Published in final edited form as:

J Chromatogr A. 2007 June 22; 1154(1-2): 295–307.

Ultratrace liquid chromatography/mass spectrometry analysis of large peptides with post-translational modifications using narrow-bore poly(styrene-divinylbenzene) monolithic columns and extended range proteomic analysis

Jian Zhang^{†,++}, Shiaw-Lin Wu[†], Jeongkwon Kim, and Barry L. Karger^{*}

Barnett Institute and Department of Chemistry and Chemical Biology, Northeastern University, Boston, MA 02115, USA.

Abstract

This paper describes approaches to optimize the chromatographic performance for our recently developed LC-MS platform, extended range proteomic analysis (ERPA), for comprehensive protein characterization at the ultratrace level. Large digested peptide fragments up to 10 kDa (*e.g.* from lysyl endopeptidase digestion) with or without modifications were well separated with high resolution using narrow bore (20 and 50 μ m I.D.) poly(styrene-divinylbenzene) (PS-DVB) monolithic columns constructed by *in situ* solution polymerization. Importantly, the macroporous structure of the monolithic columns facilitated mass transport of large peptides with improved recovery relative to small pore size reversed-phase packings. High sequence coverage (>95 %), including identification of phosphorylated and glycosylated particles was achieved for β -casein and epidermal growth factor receptor (EGFR) at the 4 fmol and 20 fmol levels per injection, respectively, using the 20 μ m I.D. PS-DVB monolithic column. For peptides with greater ionization efficiency, the detection limit could be lowered to $\sim$ 400 zmol. Typically, the separation system produced a peak capacity of $\sim$ 200 for a 10 cm column. This paper demonstrates that narrow-bore monolithic columns are suitable for high sensitivity and high resolution separation of large peptide fragments by LC-MS analysis.

Keywords

Extended range proteome analysis (ERPA); Protein characterization; Monolithic column; Ultranarrow column

1. Introduction

Recently, we introduced a new LC-MS strategy, extended range proteomic analysis (ERPA), which combines the advantages of reduced sample complexity and improved mass ionization efficiency of modified peptides for enhancement of low level detection of digest products of proteins [1]. ERPA is an intermediate approach between “top-down” and “bottom-up” proteomics. It has as its goal an improvement over the limited sensitivity and limited protein

* To whom correspondence should be addressed: b.karger@neu.edu

[†]: These authors contributed equally to this work.

⁺⁺: Current address: Bristol-Myers Squibb Co., Hopewell, NJ 08534, USA.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

size and heterogeneity of the “top-down” approach [2,3]. In addition, relative to the bottom-up approach, ERPA can generate higher sequence coverage and readily detect post-translational modifications (PTMs) [4,5].

The ERPA platform is focused on the analysis of, on average, larger peptide fragments (up to 10 kDa) than peptide fragments from tryptic digests using proteolytic enzymes such as Lysyl endopeptidase (Lys-C), C-terminal K. Separation of peptide fragments of 3 kDa and higher, including post-translational modifications (*e.g.* glycosylation), results in the need to select an LC column different from that typically used for tryptic digests. In particular, the pores in the separation media must be sufficiently open and large to facilitate free transport of higher molecular weight peptides in order to minimize band broadening. Furthermore, it is well known that the recovery of such peptides on a column is related to the pore diameter, pore size distribution and pore shape, along with the hydrophobicity of the stationary phase [6]. Thus, reversed-phase packings that are typically used for tryptic peptides, *e.g.* 20 nm pore packing materials, will generally not be optimum for analysis of large peptide digest fragments. In addition, large peptides tend to be retained longer on C-18 reversed-phase columns than small peptides because of their generally greater hydrophobicity, and too strong interaction with the stationary phase can lead to poor recovery from the column.

Organic polymer-based monolithic columns as the separation media for large peptides would appear to offer favorable structures for high recovery of large peptides. First, the monolithic column contains macropores, which facilitate free transport of large peptide fragments. Second, by tuning the phase ratio ϕ in the monolithic column structure, the retention of large peptides can be modulated to balance retention and recovery.

The concentration of clinically important proteins is often at trace levels. To achieve high sensitivity analysis of biologically important proteins, narrow-bore column technology has been investigated by a number of groups [7-11], in part because electron spray ionization (ESI) MS is concentration sensitive over a wide flow rate range [12,13]. Additionally, very low mobile phase flow rates enhance electrospray ionization efficiency because of the formation of small electrospray droplet sizes [14-16]. However, construction of narrow bore columns using conventional slurry packing techniques with particulate packings imposes technical challenges because of the ultra-high pressures (>10 000 psi) needed to pack such columns. An alternative approach to prepare narrow-bore columns is *in situ* polymerized monolithic columns, which appear as a single piece of porous matrix within the column tubing.

Polymer and silica based monolithic columns have been reported in capillary electrochromatography (CEC) and micro-HPLC for the analysis of a variety of biological species such as proteins, peptides [17-19] and nucleotides [20]. An excellent review on monolithic columns has been recently written [21]. Silica based monolithic columns are prepared by sol-gel process [18-21]. The porous structure has a bimodal pore size distribution, consisting of through pores (1.5-5 μm) and mesopores (~ 10 to 30 nm). These mesopores can hinder free transport of relatively large peptide fragments in the siliceous matrix, which could lead to poor recovery and loss in efficiency. In addition to the mesopores, the pore size distribution and pore shape in the silica monolithic matrix can also affect the efficiency and recovery of large peptide fragments. On the other hand, polymer monolithic columns contain large through pores ($\sim$ a few μm) but little if any mesopores. For tryptic (small) peptide analysis, successful application of 10 and 20- μm I.D. silica based monolithic column has been demonstrated [10,11]. For large peptide separation (Lys-C), as described in this work, we have selected polymer-based monolithic columns. Polymer-based monolithic columns are prepared by *in situ* thermal or UV initiated solution polymerization. The column structure can be tuned by changing the ratio of monomer-to-porogen, the type of porogen and the ratio of porogenic solvent mixtures. It has been reported [22] that the polymer matrix retains high mechanical

stability in commonly used reversed-phase solvent systems such as ACN/H₂O, in part due to the covalent attachment of the porous polymer to the column wall.

