Journal of Chromatography A, xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

Journal of Chromatography A

journal homepage: www.elsevier.com/locate/chroma

High-resolution ultrahigh-pressure long column reversed-phase liquid chromatography for top-down proteomics

Yufeng Shen[a,*], Nikola Tolić[b], Paul D. Piehowski[a], Anil K. Shukla[a], Sangtae Kim[a], Rui Zhao[b], Yi Qu[a], Errol Robinson[b], Richard D. Smith[a], Ljiljana Paša-Tolić[b]

[a] Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA 99354, USA
[b] Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA 99354, USA

ARTICLE INFO

Article history:
Received 18 July 2016
Received in revised form
29 November 2016
Accepted 3 January 2017
Available online xxx

Keywords:
Top-down proteomics
Intact proteins
Proteoforms
UPLC
Mass spectrometry
Columns and stationary phases

ABSTRACT

Separation of proteoforms for global intact protein analysis (i.e. top-down proteomics) has lagged well behind what is achievable for peptides in traditional bottom-up proteomic approach and is becoming a true bottle neck for top-down proteomics. Herein, we report use of long (≥1 M) columns containing short alkyl (C1-C4) bonded phases to achieve high-resolution RPLC for separation of proteoforms. At a specific operation pressure limit (i.e., 96.5 MPa or 14 K psi used in this work), column length was found to be the most important factor for achieving maximal resolution separation of proteins when 1.5–5 μm particles were used as packings and long columns provided peak capacities greater than 400 for proteoforms derived from a global cell lysate with molecular weights below 50 kDa. Larger proteoforms (50–110 kDa) were chromatographed on long RPLC columns and detected by MS; however, they cannot be identified yet by tandem mass spectrometry. Our experimental data further demonstrated that long alkyl (e.g., C8 and C18) bonded particles provided high-resolution RPLC for <10 kDa proteoforms, not efficient for separation of global proteoforms. Reversed-phase particles with porous, nonporous, and superficially porous surfaces were systematically investigated for high-resolution RPLC. Pore size (200–400 Å) and the surface structure (porous and superficially porous) of particles was found to have minor influences on high-resolution RPLC of proteoforms. RPLC presented herein enabled confident identification of ∼900 proteoforms (1% FDR) for a low-microgram quantity of proteomic samples using a single RPLC–MS/MS analysis. The level of RPLC performance attained in this work is close to that typically realized in bottom-up proteomics, and broadly useful when applying e.g., the single-stage MS accurate mass tag approach, but less effective when combined with current tandem MS. Our initial data indicate that MS detection and fragmentation inefficiencies provided by current high-resolution mass spectrometers are key challenges for characterization of larger proteoforms.

© 2017 Published by Elsevier B.V.

1. Introduction

Proteomics is now advancing from measurements of protein expressions (e.g. using shotgun bottom-up approaches for analysis of enzyme digests [1]) to characterizing actual proteoforms more directly tied to specific biological processes, based on intact (or top-down) protein characterization approaches. The functional states of proteins include proteoforms arising from hydrolysis [2], signal peptide cleavages, proteolytic processing, site mutations, etc., and broad range of (combinatorial) post-translational modifications (PTMs), which largely go undetected in traditional bottom-up proteomic analyses. But proteoform characterization is technically challenging, and bottom-up approaches still dominate the field. For proteome degradomic activities, which typically involve smaller protein fragments (e.g., <15 kDa) [3], high-resolution RPLC–MS/MS approach has been successfully demonstrated [4,5]. However, the quality of separations greatly degrades as the protein mass increases. Even though proteins with mass above 100 kDa have been successfully analyzed [6], top-down proteomics remain far inferior to bottom-up approaches in terms of coverage, sensitivity, and throughput. As a recent example, assignment of ∼3000 proteoforms used a complex four-dimensional separation platform resulting in low sensitivity (mg sample amounts required) and low throughput (long analysis times) [6]. These requirements, together with lack of robust computational tools for proteoform

* Corresponding authors.
E-mail addresses: Yufeng.shen@pnnl.gov (Y. Shen), ljiljana.pasatolic@pnnl.gov (L. Paša-Tolić).

http://dx.doi.org/10.1016/j.chroma.2017.01.008
0021-9673/© 2017 Published by Elsevier B.V.

characterization, greatly limit the use and applicability of top-down proteomics.

Quality separation improves proteome coverage, enhances sensitivity and simplifies analysis. While no single suitable separation is yet available, the last stage of separation applied on-line with MS analysis, typically RPLC, is of critical importance for these aspects. A major obstacle for development of high-resolution RPLC of proteins arises from the traditional concept that the resolution of gradient RPLC of proteins is unrelated to column length [7,8]. Gradient RPLC of proteins is considered as a single adsorption/desorption event near the top of the column, rather than protein partitioning along the column, which leads to a claim that protein separation efficiency is irrelevant to column length [7,8] and the inability to predict RPLC protein resolution using classical chromatographic parameters such as peak capacity expression [9]. This is also the reason why majority of efforts for improving gradient RPLC resolution are focused on optimization of separation media using large pore particles [6,10], nonporous particles [11,12], core-shell particles [13,14], and monoliths [15,16] in conjunction with short (e.g., 2–20 cm) columns. It is thus not surprising that strongly hydrophobic reversed phases (e.g. C8, C8-similar polystyrene-divinylbenzene, and C18 phases) have been traditionally selected for protein separation and top-down proteomic applications [6,10–18]. Such approaches provide moderate quality resolution for proteins, evidenced by absence of reports of complex chromatograms, regardless of the magnitude of separation power (e.g., peak capacity) estimated [6,9–18].

Herein, we report a new approach for achieving high-resolution RPLC of proteins. In this approach, we utilized C1-C4 reversed phases to minimize the effect of hydrophobic adsorption on RPLC resolution observed for C8 and C18 for proteins [7,8] and thus enhance the role of column length in separation resolution. This approach is hereby referred as weak interaction RPLC (WI-RPLC) in contrast to conventional strong interaction RPLC (SI-RPLC) employing C8 and C18 phases. With use of long columns containing such weak reversed phases, complex chromatograms displaying over 200 countable peaks and separation peak capacities of ~400 have been achieved. We systematically investigated various reversed phase particles with size of 1.5–5 μm and porous, nonporous, and superficially porous surfaces for high-resolution RPLC, and demonstrated that WI-RPLC coupled with FT-MS/MS enabled identifying ~900 proteoforms (at 1% FDR) from a single analysis of a few micrograms of a simple microbial lysate.

2. Experimental

2.1. Sample preparation

Both microbial and mammalian proteomic samples were used for examination of RPLC–MS performances in this work. For microbial sample, *S. oneidensis* cells were grown to steady state in fed-batch mode using a Bioflow 3000 model fermenter (New Brunswick, Enfield, CT). The cells were pelleted (11900 g for 8 min at 4 °C) and frozen at −80 °C until lysis. Cells were lysed by homogenizing the cells with 0.1 mm zirconia/silica beads in the Bullet Blender (Next Advance, Averill Park, NY) speed 8 for 3 min. Protein concentration was determined using BCA assay (Thermo Scientific, San Jose CA). An aliquot containing 500 μg of protein was diluted 20-fold using extraction buffer (1 mM phenylmethanesulfonylfluoride (PMSF)), 8 M Urea, 0.5% Sigma (Saint Louise, MI) phosphatase inhibitor cocktail 2, 0.5% Sigma phosphatase inhibitor cocktail 3, 100 mM ammonium bicarbonate and incubated for 30 min to facilitate protein solubilization and inhibit proteolysis. The resulting solution was spun at 100,000 rpm for 15 min at 4 °C to remove particulates. The supernatant was then buffer exchanged into Milli-Q water (EMD Millipore, Billerica, MA) by washing with 3 × 450 μL aliquots of Milli-Q in an Amicon Ultra 10 K MWCO centrifugal filter (EMD Millipore). Final protein yield was determined by BCA assay and the solution was diluted to a final concentration of 0.5 μg/μL.

