl of NH_4HCO_3 buffer to obtain a trypsin to protein mass ratio of 1:50 in the final solution and the trypsin solution was added to the sample, which was then incubated overnight (16 h) at 37 °C. The digestion was stopped by the addition of 15 µl of 50% FA and the digest was stored at -20°C until further use. Before injection, the digest was diluted 5 times by addition of mobile phase A (water + 0.5% FA), resulting in a 100-fold dilution of the original protein stock concentrations.

Table 1: Composition of the protein stock solutions

Protein	*Concentration (µM)*
Cytochrome c	161.8
Bovine serum albumin	75.2
β-Lactoglobulin A	108.9
Carbonic anhydrase	68.9
Catalase	34.7
Lysozyme	139.8
Myoglobin	118.0
Ribonuclease A	146.2
α-Lactalbumin	141.0

2.2 Apparatus and LC columns

All analyses were performed using an Agilent 1100 nanoLC system (Waldbronn, Germany), consisting of a vacuum degasser, a binary Nano-Pump, a μ-well plate sampler and a thermostatted column compartment, which was maintained at 60°C throughout the experiments. Detection was performed using an Agilent LC/MSD trap XCT-ultra (Waldbronn, Germany) ion-trap mass spectrometer, equipped with an orthogonal ESI interface. PS-DVB monolithic capillary columns were obtained from Dionex Corporation (Amsterdam, The Netherlands), the short column was 50 mm x 200 μm id and the long column 250 mm x 200 μm id.

2.3 Method and data analysis

LC solvent A was water + 0.5% formic acid (v/v), solvent B was water/ACN (1:1) + 0.5% FA (v/v). After injection of 1.0 μl of the digest, the sample was trapped on the top of the column and the gradient was started. All separations were performed in duplicate in full-scan mode for peak capacity determination and a third time in MS fragmentation mode for protein identification. The gradient was run from 2-62% solvent B (1-31% ACN) with gradient times varying from 5-300 min at a flow rate of 3.0 μl/min, resulting in a backpressure of approximately 40 bar on the 50 mm column and 130 bar on the 250 mm column. An electrospray voltage of -4.00 kV was used for all experiments, MS spectra were acquired at an *m/z* range of 400-2000 and MS/MS-fragment spectra were acquired over a 100-2200 *m/z* range using the peptide fragmentation settings of the mass spectrometer.

Peak capacities were calculated according to Eq. 1. The average peak width at 13.4% of the peak height (4σ) was calculated using peak widths determined from extracted ion chromatograms of ten peptides from different parts of the gradient, representing seven of the nine proteins in the mixture. The peptide masses used are shown in Table 2. The

Table 2: Peptides used for peak capacity calculations

Peptide	*m/z*	*Protein*
EDLIAYLK	482.77 (2+)	Cytochrome c
TGQAPGFSYTDANK	728.84 (2+)	Cytochrome c
LVNELTEFAK	582.32 (2+)	BSA
HLVDEPQNLIK	653.36 (2+)	BSA
LSFNPTQLEEQCHI	858.40 (2+)	β-Lactoglobulin A
VLDALDSIK	487.28 (2+)	Carbonic anhydrase II
GTDVQAWIR	523.27 (2+)	Lysozyme
FESNFNTQATNR	714.83 (2+)	Lysozyme
VEADIAGHGQEVLIR	803.93 (2+)	Myoglobin
LDQWLCEK	546.26 (2+)	α-Lactalbumin

effect of separation efficiency on protein identification was evaluated using the MS/MS ions search feature of the MASCOT search engine (www.matrixscience.com) [22]. LC-MS/MS data files were converted to MASCOT generic format (.mgf) using the XCTs data analysis software. The peak lists were searched against the SwissProt database for tryptic peptides from all Chordata in the database, allowing two missed cleavages per peptide and containing carbamidomethyl cysteine as a fixed modification and Met-oxidation and Ser/Thr/Tyr-phosphorylation as variable modifications. Peptide mass tolerance was set to ±1.2 Da, MS/MS tolerance to ± 0.9 Da.

3. RESULTS AND DISCUSSION

3.1 Chromatographic efficiency

Choice of the right gradient conditions is an important step in optimisation of the peak capacity. Therefore it is necessary to gain insight in the influence of gradient slope on peak capacity for different column lengths. Base peak chromatograms for the 30 min gradient on both columns and the 150 min gradient on the long column are shown in Figures 1A-C. The peak capacities for these analyses are 79 and 99 for the short gradient on the short and long columns, respectively, and 266 for the long gradient. Comparison of peak capacities for the short (30 min gradient) and long (150 min gradient) columns results in a peak capacity ratio of 3.36, which is about 50% higher than the expected value of √5. This number is comparable to results obtained in the comparison of C18-silica monoliths of different length [14].

Higher peak capacities were obtained for the long column for all gradient lengths. *PC* increases rapidly for short gradients and appears to level off, approaching a maximum at longer gradient times (Figure 2). This is related to the observed linearity between peak width (4σ) and gradient time over the gradient range used in these experiments (Figure 3). The smaller peak widths obtained for the long column, result in higher peak capacities. The peak capacity appears to approach a plateau at long gradient times, this PC_{max} can be estimated under the assumption that w_{av} vs. t_G remains linear at very long gradient times. Where w_{av} vs. t_G is linear, the relationship is described by the following equation:

$$w_{av} = a \cdot t_G + b \tag{3}$$

where *a* is the slope and *b* represents the y-intercept. Combining this with Equation 1 gives

$$PC = \frac{t_G}{a \cdot t_G + b} \tag{4}$$

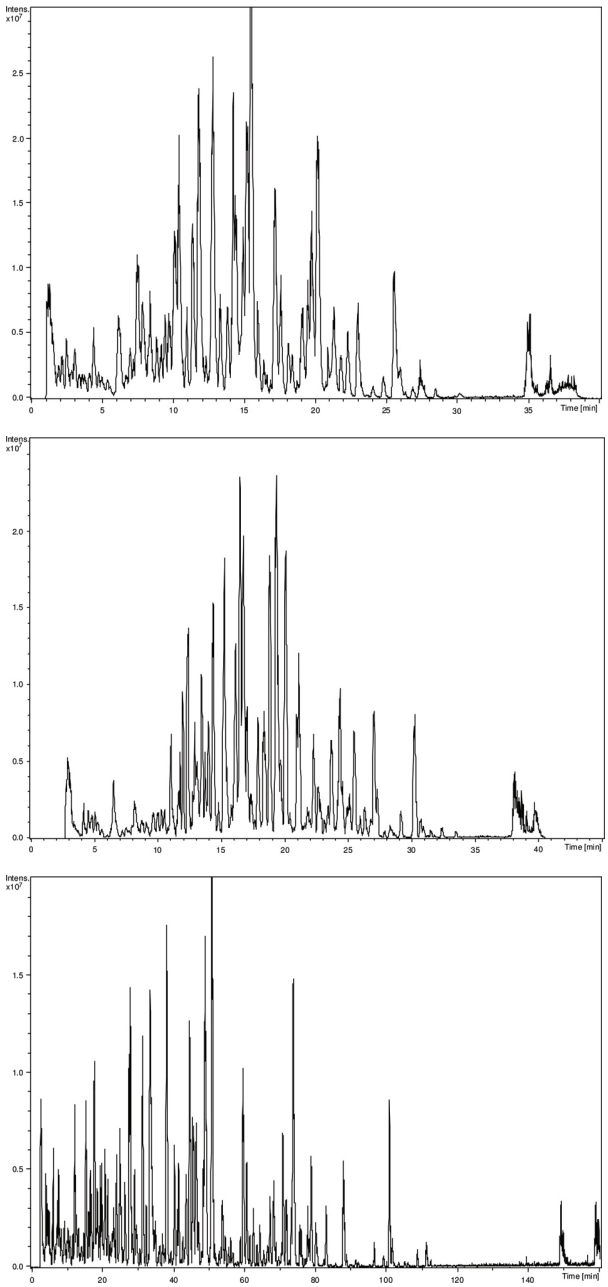

Figure 1

Base peak chromatograms of a 9-protein digest mixture, separated on PS-DVB monolithic columns of 50 x 0.2 mm ID (A) and 250 x 0.2 mm ID (B and C), using a gradient of 1-31% ACN in water (both + 0.5% FA). A and B: 30 min gradient, C: 150 min gradient.

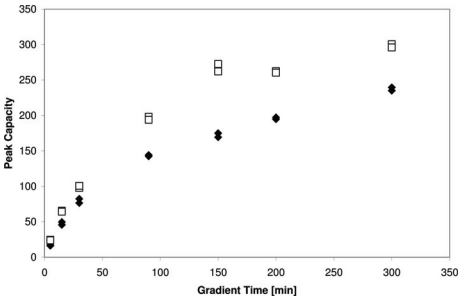

Figure 2

Peak capacity vs. gradient time for long and short PS-DVB columns, 1-31% ACN gradient in water (both + 0.5% FA). (♦) 50 x 0.2 mm ID column, (□) 250 x 0.2 mm ID column.

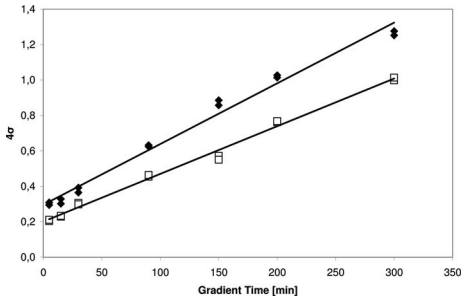

Figure 3

Peak width (4σ) vs. gradient time for long and short PS-DVB columns, 1-31% ACN gradient in water (both + 0.5% FA). (♦) 50 x 0.2 mm ID column, (□) 250 x 0.2 mm ID column.

For very long gradient times, b becomes negligible compared to $a \cdot t_G$ and thus

$$PC_{\max} \approx \frac{t_G}{a \cdot t_G} = \frac{1}{a} \qquad (5)$$

Therefore, PC_{max} for a given separation system can be estimated from the slope of the w_{av} vs. t_G plot (Fig. 3). Maximum peak capacities for the two columns are 294 and 370 for the short and long column, respectively.

Direct comparison of the performance of polystyrene and silica-based monoliths is difficult since column lengths usually differ, the two materials display different retention properties and are typically run under different temperatures (ambient temperature for C18-silica, 60°C for PS-DVB). However, there does not seem to be any indication that C18-silica monoliths outperform the PS-DVB columns, as suggested in literature [24]. While the peak capacities obtained for the C18-silica monolith in ref. 24 are similar to those we found earlier [14], the authors find much lower values for the PS-DVB column. A peak capacity of 65 was reported for a gradient of 0.25% ACN min^{-1}, whereas the peak capacity at such a gradient would be around 170 in our system. The difference in peak capacities could be the result of using different temperatures. The peak capacity in ref. 24 was obtained at ambient temperature, while our experiments were carried out at 60°C. A higher temperature will increase separation efficiency as it reduces the resistance to mass transfer through higher diffusivity of the analytes. This is in accordance with observations in literature [25,26]. However, pilot experiments using a BSA tryptic digest and UV detection indicated that increasing the temperature from ambient to 60°C will only yield a gain in *PC* of approximately 15% (data not shown). A second explanation for the observed difference may be differences in flow rate, they used 300 nl/min for a 100 µm ID column while the flow for our 200 µm ID column was 3

µl/min. It is difficult to predict the exact effect of using a 2.5-fold higher linear flow rate on peak capacity as changing the column temperature might also change the optimum flow rate for the column.

Another way to judge LC performance is by evaluating the number of peaks analysed per unit of time. This is commonly called productivity [27] and is most easily calculated by dividing peak capacity by total analysis time. The relative performance of completely different columns can be judged by comparing the productivity of the columns within a fixed analysis time, e.g. 20 min for a fast analysis and 300 min for an analysis with high peak capacity. The gradient time is easily calculated by subtracting the system dead time (t_0) and the equilibration time (t_{eq}). For longer columns, t_0 and t_{eq} will be longer and thus limit the available gradient time. As monolithic columns contain only shallow mesopores, they require less equilibration time than packed columns. Five times the column volume seems a reasonable estimate. System t_0 was determined to be 1.06 min for the short column and 2.72 min for the long column. Column dead times were 0.42 min for the short column and 2.08 min for the long column and void volumes of 1.2 µl and 6.0 µl, respectively, were calculated from these values. Peak capacities can be calculated from t_G using Eq. 4 and by dividing this peak capacity by the total analysis time the productivity or peak production rate (PPR) is obtained (Table 3). The highest PPR is obtained for fast analysis using the short column. For longer analysis times the relative contribution of t_0 and t_{eq} to the total analysis time is reduced and the PPR of the long column exceeds that of the short column as a result of the higher peak capacity. This effect already becomes apparent after a total analysis time of approximately 30 min where peak capacities for both columns are the same despite the 10 min difference in gradient time.

Table 3: Peak capacity and productivity of digest separation on PS-DVB capillary monolithic columns for different analysis times.

Total analysis time (min)	*L* (mm)	t_0 (min)	t_G (min)	t_{eq} (min)	*PC*	*PC/time* (min^{-1})
20	50	1.1	16.9	2.1	48	2.4
20	250	2.7	6.9	10.4	31	1.6
100	50	1.1	96.9	2.1	155	1.5
100	250	2.7	86.9	10.4	199	2.0
300	50	1.1	296.9	2.1	227	0.8
300	250	2.7	286.9	10.4	294	1.0

3.2 Mass Spectrometry

Datafiles from the LC-MS/MS experiments were searched against the SwissProt database using the MS/MS ions search feature of the MASCOT search engine. Identification scores for both columns using various gradient times are summarised

in Table 4. Both the cumulative score and the number of identified peptides generally increase with increasing gradient length, reaching a plateau at longer gradient times (90 and 150 min) where the peak capacity is higher than the number of peptides in the digest. For all gradient lengths, more peptides are identified using the long column than using the short column, which is expected since peak capacities are also higher for the long column. Mascot scores are also generally higher for the long column and increase with longer gradients. The highest identification scores are obtained for BSA, the largest protein in the mixture, yielding the largest number of peptides after tryptic digestion. The highest sequence coverages are obtained for cytochrome c, 60% for the 30 and 90 min gradients using the long column. This can be explained by the fact that cytochrome c is easily digested using trypsin, with most resulting peptides falling within the 500-2500 Da mass range.

Table 4: Protein identification data

Column	*50 x 0.2 mm*		*250 x 0.2 mm*	
t_G (min)	*Score[a]*	*Peptides[b]*	*Score[a]*	*Peptides[b]*
5	1372	32	1641	37
15	2141	44	2681	55
30	2115	45	2968	65
90	2832	58	3385	78
150	2794	58	3491	78

[a] *cumulative Mascot score over all identified proteins*

[b] *total number of identified peptides*

For all gradients and both columns, a minimum of eight out of the nine proteins has been identified. Ribonuclease A was identified by only one or two peptide matches for the analyses using the long column and not identified at all for 4 out of 5 gradients for the short column. The fact that ribonuclease is only identified in analyses where separation performance is high indicates that only a few ribonuclease peptides are present in the digest. Myoglobin has also not been identified in some of the analyses, this is probably the result of the known resistance of myoglobin to tryptic digestion [28].