This paper reports the combination of narrow bore polymer monolithic columns and the ERPA platform [1]. The effectiveness of polymer-based narrow-bore monolithic columns is demonstrated in terms of enhanced detection limits, improved peptide recoveries, while maintaining high sequence coverage, using two model proteins, β -casein and epidermal growth factor receptor (EGFR). In this study, we obtained > 95% sequence coverage including PTMs at the 4 fmol and 20 fmol levels per injection, respectively. The detection of low-abundant PTM peptides, such as a tetra-phosphorylated β -casein peptide and a partially phosphorylated EGFR peptide at these levels is demonstrated.

2. Experimental

2.1. Materials and reagents

Polyimide-coated fused-silica (FS) (20, 50 and 75 μ m I.D., 360 μ m O.D.) capillary tubing was purchased from Polymicro Technologies (Phoenix, AZ, USA). PTFE tubing (0.3 mm I.D., 0.75 mm O.D.) was from Cole-Palmer Instrument (Vernon Hills, IL, USA). Electrospray emitters (distal coated SilicaTips™, 5 μ m and 10 μ m I.D. tip, pulled from 20 μ m I.D., 360 μ m O.D. FS capillary) and pre-packed Biobasic columns (C-18 and C-4) were purchased from New Objective. (Woburn, MA, USA). Nanovolume fittings (tee, union), 2-position Nanovolume valves (4 nL internal loop model CN4-4344-.004 and 250 nL external loop model CN2-4346EH) were from VICI (Houston, TX, USA). Formic acid, 2-propanol, tetrahydrofuran (THF), 1-decyl alcohol, 3-(trimethoxysilyl)propyl methacrylate (3-TMSPMA) (98%), styrene (99%+), divinylbenzene (80%), *N,N*-dimethylformamide (DMF), 2,2'-diphenyl-1-picrylhydrazyl hydrate (DPPH) and 2,2'-azobisisobutyronitrile (AIBN) (98%) were from Aldrich (Milwaukee, WI, USA). Bovine β -casein, human-epidermal growth factor receptor (EGFR), rabbit IgG, sequencing-grade TPCK-treated trypsin were purchased from Sigma (St. Louis, MO, USA). HPLC-grade methanol, acetone, acetonitrile, 1.0 M sodium hydroxide and 1.0 M hydrochloric acid solutions were from Fisher (Fair Lawn, NJ, USA). Deionized water was prepared using a Milli-Q system from Millipore (Bedford, MA, USA). Lysyl endopeptidase (Lys-C) was purchased from Wako Chemicals (Richmond, VA, USA). Standard tryptic digests (bovine serum albumin, catalase) and Magic C-18 packings (200 Å, 5 μ m) were purchased from Michrom BioResources (Auburn CA, USA).

2.2. Preparation of monolithic columns

The overall preparation procedure followed that previously published [7]. Typically, 1.0 meter of 20 μ m or 50 μ m I.D. FS capillary tubing was washed for 30 min with 1.0 M NaOH solution and flushed with deionized water for 30 min. Then, the tubing was washed with 1.0 M HCl solution for 30 min, followed by flushing with deionized water and acetone. Subsequently, the tubing was placed in an oven at 120°C and purged with nitrogen for 1 h to remove residual water. After cooling, the tubing was filled with the silanization solution containing 30% (v/v) 3-TMSPMA in dry DMF with 0.1% (w/v) DPPH used as inhibitor. With both ends sealed, the tubing was placed in the oven at 110°C for 6 h. After cooling, the capillary was washed extensively with acetone and blown dry under nitrogen. The silanized capillary tubing was cut into ~30 cm a piece. The polymerization solution consisted of monomers of styrene (200 μ L) and DVB (200 μ L), dual porogenic solvent (600 μ L) of THF and 1-decanol with 0.1% (w/v) AIBN as initiator. The volume ratio of the microporogen (THF) to the macroporogen (1-decanol) was 70/530 for the 50 μ m I.D. column and 60/540 for the 20 μ m I.D. column. The surface-area-to-volume ratio increased as column I.D. decreased, and the porogen composition was altered accordingly. The polymerization solution was cleaned using a cellulose membrane filter with 0.22 μ m pores. The filtrate was placed in a 1.5 mL glass vial, briefly aerated with

helium, then capped with double septa (Alltech, Deerfield, IL, USA). A piece of silanized tubing was inserted through the septa into the vial, and the tubing was filled with the polymerization solution under a N₂ pressure of ~50 psi. The outlet end of the tubing was inserted back into the vial, which was then placed in the oven at 70°C. Polymerization was conducted for 16 h under a N₂ pressure of ~50 psi applied in the vial. Upon completion, the synthesized monolithic column was cut off from the vial, flushed for several hours with acetonitrile and blown dry with nitrogen.

2.3. Preparation of samples

A stock solution of 50 pmol/μL bovine β-casein was prepared as follows. First, 50 nmol of bovine β-casein was dissolved in 1 mL of 100 mM NH₄HCO₃ (pH 7.2), and 10 μg endoproteinase Lys-C was added for digestion. After 4 h incubation, the sample was stored at -20 °C. A stock solution of 50 pmol of EGFR was dissolved in a solution containing 6 M guanidine-HCl. The sample solution was reduced with 20 mM dithiothreitol for 30 min at 37° C and then alkylated with 50 mM iodoacetamide for 1 h 30 min at room temperature. A microcon YM-10 filter (Millipore, MA, USA) was used to remove small molecules and buffer exchange (desalt) to 100 mM NH₄HCO₃ (pH 7.9) as follows. The filter was first conditioned with 100 μL of 100 mM NH₄HCO₃ by centrifugation at 12 000 g for 10 min. The sample solution was then loaded and centrifuged at 12 000 g for 10 min. The filtrate was discarded, and 100 μL of 100mM NH₄HCO₃ was added to the retentate for an additional centrifugation of 10 min. The above step was repeated three times to complete the buffer exchange process. After desalting, digestion was performed with endoproteinase Lys-C (0.6 μg) at 37°C for 4 h. The final concentration of the sample was ~1 pmol / μL. The sample was diluted with a desired buffer solution for further analysis or frozen at -80 °C. The Lys-C digestion procedure of IgG was similar to that of EGFR. For the sensitivity study, diluted sample solution was directly made from the stock solution (50 pmol / μL) instead of performing a serial dilution in order to prevent sample loss during the sample preparation steps, particularly, for the analysis of large peptides and phosphorylated peptides.