For mammalian sample, mouse RAW264.7 macrophage-like cells (ATCC #TIB-71, Manassas, VA) were cultured in Dulbecco's Modification of Eagle's Medium (Sigma, St Louis, MO) supplemented with 10% fetal bovine serum, 100 I.U./mL penicillin, and 100 μg/mL streptomycin. Cells were maintained below passage 15 in 95% humidity and 5% CO_2 atmosphere at 37 °C. Cells were washed with phosphate buffered saline (PBS, Sigma) once before scraping cells with a rubber policeman in PBS. Cells were collected in a microcentrifuge tube, pelleted 10,000 × g for 15 min at 4 °C, snap frozen in liquid N_2, and stored at −80 °C until processing. The cell pellet was homogenized with a pellet pestle in 1 mL of homogenization buffer (8 M Urea, 1 mM PMSF, 0.5% Sigma Phosphatase inhibitor II, 0.5% Sigma Phosphatase inhibitor III, in 100 mM Ammonium Bicarbonate, pH 8.0). The solution was then incubated for 30 min at 37 °C to extract and denature the proteins. Samples were then transferred to Amicon ultra 100 K MWCO filters (EMD Millipore) that were pre-rinsed with 500 uL homogenization buffer. All centrifugation steps were carried out at 4 °C and 15000 rpm. Samples were then centrifuged for 30 min obtain minimum volume. A 300 uL aliquot of homogenization buffer was added to the filter and spun for an additional 30 min to maximize protein recovery. The filtrate was then transferred to an Amicon ultra 10 K MWCO filter and spun for 20 min obtain the minimum volume. Buffer exchange was achieved using 3 washes with 450 uL of separation buffer A (3% acetonitrile (ACN), 0.2% formic acid (FA) in Milli-Q water). Final protein concentration was determined using a Coomassie assay (Thermo Fisher, Waltham, MA), and samples were normalized to a concentration of 0.5 ug/uL for analysis.

2.2. Ultra-high pressure RPLC

RPLC was perform on a home-built 20 K psi LC system [19] with an operation pressure limit of 96.5 MPa (14 K psi), such that the approach developed here can be readily implemented on commercial ultrahigh-pressure LC platforms. Various columns (Table 1) were manufactured using previously reported method [20].The packing particles, including C1-C18-bonded porous, core-shell (superficially porous), and non-porous silica particles, were from Phenomenex (Terrence, CA), Sepax (Newark, DE), Separation Methods Technologies (Newark, DE), Waters (Milford, MA), Eichrom Technologies (Darien, IL), and Advanced Material Technology (AMT, Wilmington, DE). At the operation pressure limit, the different lengths of columns generated similar flow rates, i.e., 0.7–0.9 μL/min (measured for mobile phase A) for porous and core-shell particle-packed columns and ~0.5 μL/min for non-porous particle packed columns. The cell lysate containing 2.5 μg of protein was loaded onto the column using a 5-μL sample loop, and separated with a gradient from mobile phase A (5% ACN/95% H_2O/0.1% FA, v/v/v) to 80% B (0.1% FA in ACN, v/v) in 1000 min except for specifications. For RPLC of mouse cell lysate, RPLC gradient was performed from solvent A (10% ACN, 0.1% FA) to 75% B (70%ACN, 30% isopropyl alcohol (IPA), 0.1%FA) in 600 min. The column outlet was connected to a 3 cm × 20 μm I.D. emitter (with 10 ± 1 μm orifice; New Objective, Woburn, MA) for electrospray ionization.

2.3. Mass spectrometry

Exactive and Orbitrap Elite MS (Thermo Fisher, San Jose, CA) were used. The Exactive was used for evaluation of RPLC performance and operated under the following conditions: mass range $600 < m/z < 2000$, resolution 100,000 (at m/z 400), AGC of 3E6, two microscans. Orbitrap Elite was used for identification of proteo-

Table 1
Capillary RPLC columns produced in this work to explore high-resolution separation of intact lysate proteoforms at ultrahigh-pressure.

Column manufactured	Stationary phase (SP)	SP vendor	C_p [a]
200 cm L × 100 μm i.d.	5 μm C4-porous silica (300 Å)	Phenomenex	350
100 cm L × 100 μm i.d.	3 μm C1-porous silica (300 Å)	Separation Methods	350
100 cm L × 100 μm i.d.	3 μm C2-porous silica (300 Å)	Separation Methods	400
100 cm L × 100 μm i.d.	3 μm C4-porous silica (300 Å)	Sepax	450
100 cm L × 100 μm i.d.	3 μm C8-porous silica (300 Å)	Sepax	– [b]
100 cm L × 100 μm i.d.	3 μm C18-porous silica (300 Å)	Phenomenex	–
30 cm L × 100 μm i.d.	1.7 μm C4-porous silica (300 Å)	Waters	300
120 cm L × 100 μm i.d.	3.6 μm C4-core-shell silica (200 Å)	Phenomenex	450
120 cm L × 100 μm i.d.	3.6 μm C8-core-shell silica (200 Å)	Phenomenex	–
120 cm L × 100 μm i.d.	3.6 μm C18-core-shell silica (200 Å)	Phenomenex	–
100 cm L × 100 μm i.d.	3 μm C18-nonporous silica	Micra	–
25 cm L × 100 μm i.d.	1.5 μm C18-nonporous silica	Micra	–
120 cm L × 100 μm i.d.	3.4 μm C4-core-shell silica (400 Å)	AMT	450

[a] The peak capacity (C_p) was estimated for 800 min separations.
[b] Not estimated due to significant peak tailing.

forms eluted from the high-resolution RPLC, and operated under following conditions: 240,000 resolution (at m/z 400) for precursors ($600 < m/z < 2000$), AGC of 1E6 and 4 micoscans; 120,000 resolution (at m/z 400) for fragments ($100 < m/z < 2000$) with AGC of 5E4 and 1 microscan; top 10 most intense ions were selected for HCD dissociation at 30% collision energies with dynamic exclusion duration of 120 s. A Thermo LTQ ion trap mass spectrometer was applied for examination of long column RPLC for elution of large proteins and used under conditions of $600 < m/z < 3800$ with three microscans and AGC of 5E4.

2.4. Data analysis

The FT MS/MS data were processed using an in-house developed database search engine MSPathFinder (http://omics.pnl.gov/software/mspathfinder). MSPathFinder takes a fasta file, a spectrum file, and a list of modifications as an input and reports proteoform spectrum matches (PrSMs) and their scores. The spectra were searched against the *S. oneidensis* database containing ~4300 proteins [21]. In addition to the intact protein sequences, extra sequences derived by N-terminal Met cleavage and a single internal cleavage were included. Acetylation of protein N-terminus, Oxidation of Met, and dehydration of Cys were considered as dynamic modifications. The database search parameters were set as follows: precursor and product ion mass error tolerance of 10 ppm, sequence length range of 21–500 residues, precursor ion charge range of 2–30 and product ion charge range of 1–15. A decoy database was constructed by reversing the protein sequences and applying 3 amino acid mutations at random positions and assignment was achieved by cutting off the dataset with specific false discovery rate (FDR, at the proteoform level) using target-decoy approach [22].