4. CONCLUSIONS

Our experiments have shown that monolithic columns are very well suited for the analysis of protein digest-mixtures. We compared both chromatographic efficiency and protein identification for poly(styrene-divinylbenzene) monoliths of different length. As expected, the longer column produces higher peak capacities than the short column, which was also observed in earlier experiments using C18-bonded silica monoliths [13,14]. For the (relatively short) PS-DVB monoliths higher peak capacities have been obtained than for the silica-based columns if comparable gradient times were used, but

this difference is partly attributable to difference in column temperature. As was also observed in our previous experiments, the gain in peak capacity for analyses on both columns with equal gradient steepness (30 min for the short column and 150 min for the long column) is higher than the expected value of √5. This may be explained by a slight difference in average macropore size of the two monolithic columns. The higher peak capacities obtained with the long column are accompanied by better protein identification with higher mascot-scores and a larger number of identified peptides for all gradients.

For short analysis times, the short PS-DVB column achieves a higher peak production rate than the long column. For analysis times above 30 min the higher peak capacity for the long column compensates for the t_G limiting effect of the longer equilibration time. For high throughput it is advisable to employ fast analyses using the 50 mm column, thus exploiting the high peak production rate at short analysis time. For high efficiency separations, where analysis time is less important, the 250 mm column offers a high peak capacity at a higher PPR than the 50 mm column.

ACKNOWLEDGEMENTS

This work was supported by The Netherlands Proteomics Center (http://www.netherlandsproteomicscenter.nl). The authors would like to thank Dionex Corporation for generously providing the PS-DVB monolithic capillary columns.

REFERENCES

[1] Y. Shen, R. Zhao, M.E. Belov, T.P. Conrads, G.A. Anderson, K. Tang, L. Pasa-Tolic, T.D. Veenstra, M.S. Lipton, H.R. Udseth, R.D. Smith, *Anal Chem* **73** (2001) *1766*.

[2] Y. Shen, R. Zhao, S.J. Berger, G.A. Anderson, N. Rodriguez, R.D. Smith, *Anal Chem* **74** (2002) *4235*.

[3] J.E. MacNair, K.D. Patel, J.W. Jorgenson, *Anal Chem* **71** (1999) *700*.

[4] Y. Shen, R.D. Smith, K.K. Unger, D. Kumar, D. Lubda, *Anal Chem* **77** (2005) *6692*.

[5] L. Tolley, J.W. Jorgenson, M.A. Moseley, *Anal Chem* **73** (2001) *2985*.

[6] T.B. Tennikova, M. Bleha, F. Svec, T.V. Almazova, B.G. Belenkii, *J Chromatogr A* **555** (1991) *97*.

[7] F. Svec, J.M.J. Frechet, *Anal Chem* **64** (1992) *820*.

[8] Q.C. Wang, F. Svec, J.M.J. Frechet, *Anal Chem* **65** (1993) *2243*.

[9] H. Minakuchi, K. Nakanishi, N. Soga, N. Ishizuka, N. Tanaka, *Anal Chem* **68** (1996) *3498*.

[10] T. Ikegami, E. Dicks, H. Kobayashi, H. Morisaka, D. Tokuda, K. Cabrera, K. Hosoya, N. Tanaka, *J Sep Sci* **27** (2004) *1292*.

[11] Q. Luo, Y. Shen, K.K. Hixson, R. Zhao, F. Yang, R.J. Moore, H.M. Mottaz, R.D. Smith, *Anal Chem* **77** (2005) *5028*.

[12] V.V. Tolstikov, A. Lommen, K. Nakanishi, N. Tanaka, O. Fiehn, *Anal Chem* **75** (2003) *6737*.

[13] M.H.M. Van De Meent, G.J. De Jong, *Anal Bioanal Chem* **388** (2007) *195*.

[14] M.H.M. Van De Meent, G.J. De Jong, *J Sep Sci* **32** (2009) *487*.

[15] M. Batycka, N.F. Inglis, K. Cook, A. Adam, D. Fraser-Pitt, D.G.E. Smith, L. Main, A. Lubben, B.M. Kessler, *Rapid Commun Mass Spectrom* **20** (2006) *2074*.

[16] A. Premstaller, H. Oberacher, W. Walcher, A.M. Timperio, L. Zolla, J. Chervet, N. Cavusoglu, A.v. Dorsselaer, C.G. Huber, *Anal Chem* **73** (2001) *2390*.

[17] W. Walcher, H. Toll, A. Ingendoh, C.G. Huber, *J Chromatogr A* **1053** (2004) *107*.

[18] J.C. Giddings, *Anal Chem* **39** (1967) *1027*.

[19] C.G. Horvath, S.R. Lipski, *Anal Chem* **39** (1967) *1893*.

[20] X. Wang, W.E. Barber, P.W. Carr, *J Chromatogr A* **1107** (2006) *139*.

[21] J.K. Eng, A.L. McCormack, I. Yates, John R., *J Am Soc Mass Spectrom* **5** (1994) *976*.

[22] D.N. Perkins, D.J.C. Pappin, D.M. Creasy, J.S. Cottrell, *Electrophoresis* **20** (1999) *3551*.

[23] A. Schmidt, M. Karas, T. Dulcks, *J Am Soc Mass Spectrom* **14** (2003) *492*.

[24] V. Guryča, S. Kieffer-Jaquinod, J. Garin, C.D. Masselon, *Anal Bioanal Chem* **392** (2008) *1291*.

[25] X. Wang, D.R. Stoll, A.P. Schellinger, P.W. Carr, *Anal Chem* **78** (2006) *3406*.

[26] S. Eeltink, S. Dolman, R. Swart, M. Ursem, P.J. Schoenmakers, *J Chromatogr A* **1216** (2009) *7368*.

[27] M. Gilar, A.E. Daly, M. Kele, U.D. Neue, J.C. Gebler, *J Chromatogr A* **1061** (2004) *183*.

[28] J.R. Freije, P.P.M.F.A. Mulder, W. Werkman, L. Rieux, H.A.G. Niederlander, E. Verpoorte, R. Bischoff, *J Proteome Res* **4** (2005) *1805*.

INTEGRATED SYSTEMS FOR PROTEIN ANALYSIS

Chapter 6

COUPLING OF A CATION-EXCHANGE PRECOLUMN WITH AN ON-LINE IMMOBILISED TRYPSIN-LC-MS SYSTEM FOR TRAPPING AND IDENTIFICATION OF PROTEINS

M.H.M. van de Meent, G. Wiese and G.J. de Jong

submitted (Analytical and Bioanalytical Chemistry)

Ch6

ABSTRACT

A system was developed for preconcentration of proteins on a strong cation exchange (SCX) precolumn, followed by on-line tryptic digestion and protein identification. Several buffers were evaluated for both desorption of proteins from the SCX column and digestion by a poroszyme immobilised enzyme reactor (IMER, 40 x 0.25 mm ID). Ammonium acetate buffer (0.5 M, pH 8) was found to be the optimal buffer for the desorption of both acidic and basic proteins followed by on-line digestion. The IMER was found to produce digests sufficient for protein identification after only 60 seconds digestion in 10% acetonitrile. Trapping and desorption of β-lactoglobulin A on the SCX trap was shown to be linear for both injection volume (0.25-5.0 µl) and concentration (10-2000 ng/µl). Analysis of β-lactoglobulin A, cytochrome c and BSA resulted in positive identification of the proteins with sequence coverages of 24%, 48% and 9%, respectively. The total system was used to trap, digest and identify proteins in approximately 45 min.

1. INTRODUCTION

Ion-exchange chromatography (IEC) is a powerful tool for the separation of peptides and proteins due to its reproducibility and high resolution. These days, the use of IEC in proteomics is strongly focused on the application of strong cation exchange (SCX) columns in bottom-up proteomics [1-3]. In these systems SCX is employed to fractionate a complex peptide mixture prior to reversed phase (RP) separation and (tandem-)mass spectrometric (MS) detection. IEC is also used for protein fractionation prior to SDS-PAGE [4] or RPLC [5,6] or for the removal of high abundant proteins [7]. Reversed phase columns are most often applied for protein trapping [8] as desalting is usually a major objective. However, ion-exchange materials are also very suitable for trapping and preconcentration of polypeptides [9].

Peptides and proteins, due to their zwitterionic nature, can be trapped or separated using both cation-exchange and anion-exchange materials. Elution of both proteins and peptides from an ion-exchange resin can be accomplished in two ways. The first and most widely used elution mode is employing either a linear or stepwise gradient of increasing salt concentration. A second method is gradually changing the pH of the eluent, this is commonly referred to as chromatofocusing [10,11]. This method makes use of the zwitterionic properties of the analytes, which desorb from the ion-exchanger when the pH of the eluent buffer increases (cation exchange) or decreases (anion exchange) beyond the pI of the analyte. The applications of chromatofocusing as a chromatographic technique are limited by the difficulty of generating suitable pH-gradients [12].

Tryptic digestion of proteins is probably the most widely employed tool in proteomics. Despite the availability of other proteolytic enzymes, each with their own advantages and disadvantages, trypsin is still the most widely used proteolytic enzyme. In recent years, much effort has been put into making tryptic digestion faster and more efficient [13-15]. Traditional digestion protocols, which use an in-solution digestion step, are hampered by trypsin autolysis yielding fragments that can interfere during MS analysis. In order to reduce autolysis, trypsin concentrations need to be kept low, typically at 1-2% of protein weight, which leads to low reaction rates and long incubation times [16]. Immobilisation of trypsin onto a solid support is the preferred strategy to reduce autolysis and increase the trypsin-to-protein ratio during digestion. Digestion times in the range of minutes [17] or even seconds [15,18] have been achieved with satisfactory results but some groups also employ a stop-flow method, stopping the protein-plug in the immobilised enzyme reactor (IMER) for as long as 30 min [19]. Immobilised trypsin materials are commercially available from several manufacturers with trypsin immobilised on supports like sepharose or polystyrene beads. Alternatively, many groups also employ home-made reactors using particles, monoliths or the inner wall of

a fused silica capillary as a support for immobilisation [20-23].

In this paper several buffers are compared for preconcentration and desorption of proteins using an SCX trapping column as well as on-line digestion of the proteins by an immobilised trypsin reactor. Reaction times between 0.5 and 10 min were evaluated to optimise the reaction time. The SCX-IMER set-up has been coupled on-line to RPLC-MS for evaluation of digestion efficiency and protein identification.

2. EXPERIMENTAL

2.1 Materials and reagents

BSA, β-lactoglobulin A, cytochrome c and trypsin (porcine, type IX-S, EC 3.4.21.4) were purchased from Sigma (St.Louis, MO, USA), MS grade formic acid (FA) and ammonium hydrogen carbonate from Fluka (Buchs, Switzerland). Acetonitrile (ACN) was HPLC gradient-grade (Fisher Scientific, Loughborough, UK), Tris and biochemistry grade trifluoroacetic acid (TFA) were obtained from Acros Organics (Geel, Belgium). Ammonium acetate was obtained from Merck (Darmstadt, Germany) and calcium chloride was purchased from JT Baker (Deventer, The Netherlands). All solutions were prepared using water from a SynergyUV water purifying system (Millipore, Bedford, MA, USA).

The buffers used for desorption and digestion were prepared as follows: for the Tris buffer a 1.0 M Tris solution was set to pH 8 using hydrochloric acid. The resulting buffer was diluted to 0.05 M and calcium chloride was added to 0.01 M concentration. For the Phosphate buffer a solution of 0.2 M Na_2HPO_4 was set to pH 8 by addition of 0.2 M NaH_2PO_4 (total Na^+ concentration: ca. 385 mM). For the ammonium acetate buffers a 0.5 M ammonium acetate solution was set to pH 8 using 10% ammonia solution and this buffer was diluted 2.5-fold to prepare the 0.2 M buffer.

2.2 Apparatus and methods

2.2.1 System design

The SCX-IMER-RPLC system (figure 1) consisted of an Agilent 1100 nanoLC system (Agilent technologies, Waldbronn, Germany) containing a vacuum degasser, a binary Nano-Pump, a μ-well plate sampler and a thermostatted column compartment with two separately operable heating zones. Trapping and desorption were performed using Shimadzu LC-10ADvp LC pumps (Shimadzu, Kyoto, Japan): pump A 0.5% (v/v) formic acid, pump B 0.5 M ammonium acetate, both at 5 μl/min. The SCX trap was prepared by packing POROS® S 10 particles (Applied Biosystems, Nieuwerkerk a/d IJssel, The Netherlands) into a 0.25 mm ID fused silica capillary. A 40 mm piece of fused silica was cut off and fitted inside a PEEK sleeve between two Inline MicroFilter assemblies with 0.5 μm PEEK frits (M-520, Upchrurch Scientific, Oak Harbor, WA, USA). The IMER

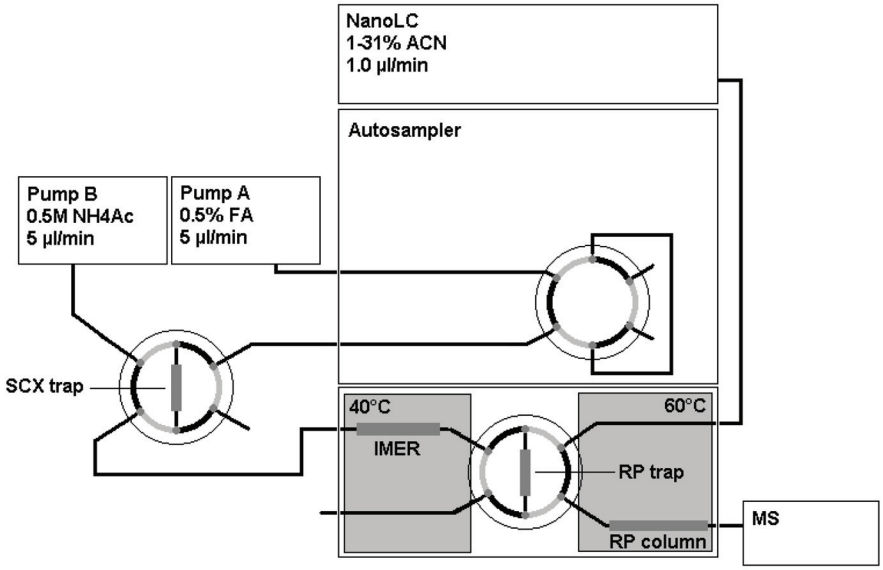

Figure 1

Set-up of the SCX-IMER-RPLC-MS system. SCX trap: 40 x 0.25 mm ID; IMER: 100 x 0.25 mm ID; PS-DVB trap: 5 x 0.2 mm ID; PS-DVB column: 50 x 0.1 mm ID. Protein was injected onto the SCX trap using pump A, after 5 min the protein was desorbed, digested and washed to the PS-DVB trap using pump B. After another 10 min the digest was analysed using a 1-31% acetonitrile gradient containing 0.5% (v/v) FA.

was prepared in the same way and it was cut at a length of 100 mm and fitted with separate PEEK sleeves at both ends. The material used for the IMER was Poroszyme® immobilized Trypsin (Applied Biosystems, Nieuwerkerk a/d IJssel, The Netherlands). The RPLC system consisted of a 5 x 0.2 mm polystyrene-divinylbenzene (PS-DVB) monolithic trap and a 50 x 0.1 mm PS-DVB monolithic column (Dionex, Amsterdam, The Netherlands). Sub-sections of this system were used for optimisation of desorption and digestion buffers as well as evaluation of the SCX trap.