2.4. Chromatography

Monolithic columns were cut to their final length (~10 cm) with a carbide cutter (Upchurch, Oak Harbor, WA, USA) and inspected under a stereomicroscope to ensure removal of debris. An ESI emitter was cut to ~3.5-cm length and carefully connected butt-to-butt to the monolithic column through a ~3 cm PTFE tubing (0.3 mm I.D., 0.75 mm O.D.) to minimize dead volume. A quaternary gradient pump (Ultimate, Dionex, Sunnyvale, CA, USA) with a laboratory-made splitter was used to obtain the desired flow rates from 20 to 100 nL/min. A 6-port narrow-bore injector (100 μm I.D. bore, Valco, Plainview, TX, USA) with an external 250 nL loop was used for injection on 50 μm I.D. monolithic columns. A 4-port injector, having only a 4-nL internal loop, was employed for loading on the 20 μm I.D. columns. Mobile phase A was 0.1% formic acid in water, and mobile phase B was 0.1 % formic acid in acetonitrile. A shallow gradient was used for EGFR and β-casein analyses: (i) 30 minutes at 0 % B for sample loading, (ii) linear gradient to 35% B over 40 min, then (iii) to 80% B over 25 min, and finally (iv) constant 80% B for 20 min.

2.5. Mass spectrometry

A hybrid LTQ-FT mass spectrometer (Thermo Electron, San Jose, CA, USA) was used in all experiments. On-line ESI-MS was performed in the positive-ion mode with the ESI voltage set at 1.2-1.3 kV and the ion transfer capillary maintained at 245 °C. The normalized collision energy was 28% for MS² and 20% for MS³. Briefly, the mass spectrometer was operated in the data-dependent mode to switch automatically between MS, MS² and MS³ acquisition. Survey MS scan spectra with 2 microscans (m/z 400-2000) were acquired in the FTICR cell

with a mass resolution of 100 000 at m/z 400 (target value of 2×10^6 ions), followed by 4 pairs of sequential MS^2 and MS^3 scans. Further details are described elsewhere [1].

2.6. Peptide identification

The methods for assignment of tryptic peptides (for charge state $\leq 3+$), large peptides (for charge state $\geq 4+$), phosphopeptides and glycopeptides were similar to our previous report [1]. Briefly, the Sequest algorithm in the Bio Works software (version 3.2, Thermo Electron Corp.) was used to search all MS^2 and MS^3 spectra of theoretical fragmentations against a human proteomic database (ncbi.nih.gov/blast.db/FASTA update on 01/15/2006) with a mass tolerance ± 1.4 Da. Peptide ions ($\leq 3+$ ions) were assigned by the BioWorks 3.2 software, which identified the peptides with a double filter, (i) a peptide probability greater than 95% confidence, and (ii) peptide X_{corr} scores above the following thresholds: ≥ 2.5 for 3+ and higher charge state ions, ≥ 2.0 for 2+ ions, and ≥ 1.5 for 1+ ions, with semi-Lys-C specificity and up to 3 internal missed cleavages. Identification for each individual peptide was further confirmed by high mass accuracy (< 5 ppm). For high charge state ions ($\geq 4+$ charge), Sequest was used to assign the most likely peptide, and the top assignment was further confirmed by (i) mass accuracy (< 5 ppm), and (ii) preferred fragmentation patterns in the observed MS^2 and MS^3 spectra. Manual inspection was performed on all glyco- and phosphopeptides.

3. Results and Discussion

The ERPA platform is focused on targeted proteomic analysis of biologically important multi-post-translationally modified large proteins. The complexity of the sample is reduced using Lys-C as the digesting enzyme, which, for most proteins, generates a lower number of well retained peptides in reversed-phase LC, relative to trypsin. Moreover, additional positively charged basic residues (K and R) in the Lys-C digest fragments enhance the sensitivity of post-translationally modified peptides [1]. Using this approach, we have demonstrated $>95\%$ sequence coverage in the analysis of two heavily phosphorylated and glycosylated proteins, β -casein at the 50 fmol level and the epidermal growth factor receptor (EGFR) at 1 pmol (5 runs of 200 fmol per run) when using a 75 μm I.D. packed column.

As a consequence of the usefulness of ERPA, it has been our goal to develop separation technologies tailored to large peptide fragment analysis that are readily adaptable to bench-top HPLC systems (maximum pressure limit ~ 400 bar, or 6000 psi), available in most research laboratories. It is generally known that narrow bore columns such as 20 or 50 μm I.D., operating at flow rates ~ 20 –100 nL/min (linear velocity ~ 1.0 mm/s), deliver higher sensitivity and lower detection limits than 75 μm I.D. columns operating at ~ 200 nL/min for a given amount of sample injected under otherwise identical conditions. The higher sensitivity is attributed in part to the lower peak dilution factor (for a fixed amount injected) and the higher ionization efficiency under low flow rates [14–16]. We have previously reported the construction of a 20 μm I.D. PS-DVB monolithic column [7] and its applications for high sensitivity (~ 10 amol) analysis for tryptic peptides. A similar preparation procedure was applied here with the only change being the use of an organic phase silanization method, which was reported in ref. [18]. The non-aqueous phase procedure enhanced the coverage of the coupling silane 3-TMSPMA to the inner wall surface, which is beneficial for column stability and reproducibility [23].

3.1. Column permeability and loadability

It is desirable to optimize the column permeability in order to obtain proper flow rates within the limit of normal HPLC pump pressures (*i.e.* < 6000 psi). The permeability of the polymeric matrix is typically investigated with HPLC solvents, *e.g.* acetonitrile and water, by measuring the pressure drop and flow rate. For the 20 μm I.D. monolithic column, the column permeability

K was calculated as $\sim 4.2 \times 10^{-15} \text{ m}^2$ for water and $\sim 3.6 \times 10^{-15} \text{ m}^2$ for acetonitrile. These levels are consistent with what was recently reported for a similar PS-DVB based monolithic column [22]. Assuming a column porosity $\varepsilon \sim 0.75$, as reported in ref. [22], the microglobule diameter d_p is estimated to be $\sim 0.33 \text{ }\mu\text{m}$, according to the Kozeny-Carman equation. This estimation is also consistent with what is observed by SEM, see Fig. 1 ($\sim 0.5 \text{ }\mu\text{m}$ of microglobule diameter). Such small size microglobules enhance the specific surface area (SSA), leading to a higher loading capacity. As reported in ref. [22], the SSA for monolithic columns of similar permeability to that in the current work is estimated to be $\sim 30\text{--}40 \text{ m}^2/\text{g}$, which is ~ 10 -fold higher than earlier reported polymer based monolithic columns [17,18].