3. Results

3.1. High-resolution RPLC of proteoforms using long columns packed with short alkyl bonded particles

A 120 cm capillary column containing the 3.6-μm C4-core-shell particles with 200 Å pores was examined for separation of *S. oneidensis* lysate, and the resultant RPLC–MS chromatogram is shown in Fig. 1. More than 200 peaks were readily evident (manually counted peaks with S/N >3). Such resolution (see extended chromatograms, Fig. 1B and C) is significantly higher than previously reported any liquid phase separation approach compatible to MS detection. Although the elution order was primarily determined by hydrophobicity, it was also significantly correlated with molecular mass, e.g., <10 kDa species were dominant in the early (<220 min), 10–20 kDa species in the middle (~220–480 min), and >20 kDa species in the late (~480–800 min) elution zones (see spectra inserted in Fig. 1A). Different sizes of proteoforms displayed varied peak widths across the separation, and hence we estimated the total peak capacity of ~450 for this RPLC separation (listed in Table 1), by summing peak capacities contributed from early (~200 with average peak width of ~1 min at 10% peak height), middle (~150 with peak width of ~1.5 min) and late (~100 with peak width of ~3 min) zones, and excluding contributions from highly intense broad peaks. The largest proteoform (intact degQ confirmed by using UStags approach [23,24], see below) detected by the Exactive mass spectrometer had a molecular mass of 44 kDa.

3.2. Benefits of use of long columns for RPLC resolution of proteoforms

We investigated the impact of column length on RPLC separation and MS detection of *S. oneidensis* proteoforms. Fig. 2 shows separations obtained using 120 cm and 20 cm columns packed with C4-core-shell particles (200 Å pores). A gradient of 180 min was applied, which benefits use of short columns. The 20 cm column provided peak capacity of ~150, about half of what was achieved using a 120 cm column (i.e., ~300 in the same separation time window). Broad peaks were observed for both small (e.g., ~10 kDa) and large (e.g., >20 kDa) proteoforms, which degraded the separation resolution and thus simplify the chromatogram. We examined MS detection of >20 kDa proteoforms (spectra inserted in the figure) and found that of the twelve proteoforms identified with the 120 cm column, only four were observed using the 20 cm column. The intensity of the spectra obtained using the short column was roughly half of the intensity obtained using the long column, an observation we attribute to peak broadening, which impeded the detection of lower abundance proteoforms. Extending the separation to 800 min by slowing the gradient improved the separation on the short column (Supporting Fig. 1), but the resolution was still significantly worse than in the case of the long column.

3.3. High-resolution RPLC of proteoforms using various sizes of packing particles

In principle, reducing the packing particle diameter for a specific column length improves RPLC separation efficiency. However, given an operation pressure limit, decreasing particle size will reduce the maximum useful column length since the pressure required is proportional to the column length and the inverse square of the particle size; this in turn reduces achievable resolution (Fig. 2). To investigate the net gain obtained by reducing the particle size at a specific operation pressure (i.e.,

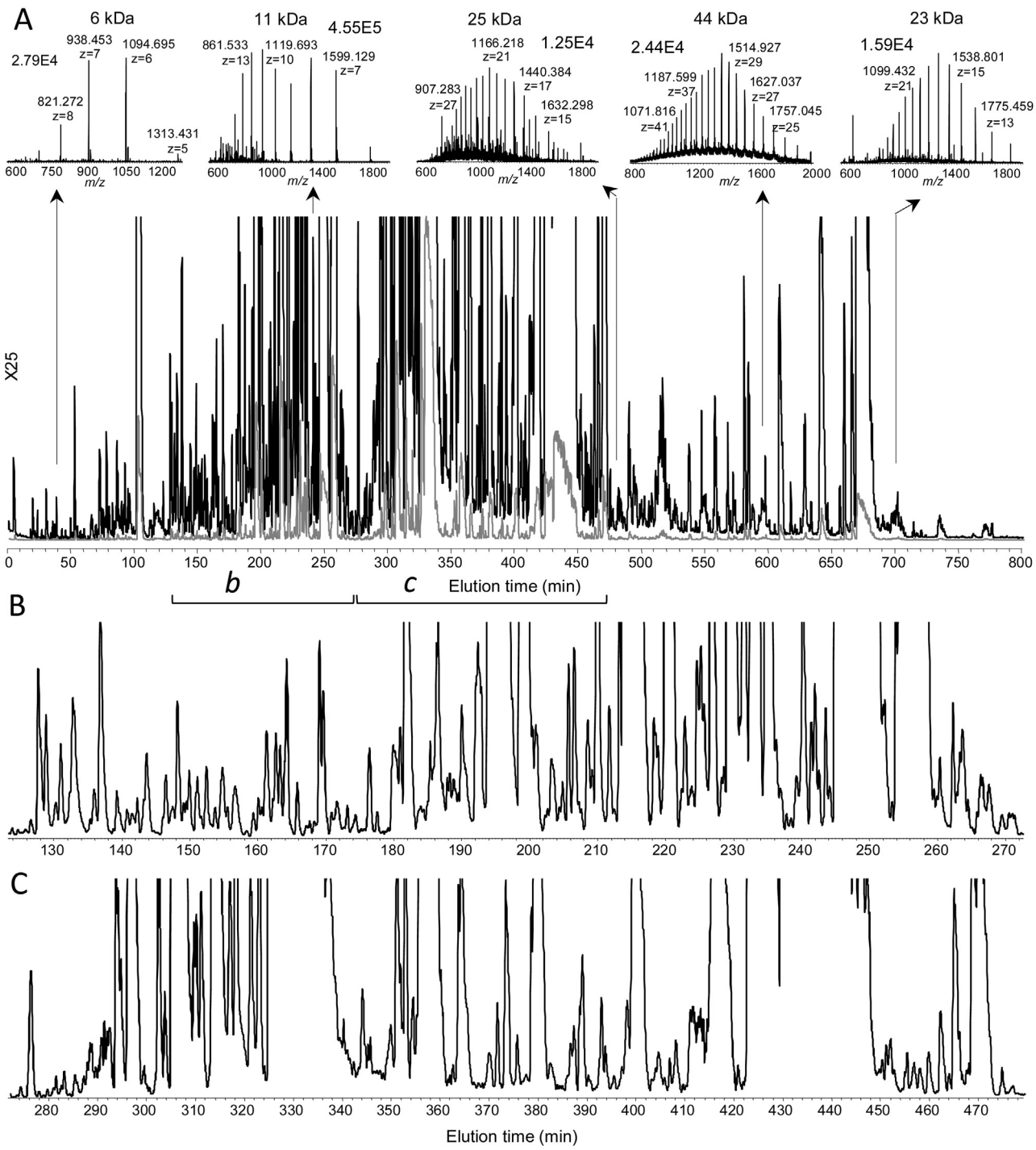

Fig. 1. High-resolution RPLC–MS chromatogram of global proteoforms in *S. oneidensis* lysate on a long column. Conditions: 120 cm × 100 μm i.d. column packed with 3.6 μm C4-core-shell particles (200 Å pores); 2.5 μg of the sample; an Exactive mass spectrometer for detection. Detailed conditions are given in text. A: Base peak chromatogram and mass spectra along the separation (the base peak profile was inserted with grey line). B and C: Expansions of the chromatogram for dense peak zones *b* and *c*, respectively. Numbers xxxEx represent mass peak intensity (scaled by Thermo Xcalibur™ software) (same below).

14 K psi), we packed 30–200 cm columns with 5, 3, and 1.7 μm porous C4-bonded porous (300 Å pores) particles so as to generate similar mobile phase linear velocities (~0.1 cm/s) or flow rates (~0.7 μL/min). Fig. 3 shows chromatograms obtained using a 200 cm column packed with 5 μm C4 particles and a 30 cm column packed with 1.7 μm C4 particles. While both columns provided dense arrays of peaks, the 5 μm particle packed column generated greater peak capacity (~350) than the 1.7-μm particle packed column (~300). The 1.7 μm particle-packed 30 cm column provided better resolution than the short (i.e. 20 cm) column packed with larger (3.6 μm) particles (Fig. 3B and Supporting Fig. 1), indicating that use of small particles has a potential for further improvement of RPLC resolution for separation of proteoforms when higher pressure LC systems become practical.

