Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Chromatography A, 1053 (2004) 27–36
JOURNAL OF CHROMATOGRAPHY A
www.elsevier.com/locate/chroma

Review

The role of liquid chromatography in proteomics

Yang Shi[1], Rong Xiang[1], Csaba Horváth✠, James A. Wilkins*

Department of Chemical Engineering, Yale University, 205 Mason Laboratory, P.O. Box 208286, New Haven, CT 06520-8286, USA

Available online 27 August 2004

Abstract

Proteomics represents a significant challenge to separation scientists because of the diversity and complexity of proteins and peptides present in biological systems. Mass spectrometry as the central enabling technology in proteomics allows detection and identification of thousands of proteins and peptides in a single experiment. Liquid chromatography is recognized as an indispensable tool in proteomics research since it provides high-speed, high-resolution and high-sensitivity separation of macromolecules. In addition, the unique features of chromatography enable the detection of low-abundance species such as post-translationally modified proteins. Components such as phosphorylated proteins are often present in complex mixtures at vanishingly small concentrations. New chromatographic methods are needed to solve these analytical challenges, which are clearly formidable, but not insurmountable. This review covers recent advances in liquid chromatography, as it has impacted the area of proteomics. The future prospects for emerging chromatographic technologies such as monolithic capillary columns, high temperature chromatography and capillary electrochromatography are discussed.

Keywords: Review; Proteomics; Quantitative

Contents

Abbreviations: AMT, accurate mass tag; CEC, capillary electrochromatography; CE, capillary electrophoresis; DIGE, difference gel electrophoresis; EMDC, elution-modified displacement chromatography; ESI, electrospray ionization; FTICR, Fourier transform ion cyclotron resonance; HPLC, high-performance liquid chromatography; IAC, immunoaffinity chromatography; ICAT, isotope-coded affinity tags; IEX, ion-exchange chromatography; IMAC, immobilized metal affinity chromatography; LC, liquid chromatography; MALDI, matrix-assisted laser desorption/ionization; MicroSPE, micro-solid-phase extraction; MMA, mass measurement accuracy; MS, mass spectrometry; MudPIT, multi-dimensional protein identification technology; PAGE, polyacrylamide gel electrophoresis; PS–DVB, polystyrene divinylbenzene; Q-TOF, quadrupole time-of-flight; rhGH, recombinant human growth hormone; RP, reversed phase; SEC, size-exclusion chromatography; SCX, strong cation-exchange; 1D, one-dimensional; 2D, two-dimensional

* Corresponding author. Tel.: +1 203 432 4373; fax: +1 203 432 4360.
E-mail address: james.wilkins@yale.edu (J.A. Wilkins).
[1] These authors contributed equally to the work.
✠ Author deceased.

0021-9673/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.chroma.2004.07.044

1. Introduction

"Proteomics" is sometimes defined as the analysis of part or all of the protein complement of a complex biological system at any given moment. With the elucidation of the human genome sequence, attention has focused on gaining an understanding of the gene products themselves, the proteins. However, the understanding of the diverse structural characteristics and interactions of proteins represents a significantly greater analytical challenge than that posed by nucleic acid sequence analysis required for acquisition of the genome sequence.

It is well known that the basic analytical requirements in proteome analysis are high sensitivity, high resolution, and high throughput along with high-confidence protein identification. Proteins must also be quantified and post-translational modifications identified [1]. To reduce sample complexity prior to mass spectrometry, one of two approaches is usually taken prior to protein identification: (1) proteins are first separated, then digested (this may be characterized as "top-down" proteomics [2]); (2) in "shotgun" proteomics, a complex protein mixture is first digested. Peptides are then chromatographically resolved (characterized as "bottom-up" proteomics [3]). In both cases, separation technologies play a critical role in protein identification and analysis.

Two-dimensional (2D) polyacrylamide gel electrophoresis (PAGE) is currently widely used to separate and to quantify intact proteins [4,5] because thousands of proteins can be separated in a single analytical run. 2D-PAGE offers the advantage that radioactive or fluorescence tagging can be used for the detection of post-translational modifications and for measurement of differential protein expression in cellular populations that differ in their physiological, metabolic or disease states [6]. Recently, two-color fluorescent labeling techniques analogous to those used for detection of mRNA levels in microarray format, termed as 2D difference gel electrophoresis (DIGE), have been introduced for differential quantitation in 2D-PAGE [7]. 2D-PAGE analysis has enabled the identification of many proteins from a vast variety of sources and has continued to serve as the "workhorse" technique in proteomics [7,8].

However, it is widely recognized that 2D-PAGE-based methods also suffer from several technical limitations. Although advances have been made, reproducibility remains a concern. Highly acidic and basic proteins and hydrophobic proteins are generally difficult to detect in 2D-PAGE separations [9]. Sensitivity and dynamic range, i.e. the absolute amount of proteins that can be loaded onto the gel are limited [10]. Therefore, low abundance proteins are usually not detected on 2D-PAGE. It was recently estimated on the basis of codon-bias distribution that more than half of all proteins in the yeast proteome are not detectable by 2D-PAGE analysis [9].

Because of its high resolving power, reproducibility and its compatibility with electrospray mass spectrometry (MS), high-performance liquid chromatography (HPLC) represents an attractive alternative to 2D-PAGE for the separation of both proteins and peptides. Another attractive feature of liquid chromatography (LC) is the broad selection of stationary and mobile phases. This feature makes LC a versatile and fundamentally important tool in proteomics. During the last few years, a number of LC techniques have appeared in the literature. A variety of modes have been used both alone and in combination. Reversed-phase, ion-exchange, affinity and size-exclusion chromatography [11–13] in addition to other LC methods share several important characteristics. In comparison with gel-based separation methods, sample handling and preparation are minimal. Proteins or peptides digests are separated by RP-HPLC and can be introduced directly into the mass spectrometer through electrospray ionization (ESI) for identification and analysis. Increased resolution of proteins and peptides can be achieved by orthogonal steps of chromatography (2D- or 3D-LC) [14]. Because of the high resolving power of LC, ion-suppression effects in MS (caused by overlapping signals from high and low abundance ions) can be reduced or even eliminated. It is possible to enrich low-abundance proteins or peptides presented in complex sample mixtures within a wide dynamic range of concentrations using selected LC methods.

This review will focus on the current practice of LC as applied to proteome analysis, including single and multi-dimensional chromatography coupled with mass spectrometry (MS and/or tandem MS). New LC methods and their potential to aid in enhanced identification and analysis of cellular proteins are also addressed.