It is known that SSA is approximately inversely proportional to pore diameter d_p . For typical commercial porous particulate HPLC packings, the SSA is $\sim 300 \text{ m}^2/\text{g}$ for $100 \text{ }\text{\AA}$ and $\sim 100 \text{ m}^2/\text{g}$ for $300 \text{ }\text{\AA}$ pore materials, respectively [24]. Generally, for tryptic peptide separations, where most peptides are 3000 Da or less, small particles (*i.e.* 1.5 to $3 \text{ }\mu\text{m}$) with small pores (*i.e.* 10 to 20 nm) are often used. For a commonly used $75 \text{ }\mu\text{m}$ I.D. packed column, the sample loadability is about 5 to 10 -fold higher than that of a polymeric monolithic column of similar dimension. On the other hand, for larger peptide ($\text{MW} > 5000 \text{ Da}$) separations, wider pore (*i.e.* 30 to 50 nm) particles are generally used to facilitate mass transport [24]. Thus, the loadability difference between a particulate-packed column and a monolithic column is reduced due to the smaller surface area of wider pore packings. Furthermore, for ultra-narrow bore columns such as $20 \text{ }\mu\text{m}$ I.D., the difference in loadability is even less due to the reduced tubing-I.D.-to-particle-size ratio. The narrower the column I.D., the less the amount of wide-pore packing materials that can be packed into it. In addition, packing small size particles (*e.g.* $< 3 \text{ }\mu\text{m}$) into narrow I.D. capillaries ($< 50 \text{ }\mu\text{m}$) can still be technically difficult and the operation of these packed narrow-bore columns is limited by the HPLC pump pressure as well. Thus, a monolithic packing is a useful choice to construct narrow-bore columns for large peptide separations in comparison to particulate-packed columns for readily available bench-top HPLC systems (pressure limit $< 6000 \text{ psi}$).

3.2. Peptide recovery

High sample recovery is particularly important to achieve low detection levels for trace amounts of sample and to minimize cross contamination. For reversed-phase packings, it has been shown that recovery is related to pore size and alkyl chain length [6]. We have conducted a study of the recovery of Lys-C digest peptides on particulate-packed Magic C-18, Biobasic C-18, Biobasic C-4 and PS-DVB monolithic columns. Rabbit IgG was Lys-C digested and injected into a specific column, followed by gradient elution. The initial run was then followed by three blank (solvent A) injections. Recovery was estimated by calculating the peak area ratios for a panel of peptides between the first sample injection and the first blank injection, where the specific peptide could be found to elute again.

Table 1 presents the results of this study. High recovery was found for representative peptides having molecular weights below 3 kDa on all of the four columns ($\sim 99\%$), while for larger molecular mass peptides ($\text{MW} > 3 \text{ kDa}$), the two C-18 columns delivered poorer recovery. Both the C4 and monolithic columns maintained high recovery, but the monolithic column provided the best recovery ($> 95\%$) for all of the examined peptides. The high recovery from the monolithic column is believed a result of the open macroporous column structure, resulting in good mass transport for the large peptide fragments, as well as the lower hydrophobicity than the C-18 columns

3.3. Separation performance

Fig. 2A shows a typical result of a gradient elution separation of a bovine serum albumin (BSA) tryptic digest (4 fmol) with a $50 \text{ }\mu\text{m}$ I.D. PS-DVB monolithic column. The peak widths (at half

height) of three representative peptides, which were eluted in the early, middle, and late portion of the chromatogram were measured. The estimated peak capacity (C_p) was then determined from the effective gradient time divided by the average peak width of these 3 peptides (Fig. 2B). We determined the estimated C_p to reach ~210 for the 10 cm monolithic column. In comparison, the peak capacity obtained on the 75 μ m I.D. Biobasic C18 column (5 μ m, 300 Å) was ~160 for the 10 cm column, indicating a 30% enhancement of separation performance for the monolithic column. The separation performance for large peptides on the PS-DVB monolithic column and Biobasic C4 column was also compared using an EGFR Lys-C digest as sample, see Fig. 3. Narrower peaks were observed from the monolithic column. Since the gradient delay time was not significantly different for flow rates at either 100 nL/min (for the 50 μ m I.D. monolithic column), or 200 nL/min (for the 75 μ m I.D. C-4 column), a side-by-side comparison of separation performance on both columns was performed using the same instrumentation system. Separately, we also examined separation performance on a 20 μ m I.D. monolithic column at 20 nL/min. Similar peak capacities were obtained for the 20 μ m I.D. and 50 μ m I.D. monolithic columns. Importantly, care needed to be exercised in the HPLC and nanoESI spray connections as well as the injection system for the 20 μ m I.D. column at the flow rate of 20 nL/min. The goal was to minimize delay time and chromatographic peak widths.

3.4. Reproducibility

Batch-to-batch column reproducibility is an important factor to consider in the evaluation of the 20 μ m I.D. columns. The production reproducibility of 50 μ m I.D. monolithic column was straightforward and found to be in the range of 5 – 10% (absolute gradient elution time). We focused in this work on the column-to-column reproducibility of the 20 μ m I.D. PS-DVB monolithic column by measuring the elution times for a panel of tryptic bovine catalase peptides. The calculated relative standard deviation (RSD) for each peptide is listed in Table 2. The average RSD for absolute elution times was found to be ~9%, and the average RSD for relative elution times was ~3%. Considering the variation of low flow rates at ~20-30 nL/min with the split-gradient system, and the variation of pipetting porogens and monomers during column preparation, the RSD values suggest quite satisfactory reproducibility. In this work, the three 20- μ m I.D. columns were prepared from three different batches of polymerization mixture. Since every monolithic column is made *in situ*, we did, however, find that a certain percentage of monolithic columns made with the same procedures could have inconsistent polymerization or permeability. Under carefully controlled polymerization conditions in each step, such as avoiding solvent evaporation by proper seal and applying pressure during polymerization, the successful rate reached ~70 to 80%. It is expected that further study will improve the success rate even more.