A general concern for the use of long columns may be reduced sample recovery in comparison to short columns. The observation of intense peaks in the late elution zone (550–800 min in Fig. 3A and C) and mass spectra obtained on a 200 cm column *vs.* a 30 cm

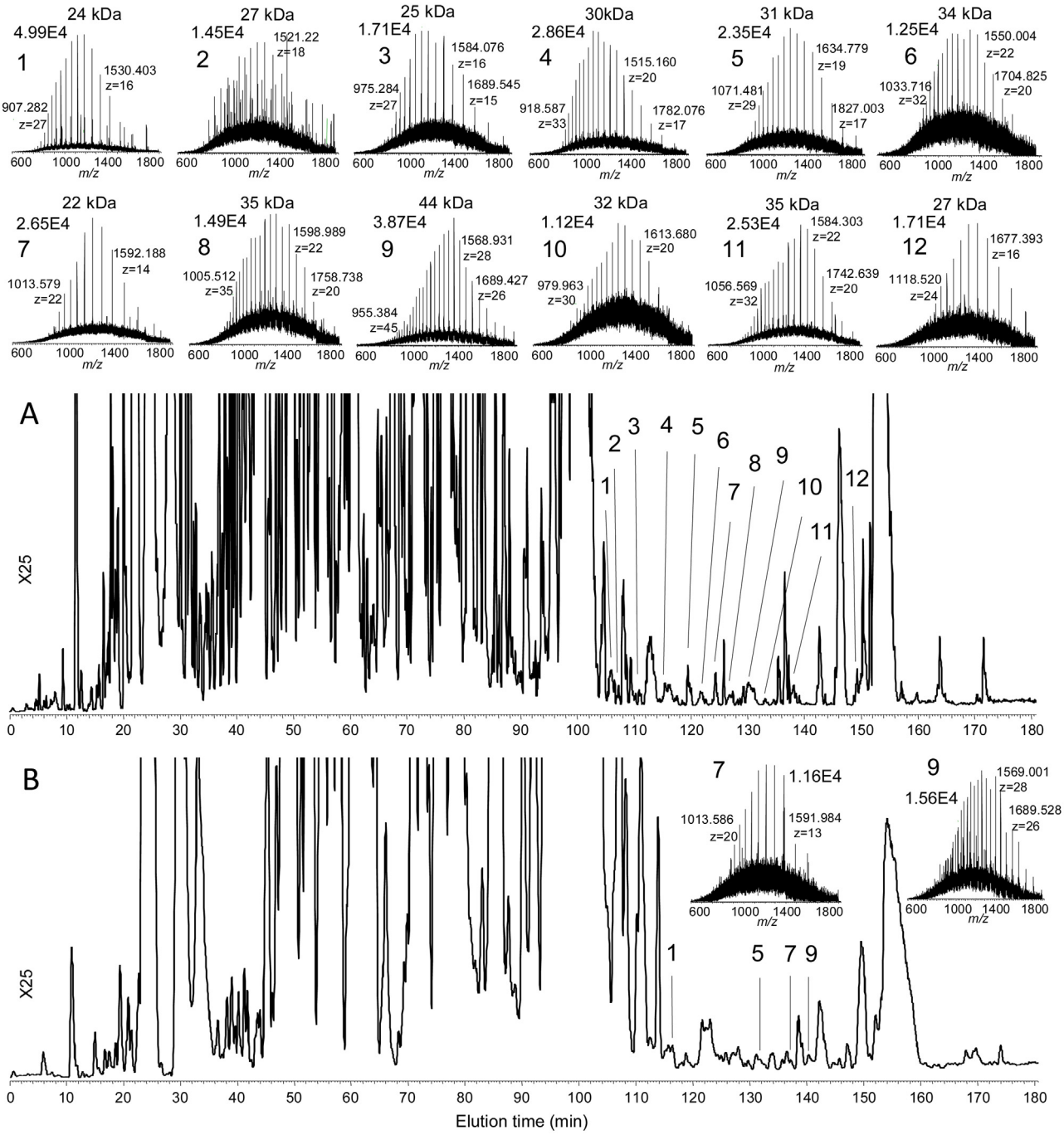

Fig. 2. Contribution of column length to RPLC resolution and MS detection of proteoforms. A: The 120 cm and B: 20 cm 100-μm i.d. columns containing 3.6-μm C4-core-shell particles; the 120 cm column operated at 96.5 MPa (14 K psi) and the 20 cm column at 17.2 (2.5 K psi). The spectra indicate detection of low intensity, >20 kDa proteoforms observed with use of short and long columns. Other conditions are the same as for Fig. 1 except for accelerating the gradient to complete separations in 180 min.

column (450–680 min in Fig. 3B and C) for proteoforms with masses above 20 kDa indicate minimal losses during the separation of a small size of samples (2.5-μg) on a long column.

3.4. Weak interaction RPLC vs. strong interaction RPLC for high-resolution separation of proteoforms

The length of alkyl chains bonded on the surface of packing particles is potentially a major factor affecting resolution and recovery of proteoforms. We compared the RPLC–MS performances columns packed with 3 μm porous (300 Å pores) particles bonded with C1, C2, C4, C8 and C18 for separation and detection of proteoforms. Using 100 cm columns packed with C2- and C4-particles (Fig. 4A and B), we counted ~75 peaks in the late elution zones (500–800 min). Furthermore, the C2 RPLC provided >5-fold greater intensity than the C4 RPLC for the 44 kDa protein (degQ) (Fig. 4E*a* and *b*); while the C4 RPLC provided densest peaks for the middle elution zone rich in 10–20 kDa proteoforms. C1 RPLC has the greatest potential in terms of elution of hydrophobic and larger proteoforms; accordingly, we observed dense peak arrays in the C1 RPLC late elution zone (Supporting Fig. 2). However, aberrant behavior was observed for elution of several hydrophilic proteoforms on the C1 column. For instance, a polypeptide (~6.6 kDa) eluted from the C1-column for over 30 min, (*vs.* ~2 min on the C2-