2.2.2 SCX-RPLC

SCX-RP experiments were performed using the system described above but without the IMER and using a 50 x 0.1 mm ID PS-DVB monolithic column (Dionex, Amsterdam, The Netherlands). Analytes were detected using an MU701 UV-Vis detector (GL Sciences Inc., Tokyo, Japan) at 215 nm.

The SCX trap was evaluated by trapping between 10-2000 ng β-lactoglobulin A in 0.25-5.00 µl injection volume, both at constant concentration and constant volume. The protein samples were injected onto the SCX column from the autosampler and after

5 min the trap was switched in-line with pump B for desorption of the protein. The protein was retained on the PS-DVB precolumn and after another 5 min the PS-DVB trap was switched in-line with the RP column, followed by a 10 min gradient of 20-50% B (A - water + 0.1% TFA, B - ACN + 0.08% TFA) at 0.75 µl/min.

2.2.3 SCX-IMER

For evaluation of the optimal buffer for both desorption and digestion an SCX-IMER system was tested using only the two LC-10ADvp pumps, a Rheodyne type 7010 manual injection valve and a UV detector (ABI 785A, Applied Biosystems, Nieuwerkerk a/d IJssel, The Netherlands). TrisHCl, sodium phosphate, ammonium hydrogen carbonate and ammonium acetate buffers were tested at various flow rates for desorption and digestion properties. Digests were collected and analysed off-line by RPLC-MS.

The LC-MS system consisted of an Agilent 1100 nanoLC system as described above, with a 5 x 0.2 mm PS-DVB monolithic trap and a 50 x 0.1 mm PS-DVB monolithic column (Dionex, Amsterdam, The Netherlands). Digests were diluted four times prior to injection to prevent overloading the column and analysed using a 15 min gradient of 5-35% B in A (A - water + 0.5% FA, B - ACN + 0.5% FA) at 0.6 µl/min with the column thermostat set to 60°C. Detection was performed using an Agilent LC/MSD Trap XCT (Agilent Technologies, Waldbronn, Germany) ion trap mass spectrometer equipped with an orthogonal nanoESI interface. The nano-electrospray voltage was -2.75 kV, mass spectra were acquired over an *m/z* range of 400-2000 and MS/MS-fragmentation spectra over a 100-2200 range using the peptide fragmentation settings of the mass spectrometer.

2.2.4 SCX-IMER-RPLC

Contrary to the experiments described in section 2.2.3, the IMER was now fitted inside the thermostatted column compartment, which allowed digestion at elevated temperatures. Several proteins (β-lactoglobulin, cytochrome c, BSA) were analysed using 0.5 M ammonium acetate as desorption/digestion buffer. The proteins were injected onto the SCX trap at 5 µl/min, after 5 min the 6-port valve was switched and the proteins were desorbed and washed through the IMER to the RP trap at 3 µl/min. After 10 min the RP trap was switched in line with the RP column and the digest was analysed using a 1-31% ACN gradient (15 min) at 1.0 µl/min and 60°C.

3. RESULTS

3.1 SCX

The proteins were loaded onto the SCX column in water containing 0.5% (v/v) formic acid (pH 2.3). At this pH most proteins have a net positive charge and are trapped on the SCX column. Preliminary experiments monitoring the breakthrough from

loading the SCX column showed that the column effluent after injection of 1.0 µl of 400 ng/µl β-lactoglobulin A or cytochrome c did not differ from blank injections. Further experiments proved that the proteins are retained on the SCX column even after washing with loading solvent for 10 min at 5 µl/min. The proteins can be desorbed rapidly after changing the eluent. Quantification of the proteins was difficult due to baseline fluctuation resulting from the eluent changing but comparison of multiple sequential elution steps showed that the second desorption step was not distinguishable from a blank injection. Four different buffers of pH 8 were used to test the desorption of two proteins from the SCX trap and the results are summarised in Table 1. The acidic protein β-lactoglobulin A (pI 5.1) was desorbed using all four buffers while the basic protein cytochrome c (pI 10-10.5) was only detected in systems employing the phosphate buffer and 0.5 M ammonium acetate. At pH 8 cytochrome c still has an overall positive charge and is therefore only desorbed by the buffers with a relatively high concentration of cations: the 0.2 M phosphate buffer, which contains 385 mM Na^+ and 0.5 M ammonium acetate, containing 500 mM NH_4^+.

Table 1: Desorption of β-Lactoglobulin A and cytochrome c from the SCX trap

Buffer (pH 8)	*β-Lactoglobulin A*	*Cytochrome C*
Tris, 0.05 M + $CaCl_2$, 0.01 M	+[a]	-[b]
Phosphate buffer, 0.20 M	+	+
Ammonium Acetate, 0.20 M	+	-
Ammonium Acetate, 0.50 M	+	+

[a] *Desorption considered complete when second desorption step was indistinguishable from blank*

[b] *Not distinguishable from blank*

The loadability of the SCX trap was evaluated by trapping β-Lactoglobulin A in various injection volumes and concentrations. Figure 2a shows the linear relationship between the amount injected and peak area for LC-UV analysis of the desorbed protein. The linearity was found to be good after injection of different concentrations (10-2000 ng in 1.0 µl) and volumes (0.25-5.0 µl for 200 ng/µl) of the protein. However, the chromatograms showed a considerable increase in peak width for larger amounts of protein (Figure 2b), suggesting that the PS-DVB column was being overloaded when injecting more than 100 ng (5.4 pmol) of protein. This corresponds with earlier experiments, in which the maximum loading capacity for proteins using a 60 x 0.2 mm ID PS-DVB monolithic column was found to be in the order of 0.5 pmol [24].

3.2 SCX-IMER

A Poroszyme® immobilized Trypsin reactor was tested for digestion of bovine cytochrome c at room temperature using a 0.05 M Tris buffer containing 0.01 M $CaCl_2$ and set to pH 8 for maximum proteolytic activity. Various flow rates and various concentrations of acetonitrile were evaluated. Reactor efficiency was expressed as the

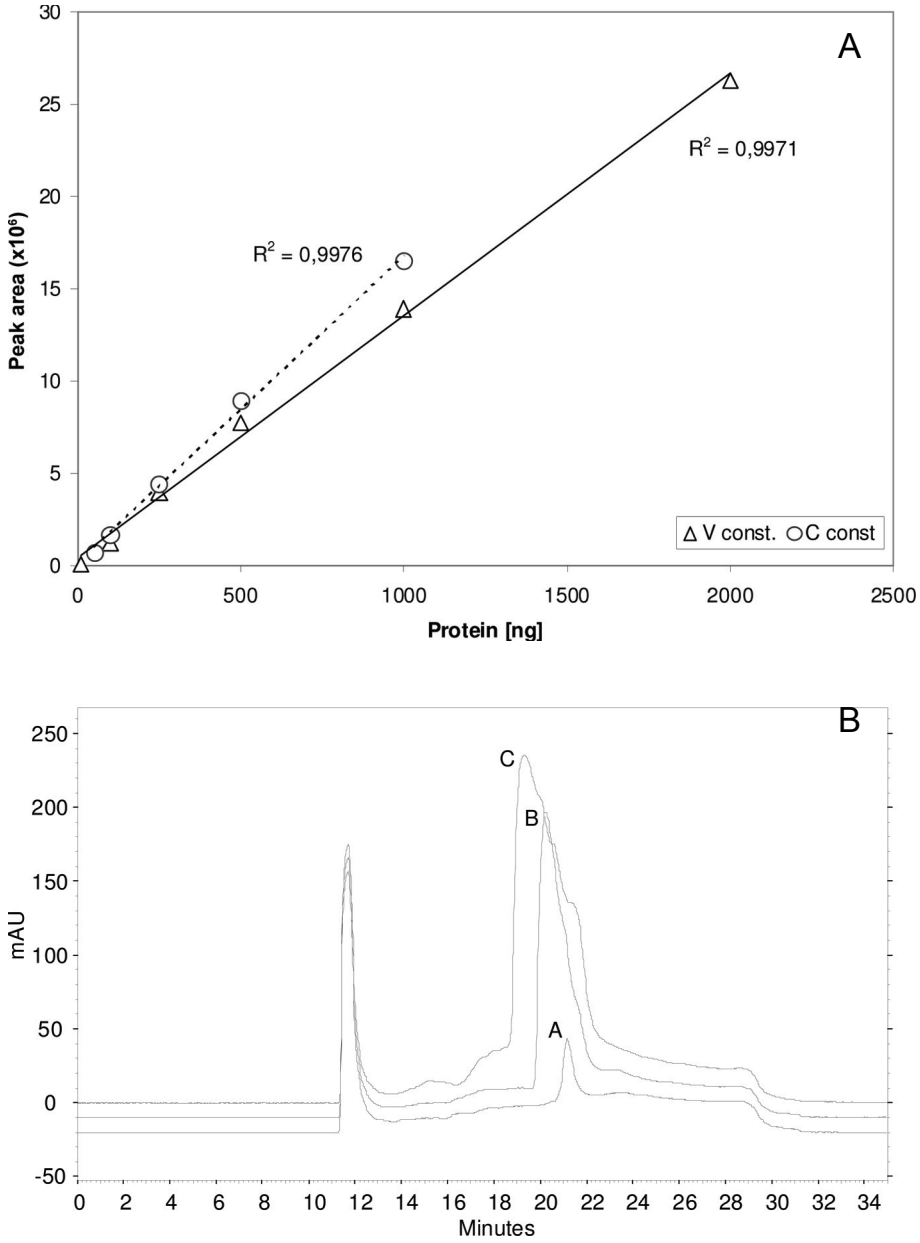

Figure 2

A: Linearity of the amount of injected β-lactoglobulin A versus peak area for constant injection volume (Δ) or constant sample concentration (○).

B: Overlaid chromatograms for increasing amounts of injected β-lactoglobulin A for a constant injection volume: (a) 100 ng, (b) 1000 ng and (c) 2000 ng protein injected.

amount of protein remaining in the digest relative to the amount injected. Digestion rate increases considerably for higher acetonitrile concentrations, as is visible from the absence of intact protein in the high-ACN (>40%) digests after 30 s digestion (figure 3). In order to obtain the same degree of digestion using a buffer without ACN the digestion time needs to be extended to about 10 minutes (data not shown).

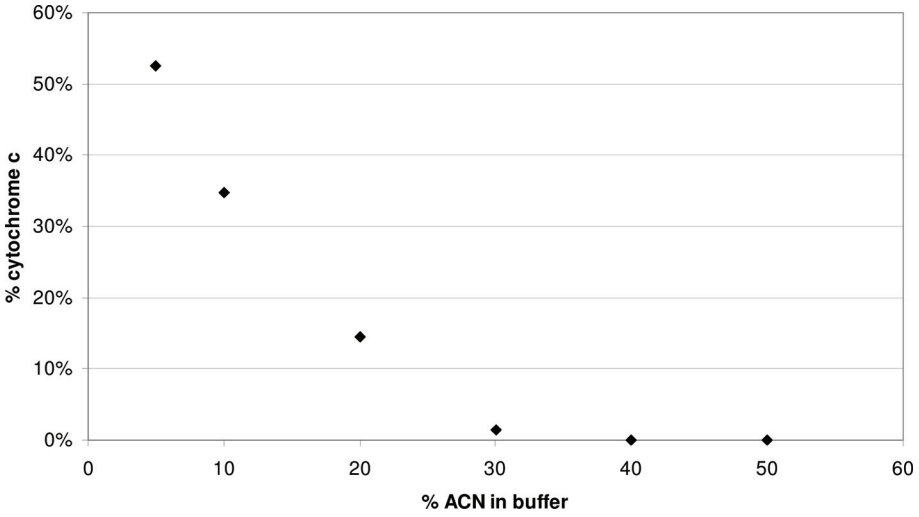

Figure 3

Percentage of remaining cytochrome c after 30 s digestion at various percentages of acetonitrile.

However, as expected the increased amount of ACN in the buffer also influences the trapping of the digest on an RP trapping column. Experiments using an on-line IMER-RPLC-UV digestion system showed that retention of both the peptides and the intact protein on a C18-silica trap was strongly reduced for buffers containing 30% ACN or higher. Only 2 peptides from a cytochrome c digest are observed at 30% ACN, compared to 7 peptides at 20% ACN, 13 peptides at 10% and 16 peptides at 5%. For 10% ACN almost the same number of peptides is retained as for 5% ACN but the amount of intact protein after 30 s digestion time is much lower (35% vs. 53% of the amount injected). Therefore digestion buffer containing 10% ACN was used in the SCX-IMER system.

The SCX-IMER system was used to evaluate some buffers for both preconcentration and digestion of proteins. The system was tested at two different flow rates using β-Lactoglobulin A as a test protein since this was desorbed by all digestion buffers. Digests were collected and analysed by LC-MS/MS. The MS/MS data were searched against the online Swiss-Prot database using Mascot MS/MS ions search (www.matrixscience.com) [25] and expressed as protein sequence coverage (Table 2). The digest profiles are clearly different for the different buffers while the differences between two flow rates for the same buffer are small. The ammonium acetate buffer

Table 2: Mascot identification results for on-line β-lactoglobulin A digestion using three different buffers and two digestion times

Peptide	*Tris*		*AmAc*		*AmCarb*	
	30 s	*60 s*	*30 s*	*60 s*	*30 s*	*60 s*
VLVLDTDYK		x[a]				
VLVLDTDYKK	x	x				
IPAVFKIDALNENK	x	x		x		
TPEVDDEALEKFDK	x	x	x	x		
LSFNPTQLEEQCHI				x		x
TKIPAVFKIDALNENK			x	x		
TPEVDDEALEKFDKALK			x	x		
VYVEELKPTPEGDLEILLQK	x	x	x	x	x	
VAGTWYSLAMAASDISLLDAQSAPLR			x			
Sequence coverage	*32%*	*32%*	*44%*	*37%*	*11%*	*7%*

[a] *x - identified by MASCOT database search*

produces the highest number of miscleaved peptides, averaging 1 missed cleavage per identified peptide. However, it also yields the highest number of peptides (6 and 5 peptides at 60 and 30 seconds reaction time, respectively) adding up to the highest sequence coverage. Tris buffer provides a more complete digestion but fewer identified peptides (5 and 4) and less protein sequence information. The ammonium hydrogen carbonate buffer produces only a single peptide identification at both flow rates, albeit without missed cleavages. Using the ammonium acetate and Tris buffers, the IMER digestion provides similar sequence information as the commonly used in-solution digestion with ammonium hydrogen carbonate buffer (5 peptides, 33% coverage) with a major reduction in digestion time (15 h vs. <1 min).