3.5. Detection Limits

We have demonstrated a ~5-10 amol detection limit for tryptic peptides with the 20 μ m I.D. PS-DVB monolithic column coupled to an LCQ Deca XP⁺ mass spectrometer [7]. The hybrid LTQ-FT-MS, a more sensitive instrument, was examined with a standard bovine catalase tryptic digest sample, as used in ref. [7]. The analysis of a 400 zmol sample (4 nL of 0.1 nM concentration) is shown in Fig. 4. Two peptides (DALLFPSFIHSQK and LGPNYLQIPVNCPYR) are clearly identified by precursor mass (m/z) and fragmentation (MS/MS) measurements. The precursor mass measurement benefited greatly from the high resolution and accurate mass measurement in the FTICR, even with S/N ratios of ~2 and 3 for the two peptides (Fig. 4B and 4C). In addition to the observation of the peptide fragment ions in the MS/MS spectra (Fig. 4D and 4E), the precise precursor mass measurement provided additional confidence for the identification. We estimated the new mass spectrometer delivered ~5 to 10 times higher sensitivity than the 3D ion trap previously used. Importantly, the signal intensities or peak areas of the detected peptides were also proportional to the amount of sample load (from 400 zmol to 4 fmol), which indicated a linear range approximately 4-orders of

magnitude. On the basis of the evaluation of column performance, peptide recovery and sensitivity, the high performance monolithic column (20 μm I.D.) was employed in the following for the ERPA analysis of proteins modified with complex PTMs.

3.6. Bovine β -casein

For demonstration of the analysis of post-translational modifications (PTMs), in particular phosphorylation, bovine β -casein has been extensively studied as a model phosphoprotein [25] as it is known to possess five fully phosphorylated sites (pS15, pS17, pS18, pS19, and pS35). Both trypsin and Lys-C digestion yield the same single phosphopeptide, but the tetraphosphopeptide from Lys-C digestion, relative to trypsin, has additional three amino acids, RELEELNVPGEIVES*LS*S*S*EESITRINK, where the underlined sequence is from tryptic digestion. We previously showed that the detection limit of the tetraphosphopeptide from Lys-C digestion was at the 50 fmol level on the LTQ-FT-MS (detected on MS, MS², and MS³ modes) with a 75- μm I.D. packed column [1]. In a recent study by others [26], using a monolithic PS-DVB column (60 $\times$ 0.20 mm I.D.), the detection limit of the tetraphosphopeptide was improved to 1 pmol or 10 fmol level in the positive or negative ion detection modes, respectively (note: only the m/z of the tetraphosphopeptide was determined in the negative mode). With a 10 cm $\times$ 20 μm I.D. monolithic PS-DVB column, we have further enhanced the detection limit, *in the positive mode*, of the tetraphosphopeptide to the 4 fmol level, as shown in Fig. 5. Note that not only the m/z but also the fragmentation of the tetraphosphopeptide in the MS² and MS³ scans were observed in this example. In addition, we achieved ~96 % sequence coverage (202 out of 209 amino acid residues, see Table 3). The S/N ratios of the non-phosphopeptides, particularly, the two large peptides (5,316 Da and 6,359 Da) suggest that the detection limit would be at the low amole level. The high sensitivity and high sequence coverage are clearly a result of the use of the narrow-bore column, new MS instrumentation and the additional basic amino acid (K) residue on the tetraphosphopeptide as a result of the Lys-C digest.

3.7. Epidermal growth factor receptor (EGFR)

EGFR is a trans-membrane protein, conducting signal transduction from the plasma membrane through the cytoplasm. This receptor is a well-known target protein for cancer treatment, since EGFR is often over-expressed in many epithelial tumors [27-29]. Previously, we have demonstrated ~95 % sequence coverage at the 1 pmol (5 runs of 200 fmol per run) of Lys-C digest EGFR on a packed 75- μm I.D. Biobasic C4 column (5 μm , 300 $\text{\AA}$ pore) [1]. Recently, using a 50 μm I.D. PS-DVB monolithic column, we achieved the same level of sequence coverage with only ~200 fmol sample (3 runs of 50–75 fmol per run).

In this work (with the known precursor ions of EGFR peptides obtained from the previous studies on a 75- μm i.d column), we tested a 20 μm I.D. monolithic PS-DVB column and found that we could reduce the sample amount to ~10 to 20 fmol per injection to achieve a similar degree of characterization. It needs to be noted multiple runs (3 to 5) may be necessary for complete characterization if *de novo* structure analysis using data dependent analysis is employed. The enhanced sensitivity is attributed to the improved recovery of large peptides, particularly, the 10 kDa peptide, reduced peak dilution with the narrow bore column compared to the 75 μm I.D. column, and operation at 20 nL/min flow rate. The same strategy as in our previous publication [1], which combines the high resolution and accurate mass measurement of peptide precursor ions with MS² and MS³ spectral measurements of peptide fragment ions, was utilized in this study. In the following, we select several informative peptide ions to demonstrate the peptide structure characterization potential at this low level.

An illustration of large peptide structure elucidation is shown in Fig. 6, in the unambiguous identification of the 10 266 Da Lys-C peptide digest fragment of EGFR at the 20 fmol level.

This peptide was eluted at 83.13 min (corresponding to ~75% ACN in the elution mobile phase) on the 20 μ m I.D. monolithic column (Fig. 6A). The hydrophobicity of this peptide is such that it is difficult for the species to elute from even a reversed-phase C4 column. This 10 kD peptide has observed charge states from 6+ to 10+ (with 7+ as the most intense), as shown in Fig. 6B. The high resolution and accurate measurement of the 7+ charge state ion ($m/z = 1468.5654$) is shown in the insert of Fig. 6B. Assuming Lys-C specificity and ± 5 ppm mass accuracy, only one peptide candidate in the Swiss-Prot database, EGFR residue 947-103, matched the spectrum in Fig 6B, given the large molecular weight. We next examined the preferred CID cleavage sites and confirmed that the high-intensity ions in the MS^2 (Fig. 6C) and MS^3 (Fig. 6D) spectra were in agreement with the predicated cleavages, as shown in the inserts of Fig. 6C and 6D. Thus, non-PTM modified high molecular weight peptide fragments can readily be identified with high confidence. Of course, special care must be exercised in handling such large peptides in order to minimize losses due to adsorption.