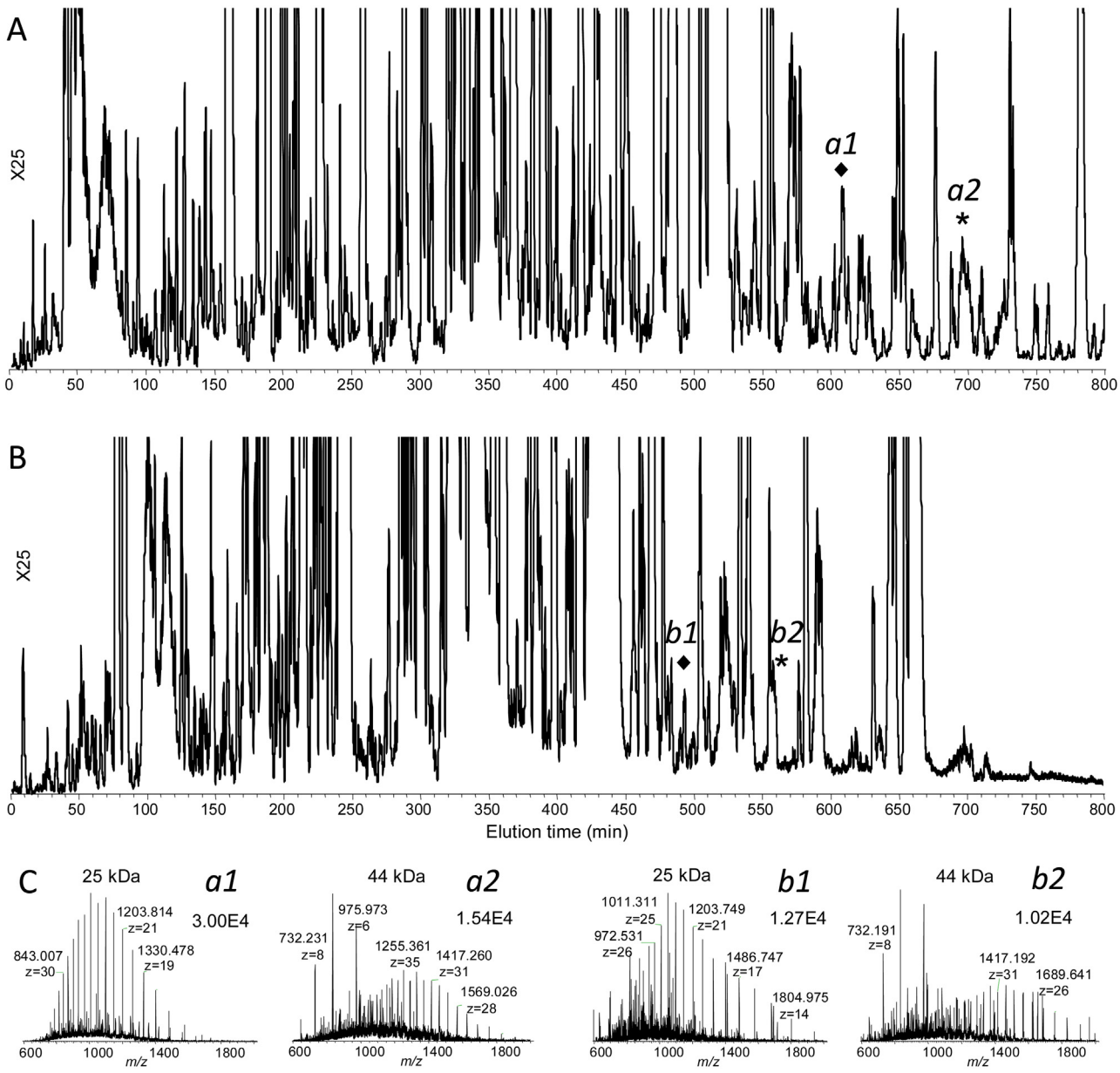

Fig. 3. RPLC–MS of global proteoforms using a large particle-packed long column and a small particle-packed short column. Conditions: A: The 200 cm × 100 μm i.d. column containing 5 μm C4-porous (300 Å) particles and B: 30 cm × 100 μm i.d. column containing 1.7 μm C4-proues (300 Å) particles; two columns were operated at the same pressure limit of 96.5 MPa (14 K psi). C: The spectra *Ca1-Ca2* show detection of two >20 kDa proteins for A, in comparison with *Cb1-Cb2* for B (♦ and * labeled respectively represent their elution). Other conditions are the same as for Fig. 1.

column) and carried more charge (e.g., z = 9, 10 and z = 6, 7 for C1 and C2 columns, respectively) (Supporting Fig. 2). It is noted that these two phases were bonded on the same type of silica particles by the same vendor and evaluated under the identical LC–MS conditions. We speculate that the C1-bonded particles may have a polar surface (e.g., ionic silanol groups) that exerts secondary interactions to broaden chromatographic peak for hydrophilic polypeptides and potentially increases the protonation of hydrophilic proteoforms.

C8 and C18 RPLC provided higher quality separations for <10 kDa proteoforms eluting in the early elution zone (i.e., <200 min) (Fig. 4C and D) than C1-C4 RPLC. However, some badly tailing peaks were observed for the late zones, preventing useful estimation of peak capacity. There were only ~25 peaks in the late elution zone (500–800 min) *vs.* ~75 observed for C2 and C4 RPLC. In addition, the largest proteoform observed was reduced to 31 kDa and 25 kDa for C8 and C18 columns, respectively (Fig. 4E*c* and E*d*). Similar behavior, in terms of chromatographic separation and MS detection, was observed for RPLC columns packed with C8- and C18-core-shell particles (Supporting Fig. 3), even though these particles have the surface structure different from that of the porous particles. We further examined RPLC of proteoforms on C18-bonded nonporous particles, which have traditionally been used for fast separation of large, hydrophobic proteins [11,12,25]. However, we found hydrophobic proteoforms, eluting in the late elution zone (after 500 min), had poor peak shapes on columns packed with this type of particles (Supporting Fig. 4). Sharp peaks (peak widths <1 min) were only observed for low intensity species (not shown), while majority of the moderate and high intensity species exhibited broad peaks (peak widths >10 min) across the entire separation. This was in striking contrast to porous and superficially porous particle-packed C18 columns and may be attributed to the limited specific surface area of nonporous particles and thus column