2. The role of liquid chromatography in proteomics

2.1. Liquid chromatography–mass spectrometry

Currently, most LC separations in proteomics are done in RPLC mode, because of its compatibility with MS [10]. The mobile phase in RPLC normally contains a mixture of water and a water miscible organic solvent such as acetonitrile. Acid (formic, acetic or trifluoroacetic) is added to the mobile phase to render all of the component proteins and peptides positively charged and denatured and to reduce unwanted

ionic interactions with the stationary phase. Trifluoroacetic acid concentrations are limited because of its ion suppressive effects. The chromatography step could be viewed as a "desalting" step prior to introducing a protein or peptide sample into the MS. However, the increase in resolution offered by LC separation greatly enhances MS detection of sample components. High-resolution separation of components reduces the number of co-eluting analytes and therefore reduces ion suppression in MS [1]. Although relatively complex mixtures can be well separated in RPLC, the analysis of mixtures in proteomics experiments containing thousands of peptides or proteins has demanded development of separation techniques with higher resolution.

2.1.1. "Ultra" high pressure capillary RPLC–MS

It has been demonstrated that "ultra" high pressure HPLC can dramatically increase separation speed and hence decrease analysis time [15,16]. At the back pressures of 20 000 psi (1 psi = 6894.76 Pa), efficiencies up to 570 000 plates m^{-1} were obtained using a column packed with 1.5 μm non-porous isohexylsilane-modified (C_6) silica beads [15]. "Ultra" high pressure HPLC was combined with a hybrid-quadrupole time-of-flight (Q-TOF) MS or a Fourier transform ion cyclotron resonance (FTICR)-MS through a nano-ESI interface [17–19]. A 87 cm × 14.9 μm i.d. long fused silica capillary was packed with 3 μm C_{18} porous silica particles [19]. Separation of a proteolytic digest of soluble yeast proteins gave peak capacities up to about 1000 at a back pressure of 10 000 psi. Recent efforts to increase sensitivity have applied a micro-solid-phase extraction (MicroSPE) step followed by capillary LC coupled to an FTICR-MS through a nanoscale ESI interface [20]. Fig. 1 depicts MicroSPE–nanoLC–ion trap-MS–MS spectrum from 0.25 ng (A) and 10 ng (B) injections of a *Deinococcus radiodurans* lysate tryptic digest sample. In this report, some proteins were identified at a level of 0.5 pg. This technique may therefore enhance the sensitivity and expand the scope of analysis to smaller cell populations and possibly to single mammalian cells [10]. A fully-automated capillary LC–9.4 T FTICR was recently developed to improve the throughput of proteome analyses [21]. This high efficiency RPLC–MS strategy has also been successfully extended to 2D-LC to increase the peak capacity further [22].

The "ultra" HPLC approach offers a wide range of applications, and is highly attractive for those situations which require ultra-high sensitivity such as those done with limited sample quantities [19]. The limitations of this approach are that it requires special high pressure pumping equipment and that its peak capacity is still too low to adequately resolve complex mixtures often encountered in proteomics.

2.1.2. Protein identification by accurate mass tags using LC–Fourier transform ion cyclotron resonance MS

A unique protein is theoretically identified by discovery of several peptides in tandem MS (MS–MS) experiments. In some cases, identification of a single peptide has proven adequate for the identification of a protein [23]. On the basis of these observations, a new concept called "accurate mass tag" (AMT) was introduced for proteome-wide protein identification to obviate the use of MS–MS in protein identification [23]. AMT uses identification of a single peptide whose mass can be measured with high-mass accuracy using FTICR-MS and is unique among all of the possible peptides predicted from a protein. This approach strongly depends on the accuracy of high mass measurement and employs a single high-resolution capillary LC separation combined with FTICR-MS to validate polypeptide AMTs [23]. In this pioneering work, a large fraction of the measured peptide masses can be designated as AMTs at extremely high mass accuracy (ca. 1 ppm) according to the analysis results of the proteins and their tryptic peptides predicted from *Saccharomyces cerevisiae* and *Caenorhabditis elegans* genomes. For example, at 0.1 ppm mass accuracy, for predicted peptides with a mass of about 2000 units, approximately 65% for *S. cerevisiae* have unique masses and therefore can act as AMTs to identify a unique protein. This approach was also applied to the global analysis of *Deinococcus radiodurans*. The authors claimed a high (61%) coverage of the predicted proteins from this organism [24].

The AMT strategy extends the sensitivity, dynamic range, comprehensiveness, and throughput of proteomic measurements [23,25] and also has suggested a possible way for polypeptide identification from the largest genomic databases yet available [26]. Current limitations to this approach include the necessity of expensive and complex instrumentation and the continued difficulty in the detection of low-abundance proteins in the presence of proteins at significantly higher concentrations.

2.1.3. Elution-modified displacement chromatography–MS

Most chromatographic separations are carried out using linear elution chromatography. That is, components are loaded onto the stationary phase in the linear portion of their binding isotherms and are eluted in (more or less) symmetrical "Gaussian" distributions. Although linear elution chromatography can lead to the concentration of peptides and proteins, the final relative concentrations of the separated components remain the same as they were in the original mixture. That is, linear elution chromatography does not allow for selective enrichment of low-abundance components in complex mixtures.

Non-linear chromatography (displacement chromatography) [27], which has received relatively little attention since its invention more than 60 years ago, was shown to allow for enhancement of the concentrations of low abundance peptides in a tryptic digest of recombinant human growth hormone (rhGH) [28]. Using displacement chromatography, it was shown that trace peptides in rhGH tryptic digests representing chemically modified peptides present at less than 0.1% were enhanced in concentration relative to the major tryptic peptides. The enhanced concentration of these