3.3 SCX-IMER-LC-MS system

The 0.5M ammonium acetate buffer was further evaluated in an on-line SCX-IMER-LC-MS system for three different proteins. The test proteins (200 ng/µl) were injected onto the SCX trap and after 5 min the proteins were desorbed via the IMER to an RP trap using the desorption/digestion buffer. The flow rate through the IMER was reduced to 3 µl/min (60 s reaction time) to avoid excessive backpressure due to the PS-DVB trap. The digestion temperature was set at 55°C to increase diffusion of the protein to the immobilised trypsin to increase the rate of proteolysis. The digests were separated using a PS-DVB monolithic column and detected by MS/MS. The MS/MS data were searched against the Swiss-Prot database using Mascot and data are summarised in Table 3. All three test proteins, cytochrome c (12.4 kDa, pI 9.6), β-lactoglobulin A (18.4 kDa, pI 5.1) and bovine serum albumin (66.4 kDa, pI 5.6), were identified from MS/MS data. Sequence coverage and number of identified peptides for β-lactoglobulin were lower than for the SCX-IMER experiments with off-line LC-MS analysis. This is probably the result of a

Table 3: Mascot data for on-line SCX-IMER-LC-MS/MS analysis of proteins

Protein	*Score*	*Peptides*	*Coverage*
β-Lactoglobulin A	105	3	24%
Cytochrome C	331	10	48%
BSA	264	6	9%

higher ACN content during peptide trapping in the on-line system. The digests in the off-line analysis were diluted 4 times prior to injection to prevent overloading of the RP column, reducing the ACN concentration from 10% to 2.5%. The higher ACN content in the on-line system has likely caused breakthrough of relatively polar peptides. A second parameter influencing peptide identification may be the LC flow rate. For off-line LC-MS analysis the attainable LC flow rate was limited due to high backpressure and this lower flow rate may result in increased ionisation efficiency. The LC-MS base peak chromatogram for the cytochrome c analysis is shown in figure 4, while the MS and MS/MS spectra for the identified peptide GITWGEETLMEYLENPK, indicated by the arrow in fig. 4, are shown in figure 5a and 5b, respectively.

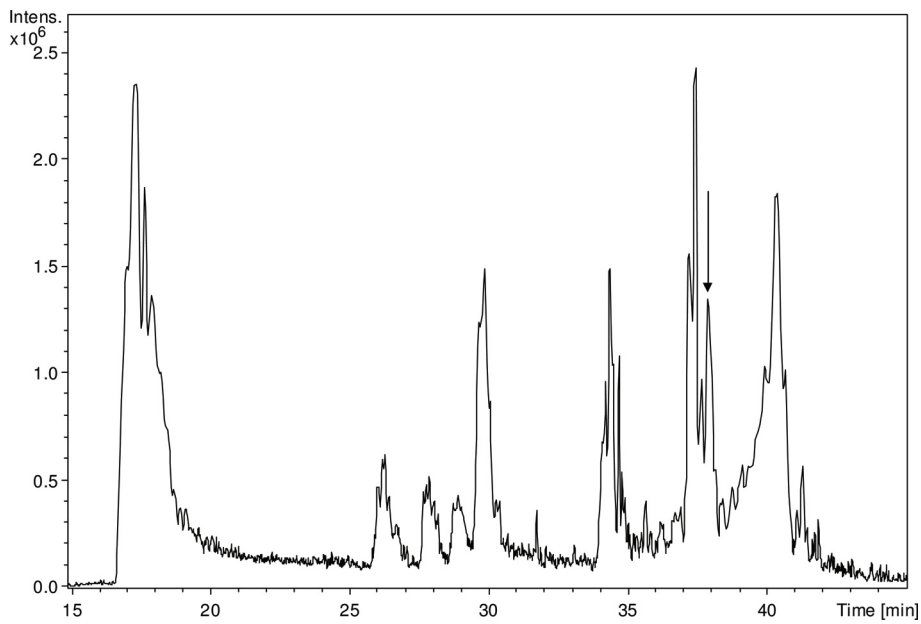

Figure 4

Base peak chromatogram of bovine cytochrome c using the SCX-IMER-RPLC-MS system; 0-5 min - injection (water + 0.5% FA), 5-15 min - digestion (0.5M NH_4Ac, pH8), 15-35 min - 1-31% ACN gradient (water/ACN + 0.5% FA). Mass spectra of the peak indicated by (↓) are shown in figure 5.

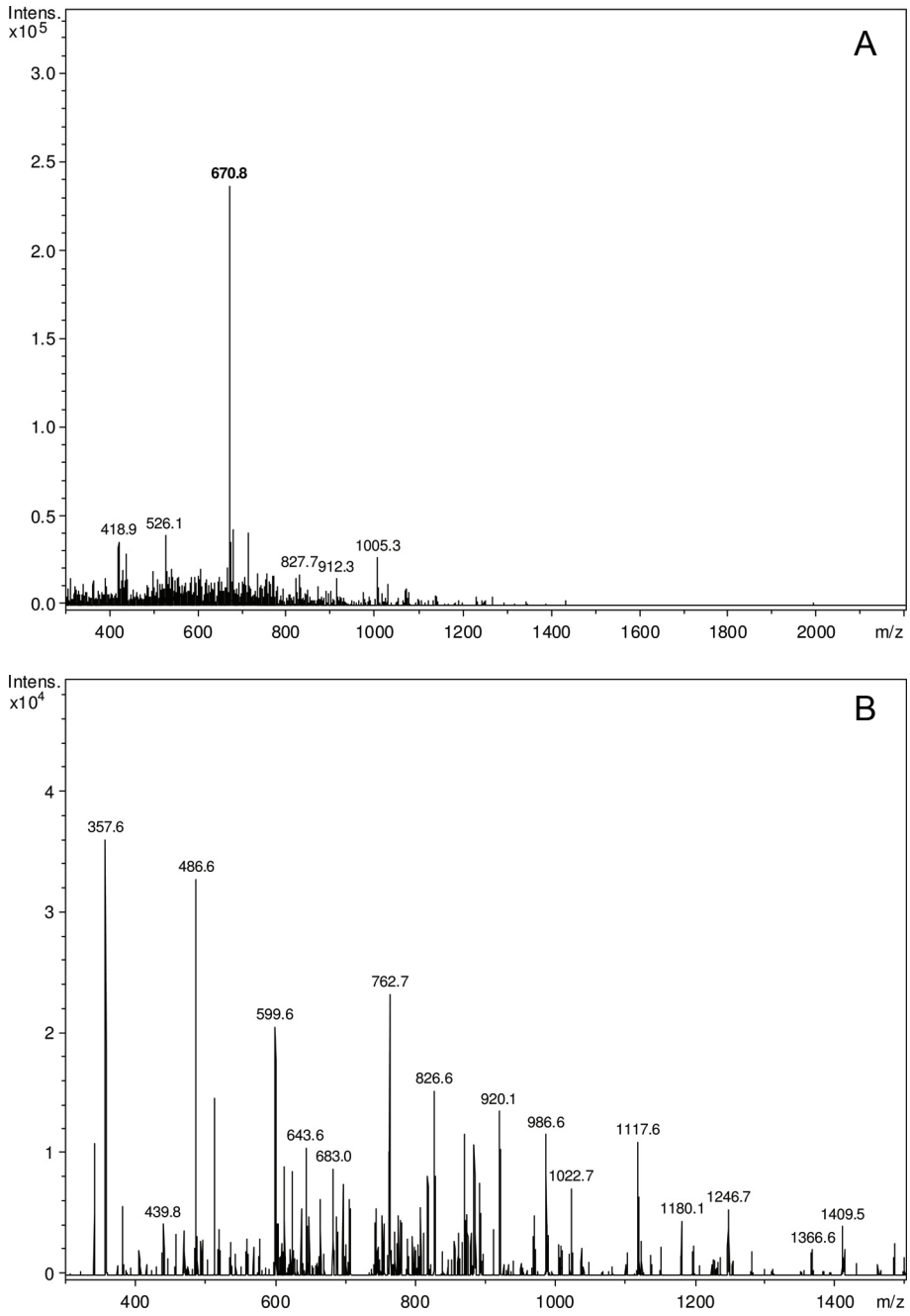

Figure 5

A: MS spectrum of peptide GITWGEETLMEYLENPK (m/z 670.8 - MH_3^{3+})

B: MS/MS spectrum of peptide GITWGEETLMEYLENPK

4. CONCLUSION

Some proteins with different pI and molecular mass were trapped and preconcentrated using a strong cation-exchange precolumn. A change in eluent pH from 2 to 8 was sufficient to desorb acidic proteins from the SCX trapping column. For desorption of basic proteins, a relatively high cation concentration is required. 0.5 M ammonium acetate buffer was found to be the optimal buffer for combining desorption and on-line digestion as it desorbs both acidic and basic proteins and was also shown to be as effective for digestion as the standard Tris buffer.

The SCX-IMER system was succesfully coupled to RPLC-MS/MS for peptide analysis and an integrated system for trapping, digestion and identification of proteins was developed. The test proteins β-Lactoglobulin A, cytochrome c and BSA were successfully identified using the SCX-IMER-RPLC-MS/MS system in a total analysis time of 45 minutes.

This system seems promising for preconcentration and clean-up combined with tryptic digestion and LC-MS based protein identification. It may be used to capture and identify proteins from various matrices, like samples containing detergents, urea or reagents for e.g. reduction and alkylation of disulfide bonds. The analysis of biological samples containing high concentrations of inorganic ions will be more difficult as these ions will interfere with retention of the proteins on the SCX column. Similar systems, based on anion-exchange for preconcentration and/or another type of enzyme in the IMER also seem very promising.

ACKNOWLEDGEMENTS

This work was supported by The Netherlands Proteomics Center (http://www.netherlandsproteomicscenter.nl).

REFERENCES

[1] Q. Luo, Y. Gu, S. Wu, T. Rejtar, B.L. Karger, *ELECTROPHORESIS* **29** (2008) *1604*.

[2] F. Wang, J. Dong, M. Ye, X. Jiang, R.x. Wu, an, H. Zou, *J Proteome Res* **7** (2008) *306*.

[3] M.P. Washburn, D. Wolters, J.R. Yates III, *Nat Biotechnol* **19** (2001) *242*.

[4] S. Kovac, X. Yang, F. Huang, D. Hixson, D. Josic, *J Chromatogr A* **1194** (2008) *38*.

[5] B.F. Assiddiq, A.P.L. Snijders, P.K. Chong, P.C. Wright, M.J. Dickman, *J Proteome Res* **7** (2008) *2253*.

[6] B.F. Assiddiq, J.C. Williamson, A.P.L. Snijders, K. Cook, M.J. Dickman, *Proteomics* **7** (2007) *3826*.

[7] J.H. Ringrose, W.W. Van Solinge, S. Mohammed, M.C. O'Flaherty, R. Van Wijk, A.J.R. Heck, M. Slijper, *J Proteome Res* **7** (2008) *3060*.

[8] C. Schley, R. Swart, C.G. Huber, *J Chromatogr A* **1136** (2006) *210*.

[9] F. Wang, X. Jiang, S. Feng, R. Tian, X. Jiang, G. Han, H. Liu, M. Ye, H. Zou, *J Chromatogr A* **1171** (2007) *56*.

[10] T. Andersen, M. Pepaj, R. Trones, E. Lundanes, T. Greibrokk, *J Chromatogr A* **1025** (2004) *217*.

[11] B.Q. Tran, M. Pepaj, E. Lundanes, T. Greibrokk, *Chromatographia* **66** (2007) *709*.

[12] L.I. Tsonev, A.G. Hirsh, *J Chromatogr A* **1200** (2008) *166*.

[13] J. Ma, L. Zhang, Z. Liang, W. Zhang, Y. Zhang, *Anal Chim Acta* **632** (2009) *1*.

[14] G. Massolini, E. Calleri, *J Sep Sci* **28** (2005) *7*.

[15] G.W. Slysz, D.C. Schriemer, *Rapid Commun Mass Spectrom* **17** (2003) *1044*.

[16] N. Lundell, T. Schreitmüller, *Anal Biochem* **266** (1999) *31*.

[17] R.K. Blackburn, R.J. Anderegg, *J Am Soc Mass Spectrom* **8** (1997) *483*.

[18] G.W. Slysz, D.C. Schriemer, *Anal Chem* **77** (2005) *1572*.

[19] B.Q. Tran, H. Loftheim, L. Reubsaet, E. Lundanes, T. Greibrokk, *J Sep Sci* **31** (2008) *2913*.

[20] E. Calleri, C. Temporini, E. Perani, C. Stella, S. Rudaz, D. Lubda, G. Mellerio, J.L. Veuthey, G. Caccialanza, G. Massolini, *J Chromatogr A* **1045** (2004) *99*.

[21] D.S. Peterson, T. Rohr, F. Svec, J.M.J. Fréchet, *J Proteome Res* **1** (2002) *563*.

[22] J. Samskog, D. Bylund, S.P. Jacobsson, K.E. Markides, *J Chromatogr A* **998** (2003) *83*.

[23] E. Stigter, G. de Jong, W. van Bennekom, *Anal Bioanal Chem* **389** (2007) *1967*.

[24] A. Premstaller, H. Oberacher, W. Walcher, A.M. Timperio, L. Zolla, J. Chervet, N. Cavusoglu, A.v. Dorsselaer, C.G. Huber, *Anal Chem* **73** (2001) *2390*.

[25] D.N. Perkins, D.J.C. Pappin, D.M. Creasy, J.S. Cottrell, *Electrophoresis* **20** (1999) *3551*.

Chapter 7
DEVELOPMENT OF AN INTEGRATED SYSTEM FOR PROTEIN SEPARATION AND IDENTIFICATION BY COUPLING PROTEIN RPLC WITH AN ENZYME REACTOR AND LC-MS USING A CATION-EXCHANGE TRAPPING COLUMN

M.H.M. van de Meent, G. Wiese and G.J. de Jong

manuscript in preparation

ABSTRACT

An integrated liquid-chromatography based system was developed for the on-line separation, digestion and identification of proteins. A protein mixture was separated using a poly(styrene-divinylbenzene) (PS-DVB) monolithic column and an acidic water-acetonitrile gradient. A fraction was trapped from the column effluent using a strong cation-exchange column. The trapped protein fraction was then eluted using a high-pH ammonium acetate buffer to an immobilised trypsin reactor for proteolysis, after which the peptides were trapped on a PS-DVB precolumn. The digest was analysed by reversed-phase liquid-chromatography mass spectrometry using the PS-DVB column and the protein was identified from MS/MS data. Using this system, 100 ng β-lactoglobulin A was trapped after separation of a 9-protein mixture, digested and identified. The total run time for the experiment was 70 minutes.