In the above example, multiple injections (5 consecutive loadings of 4 fmol each) was employed to inject the EGFR sample with relatively high concentration ($\sim 1 \mu$ M) on the column prior to the gradient elution run. This approach was used due to the need for a small loop volume (4 nL), in order to minimize delay time and peak broadening. We have recently implemented a high pressure T-design for use with a pre-column for injection of μ L volumes of sample [30].

As an example of PTM analysis, in the same LC-MS run as in Fig 6 (20 fmol of Lys-C digest of EGFR), we demonstrate the characterization of a relatively long glycopeptide (residues 337-372), Fig 7. The base ion chromatogram is shown in Fig. 7A, and it can be seen that the glycopeptide was eluted relatively late in the gradient (corresponding to high organic content in the mobile phase). Elution under high organic content can lead to enhancement in signal relative to elution of shorter glycopeptide fragments at lower organic content, such as would normally occur with the corresponding tryptic peptide fragment. In addition, the signal will often be enhanced for the Lys-C fragment as a result of additional basic amino acid residues [1]. The heterogeneity of glycosylation, such as Man 8, Man 7, and Man 6 of the glycopeptide isoforms, is shown in Fig. 7B, which displays the precursor mass measurement of the glycopeptide. As expected, the CID fragmentation of the glycopeptide, modified with a heavily glycosylated form (Man 8), generated mainly glycosidic bond cleavages, such as sequential loss of mannose, as indicated in Fig. 7, rather than backbone peptide cleavages. The further CID fragmentation (MS^3) of the glycopeptide with only one remaining monosaccharide residue (N-acetyl glucosamine), $[(337-372) + (\text{GlcNAc})]^{4+}$, led to mainly peptide backbone cleavages, which provided the backbone sequence of the peptide (Fig. 7D). It needs to be noted that predominant glycosidic bond cleavage can still occur on a complex-type glycopeptide even using CID for fragmentation in MS^3 mode [1]. In this case, a de-glycosylation step (*i.e.* treated with N-glycanase) is often needed in order to obtain the peptide sequence information [1]. Alternatively, ETD fragmentation, which preserves labile post-translational modifications, can be implemented to obtain the peptide bone sequence to complement CID fragmentation for analysis of such complex-type glycopeptides [31].

As a further example of PTM analysis from the same LC-MS run, Fig 8 illustrates the characterization of a phosphopeptide (residues 1038-1075) from EGFR. In contrast to glycopeptide fragmentation, the CID fragmentation of the large phosphopeptide generated mainly the peptide backbone cleavages, which provided not only the peptide sequence but also the phosphorylation site information. As shown in Fig. 8C, the observation of both b_8^+ and y_{29}^{2+} ions indicated that the phosphorylation site is at S1046. Previous reports by others could not distinguish the phosphorylation between the two serine residues (pS1046 or pS1047) on the CID fragmentation of the corresponding smaller tryptic fragment [32,33]. The predominant neutral loss fragmentation, which is typically observed in CID for smaller tryptic fragments,

is much less prevalent for larger Lys-C fragments. The alternative ETD fragmentation method [31] could also be employed to identify or confirm EGFR phosphorylation sites, as was shown in a separate report [34].

4. Conclusions

Narrow bore monolithic column technology together with the ERPA platform has been shown to be a powerful LC-MS tool for large peptide analysis at the trace level. As a result of the favorable mass transport possible for large biomolecules, the monolithic columns deliver ~ 30% higher separation performance than granular packed columns as well as provide good recovery and capacity for large peptides. Given that high sequence coverage was achieved at the low fmol level, it is clear that the monolithic column works well over the whole range of peptide molecular sizes up to 10 kDa. The simple column preparation procedure (*in situ* polymerization, no post-preparation on-column alkylation and end-capping) with suitable retention strength for large peptides renders the monolith columns a valuable alternative to particulate packed columns, especially in the domain of narrow bore columns. The integrated column structure makes column manipulation and maintenance easy without the concern of packing disturbance, which leads to decreased separation performance.

In addition to achieving the high-performance separation of large peptides, the polymer, PS-DVB, monolithic column can further improve detection sensitivity for the ERPA platform. We were able to achieve > 95% sequence analysis for β -casein and EGFR at low fmol levels. This methodology opens the way to achieve the goal of comprehensive characterization of biologically important proteins at the trace level and facilitate biomarker discovery. Of course, great care needs to be exercised in the handling of such small quantities of sample prior to injection into the chromatographic column.

Finally, in a recent report the 50 μ m I.D. monolithic column was used for the LC-MS analysis of large peptides with a newly developed ion fragmentation technique-electron transfer dissociation (ETD) [31], which is a non-ergodic fragmentation method that preserves labile post-translational modifications. ETD can be a part of the ERPA platform using the narrow-bore monolithic columns for comprehensive characterization of these large peptides with complex modifications, as demonstrated in our recent paper [34].

Acknowledgements

The authors thank NIH for supporting of this work under GM 15847. Contribution Number 872 from the Barnett Institute.

References

1. Wu S-L, Kim J, Hancock WS, Karger BL. *J Proteome Res* 2005;4:1155. [PubMed: 16083266]
2. Kelleher NL, Taylor SV, Grannis D, Kinsland C, Chiu H-J, Begley TP, McLafferty FW. *Protein Sci* 1998;7:1796. [PubMed: 10082377]
3. McLafferty, FW.; Fridriksson, EK.; Horn, DM.; Lewis, A.; Zubarev, RA. *Science*. 284. Washington, D.C: 1999. p. 1289
4. Link AJ, Eng J, Schieltz DM, Carmack E, Mize GJ, Morris DR, Garvik BM, Yates JR III. *Nat Biotechnol* 1999;17:676. [PubMed: 10404161]
5. Delahunty C, Yates JR III. *Methods* 2005;35:248. [PubMed: 15722221]
6. Tanaka N, Kimata K, Mikawa Y, Hosoya K, Araki T, Ohtsu Y, Shiojima Y, Tsuboi R, Tsuchiya H. *J Chromatogr* 1990;535:13. [PubMed: 1965193]
7. Ivanov AR, Zang L, Karger BL. *Anal Chem* 2003;75:5306. [PubMed: 14710807]
8. Shen Y, Tolic N, Masselon C, Pasa-Tolic L, Pasa-Tolic L, Camp DG II, Hixson KK, Zhao R, Anderson GA, Smith RD. *Anal Chem* 2004;76:144. [PubMed: 14697044]