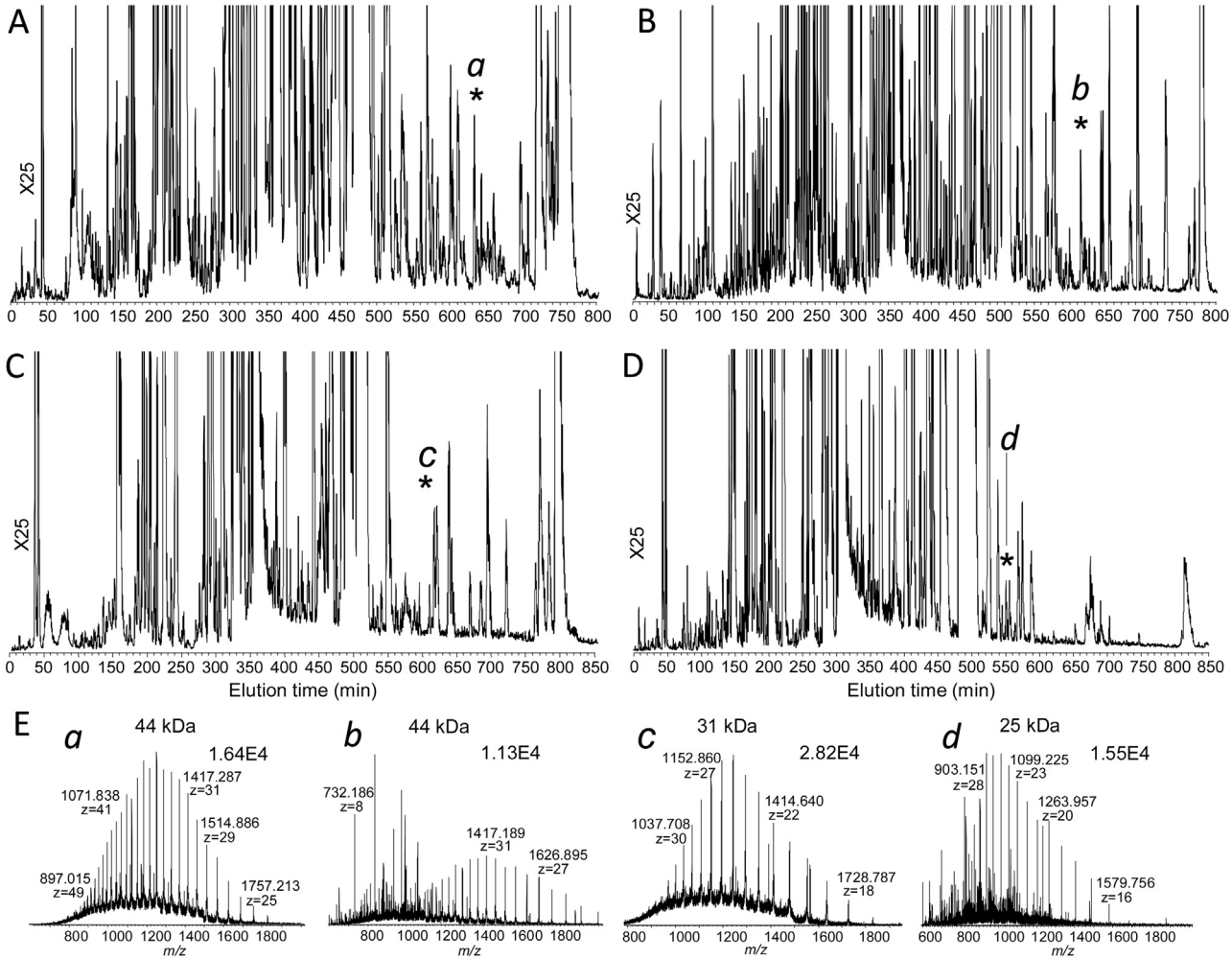

Fig. 4. RPLC–MS performances for separation and detection of global proteoforms on columns containing particles bonded with different lengths of alkyl chains. Conditions: 100 cm × 100 μm i.d. columns packed with 3 μm porous (300 Å) particles bonded with C2 for A, C4 for B, C8 for C, and C18 for D. The spectra E*a*, E*b*, E*c*, and E*d* show the largest proteoforms detected respectively for A, B, C, and D (* labeled represents its elution). Other conditions are the same as for Fig. 1.

sample capacity, even though only 2.5 μg of the proteomic sample was loaded on the large volume (e.g., 100 cm × 100 μm i.d.) column. Reducing the particle size from 3 to 1.5 μm, while shortening the column by 4-fold (limited by the operation pressure), further lower the sample capacity and thus degraded the separation.

3.5. Elution of large proteoforms on long column weak interaction RPLC

Larger than 50 kDa proteoforms in the *S. oneidensis* lysate were not detected in our RPLC-Exactive FT MS experiments described above. We collected the RPLC eluent of the sample and confirmed existence of >50 kDa proteins using gel electrophoresis (not shown), suggesting that the mass spectrometer used was less effective for detection of large proteins. We utilized an ion trap LTQ MS for detection and observed large proteoforms from the sample, as shown in Fig. 5. Proteoforms with masses of 50–90 kDa were chromatographed on a C4-core-shell particles (200 Å pores)-packed column. (The proteome masses were estimated from the charge envelope). These >50 kDa proteoforms were observed for low intensity RPLC-ion trap MS peaks (Fig. 5B), but totally absent by using Orbitrap-based mass spectrometers (i.e., Exactive here and Elite below). A mouse cell lysate was tested for further examination of the capability of the long columns to elute large proteoforms. As shown in Fig. 6, proteoforms with masses up to 110 kDa were detected from RPLC–MS with use of a 120 cm column packed with C4-core-shell particles (200 Å pores). Similar to the observed for *S. oneidensis* cell lysate (Fig. 5B), large proteoforms in the mouse sample were also detected from low intensity peaks. High-resolution RPLC benefits detection of such low intensity peaks, as described above for Fig. 2.

The large pores on the packing particle surface may improve mass transfer of large protein molecules and thus the separation efficiency. Our experiment demonstrated that 57–110 kDa mouse proteins were again detected from low intensity RPLC–MS peaks with use of a column packed with 400 Å pore size C4-core-shell particles, but limited improvements in the protein charge envelopes and mass peak intensities detected were observed (Supporting Fig. 5) in comparison with use of the 200 Å pore size C4-core-shell particle-packed column (Fig. 6). The 400 Å pore C4-core-shell particle-packed column was estimated to generate peak capacity of 450 (Supporting Fig. 6) and resolved proteoforms in the lysate similarly as the 200 Å pore particle-packed column. The similar RPLC–MS performances of the 400 Å and 200 Å pore C4-core-shell particle-packed columns were also observed for the *S. oneidensis* lysate cell sample (Fig. 5 and Supporting Fig. 7). These suggest that under conditions of RPLC where proteins are denatured (in organic solvents) the pore size in a range of 200–400 Å on the packing particles' surface has a limited influence on high-resolution separation

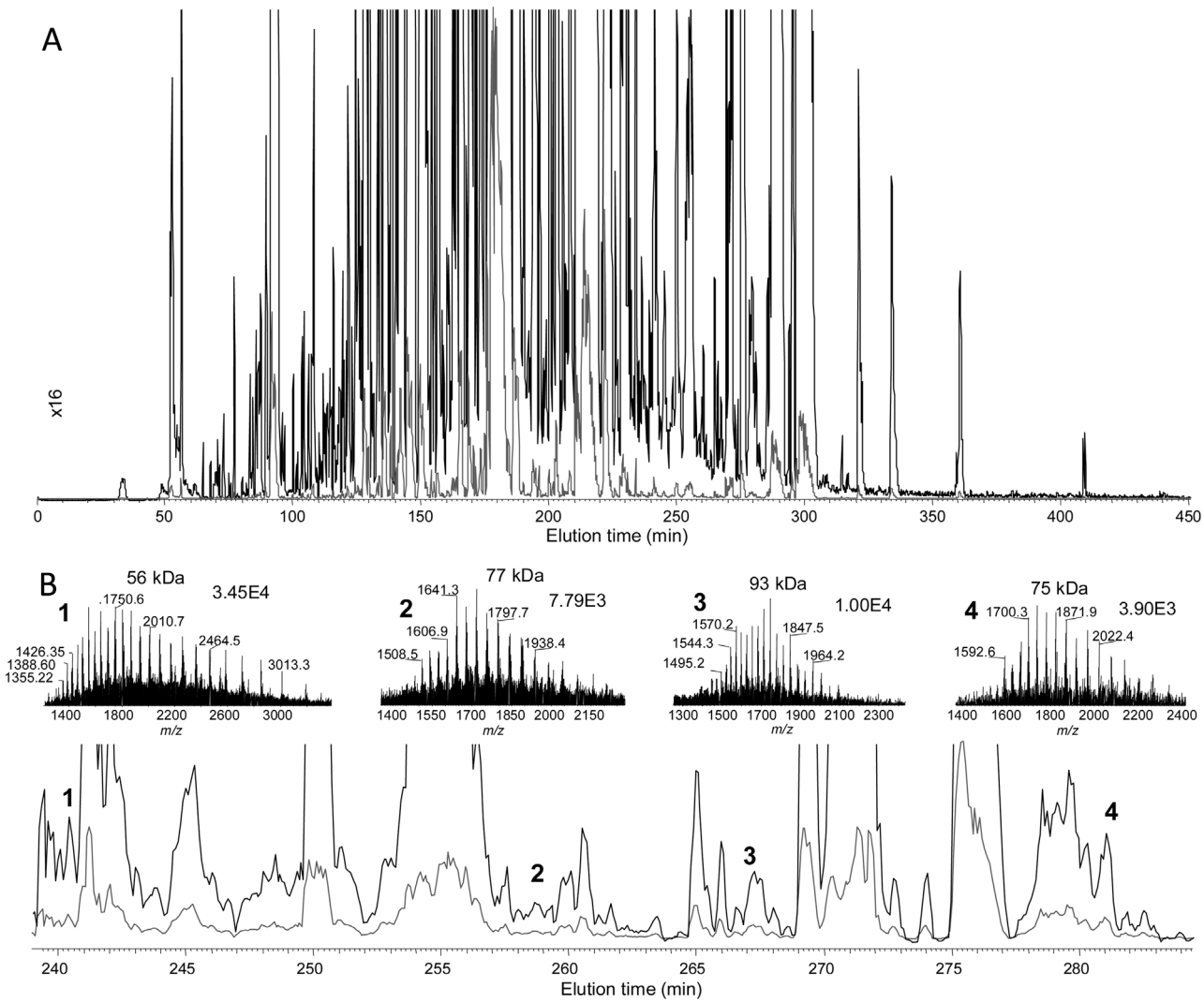

Fig. 5. RPLC elution and ion trap MS detection of >50 kDa proteoforms in *S.oneidensis cell* lysate using a column packed with C4-core-shell porous (200 Å pores) particles. A: RPLC was performed under the same conditions as for Fig. 1 except for use of 500 min gradient; ion trap LTQ MS detection; 1.5 μg of the sample. B: The expanded RPLC–MS chromatogram shows the evidence for MS detection of 50–90 kDa proteoforms from low intensity peaks. Other conditions are detailed in text.

of proteoforms with masses up to 110 kDa for both microbial and mammalian proteomic samples.