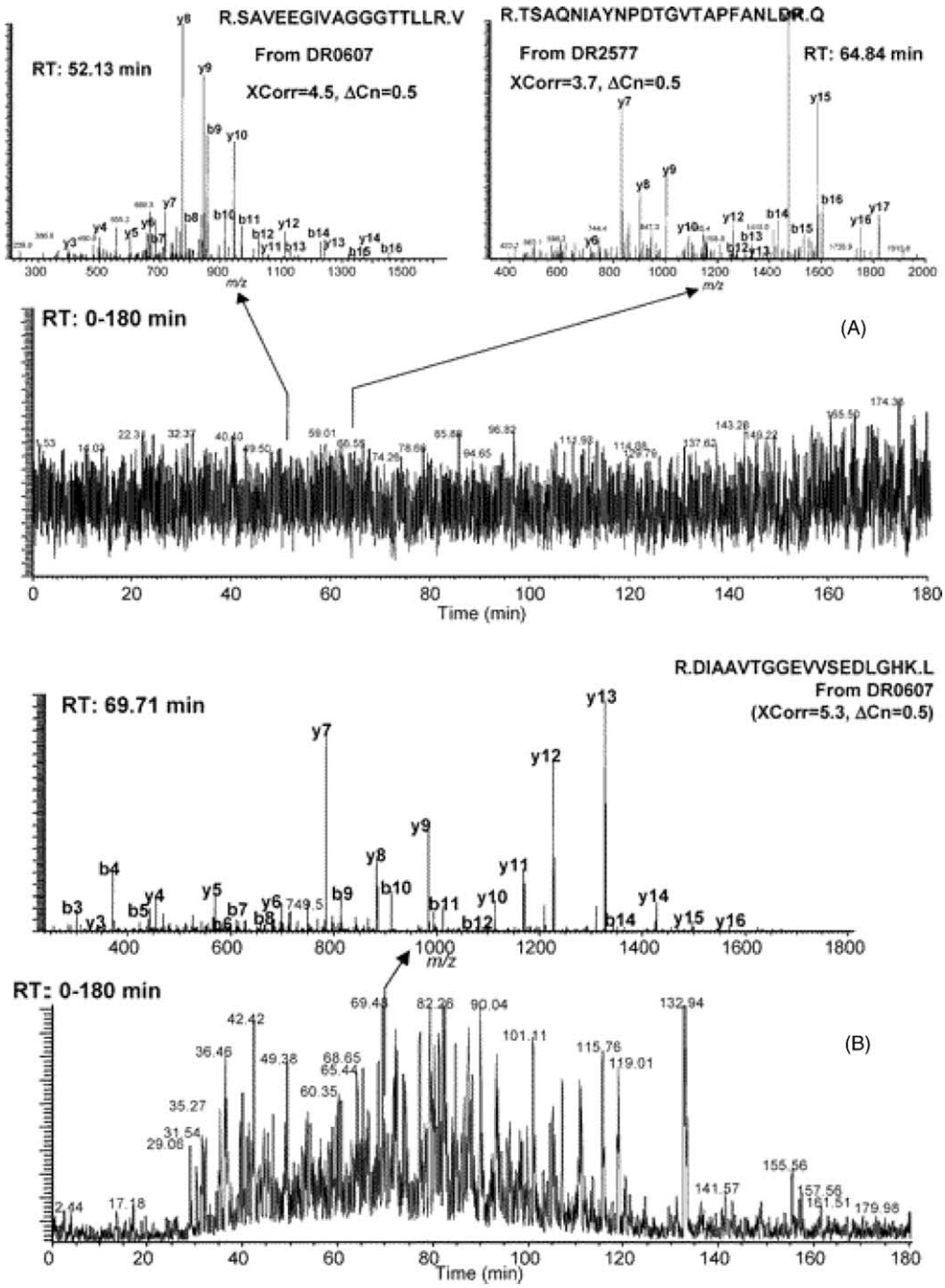

Fig. 1. Total ion chromatograms and selected MS–MS spectra of MicroSPE–nanoLC–ion trap-MS–MS from 0.25 ng (A) and 10 ng (B) injections of a *D. radiodurans* lysate tryptic digest sample. Reprinted with permission from [20].

modified peptides allowed their isolation and further characterization by MS methods.

More recently, we showed the application of elution-modified displacement chromatography (EMDC) to the analysis of complex peptide mixtures such as a tryptic digest of serum proteins [29,30]. EMDC improves upon traditional displacement by reducing the time necessary for the experiment from several hours to minutes and it allows selective enrichment of low-abundance components in complex mixtures. EMDC therefore represents a novel approach to displacement separations in which elements of elution and displacement chromatography are combined to offer a new one-dimensional (1D)-LC technique for the analysis of complex peptide mixtures [29,30]. Rapid and selective trace enrichment and high-resolution separations were achieved in a single chromatographic step using EMDC. This non-linear chromatographic technique is a hybrid technique combining features of elution and displacement chromatography.

Techniques that allow for enrichment of trace peptides and proteins are critically important in view of the wide dynamic concentration range of the proteome, which is estimated to be as high as 10^{12} in blood serum for example [29]. EMDC allowed detection and identification by MS–MS of 5 fmol of a marker peptide mixed with a tryptic digest of rhGH at weight ratios between $1{:}10^5$ and $1{:}10^6$. In the same study, the marker peptide was detected when mixed at low femtomol levels with a tryptic digest of crude bovine serum proteins. In another experiment, a trace amount (20 fmol) of rhGH digest was added to a tryptic digest of bovine serum proteins. Five tryptic peptides were detected in a single chromatographic step by EMDC performed on a conventional 15 cm × 75 μm i.d. RP-HPLC column coupled through a nano-electrospray interface with an ion trap MS [29]. Since enrichment and separation could be attained simultaneously in a single run, the analysis speed was increased dramatically over 2D-LC techniques. This data compared favorably to a study [31] in which a 2D-LC technique was used to separate a similar mixture. The EMDC separation was performed in less than 1 h while the 2D-LC experiment required more than 1 day to complete while identifying a comparable number of peptides.

Limitations of EMDC are that the column must be regenerated after each run and that the separation requires much more sample (up to 1000 times as much) than a corresponding elution separation because of the requirement of the displacement chromatography operating regime. Several columns may be used, however in a series of EMDC runs that can cover the entire range of a conventional elution separation while affording the opportunity to achieve low abundance peptide detection.

2.1.4. Limitations of one-dimensional approaches

Although 1D-LC has been proved to be an economic and effective way for protein and peptide identification, its application in proteomics is relatively restricted by the complexity of the samples. Samples in proteomic analyses often contain thousands of proteins. After proteolytic digestion, peptides numbering in the hundreds of thousands must be separated. This exceeds the analytical range of most 1D-LC methods because of insufficient peak capacity [32]. Therefore, multi-dimensional separations are often required.