1. INTRODUCTION

Proteomics methods can generally be divided into two categories. Firstly, methods where the separati on is performed at the intact protein level and the separated proteins are identified after proteolytic digestion and mass spectrometry (MS). Two-dimensional gel electrophoresis (2D-GE) is now the method of choice for separating complex protein mixtures, proteins are separated according to their isoelectric point and, subsequently according to size on an SDS-polyacrylamide gel. While 2D-GE offers outstanding resolution for separation of intact proteins, limitations like long separation times, small dynamic range and difficulty in analyzing certain protein classes (very small proteins, very large proteins, proteins with extreme pI, hydrophobic proteins) have fuelled the development of alternative methods. Today, GE-based methods are complemented more and more by liquid chromatography (LC), especially for peptide separation after digestion of protein mixtures. Some strategies combine (1D-) gel electrophoresis for protein separation with (multidimensional) chromatographic analysis of digests obtained from excised gel bands. Other methods completely work without separation at the protein level and apply (multidimensional) LC methods to the separation of complex digests. While both 2D-GE and multidimensional LC produce excellent results, they suffer from long analysis times as a result of either long electrophoresis runs or complex gradient schemes and, in both cases, long incubation times for tryptic digestion.

Traditional digestion protocols, which use in-solution digestion, are hampered by trypsin autolysis yielding fragments that interfere during MS analysis. In order to reduce autolysis, trypsin concentrations need to be kept low, typically at 1-2% of protein weight, which leads to low reaction rates and long incubation times [1]. In recent years, much effort has been put into making tryptic digestion faster and more efficient [2-4]. Chemical modification of trypsin molecules has been shown to increase enzymatic activity [5], while immobilisation of trypsin onto solid supports reduces autolysis. This makes it possible to use much higher enzyme/substrate ratios, which has lead to reduction of digestion time to minutes [5,6] or even seconds [4,7]. The use of flow-through enzyme reactors with immobilised trypsin has also made it possible to couple proteolytic digestion directly to chromatographic separation methods [8-11].

While chromatographic separation of peptides is often used in proteomics, chromatography of intact proteins is less common. Ion-exchange and size exclusion chromatography are applied for protein separation because these can be run under non-denaturing conditions. However, neither of these LC modes can match the resolution obtained by 2D-GE. Reversed-phase LC could provide a higher chromatographic resolution but the application of RPLC in protein analysis is not straightforward. Relatively high organic modifier concentrations in RP gradients can lead to protein denaturation and irreversible adsorption on the stationary phase can occur. In order

to keep organic content low, C18-bonded stationary phases are often replaced by less hydrophobic phases (e.g. C4-bonded silica). Polystyrene-based materials are comparable in polarity to C8 materials [12] and are used for protein chromatography both as particulate and monolithic phases [13-15]. Maa and Horváth first described the use of poly(styrene-divinylbenzene) (PS-DVB) particles for the separation of proteins [16] and PS-DVB monoliths have been used for protein chromatography for over a decade [13].

In this paper we describe a two-step system for the rapid capture and identification of proteins from protein mixtures. As a first step, the system runs a chromatographic protein separation followed by trapping of a protein (fraction) of interest on a strong cation-exchange column. The second step consists of desorption of the trapped protein using a high-pH buffer suited for the subsequent on-line tryptic digestion followed by LC-MS/MS analysis and identification of the digested protein fraction.

2. EXPERIMENTAL

2.1 Materials and reagents

Cytochrome C (bovine), serum albumin (bov.), β-lactoglobulin A (bov.), carbonic anhydrase (bov.), lysozyme (chicken), myoglobin (horse), ribonuclease A (bov.) and lactoferrin (bov.) were purchased from Sigma (St.Louis, MO, USA), catalase (bov.), ovalbumin (chicken) and MS grade formic acid (FA) from Fluka (Buchs, Switzerland). Acetonitrile (ACN) was HPLC gradient-grade (Fisher Scientific, Loughborough, UK), biochemistry grade trifluoroacetic acid (TFA) was obtained from Acros Organics (Geel, Belgium). Ammonium acetate was obtained from Merck (Darmstadt, Germany) and calcium chloride was purchased from JT Baker (Deventer, The Netherlands). All solutions were prepared using water from a SynergyUV water purifying system (Millipore, Bedford, MA, USA).

The ammonium acetate buffer was prepared by adjusting the pH of a 0.5M ammonium acetate solution to pH 8 using a 10% ammonia solution.

2.2 Apparatus and columns

Protein separation was optimised using an Agilent 1100 nanoLC system (Agilent technologies, Waldbronn, Germany), consisting of a vacuum degasser, a binary Nano-Pump, a μ-well plate sampler and a thermostatted column compartment. The system was equipped with a 5 x 0.2 mm polystyrene-divinylbenzene (PS-DVB) monolithic trap and a 50 x 0.2 mm PS-DVB monolithic column (Dionex, Amsterdam, The Netherlands). Injection was done from the autosampler using a Shimadzu LC-10Advp LC pump (Shimadzu Benelux, 's Hertogenbosch, The Netherlands) and detection was performed using UV detection. The UV detector was an MU 701 UV-VIS detector (ATAS GL International, Veldhoven, the Netherlands), equipped with an external optical-fiber

flow cell (6 nl, 3mm light path), proteins were detected at 215 nm.

The total analysis system was composed of the same Agilent 1100 nanoLC system as described above. The RPLC system again consisted of a 5 x 0.2 mm polystyrene-divinylbenzene (PS-DVB) monolithic trap and a 50 x 0.2 mm PS-DVB monolithic column (Dionex, Amsterdam, The Netherlands). Desorption and digestion was performed using a Shimadzu LC-10Advp LC pump (Shimadzu). Detection was performed using an LC/MSD Trap XCT (Agilent Technologies, Waldbronn, Germany) ion trap mass spectrometer, equipped with an orthogonal ESI interface.

The SCX trap was prepared by packing POROS® S 10 particles (Applied Biosystems, Nieuwerkerk a/d IJssel, The Netherlands) into a 0.25 mm ID fused silica capillary. A 40 mm piece of fused silica was cut off and fitted inside a PEEK sleeve between two inline microfilter assemblies with 0.5 µm PEEK frits (Upchrurch scientific, Oak Harbor, WA, USA). The IMER was prepared in the same way as described for the SCX trap, but it was cut at a length of 100 mm and fitted with separate PEEK sleeves at both ends. The material used for the IMER was Poroszyme® immobilized Trypsin (Applied Biosystems).

2.3 Protein chromatography

Reversed phase chromatography of proteins using PS-DVB monolithic columns was optimised using an LC-UV system. Several parameters were investigated for their effect on protein chromatography, including flow rate, injection volume, amount injected and mobile phase additives. The protein mixture was analysed using an acidic gradient of 20-50% ACN in water at flow rates between 1.0-2.5 µl/min. Injection volumes varied between 0.2-8.0 µl and TFA and FA were tested in various concentrations and combinations as mobile phase additives. The system was operated at ambient temperature.

2.4 Integrated system design

The experiments were performed in two steps, both using the same Agilent nanoLC system. Both steps are described in figure 1. During step I, 1.0 µl of a 100 ng/µl protein sample was injected directly onto the top of the RPLC column. After 5 min, a 15 min protein gradient of 20-50% ACN was started, at the appropriate time (determined from a test run) the 10-port micro-valve (type V-472-1, Upchurch Scientific) was switched, placing the SCX trap in line with the RPLC column to trap the protein of choice.

After the protein gradient was finished, the SCX trap was placed in a 6-port valve (model 7010, Rheodyne, Rhonert Park, CA, USA) located between the autosampler and the column compartment and phase II was started. The 6-port valve was switched to place the SCX trap in the path of the desorption/digestion buffer flow. After allowing 10 min for desorption, digestion and trapping on the RP trap, the valve in the column

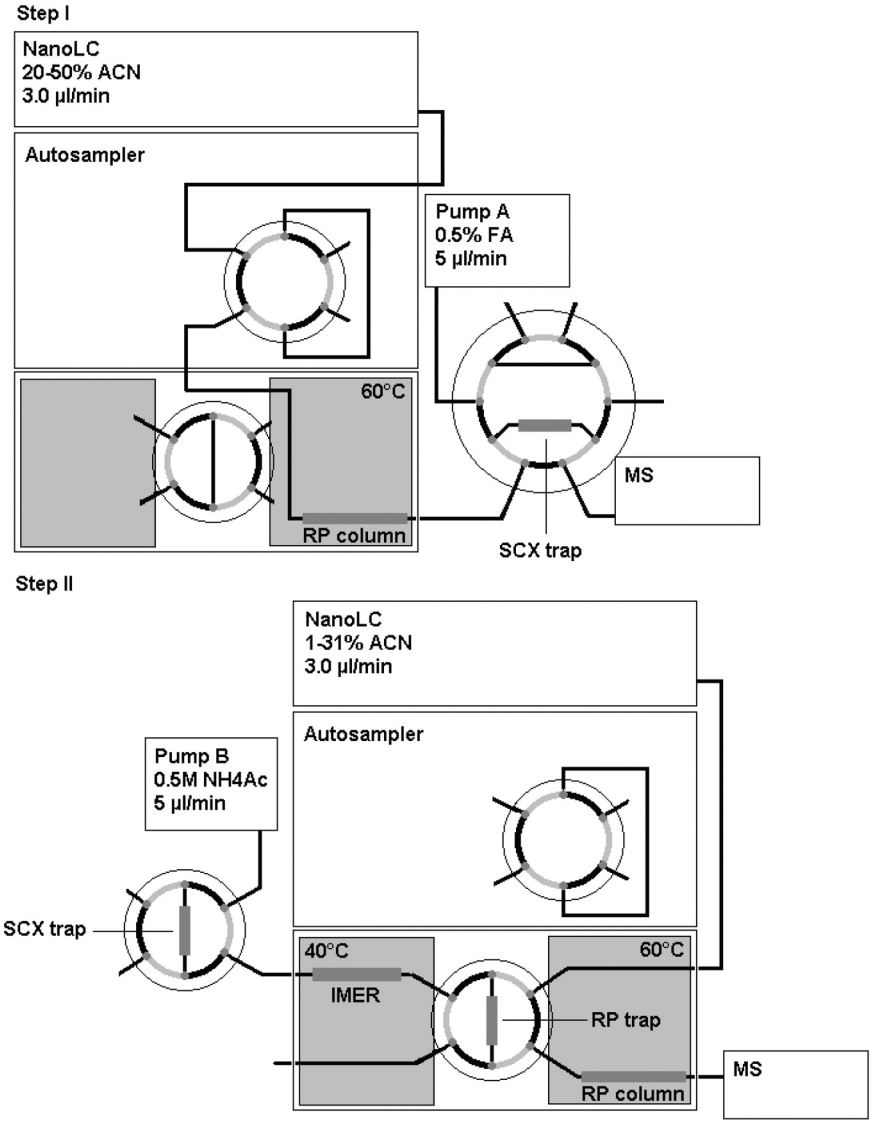

Figure 1

Chromatographic set-up of the system. Column dimensions; SCX trap: 40 x 0.25 mm ID, IMER: 100 x 0.25 mm ID, PS-DVB trap: 5 x 0.2 mm ID, PS-DVB column: 50 x 0.2 mm ID.

compartment was switched and a 15 min 1-31% ACN gradient was started.

During both steps the effluent from the RP column was analysed by mass spectrometry. An electrospray voltage of -4.0 kV was used for ionisation, MS spectra were acquired at an *m/z* range of 400-2000 and MS/MS-fragmentation spectra (step II only) were acquired over a 100-2200 *m/z* range using the peptide fragmentation settings of the mass spectrometer. LC-MS/MS data were searched against the online SwissProt database using Mascot MS/MS ions search (http://www.matrixscience.com) to verify if the target protein could be identified.

3. RESULTS AND DISCUSSION

3.1 System set-up

The SCX-IMER system developed in our group [17] for protein trapping and digestion is ideally suited for coupling with RPLC. The typically low pH of an RP mobile phase (< 3) makes it impossible to transfer the proteins directly to a tryptic digestion system. Moreover, the high ACN concentration in the eluent interferes with trapping of the digest on a peversed-phase trap. Therefore a system is needed to transfer the proteins from the low pH of the protein separation to pH 8, the optimum pH for trypsin activity, while simultaneously removing ACN. As the proteins are net positively charged under RP conditions, this enables trapping of the proteins on the SCX column. Trapping of proteins in various ACN concentrations ranging from 5-50% (v/v) showed no adverse effects of increased ACN content on trapping. As is shown in Figure 2, β-lactoglobulin A can be completely removed from the column effluent. Similar results were obtained for other proteins using UV detection (data not shown). Flushing the SCX column with a high-pH buffer will change the charge of most proteins from positive to negative and desorb them almost instantly from the SCX resin. At the same time the proteins are transferred to a high-pH, low-ACN solvent, well suited for tryptic digestion and RP trapping of the digest, followed by straightforward LC-MS analysis using the same PS-DVB monolithic column as is used for the protein separation.

3.2 Protein separation

Separation of proteins on the PS-DVB column is very efficient as is shown in figure 3, which shows the separation of a 9-protein mixture (20 ng each). Repeatability is demonstrated by determination of retention time RSDs, which range from 0.1% to 0.5% (n=6). Peaks widths are between 42 and 77 sec and are comparable to those reported in literature for analyses performed under similar conditions [18]. Detection limits (S/N=5) are 90 pg to 320 pg for a 1.0 µl injection. Retention time RSD and peak shape were found not to be affected when 4.0 ng per protein was injected in injection volumes ranging from 0.2 µl to 8.0 µl. The relatively small and wide ovalbumin peak (#9) is similar to that published before [14].

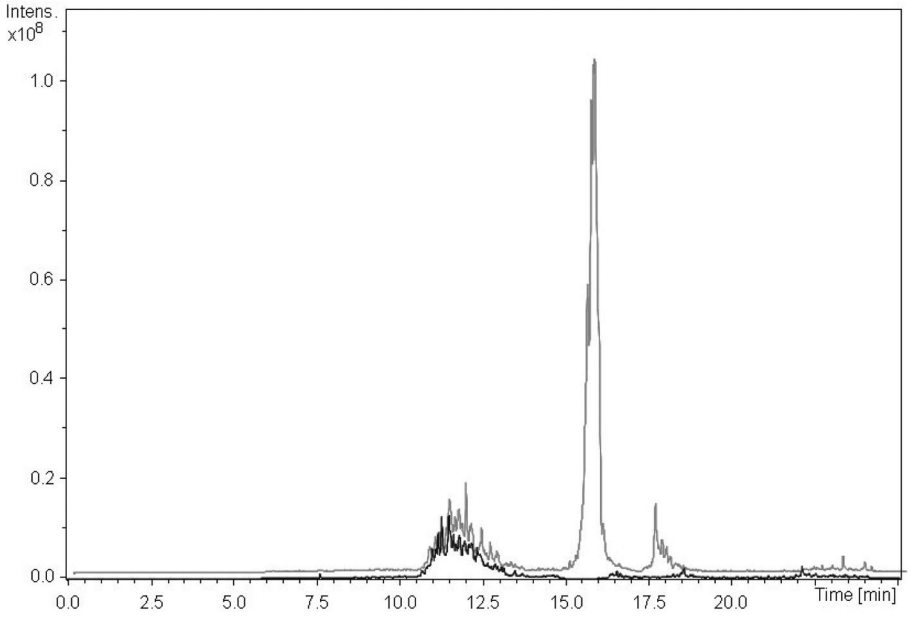

Figure 2

Chromatogram of two consecutive injections of 100 ng β-lactoglobulin in the protein trapping system. Upper trace; no trapping on SCX column, lower trace: protein trapping t= 14-17 minutes.