The capability of totally porous particles for separation of >50 kDa proteoforms were examined (Supporting Figs. 8 and 9). These large proteoforms can also be eluted from the totally porous (300 Å pores) particle-packed column. In comparison with core-cell superficially porous particles, totally porous particles provided low RPLC–MS intensities for the last eluted three large proteoforms (76–110 kDa) (Supporting Fig. 8), although less hydrophobic C2 phase was applied. Potential existences of deep pores underneath the surface of the totally porous particles should be responsible for the low recovery observed. Such situation was consistently observed for the last eluted 75 kDa proteoform for the *S.oneidensis* cell lysate sample (Supporting Fig. 9B).

3.6. Initial investigation of the Orbitrap FT MS/MS capability for assignment of proteoforms eluted from the high-resolution RPLC

FT MS/MS was employed for identification of *S. oneidensis* proteoforms eluted from the C2 RPLC. Under the chromatographic conditions as described for Fig. 4A, 1745 precursor spectra and 17449 HCD tandem mass spectra were collected in a single RPLC-FT MS/MS experiment for the sample. The collected dataset was searched against *S. oneidensis* database using MSPathFinder (see Experimental section) for proteome assignment. The number of different proteoforms obtained on various levels of FDR (based on proteoforms) is given in Fig. 7A. Increasing FDR from 1% to 5% and 10% respectively resulted in 13% and 20% increases for proteoforms assigned and 20% and 33% increases, respectively, for proteins. New proteins brought in with relaxed confidence (or high FDR) had less proteoforms assigned, which causes the higher percentage increase in protein assignments. Proteoforms with masses below 12 kDa dominate the list of the assigned species (i.e., ~80% of the total) regardless of the FDR cutoff (Fig. 7B). The FDR cut-off value more significantly affected assignments for large proteins than for small ones; e.g., loosening FDR from 1 to 5 and 10% increased the number of >20 kDa proteoforms by 2.5- and 3.3-fold, respectively, in contrast to the limited increases for <18 kDa proteoforms. Identification of larger proteoforms was found to require higher intensity precursors than in the case of small proteoforms; e.g., the minimal precursor ion intensity required for confident (at 1% FDR) MS/MS assignment increased an order of magnitude as the proteoform mass increased from ~5 kDa to >20 kDa (Fig. 7C). We speculate this is one of reasons for diminishing coverage of proteoforms with masses above 20 kDa in RPLC–MS/MS analyses

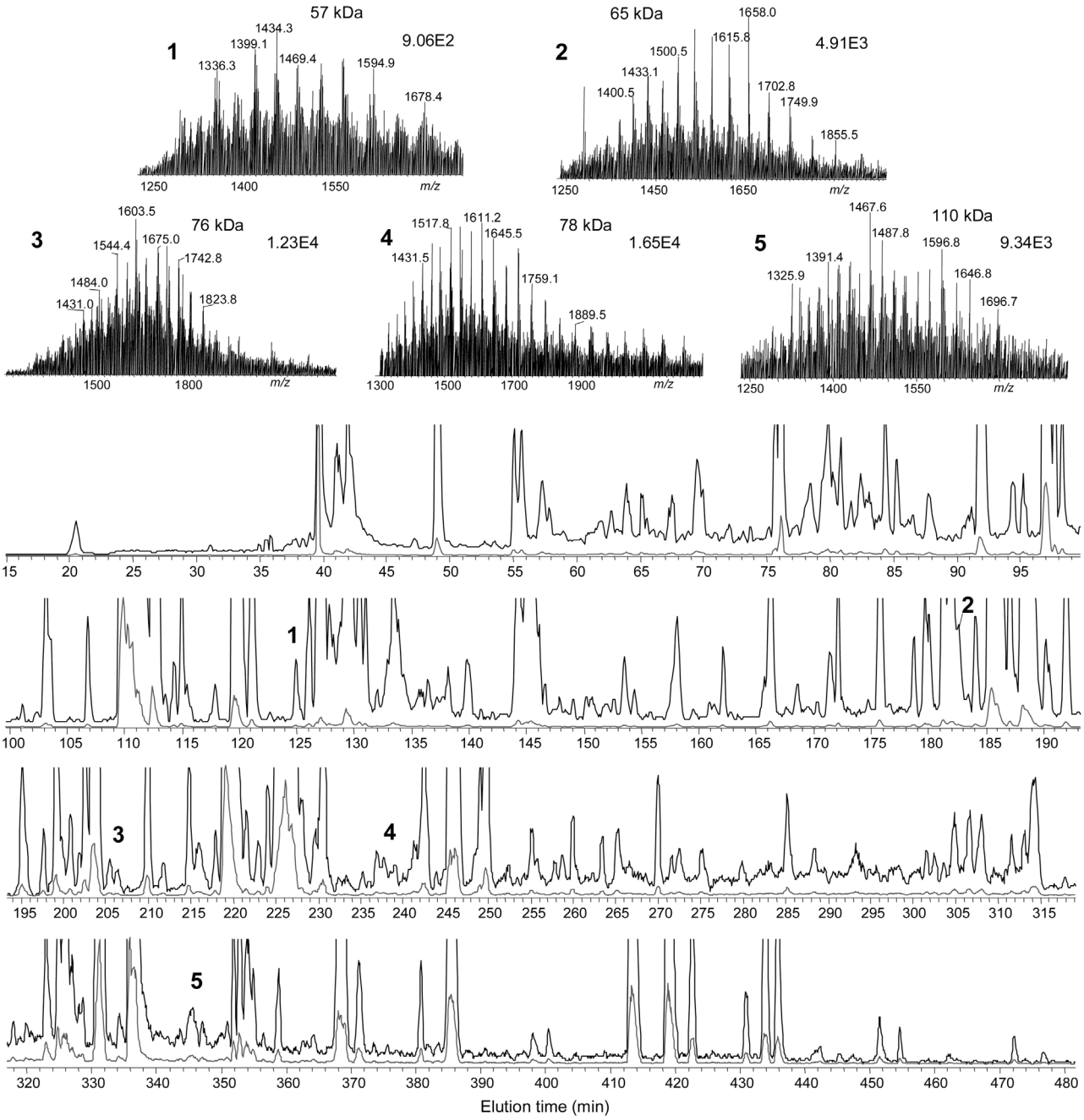

Fig. 6. RPLC elution and MS detection of >50 kDa proteoforms in mouse cell lysate using a column packed with C4 core-shell porous (200 Å pores) particles. RPLC was performed under the same conditions as for Fig. 1 except for use of mobile phase gradient from solvent A (10% ACN, 0.1%FA) to 75% solvent B (70%ACN, 30%IPA, 0.1%FA) in 600 min for separation; ion trap LTQ MS detection; 1.5 μg of the mouse sample. The expanded chromatograms and spectra evidence that the 50–110 kDa mouse proteoforms were detected from low intensity peaks. Other conditions are detailed in text.

of the global cell lysates, since most of these proteoforms were detected with relatively low MS-peak intensity (see spectra inserts, Figs. 1–6). The largest protein assigned was 44 kDa intact degQ and the assignment was confirmed with the UStags approach (Fig. 7D). For the precursor measured (Fig. 7D*a*, $m/z = 94.282$, $z = 47$), its HCD FT MS/MS spectrum (Fig. 7D*b*) led to obtain of sequence LLKDQEGAVALK sequenced with absolute mass errors of <0.005 Da from the fragment peaks determined with relative mass errors of <10 ppm. This sequence uniquely belongs to *S. oneidensis* degQ (SO_3942; a peptidase) among other database entries and was referred as a unique sequence tag (UStag) of *S. oneidensis* degQ. The UStag were then located in degQ protein (Fig. 7D*c*) and the proteoform was determined (with a mass error of 2.7 ppm) as degQ Ala27-Arg450 according to the precursor intense peak ($z = 30$). Sequence of Ala27-Arg450 was finally determined as intact *S. oneidensis* degQ without hydrolytic truncations, as the head sequence Met1-Ala26 was determined as the signal peptide by SignalPeP (Fig. 7D*d*). The >50 kDa proteoforms shown in Fig. 6 cannot be detected/identified yet by the Orbitrap instrument applied in this work.