2.2. Orthogonal chromatographic techniques in proteomics

To improve resolution, multi-dimensional liquid chromatography is routinely used and has developed rapidly. The most widely used methods involve two orthogonal steps, with the second RPLC column directly coupled with mass spectrometry (2D-LC–MS). In these methods, RPLC is almost always preceded by ion-exchange chromatography. In specific applications, such as the detection of phosphopeptides, affinity techniques are used in the first chromatographic dimension. Size exclusion has also been used in the first dimension [3].

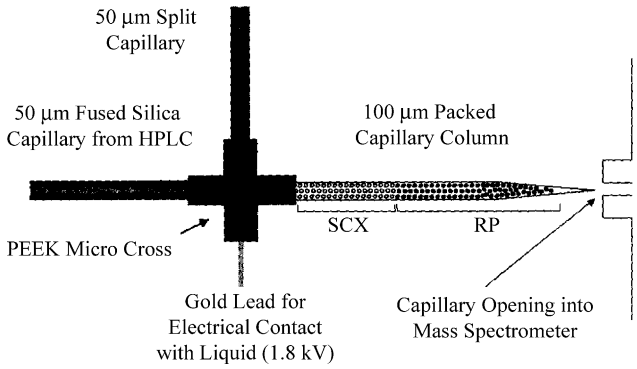

Fig. 2. MudPIT electrospray interface including a biphasic micro-capillary column, packed with SCX and RP packing material connected to a micro-cross. Reprinted with permission from [3].

2.2.1. Ion-exchange chromatography–RPLC

Ion-exchange chromatography (IEX) has long been used for protein and peptide analysis and is often used as the first dimension in 2D-LC separations. An example of orthogonal 2D-LC–MS analysis of peptides was described recently [33]. Peptides were separated in the first dimension using gradient elution cation-exchange chromatography. Fractions were collected and aliquots of each were separated using RP-HPLC coupled directly to ESI-MS. The combination of IEX–RPLC has the following characteristics: (1) IEX offers high capacity; (2) both IEX and RPLC offer high resolution; (3) most 2D-LC systems use RPLC as the second dimension because of its compatibility with ESI-MS; (4) both chromatographic modes provide high separation efficiency [10,33]. The total peak capacity in this 2D separation can be greater than 5000 [33] and high-sensitivity peptide identification may therefore be achieved because of increased resolution and the resultant decrease in peptide overlap. Using this approach, 213 and 280 proteins were identified from human brain glioma cells [34] and A431 epidermal cancer cell line [35], respectively. The limitation of 2D-LC techniques is primarily related to the time required to achieve the separations. Since many individual first dimension fractions must each be fractionated in long (e.g. 2 h) RPLC gradient elution runs, complete 2D-LC runs may require several days to complete.

An innovative approach to 2D chromatography termed “multi-dimensional protein identification technology” (MudPIT) (Fig. 2) was introduced [3,4]. In this approach, a protein mixture is digested with trypsin and peptides are separated using a “hybrid” 2D-LC column coupled to a tandem mass spectrometer. The “hybrid” column used in the MudPIT technique contains zones of strong cation-exchange (SCX) and octadecyl silica stationary phases both packed into a fused silica nanospray tip. The effluent from the column is sprayed directly into the mass spectrometer. MudPIT was applied to profile the whole proteome of the *S. cerevisiae* strain BJ5460 grown to mid-log phase. Some basic proteins from this organism with high isoelectric points are very difficult to separate and identify by 2D-PAGE [36]. In the MudPIT experiment, 5540 peptides were assigned to MS spectra, leading to the

identification of a total of 1484 proteins. Peptides were identified employing search algorithms, such as SEQUEST [37] or similar search protocols that interface directly with databases such as NHLBI or SWISSPROT [4]. MudPIT was also used to identify soluble and membrane proteins from crude rat brain homogenate [38]. The membrane was treated at high pH, and then digested with proteinase K, a robust non-specific protease. The peptide mixture produced was analyzed by using MudPIT, resulting in the identification of 1610 proteins, 454 of which were predicted membrane proteins. This strategy appears very appealing in the comprehensive identification of protein components from various sources [11].

Most recently, an ultra-high-efficiency SCX–RPLC–MS–MS (measured chromatographic peak capacities were greater than 10 000) was introduced [22] for characterization of the human plasma proteome. High-efficiency nanoscale RPLC with chromatographic peak capacities of about 1000 [20], as well as SCX LC was used to obtain ultra-high-efficiency separations in conjunction with MS–MS. Using this method, from a total of 365 μg of human plasma, up to 1682 human proteins were identified.

2.2.2. Other two-dimensional chromatography techniques

2.2.2.1. Affinity chromatography–RPLC. Post-translational modifications such as phosphorylation and glycosylation and proteolytic processing play important roles in the behavior and interactions of cellular proteins. One of the central goals in proteomics is to understand the location, timing and extent of these modifications. However, modified proteins are sometimes present at very low concentrations and are therefore difficult to detect and characterize in complex mixtures containing high concentrations of other cellular proteins. Protein phosphorylation is often substoichiometric, with individual proteins modified to levels of 10% or less on a molar basis [39]. In addition, phosphorylated peptides undergo neutral loss of phosphate during MS, which can further limit their detectability [39]. Affinity chromatography is an effective and reproducible method to specifically isolate targeted low-abundance proteins and peptides. Immobilized metal affinity chromatography (IMAC) [40–42] and immunoaffinity chromatography (IAC) [43–45] have proven particularly effective in the capture of phosphopeptides. Lectin affinity chromatography [46,47], and inhibitor affinity chromatography [48] have also received some attention in the analysis of complex proteomes. IMAC has been successfully used for the isolation of phosphorylated proteins or peptides from complex mixtures. For phosphopeptide analysis, IMAC utilizes immobilized Fe(III) or Ga(III) ions to selectively retain phosphorylated proteins and peptides [49]. This method relies on high affinity interaction between transition metal ions and phosphorylated side chains of serine, threonine or tyrosine [40,50–53]. In this technique, a stationary phase with an immobilized transition metal binding functionality (e.g. imidodiacetic acid) is first "charged" with either Fe(III) or Ga(III). The transition metals therefore form tight complexes with the stationary phase. Peptide digests of phosphorylated proteins are loaded onto the column followed by washing and selective elution.

On-line micro-IMAC–ESI-MS for the analysis of low levels of bovine casein (less than 30 pmol) was reported [40]. IMAC capture was followed by RPLC with on-line detection of eluted peptides by ESI-MS. Sensitivity was shown to reach 250 fmol of injected peptide [41].