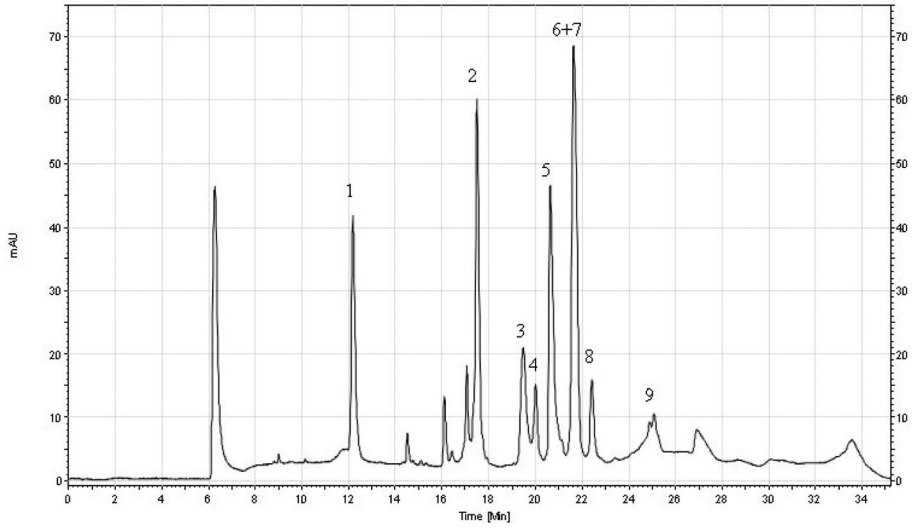

Figure 3

LC-UV chromatogram of the separation of a 9-protein mixture (20 ng each) using a 50 x 0.2 mm ID PS-DVB column. Gradient; 20-50% B in 20 min. Solvent A: water + 0.1% TFA, solvent B: ACN + 0.08% TFA. Peaks; 1 - ribonuclease A, 2 - lysozyme, 3 - BSA, 4 - lactoferrin, 5 - myoglobin, 6 - β-lactoglobulin A, 7 - carbonic anhydrase, 8 - catalase and 9 - ovalbumin.

The trapping of proteins after LC separation was evaluated using the RPLC-SCX-MS set-up (Fig.1, step I). Some of the separation efficiency achieved using LC-UV is lost in this system as a result of extra-column band broadening caused by addition of the SCX trapping set-up and by changing the mobile phase additive from 0.05% TFA to 0.5% FA to reduce ion suppression. The peak width for β-lactoglobulin increased from 48 to 54 sec when changing the mobile phase additive to FA for MS detection and further to 69 sec when adding a valve for the SCX trapping system. Limiting the delay volume between the RP column and the mass spectrometer is essential to minimise band broadening and thus obtain optimal MS data. Therefore all post-column connections were made using 25 μm ID fused silica capillaries and the SCX trapping column was mounted in a 10-port micro-valve with a 44 nl port-to-port volume.

3.3 Protein identification

During the first stage of the procedure, proteins were separated on the PS-DVB column and the target protein was trapped on the SCX column after eluting from the RP column. After trapping of the protein, the SCX column was removed from the 10-port micro-valve and mounted in the 6-port valve. Switching this valve placed the SCX column in the desorption/digestion buffer flow and the identification gradient program was started. Digestion was carried out at 40°C and at a flow rate of 3.0 μl/min, which resulted in a digestion time of approximately 1 min. Both analysis of the protein mixture and the single-protein analysis led to MS/MS-based identification of β-lactoglobulin by a single peptide. Figure 4 shows the base peak chromatogram for the analysis of the digest after trapping of β-lactoglobulin A from the protein mixture. Peaks 1 and 2 represent trypsin autolysis peptides, peak 3 is the identified β-lactoglobulin peptide and peak 4 is intact β-lactoglobulin. No peptides originating from other proteins in the mixture were identified, illustrating the selectivity of the trapping procedure. Carbonic anhydrase, which co-elutes with β-lactoglobulin A when using TFA as a mobile phase additive, was separated from β-lactoglobulin A after switching to FA as additive.

In previous experiments [17], the IMER temperature was kept at 55°C to increase the digestion rate. However, this high temperature was found to severely limit the lifetime of the reactor, therefore the temperature was reduced to 40°C for the present experiments. The fact the some intact protein (about 6% of the amount injected, based on base peak chromatogram peak areas) was found in the digest suggests that digestion efficiency is reduced as a result of the lower digestion temperature. Further optimisation of the digestion conditions is needed to maximise digestion efficiency while maintaining sufficient IMER lifetime. Despite this, it was still possible to make a positive identification after trapping and digesting a protein from the RPLC column effluent.

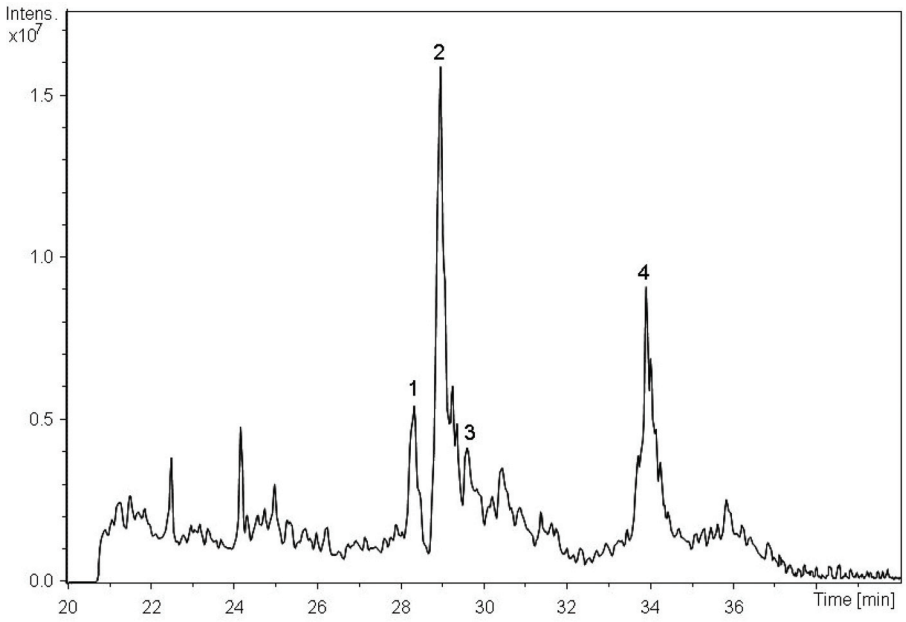

Figure 2

Base peak chromatogram of the tryptic digest of 100 ng β-lactoglobulin A after trapping from a protein mixture, tryptic digestion and digest analysis. Digest analysis: 50 x 0.2 mm ID PS-DVB monolithic column, 3.0 µl/min, 1-31% ACN (15 min).

4. CONCLUSION

We developed a system for protein separation, tryptic digestion and identification of proteins, including trapping of target proteins on an SCX column and digestion by immobilised trypsin. Using this system, proteins were separated, trapped, digested and identified in about 70 min, which is faster than by traditional 2D-GE based methods. An interesting aspect of the present system is that the same monolithic column is used for both protein and peptide separation. However, the analysis time could be limited further by using two separate LC systems with separate columns for protein and peptide analysis. In this way transfer of the SCX column can be avoided and peptide analysis can be started while the protein gradient is still running. By using multiple SCX columns in parallel, multiple protein fractions could be trapped from the RPLC eluent. Moreover, the use of two LC systems would also remove the need for manual handling of the SCX column, thus making the system more suitable for automation.

The integrated system as presented here could be used to complement traditional proteomics methods, e.g. for the analysis of protein classes that fall outside the range of

2D-GE analysis (hydrophobic proteins, proteins with extreme pI). It could also be used for direct identification of proteins that appear to be of interest in chromatographic comparison of protein samples. In combination with a protein prefractionation like 1D-GE or a second chromatographic dimension the system could be a promising alternative to 2D-GE systems.

ACKNOWLEDGEMENT

This work was supported by The Netherlands Proteomics Center (http://www.netherlandsproteomicscenter.nl).

REFERENCES

[1] N. Lundell, T. Schreitmüller, *Anal Biochem* **266** (1999) *31*.

[2] J. Ma, L. Zhang, Z. Liang, W. Zhang, Y. Zhang, *Anal Chim Acta* **632** (2009) *1*.

[3] G. Massolini, E. Calleri, *J Sep Sci* **28** (2005) *7*.

[4] G.W. Slysz, D.C. Schriemer, *Rapid Commun Mass Spectrom* **17** (2003) *1044*.

[5] J.R. Freije, P.P.M.F.A. Mulder, W. Werkman, L. Rieux, H.A.G. Niederlander, E. Verpoorte, R. Bischoff, *J Proteome Res* **4** (2005) *1805*.

[6] R.K. Blackburn, R.J. Anderegg, *J Am Soc Mass Spectrom* **8** (1997) *483*.

[7] G.W. Slysz, D.C. Schriemer, *Anal Chem* **77** (2005) *1572*.

[8] J. Samskog, D. Bylund, S.P. Jacobsson, K.E. Markides, *J Chromatogr A* **998** (2003) *83*.

[9] E. Stigter, G. de Jong, W. van Bennekom, *Anal Bioanal Chem* **389** (2007) *1967*.

[10] H. Yuan, Y. Zhou, L. Zhang, Z. Liang, Y. Zhang, *J Chromatogr A* **1216** (2009) *7478*.

[11] L. Lim, M. Tomatsu, T. Takeuchi, *Anal Bioanal Chem* **386** (2006) *614*.

[12] S.C. Moidoveanu, V. David, in S.C. Moidoveanu, V. David (Editors), *Sample Preparation in Chromatography, Elsevier*, Amsterdam, 2002, *p. 341*.

[13] R.E. Moore, L. Licklider, D. Schumann, T.D. Lee, *Anal Chem* **70** (1998) *4879*.

[14] A. Premstaller, H. Oberacher, W. Walcher, A.M. Timperio, L. Zolla, J. Chervet, N. Cavusoglu, A.v. Dorsselaer, C.G. Huber, *Anal Chem* **73** (2001) *2390*.

[15] W. Walcher, H. Toll, A. Ingendoh, C.G. Huber, *J Chromatogr A* **1053** (2004) *107*.

[16] Y.-f. Maa, C. Horváth, *J Chromatogr A* **445** (1988) *71*.

[17] M.H.M. van de Meent, G.J. de Jong, *Anal Bioanal Chem* (submitted).

[18] C. Schley, R. Swart, C.G. Huber, *J Chromatogr A* **1136** (2006) *210*.

Chapter 8

GENERAL CONCLUSIONS

Ch8

In this thesis, the application of monolithic column materials in the liquid-chromatographic (LC) analysis of proteomic samples has been evaluated. Traditional (two-dimensional) gel electrophoresis (2D-GE) based proteomics approaches are complemented more and more by LC based methods. Monolithic columns are considered to be ideally suited for the analysis of (bio)polymers due to their unique pore structure with large throughpores and shallow mesopores. This results in reduced resistance to mass transfer, which is especially advantageous for compounds with low diffusion coefficients like proteins and peptides.

It has been demonstrated in this thesis that monolithic columns provide a promising alternative to packed columns for high-efficiency separations of protein digests. By applying columns of different length, it is possible to obtain either fast separations of simple digests or high separation performance for complex samples. Both goals can be achieved without exceeding the pressure limits of standard LC equipment, thus offering an interesting alternative to ultra-high pressure systems. Poly(styrene-divinylbenzene) (PS-DVB) monolithic columns are also employed in an integrated system for protein separation and identification. Due to the relatively low backpressure of the monoliths, several columns can easily be connected in series to provide different functionalities like (pre-) concentration, separation, on-line digestion and trapping.

Digest separations

Both C18-bonded silica and PS-DVB monoliths were tested for protein digest separation. At present, LC-tandem mass spectrometry (MS) analysis of protein digests is the primary means of protein identification in proteomics. Next to confirmation of protein identity, it also provides insight into the occurrence and location of posttranslational modifications like phosphorylation, glycosylation or acetylation. In order to achieve an optimal coverage of all proteins constituting a sample, MS-based identification of as many peptides as possible is necessary. To achieve this, the efficiency of the preceding peptide separation needs to be maximized. The effect of column length on separation efficiency was evaluated by comparing monoliths differing in length by a factor 5 for both stationary phases. Separation efficiency was expressed in chromatographic parameters, such as peak capacity and productivity (peak production rate), as well as in quality of protein identification after MS detection. In addition to evaluating the applicability of monolithic columns for peptide separations, these studies also highlight some fundamental aspects of the reversed-phase LC analysis of protein digests.

It has been demonstrated that both silica-based and polystyrene-based monolithic columns provide efficient separations of protein digest mixtures. For both the silica-based and the PS-DVB monoliths the longer columns were found to outperform the shorter columns when gradients with equal steepness were compared. In these cases peak capacities were expected to differ by about $\sqrt{5}$ (= 2.24) as the peak capacity is

proportional to the square root of the plate number (N). However, experimental peak capacities for pairs with equal gradient steepness were found to differ by a factor between 2.7 and 4.0. This was true for both longer and shorter gradients and for both types of monolithic columns, suggesting that the use of longer columns provides some additional advantage besides the increased column length.

As a result of their bimodal pore structure, monolithic columns are an interesting alternative to traditional packed columns for the analysis of proteins and protein digests. The high peak capacities at short gradients and rapid reconditioning make the relatively short columns ideally suited for fast separations and high-throughput analyses. The highest peak production rates (PPR) have been obtained for fast analyses using the short (5 cm) PS-DVB monolith. The peak capacity obtained for the short column is lower than for the long column at the same gradient time, but this is compensated by a shorter column dead time (t_0) and re-equilibration time (t_{eq}). For longer gradients, the relative contribution of t_0 and t_{eq} to the total analysis time is smaller and the higher peak capacity leads to a higher PPR. In general, the highest peak capacities have been obtained using the long (25 cm) PS-DVB column. The peak capacities found for the long (75 cm) C18-silica monolith were similar to those for the short PS-DVB column, but PPRs are lower because of the longer t_0 and t_{eq} resulting from the larger column volume. It should be noted that while the PS-DVB columns were operated at 60°C, the silica based monoliths have been used at ambient temperature. This difference in temperature results in an increase in peak capacity of about 15% for the PS-DVB columns. In general, short monolithic columns are highly applicable to the fast analysis of relatively simple peptide mixtures while the longer columns are more useful for the high-efficiency separations of more complex samples. Higher separation efficiency is generally accompanied by a larger number of identified peptides and higher protein sequence coverage. However, identification scores reach a plateau when the peak capacity is higher than the number of peptides in the digest.