9. Shen Y, Moore RJ, Zhao R, Blonder J, Auberry DL, Masselon C, Pasa-Tolic L, Hixson KK, Auberry KJ, Smith RD. *Anal Chem* 2003;75:3596. [PubMed: 14570215]
10. Luo AQ, Shen Y, Hixson KK, Zhao R, Yang F, Moore RJ, Mottaz HM, Smith RD. *Anal Chem* 2005;77:5028. [PubMed: 16053318]
11. Luo BQ, Page JS, Tang K, Smith RD. *Anal Chem* 2007;79:540. [PubMed: 17222018]
12. Bruins AP, Covey TR, Henion JD. *Anal Chem* 1987;59:2642.
13. Hopfgartner G, Bean K, Henion JD, Henry RJ. *J Chromatogr* 1993;647:51.
14. Wilm M, Mann M. *Anal Chem* 1996;68:1. [PubMed: 8779426]
15. Smith RD, Shen Y, Tang K. *Acc Chem Res* 2004;37:269. [PubMed: 15096064]
16. Juraschek T, Dulcks M. *J Am Soc Mass Spectrom* 1999;10:300. [PubMed: 10197351]
17. Svec F. *J Sep Sci* 2005;28:729. [PubMed: 15940819]
18. Zhang S, Zhang J, Horváth Cs. *J Chromatogr A* 2001;914:189. [PubMed: 11358213]
19. Premstaller A, Oberacher H, Walcher W, Timperio AM, Zolla L, Chervet J-P, Cavusoglu N, van Dorsselaer A, Huber CG. *Anal Chem* 2001;73:2390. [PubMed: 11403277]
20. Premstaller A, Oberacher H, Huber CG. *Anal Chem* 2000;72:4386. [PubMed: 11008774]
21. Miyabe K, Guichon G. *J Sep Sci* 2004;27:853. [PubMed: 15354563]
22. Oberacher H, Premstaller A, Huber CG. *J Chromatogr A* 2004;1030:201. [PubMed: 15043270]
23. Buszewski B, Szumski M, Sus S. *LC-GC Eur Dec*;2002 :2.
24. Neue, UD. *HPLC Columns: Theory, Technology, and Practice*. Wiley-VCH; Weinheim: 1997. p. 81
25. Kim J, Camp DG II, Smith RD. *J Mass Spectrom* 2004;39:208. [PubMed: 14991691]
26. Tholey A, Toll H, Huber CG. *Anal Chem* 2005;77:4618. [PubMed: 16013881]
27. Carpenter G, Cohen S. *J Biol Chem* 1990;265:7709. [PubMed: 2186024]
28. Wells A. *Int J Biochem Cell Biol* 1999;31:637. [PubMed: 10404636]
29. Laskin JJ, Sandler AB. *Cancer Treat Rev* 2004;30:1. [PubMed: 14766123]
30. Yue E, Luo Q, Zhang J, Wu S-L, Karger BL. *Anal Chem* 2007;79:938. [PubMed: 17263319]
31. Coon JJ, Ueberheide B, Syka J, Dryhurst DD, Ausio J, Shabanowitz J, Hunt DF. *Proc Nat Acad Sci USA* 2005;102:9463. [PubMed: 15983376]
32. Heisermann GJ, Gill GN. *J Biol Chem* 1988;263:13152. [PubMed: 3138233]
33. Gamou S, Shimizu N. *J Cell Physiol* 1994;158:151. [PubMed: 7505275]
34. Wu S-L, Huhmer A, Hao Z, Karger BL. *J Proteome Res*. 2007in press

Fig 1. Scanning electron micrograph of a 20 μm I.D. PS-DVB monolithic column

The microglobules (~ 0.5 μm) are shown to be homogeneously distributed in the lumen, with the through-pores (~ 1 μm) indicated in dark color.

A

B

Fig 2. Estimation of peak capacity of the nano-LC/MS system using a 10 cm × 50 μm I.D. PS-DVB monolithic column

(A) Base peak ion chromatogram and gradient elution profile of 4 fmol BSA tryptic digest; (B) extracted ion chromatogram (XIC) of three peptides eluted from the early (I), middle(II) and late (III) portions of the chromatogram with peak width measured at half height. Gradient elution from 0-50% ACN in 50 min. Mobile phase A: 0.1% formic acid/H₂O, mobile phase B: 0.1% formic acid/ACN. Flow rate is ~ 100 nL/min.

Fig 3. Comparison of separation of EGFR Lys-C digest

(A) 50 µm I.D. PS-DVB monolithic column, 100 fmol injected; (B) 75 µm I.D. packed Biobasic C4 column, 250 fmol injected. Column length is 10 cm in each case. See Experimental for details.

Fig 4. Analysis of 400 zmol bovine catalase tryptic digest using the 20 μ m I.D. PS-DVB monolithic column

(A) Base peak ion chromatogram; (B) Extracted ion chromatogram (XIC) of m/z 751.90 $\pm$ 0.50, eluted at 59.03 min; (C) XIC of m/z 902.95 $\pm$ 0.50, eluted at 59.31 min; (D) MS/MS spectrum of m/z 751.90 $\pm$ 1.0 (at 59.03 min), DALLFPSFIHSQK, with the high resolution MS spectrum of the precursor ion shown in the insert; (E) MS/MS spectrum of m/z 902.95 $\pm$ 1.0 (at 59.31 min), LGPNYLQIPVNCYPYR, with the high resolution MS spectrum of the precursor ion shown in the insert.

Fig 5. Analysis of the tetraphosphorylated peptide from β -casein Lys-C digest using the 20 μ m I.D. PS-DVB monolithic column (4 fmol injected on column)

(A) Extracted ion chromatogram (XIC) of the tetraphosphorylated Lys-C peptide, eluted at 47.19 min; (B) FTICR mass spectrum of the tetraphosphorylated Lys-C peptide, with the most intense charge (3+) shown; (C) MS/MS spectrum of the tetraphosphorylated Lys-C peptide with its peptide sequence, fragment and neutral loss ions indicated in the figure; (D) MS³ scan of the highest intensity peak (initial neutral loss peak) in the MS/MS spectrum and its peptide sequence, with further sequential neutral loss and fragment ions indicated in the figure. In the inserts, S* and S# represent the phosphorylation site and its neutral loss, respectively.