The advantages of IMAC over other affinity technologies are ligand stability, high protein loading, mild elution conditions, simple regeneration, high applicability under denaturing conditions and low cost [42]. However, the specificity of the technique is still questionable. Furthermore, detection of low-abundance phosphopeptides is hindered in part because of interference by non-phosphorylated histidine-containing peptides at higher loading levels [30]. To prevent non-phosphorylated peptide binding to the IMAC column through carboxyl groups, a protocol of converting peptides to methyl esters prior to analysis was developed recently [54]. With this approach, more than 1000 phosphopeptides, as well as 383 phosphorylation sites were identified from *S. cerevisiae* whole cell lysate in a single analysis. This approach is also very sensitive. Phosphopeptides at low femtomol levels and phosphoproteins with low codon-bias were identified, because binding of non-phosphorylated peptides is minimized. Other potential improvements include using strong anion-exchange chromatography prior to IMAC to decrease the complexity of IMAC-purified phosphopeptides [55] or adopting a multi-dimensional affinity chromatography scheme [56].

Immunoaffinity chromatography is another powerful tool for the isolation and purification of phosphorylated peptides or proteins [43,44]. In this technique, the antibody, such as anti-phosphoserine, tyrosine or threonine, is immobilized on a resin. Peptide digests of phosphoproteins are loaded onto the column followed by washing and selective elution of the antigenic peptides. The advantage of IAC is high selectivity due to the high affinity of antigen–antibody interaction [44].

As mentioned earlier, glycosylation also plays an essential role in the functionality of many proteins and in various cellular processes, such as cell–cell recognition. Lectin affinity chromatography has proven to be one of the most powerful steps available for the isolation of specific classes of glycoproteins and can also be used to enrich glycopeptides [46,47]. Because a wide variety of lectins are available, this technique should prove quite useful in the characterization of glycoproteins and glycopeptides. Lectin affinity chromatography was combined with RPLC in a recent study of lymphosarcoma in dogs. Fractions from RPLC were analyzed by both matrix-assisted laser desorption/ionization (MALDI) TOF-MS and ESI-MS. Using this method, two proteins, CD 44 and E-selectin, which are both known to participate in cell adhesion and cancer cell migration, were identified [46].

2.2.2.2. Size-exclusion chromatography–RPLC. Size-exclusion chromatography (SEC) is occasionally used as a first

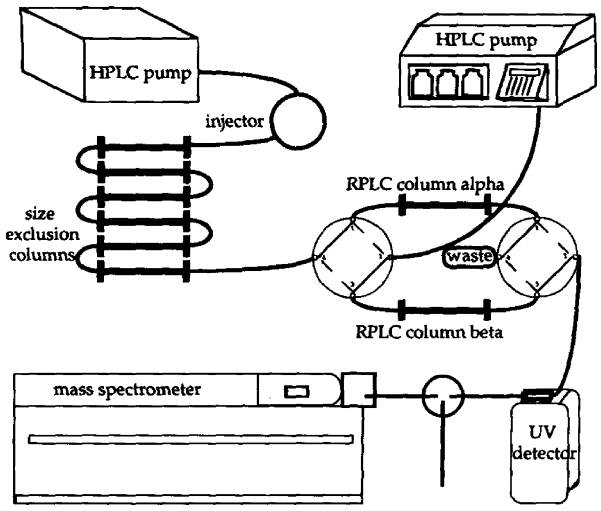

Fig. 3. Schematic of the 2D-LC–LC–MS system showing the six SEC columns, the new parallel column LC–LC interface, and the dual detection scheme of UV and mass spectrometry. Reprinted with permission from [60].

dimension in 2D-LC separations. The technique, which separates proteins based on differences in their Stokes radii, has often been ignored in proteome research because of its limited loading capacity and resolving power. This probably accounts for the fact that few applications of SEC in proteome research have appeared in the literature. However, SEC has the advantages of high reproducibility, stability and relatively short analysis time [57,58]. A few reports appeared using a comprehensive 2D SEC–RPLC technique for proteome research [59–61]. In these experiments, peptide fragments were separated by SEC followed by RPLC. The chromatographic separation system was coupled to an ESI-MS for on-line protein identification.

The experimental configuration is shown in Fig. 3. The sample was injected into the first of several (up to 12) 300 mm × 7.8 mm i.d. SEC columns connected in series. Then, the eluted sample was alternately loaded in 4-min wide plugs onto either of a pair of polystyrene–divinylbenzene (PS–DVB) 2.1 mm i.d. RPLC columns for further separation. These two RPLC columns were used in parallel. One desalts the sample from the SEC columns, and the other is used for separation; eluted peptides travel directly through a UV detector and then to the fraction collector. Unlike storage loops or fraction collection [59], this novel LC–LC interface joins two chromatographic dimensions effectively.

This comprehensive high-resolution 2D-LC–MS system with six standard SEC columns was used to map the proteolytic fragments created from the tryptic digestion of ovalbumin (M_r 44 000) and bovine serum albumin (66 000). Peak capacities up to 495 were attained. This 2D chromatographic system compared favorably to a single 25 cm × 2.1 mm i.d. 1D-RPLC system with a peak capacity of only 45 [60]. An increase in the effective column length of the first dimension by 50%, and tripling the length of the second-dimension columns gave theoretical peak capacities up to 1500 [61].

More recently, SEC–RPLC was used successfully in proteomics studies of yeast [62], immunoglobulin fusion proteins [63], cytochrome b_6f complex [64] and other complex protein or peptide mixtures [32].

3. Role of chromatography in quantitative proteomics

One of the goals in proteomics is to compare the relative amounts of different proteins in biological samples and to correlate these differences with changes in physiological state.

Many quantitative proteomics analyses have been performed using 2D-PAGE coupled with MS or MS–MS. However, the limitations of 2D-PAGE have been discussed [65]. As outlined earlier, chromatographic separation can greatly increase the sensitivity of MS detection and is a key element in quantitative proteomics. A number of reviews have been published recently [36,66–68], therefore the focus here will be placed on chromatographic techniques as applied to quantitative proteomics.

The isotope-coded affinity tag (ICAT) technique takes advantage of differential tagging of cysteine residues in proteins with stable isotopes followed by affinity and ion-exchange chromatography [65,69,70]. Peptides in individual ion-exchange fractions are then identified using online RPLC coupled to MS. Although the cysteine specificity of this technique causes it to ignore approximately 5–10% of all proteins because they contain no cysteine, the advantage offered is that the complexity of peptide mixtures to be analyzed is enormously reduced. This is because cysteine represents approximately 5% or less of the total amino acid content in cellular proteins [65].