For a 15 cm packed column, the column length is exactly between the lengths of both PS-DVB monoliths. The separation performance of the packed columns, packed with 3 µm particles, is comparable to that obtained for the short (5 cm) monolith. Peak capacities of 100-150 have been reported for gradients of about 60 min. However, the fast loading and reconditioning of the monolithic column results in a higher PPR at similar analysis times. The longer monolith achieves higher peak capacities than both the short monolith and the packed columns while maintaining the fast reconditioning, thus achieving a higher PPR for analyses of 30 min and longer. Moreover, if t_0/t_G is kept constant, t_G should be increased for comparison of the longer column. When comparing the longer PS-DVB monolith to ultra-high performance LC (UPLC) systems with similar column lengths, the peak capacities for UPLC range form 1.5-fold higher (using 1.5 µm particles) to about 3 times higher for columns packed with 0.8 µm particles. Especially

the latter system, however, suffers from relatively long loading and reconditioning times, limiting the PPR. For short analysis times (up to about 60 min) the PPR for the monolith is comparable to the UPLC system while at longer analysis times a higher PPR is obtained for the UPLC column due to the higher *PC*. In summary, the monolithic columns evaluated in thesis can match or exceed the performance of regular packed capillary columns for protein digest separations. UPLC systems, however, achieve a higher peak capacity but the long loading and reconditioning times lead to similar PPR for both systems at short analysis times. Moreover, monolithic columns have the advantage that they can be operated using standard LC equipment.

An interesting property of monolithic columns is the use of relatively high flow rates, making it possible to perform very fast, efficient analyses. However, it is a disadvantage that only relatively large internal diameters are available for commercially available columns. In order to take full advantage of the increased ionization efficiency of nanoESI at low (<500 nl/min) flow rates, column diameters should be reduced to 50 µm or less, resulting in column flow rates of around 250 nl/min. Additionally, the relatively low backpressures that resulted from the use of the monolithic columns indicate that there is a possibility to use even longer columns, producing still higher peak capacities. The maximum backpressure measured for the 250 x 0.2 mm ID PS-DVB monolith during gradient runs was around 200 bar. This implies that the column length could possibly be increased to as much as 50 cm without making the use of special high-pressure equipment necessary. Doubling the column length would lead to an increase in peak capacity of $\sqrt{2}$ or higher, making it possible to achieve peak capacities around 500 with long analysis times.

Integrated systems

The development of integrated systems for the separation and identification of proteins could be a highly promising alternative to traditional proteomics methods. Chromatographic methods for protein separation and fractionation offer a high degree of flexibility and could provide information on a number of protein properties, like molecular weight (SEC), charge (IEC) or hydrophobicity (RPLC). The use of on-line immobilized enzyme reactors reduces the time needed for protein digestion from hours to only several minutes and makes proteolysis much more amenable to automation than traditional in-solution digestion protocols. Limiting the amount of manual sample handling would also reduce the risk of sample loss which increases the chance of identification of low abundant species.

It has been demonstrated that proteins can be concentrated from an acidic environment using a strong cation-exchange (SCX) column and be subsequently eluted using a high pH buffer. This buffer provides a suitable environment for on-line proteolysis using an immobilised trypsin reactor. Combining this preconcentration method with LC-MS/

MS for digest analysis yielded a system which has been used to analyse proteins with different molecular weights and pI. This system has also been combined with RPLC-based protein fractionation to form a total analysis system for protein mixtures. The SCX column is used to trap a fraction from the LC eluate during the protein separation, this fraction is digested and the digest is analysed for protein identification. In this way, β-lactoglobulin A was separated from a protein mixture, trapped and identified within a total analysis time of 70 minutes. The development of a column based system as described in this thesis results in a highly flexible technology which can easily be adapted to incorporate different methods of protein fractionation, alternative proteolytic enzymes and (affinity-based) enrichment methods.

The high permeability of monolithic columns makes them especially useful for integrated systems as the low backpressure makes it possible to connect several columns in series without compromising optimal flow rates. Further developments of the system presented in this thesis should focus on automation of the system and applicability to real samples. In order to provide a comprehensive system for the analysis of protein mixtures, a column-switching set-up should be considered using multiple parallel columns for trapping of the protein fractions as well as the subsequent digests. In such a system, it is important to find the right balance between protein fraction size and peptide separation efficiency. The trapping of large protein fractions requires higher peptide separation efficiency as the digests will be more complex as a result of trapping multiple proteins in one fraction. Smaller protein fractions will contain fewer proteins but also allow less time for analysis of the digests. Additional developments for this system could focus on application of affinity materials at the level of either protein fractionation or peptide trapping as this type of materials could be very useful in targeting specific sub-proteomes. Such an integrated system could provide a fast and versatile alternative to traditional (gel-based) proteomics methods.

Perspectives

Future developments for the use of monolithic materials in proteomics could focus on several aspects. Firstly, column dimensions could be optimised for proteomics applications. The range of available diameters for monolithic columns from commercial sources does not extend below 100 μm ID, resulting in flow rates higher than optimal for electrospray ionisation. A reduction in column diameters would offer the possibility to operate monoliths at flow rates compatible with high-efficiency electrospray ionisation without compromising separation efficiency by use of sub-optimal flow rates. More sensitive MS detection could lead to lower sample consumption and increased identification of low-abundant proteins.

Columns of various lengths could be applied to different proteomics problems. Shorter columns can be used for fast analysis of simple digests, providing high-throughput

analysis. Longer columns are more suited for the separation of complex samples, where analysis time is less important. It should be possible to apply columns even longer than the ones used in this thesis without exceeding the pressure limits of standard LC equipment. A column length for PS-DVB monoliths of about 50 cm is possible under conditions applied in this thesis and even longer columns can be used when the flow-rate is reduced. When using long monolithic columns with very long gradients it should be possible to reach peak capacities similar to those obtained using UPLC systems, without exceeding the pressure limits of standard LC equipment. It would also be interesting to study the performance of monolithic columns under UPLC conditions, if the monoliths can withstand the high inlet. Another interesting development for fast analysis and efficient coupling with MS is the use of chip-LC. Applications of monoliths in lab-on-a-chip systems have already been reported in literature. Production of monolithic columns by in-situ polymerisation processes make these materials well suited for use in chip-LC methods, avoiding the need to pack particulate materials in chips.

A second aspect is further diversification in monolithic-column chemistry. At present, monolithic columns are mainly used for reversed-phase chromatography but especially polymeric monoliths are highly amenable to variations in stationary-phase chemistry and chromatographic mode. The high porosity of monolithic columns makes them highly suited for application in multi-column systems. Therefore, the development of functionalised monoliths could be very promising. Affinity materials and monoliths with immobilised enzymes, the first examples of which have been reported in literature, could easily be used in integrated systems for protein analysis. These integrated systems, like the one described in this thesis could provide a link between top-down and bottom-up proteomics approaches. Flow splitting after the protein separation step could offer the opportunity to obtain MS information on the intact proteins while simultaneously generating digests of the same proteins for unambiguous identification. Integrated systems based on LC and MS can offer powerful new possibilities for protein analysis.

Samenvatting

SAMENVATTING IN HET NEDERLANDS

Sam

De succesvolle opheldering van diverse genomen, inclusief het menselijke, heeft de deuren geopend voor uitgebreid onderzoek naar de functionele exponent, het proteoom. Waar het genoom de blauwdruk vormt voor alle delen van een levend organisme, vinden de belangrijkste biologische processen plaats op het niveau van het proteoom. In tegenstelling tot het genoom, dat onder vrijwel alle omstandigheden hetzelfde is, is het proteoom zeer dynamisch. Eiwitexpressieniveaus kunnen variëren tussen verschillende celtypen, weefsels en fysiologische condities. Verschillende verschijningsvormen van eiwitten en de grote variatie in concentraties maken de analyse van eiwitten zeer uitdagend. Een breder begrip van het proteoom biedt aanknopingspunten voor de identificatie van nieuwe therapeutische doelen en diagnostische en therapeutische biomarkers. Vanwege de eerder genoemde eigenschappen van het proteoom zijn zeer krachtige analytische technieken nodig om een compleet beeld van het proteoom te schetsen.

In de afgelopen jaren hebben verschillende analysetechnieken een toepassing gevonden in het proteoom-onderzoek. De toepassing van scheidingsmethoden als twee-dimensionale gelelectroforese (2D-GE) en 'zachte' ionisatietechnieken voor massaspectrometrie (MS) hebben het mogelijk gemaakt om grote aantallen eiwitten snel te identificeren. 2D-GE is tegenwoordig de eerste keus voor het scheiden van complexe eiwitmengsels, waarbij eiwitten eerst worden gescheiden op isoelektrisch punt en vervolgens op grootte met een SDS-polyacrylamide gel. Hoewel 2D-GE een uitstekende resolutie geeft voor de scheiding van intacte eiwitten hebben beperkingen als lange analysetijden, het beperkte concentratiebereik en problemen bij de analyse van bepaalde eiwitklassen (zeer kleine eiwitten, zeer grote eiwitten, eiwitten met een extreme pI, hydrofobe eiwitten) onderzoek naar de ontwikkeling van alternatieve methoden gestimuleerd. Vandaag de dag worden methoden gebaseerd op gelelectroforese meer en meer aangevuld met vloeistofchromatografie (LC), met name voor scheiding van peptiden na digestie van eiwitmengsels. Sommige strategieën combineren (1D-)gelelectroforese voor eiwitscheiding met (multidimensionale) chromatografische analyse van digesten verkregen uit uitgesneden gel-banden. Andere methoden maken geen gebruik van eiwitscheiding en gebruiken (multidimensionele) LC methoden voor de scheiding van complexe digesten.

Eiwitidentificatie vindt in het algemeen plaats door middel van MS gevolgd door data-analyse. In 'top-down' proteomics wordt MS toegepast op intacte eiwitten teneinde informatie te verkrijgen over bijvoorbeeld eiwitstructuur of interacties met liganden of andere eiwitten. Posttranslationele modificaties maken identificatie van intacte eiwitten door middel van MS tot een grote uitdaging. In de afgelopen jaren heeft de introductie van hoge-resolutie MS (Orbitrap en Fourier-transform MS) de kwaliteit van massaspectra van intacte eiwitten sterk verbeterd en dissociatie-technieken voor intacte eiwitten hebben identificatie verder geholpen. Desondanks is de MS analyse

van digesten nog altijd de primaire methode voor eiwitidentificatie. Voor zuivere eiwitten, bijvoorbeeld uitgesneden spots uit 2D gels, is het vaak voldoende om de peptidenmassa's in kaart te brengen met behulp van matrix-assisted laser desorption ionisation (MALDI-) of electrospray ionisation (ESI-)MS. Voor meer complexe monsters, die peptiden bevatten die afkomstig zijn van meerdere eiwitten, is het sequencen van peptiden een betere methode. Hierbij worden peptiden gefragmenteerd in de MS waarna de fragmenten worden gebruikt om de aminozuur-volgorde van de peptiden te bepalen. Dit kan gedaan worden door visuele controle van individuele MS/MS spectra of automatisch door de complete MS/MS datafiles te vergelijken met databases waarna de gevonden sequenties worden gebruikt om de bijbehorende eiwitten te identificeren. Deze methode van proteoom analyse, gebaseerd op scheiding en MS van gedigesteerde eiwitmengsels, wordt bottom-up proteomics genoemd.

Omdat digestie van een gemiddeld eiwit ongeveer 25-30 peptiden oplevert zullen zelfs relatief eenvoudige eiwitmengsels al snel zeer complexe digesten produceren. De hoge resolutie en eenvoud van koppeling met MS maakt (capillaire) reversed-phase (RP)LC tot de belangrijkste technologie voor scheiding van eiwitdigesten. Tengevolge van de complexiteit van de monsters vereist analyse van digesten zeer efficiente scheidingsmethoden. In de afgelopen jaren zijn nieuwe strategieën ontwikkeld om aan deze analytische uitdagingen tegemoet te komen. Er zijn pogingen gedaan om de resolutie van LC methoden te vergroten door gebruik van lange kolommen of kolommen gepakt met relatief kleine deeltjes. Beide methoden leiden echter tot hoge druk in de kolommen en maken daarmee het gebruik van speciale apparatuur noodzakelijk. Een inherent nadeel van het gebruik van gepakte kolommen voor de scheiding van eiwitten en peptiden is de trage diffusie van macromoleculen, die de massa-overdracht van de moleculen naar de stationaire fase beperkt. In tegenstelling tot gepakte kolommen hebben monolithische kolommen een tweevoudige poriestructuur met grote poriën om de doorstroming van de mobiele fase te vergemakkelijken en ondiepe mesoporiën om de massa-overdracht van moleculen naar de stationaire fase te versnellen. De toepassing van monolithische kolommen voor eiwitchromatografie kan een goed alternatief bieden voor methoden gebaseerd op gelelectroforese. Op LC gebaseerde methoden zijn sneller, gemakkelijker te automatiseren en kunnen direct gekoppeld worden aan andere analysestappen als on-line proteolyse en aan detectiemethoden als MS. Daarnaast bieden LC methoden een grote veelzijdigheid als gevolg van de mogelijkheid om verschillende typen kolommen toe te passen en te combineren. De specifieke eigenschappen van monolithische kolommen bieden ook interessante mogelijkheden voor chromatografie van peptiden. De lage tegendruk van de kolom als gevolg van de grote poriën maakt het mogelijk om hoge snelheden van de mobiele fase toe te passen voor snelle, efficiente scheidingen of om lange kolommen te gebruiken voor hoge efficiëntie bij gematigde druk. Hierdoor kunnen monolithische kolommen een lage-druk alternatief bieden voor 'ultra performance' LC (UPLC) systemen met

kolommen die gepakt zijn met zeer kleine (< 2 µm) deeltjes.

Het doel van het onderzoek dat wordt beschreven in dit proefschrift was om de toepassing van monolithische kolommen, zowel op basis van silica als organische polymeren, te evalueren voor de analyse van eiwitten. Het eerste deel beschrijft onderzoek naar de effecten van kolomlengte en stationaire-fase materiaal op de scheiding van eiwitdigesten. Het tweede deel is gewijd aan de ontwikkeling van een geïntegreerd chromatografisch systeem voor de scheiding en identificatie van eiwitmengsels.

In **Hoofdstuk 1** wordt de achtergrond van het onderzoek geschetst en wordt het doel van het onderzoek beschreven. In **Hoofdstuk 2** wordt een overzicht gegeven van recente ontwikkelingen in kolommen voor vloeistofchromatografie van peptiden. Er wordt zowel aandacht besteed aan nieuwe ontwikkelingen in de technologie van reversed-phase kolommen, met name UPLC en monolithische kolommen, als ook aan andere typen vloeistofchromatografie. De mogelijkheden van hydrofiele interactie chromatografie worden besproken, waarbij ook aandacht wordt besteed aan toepassing in multidimensionele systemen. Enkele voorbeelden van affiniteitschromatografie worden behandeld, met speciale aandacht voor de analyse van gefosforyleerde eiwitten en de verwijdering van veel voorkomende eiwitten.