Fig 6. Analysis of a large 10.3 kDa peptide from the EGFR Lys-C digest using a 20 μ m I.D. PS-DVB monolithic column (20 fmol injected on column)

(A) Extracted ion chromatogram (XIC) of the 10.3 kDa Lys-C peptide, eluted at 83.13 min; (B) FTICR mass spectrum of the 10.3 kDa Lys-C peptide, with the charge states (from 6+ to 10+) and the most intense charge (7+) shown in the insert; (C) MS/MS spectrum of the 10.3 kDa Lys-C peptide, with the peptide sequence and fragment ions indicated in the figure; (D) MS³ scan of the highest intensity peak (b_{48} of the 10.3 kDa peptide) in the MS/MS spectrum with the peptide sequence and the fragment ions indicated in the figure.

Fig 7. Analysis of a glycopeptide from an EGFR Lys-C digest using a 20 μm I.D. PS-DVB monolithic column (20 fmol injected on column)

(A) Base peak chromatogram showing the specific glycopeptide (337-372) being eluted at ~59 min; (B) FTICR mass spectrum of the glycopeptide (337-372) with different numbers of mannoses (Man 8, Man 7, and Man 6) at the N 337 site, and with the most intense charge state of the glycoforms (5+) indicated in the figure; (C) MS/MS spectrum of the glycopeptide (337-372) modified with the Man 8 glycoform, and with the fragment ions (sequential loss of mannose) at the 5+ and 4+ charge states indicated in the figure; (D) MS³ scan of the highest intensity peak (the glycopeptide remaining with only 1 GlcNAc) in the MS/MS spectrum, and with the peptide backbone fragmentation indicated in the figure. In the inserts, GlcNAc represents N-acetyl glucosamine.

Fig 8. Analysis of a large phosphorylated peptide from an EGFR Lys-C digest using a 20 μ m I.D. PS-DVB monolithic column (20 fmol injected on column)

(A) Base peak ion chromatogram showing the specific phosphopeptide (1038-1075) being eluted at ~60.55 min; (B) FTICR mass spectrum of the phosphopeptide, with the most intense charge (3+) shown; (C) MS/MS spectrum of the phosphopeptide with the peptide sequence indicated in the insert; (D) MS³ scan of the highest intensity peak (y_{10} ion of the phosphopeptide) in the MS/MS spectrum, with the peptide sequence indicated in the insert.

Table 1
Peptide recovery as a function of reversed-phase packing material

Peptide molecular weight (Da)	Recovery (%) ^a			
	Magic C-18 (200 Å pore, 5 µm particle) ^b	Biobasic C-18 (300 Å pore, 5 µm particle) ^b	Biobasic C-4 (300 Å pore, 5 µm particle) ^b	Monolithic (PS-DVB) ^c
6326	82^d	69^d	94^d	98
5502	93^d	80^d	97	95
4874	86^d	91^d	94^d	99
4166	98	80^d	99	99
3558	97	99	100	100
3121	89^d	86^d	97	100
3033	79^d	95	96	97
2734	100	97	100	100
1978	100	99	100	100
1857	99	99	100	100
1557	100	100	100	100
1319	100	100	100	100
837	100	100	100	100

^a) Recovery based on the relative amount of the highest intensity peptide observed in a blank gradient run immediately after the sample run. Sample: Lys-C digest of rabbit IgG. See Experimental section for details.

^b) 10 cm length and 75 µm I.D. column.

^c) 10 cm length and 50 µm I.D. column.

^d) Less than 95% recovery (in bold).

Table 2

Column-to-column reproducibility for 20 µm I.D. PS-DVB monolithic columns^a

Column	Ions (m/z)				
	740.8	609.6	849.8	903.7	771.2
Absolute retention time (min)					
1	21.52	23.36	28.44	35.78	38.73
2	16.79	19.08	24.35	30.33	32.94
3	18.71	20.15	24.82	32.25	35.18
RSD (%)	12.5	10.7	8.65	8.43	8.2
Average RSD (%)					
9.3					
Relative retention time					
1	0.46	0.5	0.61	0.57	0.77
2	0.42	0.48	0.61	3.9	0.76
3	0.43	0.47	0.57	0.74	0.74
RSD (%)	4.7	3.2	3.9	2	1.4
Average RSD (%)					
2.6					

^a) 4 fmol bovine catalase tryptic digest loaded on the 10 cm × 20 µm I.D. PS-DVB monolithic column. Gradient elution from 0-50% ACN in 50 min. Mobile phase A: 0.1% formic acid in H₂O, mobile phase B: 0.1% formic acid in ACN. Flow rate is ~ 20 nL/min.

Table 3Peptides identified from a 4 fmol Lys-C digest of β -casein using a 20- μ m I.D. PS-DVB monolithic column^a

Residue No.	Peptide sequence ^b	[M+H] ⁺ (Da) ^c	Identification ^d	Charge states ^e
1 - 28	RELEELNVPGEIVES*LS*S*S*EES ITRINK	3477.4879	+	3+
30 - 32	IEK	389.2395	-	-
33 - 48	FQS*EEQQQTEDELODK	2061.8284	+	2+
33 - 48	FOSEEQQQTEDELODK	1884.6560	+	2+
49 - 97	IHPFAQTQSLVYPFGPIPN SLPQNIPPLTQTPVVVPPFLQPEV MGVSK	5316.8533	+	3+, <u>4+</u> , 5+, 6+, 7+
98 - 99	VK	246.1812	-	-
100 - 105	EAMAPK	646.3228	+	1+
106 - 107	HK	284.1717	-	-
108 - 113	EMPFPK	748.3698	+	1+
114 - 169	YPVEPFTEQSLTLTDVENLHLPL PLLQSWMHQPHOPLPPTVMFPP QSVLSLSQSK	6359.2559	+	4+, 5+, 6+, 7+
170 - 176	VLPVPQK	780.4978	+	1+
177 - 209	AVPYPQDMPQIAFLLYQEPVLG PVRGPFPIIV	3721.0337	+	2+, 3+, 4+, 5+

^a) Experimental conditions see Experimental section.^b) "*" indicates phosphorylated serine.^c) Theoretical monoisotope [M+H]⁺, including phosphorylation.^d) "+" = identified, "-" = not identified.^e) The most abundant charge state underlined.