The complexity reduction produced by covalent reaction with cysteine-containing peptides is potentially a powerful method to investigate certain aspects of the proteome apart from its original use in quantitative proteomics. Because the number of peptides is substantially lower, the ability to detect low-abundance proteins should be enhanced due to the significant reduction of peak overlap and therefore of ion suppression in MS. The design of novel reagents to target other peptides with specific characteristics can greatly simplify the chromatographic tasks faced in global proteome analyses.

Despite many improvements, the disadvantages of ICAT still remain. ICAT as currently configured still requires avidin affinity and ion-exchange chromatography steps after labeling before the final LC–MS separation can be done. This leads to a loss of some peptides, especially those in low abundance, and therefore reduces the sensitivity of the method while non-cysteine-containing contaminants are still present.

Besides ICAT, many other stable isotope coding techniques applied in quantitative proteomics have appeared in the literature and have been extensively reviewed [71]. In most cases, N-terminus or C-terminus of the peptide, or a specific amino acid in the sequence, such as cysteine or lysine

is selected to label. Chromatographic steps, e.g. affinity, SCX and RPLC are usually utilized for fractionating the coded protein digests [36,71]. Separated peptides are then measured quantitatively using mass spectrometry. In the meanwhile, most research work is focused on designing isotope-coded reagents to target peptides with specific characteristics to greatly simplify the chromatographic tasks faced in global proteome analyses. Less effort has been put to improve the chromatography performance. Problems, such as separation of heavy and light sample components, co-elution of multiple peptides due to the lack of resolution, and no enrichment of low-abundance species, need to be solved in order to gain accurate and comprehensive quantitative measurement.

Since the detection and identification of all proteins seems impossible because the stable isotope coding techniques are designed to identify a subset of peptides (for example, cysteine-containing peptides), a global protein identification and quantification technology was introduced by using 2D SCX–RPLC–MS–MS of protein digests for quantification [35,72]. However, the weakness of this technology could be the inaccurate quantitation caused by overlapping of individual peaks due to the unresolved sample components.

4. Conclusions and perspectives

This paper has reviewed recent development of liquid chromatography coupled with (tandem) MS and pertinent techniques in proteome research. LC techniques have developed dramatically and their importance in proteomics has been established. Liquid chromatography plays an important role in proteomics, however improved methods will be required to advance understanding of cellular function using proteomics techniques. The objective of chromatography in proteomics is to provide high-resolution, high-speed, high-sensitivity and high-specific separation of extremely complex mixtures and to facilitate mass spectrometry detection and quantitative measurement. Technologies that have exhibited promise include novel monolithic stationary phases in capillary columns, high-temperature separation and capillary electrochromatography (CEC).

Monolithic stationary phases have been developed and used in capillary columns as an alternative to granular packed beds. Monolithic columns have attracted a great deal of interest because of their ease of preparation, reliable performance, good permeability and versatile surface chemistry [73]. Silica based capillary monolithic reversed-phase columns have been successfully applied in proteome research [74]. Because the porosity of monolith can be easily controlled by adjusting the composition and concentration of the porogens, highly porous monoliths have been made and high-speed LC separation was achieved by running at high flow rates at reasonable back pressures. The resolution of peptide mapping was very similar to that achieved by packed bed columns. In addition, monolithic columns do not require frits. They can therefore be coupled directly to the mass spectrometer and are ideal stationary phases in CEC.

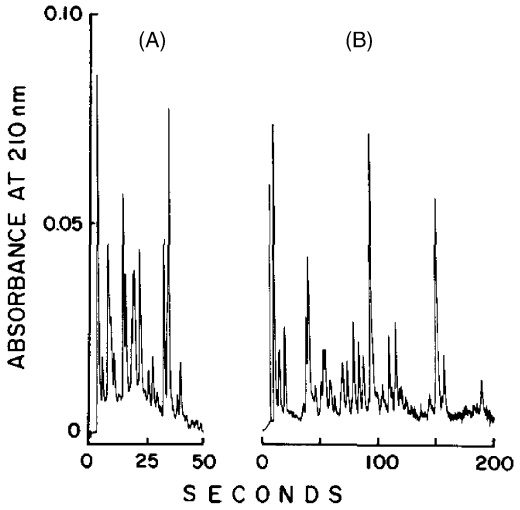

Fig. 4. Chromatographic profiles of tryptic digests of β-lactoglobulin A (A) and methionyl human growth hormone (B). Reprinted with permission from [81].

Since the task of proteomics is to analyze thousands of proteins simultaneously, high-speed and high-throughput separation and identification are required. Current protocols, such as 2D-PAGE or 2D-LC–MS requires hours to days to complete. A central theme in this area is to increase the speed of separations. High-speed chromatographic separations have been achieved using elevated temperature and non-porous pellicular stationary phases to reduce solvent viscosity and to improve column mass transfer characteristics [75–80]. In one study, a protein mixture was successfully separated at 80 °C in 20 s using a column packed with 2 μm non-porous silica microspheres [81]. The separations of the tryptic digests of β-lactoglobulin A and methionyl human growth hormone were obtained in 1 and 3 min, respectively, under the conditions above, as shown in Fig. 4. Issaeva et al. [82] reported rapid separation of a six-protein mixture using a commercially available non-porous, granular RP silica stationary phase. The complete separation was achieved in only 6 s. Elevated temperature enhances both speed and efficiency of LC separations of proteins and peptides by enhancement of kinetic and transport properties in the column [83,84].

Another emerging LC technique for proteomics is CEC. CEC is a “hybrid” technique that combines aspects of LC and capillary electrophoresis (CE) [85]. Some promising results have been shown in protein and peptide separations [73,77]. However, CEC has so far not been used successfully to separate very complex peptide mixtures. Reasons for the slow development of CEC in the area of protein and peptide separation include difficulty of column fabrication and the lack of suitable instrumentation. The development of new monolithic stationary phases is expected to facilitate the progress of CEC of proteins and peptides [86]. Separation on microchips promises to provide high-throughput separation platforms in the future. Because they are both electrokinetic separation

techniques, development of capillary-based CEC technologies will suggest ideas for the development of better separations using the microchip platform which may be used in a high throughput mode.