In **Hoofdstuk 3** wordt een monolithische C18-silica kolom van 750 x 0.2 mm diameter gepresenteerd. De scheiding van eiwitdigesten met behulp van deze kolom wordt onderzocht en de kolom wordt vergeleken met een commercieel verkrijgbare monolithische kolom van 150 mm. Een trypsine digest van bovine serum albumine (BSA) wordt gebruikt als model monster om de effecten van kolomlengte en gradiënttijd op piekcapaciteit en eiwitidentificatie te evalueren. Door het toepassen van een UV-detector met glasvezel kunnen UV en MS detectie gecombineerd worden in een systeem met minimale bandverbreding. UV chromatogrammen werden gebruikt om de piekcapaciteit te bepalen en MS/MS data werden gebruikt voor eiwitidentificatie. Met de lange kolom werden hogere piekcapaciteiten en betere scores voor eiwitidentificatie behaald en dus biedt het gebruik van langere kolommen goede mogelijkheden voor de analyse van eiwitdigesten.

Hoofdstuk 4 geeft een vergelijking tussen C18-silica monolieten van 0.1 mm diameter en verschillende lengte. Om de mogelijkheden van de monolithische kolommen zo goed mogelijk te evalueren werd het monster direct op de kolom geïnjecteerd (zonder voorkolom). Piekcapaciteit en eiwitidentificatie werden bepaald met een digest van een mengsel van BSA, α-caseine en β-caseine. Om de de invloed van extra-kolom effecten te minimaliseren werd de UV-detector uit het systeem gelaten en werden piekcapaciteiten bepaald uit MS extracted ion chromatogrammen. Dit systeem geeft een betere scheiding

dan het systeem beschreven in hoofdstuk 3, er werden voor zowel de 150 mm als de 750 mm kolom hogere piekcapaciteiten behaald bij gelijke gradiënttijden. In vergelijking met de eerder beschreven experimenten is het verschil in efficiëntie tussen de lange en korte kolom groter, bij lange gradiënttijden wordt een winst van meer dan een factor 2 behaald. Bij vergelijking van de kolommen bij gelijke steilheid van de gradiënt, waarbij de gradiënttijd evenredig met de kolomlengte wordt verhoogd, blijkt het verschil in piekcapaciteit voor de twee kolommen groter dan de verwachte factor √5.

Hoofdstuk 5 beschrijft het gebruik van PS-DVB monolithische kolommen van 50 en 250 mm lengte voor de scheiding van een digest van negen eiwitten. De prestaties van de kolommen werden vergeleken door middel van piekcapaciteit, piekproductiesnelheid en eiwitidentificatie. Enkele fundamentele aspecten van gradiëntchromatografie van peptiden worden bediscussieerd en er wordt een voorspelling gedaan voor de maximaal haalbare piekcapaciteit voor beide kolommen, deze zijn 294 voor de korte en 370 voor de lange kolom. Deze kolommen bieden uitstekende mogelijkheden voor de analyse van eiwitdigesten. Alle negen eiwitten worden al gedetecteerd bij korte gradiënttijden en zowel identificatiescores als aantallen peptiden nemen toe met oplopende gradiëntlengte, van ca. 35 peptiden voor beide kolommen bij een 5 min gradiënt tot 58 en 78 voor respectievelijk de korte en de lange kolom bij een gradiënttijd van 300 min. De piekcapaciteit van de PS-DVB kolommen ligt hoger dan bij gebruik van de silica monolieten terwijl kortere analysetijden leiden tot hogere piekproductiesnelheid.

De korte monolithische kolommen produceren piekcapaciteiten die vergelijkbaar zijn met die van gepakte capillaire kolommen. Echter, door de snelle re-equilibratie is de piekproductiesnelheid van de monolithische kolommen hoger. De lange monolithische kolommen geven een piekcapaciteit die hoger is dan voor reguliere gepakte kolommen maar lager dan voor UPLC kolommen, die vaak een veel langere regeneratietijd vergen. Hierdoor zijn de piekproductiesnelheden van UPLC en lange monolithische kolommen vergelijkbaar en bieden de monolieten een interessant alternatief voor systemen met gepakte kolommen die een zeer hoge druk genereren.

In **Hoofdstuk 6** wordt de ontwikkeling beschreven van een geïntegreerd systeem voor de preconcentratie en identificatie van eiwitten. Het systeem bestaat uit een kationenwisselings (SCX) voorkolom voor eiwit-trapping en preconcentratie, een reactor met geïmmobiliseerd trypsine voor eiwitdigestie en een RPLC-MS systeem voor analyse van het digest en eiwitidentificatie. Verschillende buffers werden vergeleken voor zowel optimale desorptie van eiwitten van de SCX kolom als digestie door het geïmmobiliseerd trypsine. Met dit systeem konden enkele eiwitten van uiteenlopende grootte en isoelectrisch punt geconcentreerd worden vanuit een zuur milieu. Met een hoge-pH buffer konden de eiwitten worden gedesorbeerd en gedigesteerd om identificatie met behulp van LC-MS/MS mogelijk te maken.

In **Hoofdstuk 7** wordt een systeem beschreven voor de analyse van eiwitmengsels. In dit systeem wordt een chromatografische eiwitscheiding met behulp van een PS-DVB monoliet gekoppeld met een kleine SCX kolom voor het trappen van interessante eiwitten. Deze SCX kolom wordt vervolgens in een tweede systeem geplaatst waar de getrapte eiwitten worden gedesorbeerd met een buffer met hoge pH en via een reactor met geïmmobiliseerd trypsine worden meegevoerd naar een RP voorkolom. Het digest wordt geanalyseerd met RPLC-MS om het eiwit te identificeren. Met dit systeem kon in slechts 70 minuten een eiwitmengsel worden gescheiden, β-lactoglobuline A worden vastgehouden op de SCX kolom en vervolgens worden geidentificeerd. Door de relatief korte analysetijd kan dit systeem een efficient alternatief bieden voor traditionele proteomics methoden als 2D-gelelectroforese.

Hoofdstuk 8 geeft enkele algemene conclusies over het onderzoek dat wordt gepresenteerd in dit proefschrift. Daarnaast worden aanbevelingen gedaan voor verder onderzoek en ook worden enkele toekomstperspectieven besproken.

Nawoord
NAWOORD

Na

Het uitvoeren van een promotieonderzoek en het aansluitend produceren van een proefschrift is een behoorlijke krachttoer. Gelukkig hoef je deze klus niet alleen te klaren.

Allereerst mijn promotor, prof. dr. Ad de Jong. Toen ik in 2004 solliciteerde op een vacature binnen het NPC was er sprake van twee AIO posities, één binnen de massaspectrometrie-groep van prof. Heck en één bij Farmaceutische Analyse. Hoewel er destijds min of meer vóór mij gekozen is heb ík in elk geval nooit spijt gehad van die keuze. De afgelopen jaren hebben we vrijwel iedere twee weken bij elkaar gezeten en omdat je mijn enige begeleider was waren dit vaak één-op-één sessies. Ik heb deze besprekingen altijd erg nuttig en constructief gevonden. Daarnaast stimuleerden onze gesprekken mij ook om regelmatig een stapje achteruit te doen en mijn onderzoek ook eens van een afstandje te bekijken en me op de grote lijn te concentreren, iets waar je me regelmatig aan moest helpen herinneren. Ook het feit dat er op zijn tijd ruimte was voor wat meer persoonlijke en minder professionele gespreksonderwerpen heb ik altijd als er prettig ervaren.

Op deze plaats wil ik ook graag prof. dr. A.J.R. Heck, prof. dr. R.M.J. Liskamp, prof. dr. R. Bischoff en prof. dr. P.D.E.M. Verhaert bedanken voor het lezen en beoordelen van dit proefschrift. Ik wil ook graag het Netherlands Proteomics Center bedanken voor het financieel mogelijk maken van dit project.

During this project, we had a very enjoyabe collaboration with GL Sciences. I would like to thank mr. M. Kurano ang mr. G. Alkema for providing the C18-silica monolithic columns. Daarnaast was er ook een zeer prettige samenwerking met Dionex Benelux, die het hoofdstuk over de PS-DVB monlieten heeft opgeleverd. Ik wil graag dr. S. Eeltink bedanken voor het beschikbaar stellen van de kolommen en de nuttige discussies over de theoretische uitwerking van de resultaten.

Een tweede persoon zonder wie dit proefschrift niet geworden zou zijn wat het nu is, is natuurlijk Gerard Wiese. Na een aantal vruchteloze pogingen om een nano-LC systeem op te zetten met gebruik van standaard HPLC pompen hebben we ons gericht op de ontwikkeling van het geïntegreerde systeem. Zonder jouw hulp bij de ontwikkeling van het SCX-reactor gedeelte zou ik zeker veel langer nodig gehad hebben om het systeem aan de gang te krijgen. Ik heb tijdens onze samenwerking veel van je geleerd en ik hoop dat ik jou, ondanks al je ervaring, op zijn tijd ook nog wat nieuws heb kunnen vertellen.

Een aantal mensen die ik ook absoluut niet mag vergeten zijn mijn kamergenoten van de leukste kamer van het Went, Linda, Remco en Bregje. Naast gesprekken over werk, wetenschap en de bijbehorende frustraties hebben we ook altijd tijd gehad voor lol en

ontspanning. Ook de koffiekamer op de 6e verdieping was altijd goede plek voor een beetje ontspanning en ik wil graag alle collega's bedanken voor de fijne momenten die ik daar heb beleefd. Dus, Govert, Wout, Joop, Johan, Jan, Jonathan, Xander, Roelof, Edwin, Paul, Rob, Rawi, Stefan, Maurice, Javier, Andrea, Dragana, Lidija, Frits, Miranda, mijn 'nieuwe' kamergenoten Yvon en Thierry en alle stagestudenten die de afgelopen vijf jaar de revue gepasseerd zijn, ontzettend bedankt!

Dan wil ik ook graag mijn paranimfen even apart noemen. Roel, ik denk dat ik zonder enig spoor van ironie kan zeggen dat dit proefschrift er zonder jou heel anders uitgezien zou hebben. Inmiddels ben jij de enige in de familie met twee proefschriften op zijn naam. Bregje, van alle collega's bij analyse ben jij waarschijnlijk degene die het het langst met mij op één kamer heeft uitgehouden. Je was altijd een geweldig klankbord voor zowel nieuwe ideeën als de frustraties wanneer die ideeën niet bleken te werken. Ik ben erg blij dat jullie mijn paranimfen willen zijn.

Omdat er ook nog een aantal mensen zijn die mij er de afgelopen jaren aan hebben helpen herinneren dat een AIO ook nog een ander leven heeft, wil ik ook nog wat andere mensen bedanken. Allereerst mijn ouders, broertje en schoonzus, die mij vanaf mijn vroegste herinnering gesteund hebben bij alles wat op mijn pad kwam. Ook al heb ik niet vaak van de mogelijkheid gebruik hoeven maken, het geeft enorm veel rust om te weten dat ik bij jullie altijd met alles terecht kan. Daarnaast mijn opa, met wie ik een speciale band heb en die op deze plaats niet mag ontbreken. Natuurlijk wil ik ook mijn schoonfouders, Bas en Nathalie bedanken voor de support in de afgelopen jaren.

De laatste woorden van dit proefschrift zijn voor de twee belangrijkste mensen in mijn leven, Dorien en Eva. Ik laat het misschien niet altijd merken maar ik had het de afgelopen jaren niet zonder jullie gekund. Nu mijn boekje eindelijk af is hoop ik ook weer wat meer tijd voor jullie te hebben wanneer ik wel thuis ben, ik hou van jullie alletwee.

Michiel

CURRICULUM VITAE

Michiel Hendrik Maarten van de Meent werd geboren op 9 juli 1978 te Alphen aan den Rijn. Na het behalen van het diploma Gymnasium aan het Groene Hart Lyceum te Alphen aan den Rijn, begon hij in september 1996 aan de studie Bio-Farmaceutische Wetenschappen aan de Universiteit Leiden. Tijdens de doctoraalfase heeft hij een hoofdvakstage gedaan op de afdeling Farmacognosie van het Leiden/Amsterdam Center for Drug Research (LACDR), onder begeleiding van dr. K. Ingkaninan en prof. dr. R. Verpoorte. De titel van het stageverslag was: Acetylcholinesterase-inhibitors from different Amaryllidaceae species as potential ant-alzheimer drugs. Aansluitend volgde hij een tweede onderzoeksstage bij de afdeling Farmacochemie van het LACDR, onder begeleiding van prof. dr. A.P. IJzerman, het veslag was getiteld: Molecular mechanics calculations of human adenosine kinase and ligands and comparison of human and *Toxoplasma gondii* adenosine kinase structures. In juni 2001 behaalde hij het doctoraaldiploma Bio-Farmaceutische Wetenschappen. Na het behalen van het doctoraaldiploma begon hij op het Laboratorium voor Analytische Chemie van het Rijksinstituut voor Volksgezondheid en Milieu te Bilthoven als wetenschappelijk medewerker, waar hij zich specialiseerde in de analyse van eiwitten en peptiden met behulp van vloeistofchromatografie en massaspectrometrie.

Van November 2004 tot en met November 2008 was hij als assistent in opleiding verbonden aan de divisie Biomedische analyse van de Universiteit Utrecht onder begeleiding van prof. Dr. G.J. de Jong, alwaar het in dit proefschrift beschreven onderzoek werd uitgevoerd. Het onderzoek werd gefinancierd door het Nederlands Proteomics Center en werd uitgevoerd in samenwerking met GLsciences en Dionex Corporation.

Sinds december 2009 is hij werkzaam als Docent bij de afdeling farmaceutische analyse van de Universiteit Utrecht.

LIST OF PUBLICATIONS

M.H.M. van de Meent, G.J. de Jong, *Novel LC columns in proteomics*, In preparation

M.H.M. van de Meent, G. Wiese, G.J. de Jong, *Development of an RPLC-SCX-IMER-RPLC-MS system for the fractionation, digestion and identification of proteins*, In preparation

M.H.M. van de Meent, S. Eeltink, G.J. de Jong, P*otential of poly(styrene-co-divinylbenzene) monolithic columns for the analysis of protein digests*, Submitted for publication

M.H.M. van de Meent, G. Wiese, G.J. de Jong, *Coupling of a cation-exchange precolumn with an on-line immobilised trypsin-LC-MS system for trapping and identification of proteins*, Submitted for publication

M.H.M. van de Meent, G.J. de Jong, *Potential of long capillary monolithic columns for the analysis of protein digests*, J. Sep. Sci., 32 (4), 2009, 487-493

M.H.M. van de Meent, G.J. de Jong, *Improvement of the liquid chromatographic analysis of protein tryptic digests by the use of long capillary monolithic columns with UV and MS detection*, Anal. Bioanal. Chem., 388 (1), 2007, 195-200

I.K. Rhee, M.H.M. van de Meent, K. Ingkaninan, R. Verpoorte, *Screening for acetylcholinesterase inhibitors from Amaryllidaceae using silica gel thin-layer chromatography in combination with bioactivity staining*, J. Chromatogr. A, 915 (1-2), 2001, 217-223