Acknowledgements

This work was supported by grant No. GM 20993 from the National Institutes of Health, US Department of Health and Human Services and a research collaboration with Thermoelectron Inc.

References

[1] P. James, Proteome Research: Mass Spectrometry, Springer, Berlin, Heidelberg, New York, 2001.
[2] J.R. Kettman, J.R. Frey, I. Lefkovits, Biomol. Eng. 18 (2001) 207.
[3] D.A. Wolters, M.P. Washburn, J.R. Yates, Anal. Chem. 73 (2001) 5683.
[4] M.P. Washburn, D. Wolters, J.R. Yates, Nat. Biotechnol. 19 (2001) 242.
[5] A. Pandey, M. Mann, Nature 405 (2000) 837.
[6] J.W. Baty, M.B. Hampton, C.C. Winterbourn, Proteomics 2 (2002) 1261.
[7] M. Unlu, M.E. Morgan, J.S. Minden, Electrophoresis 18 (1997) 2071.
[8] J.X. Yan, A.T. Devenish, R. Wait, T. Stone, S. Lewis, S. Fowler, Proteomics 2 (2002) 1682.
[9] S.P. Gygi, G.L. Corthals, Y. Zhang, Y. Rochon, R. Aebersold, Proc. Natl. Acad. Sci. U.S.A. 97 (2000) 9390.
[10] Y.F. Shen, R.D. Smith, Electrophoresis 23 (2002) 3106.
[11] H. Wang, S. Hanash, J. Chromatogr. B 787 (2003) 11.
[12] F.C. Leinweber, D.G. Schmid, D. Lubda, K.H. Wiesmuller, G. Jung, U. Tallarek, Rapid Commun. Mass Spectrom. 17 (2003) 1180.
[13] D. Figeys, Anal. Chem. 75 (2003) 2891.
[14] J.C Giddings, Dynamics of Chromatography, Marcel Dekker, New York, 1965.
[15] J.A. Lippert, B. Xin, N. Wu, M.L. Lee, J. Microcol. Sep. 11 (1999) 631.
[16] J.E. MacNair, K.C. Lewis, J.W. Jorgenson, Anal. Chem. 69 (1997) 983.
[17] Y.F. Shen, N. Tolic, R. Zhao, L. Pasa-Tolic, L.J. Li, S.J. Berger, R. Harkewicz, G.A. Anderson, M.E. Belov, R.D. Smith, Anal. Chem. 73 (2001) 3011.
[18] Y.F. Shen, R. Zhao, M.E. Belov, T.P. Conrads, G.A. Anderson, K.Q. Tang, L. Pasa-Tolic, T.D. Veenstra, M.S. Lipton, H.R. Udseth, R.D. Smith, Anal. Chem. 73 (2001) 1766.
[19] Y.F. Shen, R. Zhao, S.J. Berger, G.A. Anderson, N. Rodriguez, R.D. Smith, Anal. Chem. 74 (2002) 4235.
[20] Y.F. Shen, N. Tolic, C. Masselon, L. Pasa-Tolic, D.G.I. Camp, K.K. Hixson, R. Zhao, G.A. Anderson, R.D. Smith, Anal. Chem. 76 (2004) 144.
[21] M.E. Belov, G.A. Anderson, M.A. Wingerd, H.R. Udseth, K. Tang, D.C. Prior, K.R. Swanson, M.A. Buschbach, E.F. Strittmatter, R.J. Moore, R.D. Smith, J. Am. Soc. Mass Spectrom. 15 (2004) 212.
[22] Y. Shen, J.M. Jacobs, D.G.I. Camp, R. Fang, R.J. Moore, R.D. Smith, W. Xiao, R.W. Davis, R.G. Tompkins, Anal. Chem. 76 (2004) 1134.
[23] T.P. Conrads, G.A. Anderson, T.D. Veenstra, L. Pasa-Tolic, R.D. Smith, Anal. Chem. 72 (2000) 3349.
[24] M.S. Lipton, L. Pasa-Tolic, G.A. Anderson, D.J. Anderson, D.L. Auberry, K.R. Battista, M.J. Daly, J. Fredrickson, K.K. Hixson, H. Kostandarithes, C. Masselon, L.M. Markillie, R.J. Moore, M.F. Romine, Y.F. Shen, E. Stritmatter, N. Tolic, H.R. Udseth, A. Venkateswaran, L.K. Wong, R. Zhao, R.D. Smith, Proc. Natl. Acad. Sci. U.S.A. 99 (2002) 11049.
[25] R.D. Smith, G.A. Anderson, M.S. Lipton, L. Pasa-Tolic, Y.F. Shen, T.P. Conrads, T.D. Veenstra, H.R. Udseth, Proteomics 2 (2002) 513.
[26] C. Masselon, G.A. Anderson, R. Harkewicz, J.E. Bruce, L. Pasa-Tolic, R.D. Smith, Anal. Chem. 72 (2000) 1918.
[27] A. Tiselius, Ark. Kem. Miner. Geol. 16A (1943) 1.
[28] J. Frenz, J. Bourell, W.S. Hancock, J. Chromatogr. 512 (1990) 299.
[29] R. Xiang, Cs. Horváth, J.A. Wilkins, Anal. Chem. 75 (2003) 1819.
[30] J.A. Wilkins, R. Xiang, Cs. Horváth, Anal. Chem. 74 (2002) 3933.
[31] S.L. Wu, H. Amato, R. Biringer, G. Choudhary, P. Shieh, W.S. Hancock, J. Proteome Res. 1 (2002) 459.
[32] H. Liu, D. Lin, J.R. Yates, BioTechniques 32 (2002) 898.
[33] G.J. Opiteck, K.C. Lewis, J.W. Jorgenson, R.J. Anderegg, Anal. Chem. 69 (1997) 1518.
[34] M.T. Davis, J. Beierle, E.T. Bures, M.D. McGinley, J. Mort, J.H. Robinson, C.S. Spahr, W. Yu, R. Luethy, S.D. Patterson, J. Chromatogr. B 752 (2001) 281.
[35] D. Chelius, T. Zhang, G. Wang, R.F. Shen, Anal. Chem. 75 (2003) 6658.
[36] M.J. MacCoss, J.R. Yates, Curr. Opin. Clin. Nutr. Metab. Care 4 (2001) 369.
[37] J.R. Yates, A.L. McCormack, D. Schieltz, E. Carmack, A. Link, J. Protein Chem. 16 (1997) 495.
[38] C.C. Wu, M.J. MacCoss, K.E. Howell, J.R. Yates, Nat. Biotechnol. 21 (2003) 532.
[39] D.T. McLachlin, B.T. Chait, Curr. Opin. Chem. Biol. 5 (2001) 591.
[40] L.M. Nuwaysir, J.T. Stults, J. Am. Soc. Mass Spectrom. 4 (1993) 462.
[41] J.D. Watts, M. Affolter, D.L. Krebs, R.L. Wange, L.E. Samelson, R. Aebersold, J. Biol. Chem. 269 (1994) 29520.
[42] G.P. Vladka, V. Menart, J. Biochem. Biophys. Methods 49 (2001) 335.
[43] M.G. Weller, J. Fresenius Anal. Chem. 366 (2000) 635.
[44] R.R. Burgess, N.E. Thompson, Curr. Opin. Biotechnol. 13 (2002) 304.
[45] R. Pieper, Q. Su, C.L. Gatlin, S.T. Huang, N.L. Anderson, S. Steiner, Proteomics 3 (2003) 422.
[46] L. Xiong, D. Andrews, F.E. Regnier, J. Proteome Res. 2 (2003) 618.
[47] J. Hirabayashi, Y. Arata, K. Kasai, Proteomics 1 (2001) 295.
[48] G. Lolli, F. Thaler, B. Valsasina, F. Roletto, S. Knapp, M. Uggeri, A. Bachi, V. Matafora, P. Storici, A. Stewart, H.M. Kalisz, A. Isacchi, Proteomics 3 (2003) 1287.
[49] G.S. Chaga, J. Biochem. Biophys. Methods 49 (2001) 313.
[50] J. Porath, B. Olin, B. Granstrand, Arch. Biochem. Biophys. 225 (1983) 543.
[51] J. Porath, Protein Expr. Purif. 3 (1992) 263.
[52] M. Mann, R.C. Hendrickson, A. Bandey, Annu. Rev. Biochem. 70 (2001) 437.
[53] M.C. Posewitz, P. Tempst, Anal. Chem. 71 (1999) 2883.
[54] S.B. Ficarro, M.L. McCleland, P.T. Stukenberg, D.J. Burke, M.M. Ross, J. Shabanowitz, D.F. Hunt, F.M. White, Nat. Biotechnol. 20 (2002) 301.
[55] T.S. Nuhse, A. Stensballe, O.N. Jensen, S.C. Peck, Mol. Cell. Proteomics 2 (2003) 1234.
[56] W.C. Lee, K.H. Lee, Anal. Biochem. 324 (2004) 1.
[57] P. Lecchi, F.P. Abramson, J. Chromatogr. A 828 (1998) 509.
[58] P. Lecchi, F.P. Abramson, Anal. Chem. 71 (1999) 2951.
[59] J.W. Jorgenson, G.J. Opiteck, S.M. Ramirez, M.A. Moseley, R.J. Anderegg, Abstr. Pap. Am. Chem. Soc. 214 (1997) 157.
[60] G.J. Opiteck, J.W. Jorgenson, R.J. Anderegg, Anal. Chem. 69 (1997) 2283.
[61] G.J. Opiteck, S.M. Ramirez, J.W. Jorgenson, M.A. Moseley, Anal. Biochem. 258 (1998) 349.
[62] J. Gao, G.J. Opiteck, M.S. Friedrichs, A.R. Dongre, S.R. Hefta, J. Proteome Res. 2 (2003) 643.

[63] J.F. Nemeth-Cawley, B.S. Tangarone, J.C. Rouse, J. Proteome Res. 2 (2003) 495.
[64] J.P. Whitelegge, H.M. Zhang, R. Aguilera, R.M. Taylor, W.A. Cramer, Mol. Cell. Proteomics 1 (2002) 816.
[65] S.P. Gygi, B. Rist, S.A. Gerber, F. Turecek, M.H. Gelb, R. Aebersold, Nat. Biotechnol. 17 (1999) 994.
[66] M.B. Goshe, R.D. Smith, Curr. Opin. Biotechnol. 14 (2003) 101.
[67] R. Aebersold, H. Lee, D. Han, M. Wright, H.L. Zhou, T. Griffin, S. Purvine, D. Goodlett, FASEB J. 16 (2002) 12.
[68] R. Aebersold, J. Infect. Dis. 187 (2003) 315.
[69] W.A. Tao, R. Aebersold, Curr. Opin. Biotechnol. 14 (2003) 110.
[70] A.C. Gingras, S.P. Gygi, B. Raught, R.D. Polakiewicz, R.T. Abraham, M.F. Hoekstra, R. Aebersold, N. Sonenberg, Genes Dev. 13 (1999) 1422.
[71] S. Julka, F.E. Regnier, J. Proteome Res. (2004) in press.
[72] D. Chelius, P.V. Bondarenko, J. Proteome Res. 1 (2002) 317.
[73] I. Gusev, X. Huang, Cs. Horváth, J. Chromatogr. A 855 (1999) 273.
[74] B. Barroso, D. Lubda, R. Bischoff, J. Proteome Res. 2 (2003) 633.
[75] M.H. Chen, Cs. Horváth, J. Chromatogr. A 788 (1997) 51.
[76] H. Chen, Cs. Horváth, J. Chromatogr. A 705 (1995) 3.
[77] S.H. Zhang, X. Huang, J. Zhang, Cs. Horváth, J. Chromatogr. A 887 (2000) 465.
[78] A. Vailaya, Cs. Horváth, Ind. Eng. Chem. Res. 35 (1996) 2964.
[79] A. Vailaya, Cs. Horváth, J. Phys. Chem. B 102 (1998) 701.
[80] D. Haidacher, A. Vailaya, Cs. Horváth, Proc. Natl. Acad. Sci. U.S.A. 93 (1996) 2290.
[81] K. Kalghatgi, Cs. Horváth, J. Chromatogr. 398 (1987) 335.
[82] T. Issaeva, A. Kourganov, K. Unger, J. Chromatogr. A 846 (1999) 1373.
[83] F.D. Antia, Cs. Horváth, J. Chromatogr. 435 (1988) 1.
[84] S.H. Zhang, J. Zhang, Cs. Horváth, J. Chromatogr. A 914 (2001) 189.
[85] Cs. Horváth, J. Chromatogr. A 887 (2000) 1.
[86] Y. Li, R. Xiang, Cs. Horváth, J.A. Wilkins, Electrophoresis 25 (2004) 545.