4. Discussion

Top-down proteomics challenges every step in a traditional proteomics pipeline, including sample manipulation, separation, mass spectrometry and data analysis. Herein, we describe high-resolution protein separations using long column WI-RPLC, the

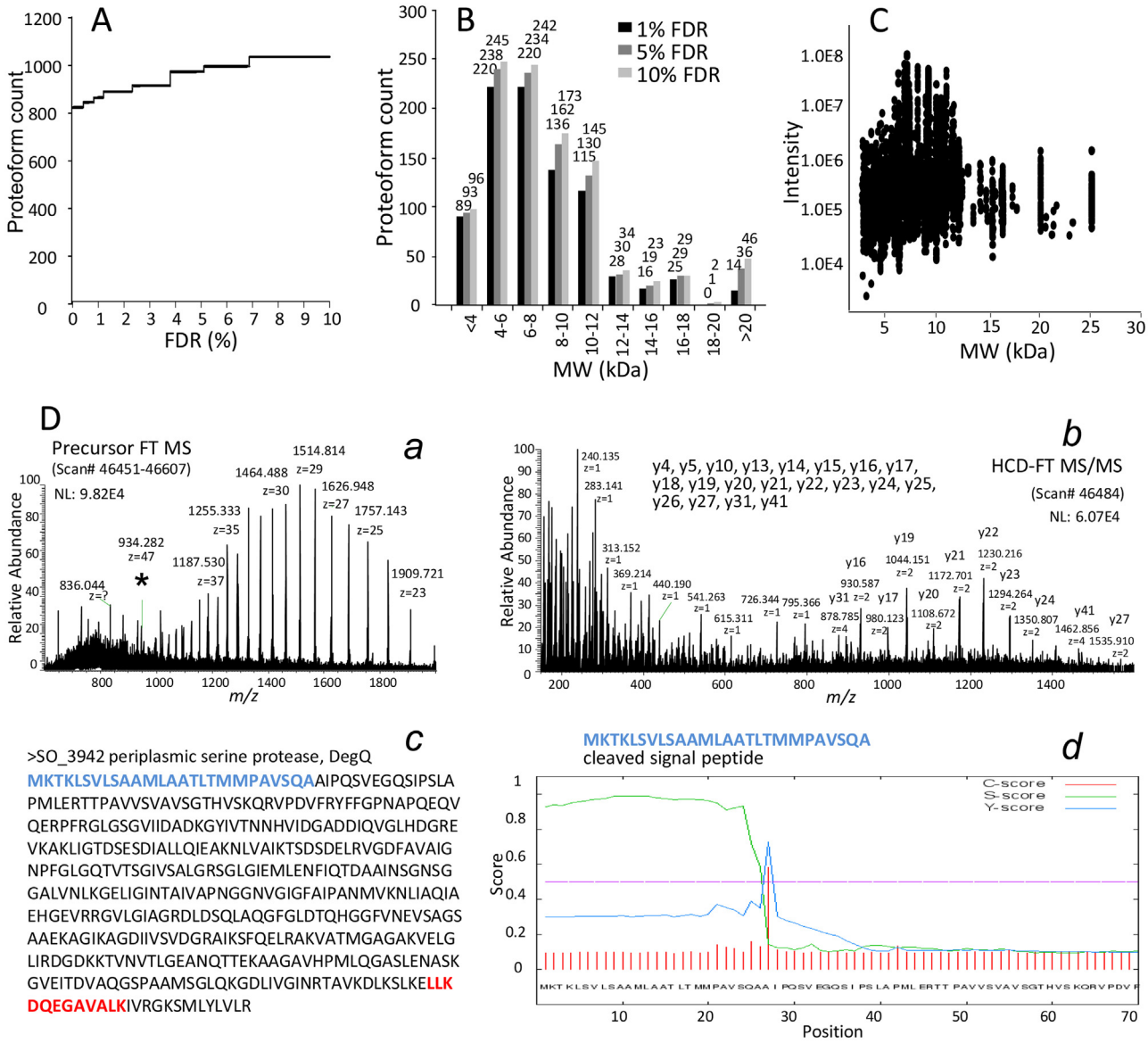

Fig. 7. Initial examination of Elite Orbitrap FT MS/MS for analysis of *S. oneidensis* cell lysate. Conditions: RPLC was performed under same conditions as for Fig. 4A and conditions of FT MS/MS on a Thermo Elite mass spectrometer are detailed in the Experimental section; 2.5 μg of the sample. A: The count (number) of different proteoforms assigned at various FDRs (proteoform level) by MSPathFinder. B: Distributions of various sizes of proteoforms assigned at FDRs of 1%, 5%, and 10%. C: The intensity of various sizes (kDa) of precursors assigned at 1% FDR. D: An example showing confident assignment of the 44 kDa protein from *S. oneidensis* cell lysate; *a*: FT MS-measured precursor, *b*: HCD FT MS/MS spectrum that leads to sequencing an UStag of degQ, *c*: localization of the UStag in degQ; *d*: signal peptide determination by using SignalPeP; detailed interpretation for the assignment are given in text.

most robust separation modality for direct coupling of LC to MS with most widely and conveniently applied methods. Strong adsorption of proteins on reversed phases is one of the major factors affecting protein chromatographic kinetics and a major obstacle for achieving high-resolution RPLC of proteins. Short alkyl bonded particles have a potential to minimize protein adsorption due to their lower hydrophobicity, but has been rarely explored in the context of top-down proteomic separation. In this work, we consistently demonstrate that RPLC resolution for proteins depends on column length when short alkyl bonded particles are used as column packings. With employing such approach, complex chromatograms exhibiting ~200 peaks were for the first time shown for separation of proteoforms (Figs. 1, 3, and 4*a*–*b*). Porous and superficially porous packing particles bonded even with C1 and C2 chains provided sufficient retention to enable high-resolution top-down proteomic separations. Retention of proteoforms appears to be more dependent on particle surface area than the length of bonded alkyl chains. For instance, proteoforms exhibited longer retention on C1 porous particles than on C18 nonporous particles (Supporting Figs. 2 and 4). In addition to improving resolution, long columns also benefit sample loading capacity, which is important for achieving greater dynamic range of measurements. Conversely, use of larger diameter columns for this purpose will lead to increased mobile phase flows and reduced ESI–MS sensitivity and dynamic range coverage [20].

In contrast to WI-RPLC described herein, SI-RPLC is widely applied for RPLC in current top-down proteomics practices [6,10–18]. For SI-RPLC, separation resolution of proteins has been claimed to less depend on column length [7,8]; hence, improved RPLC resolution can potentially be achieved only by reducing the size of column packing particles (and thus improving protein mass transfer). For fair comparison, we applied various SI-RPLC columns (e.g., 25–120 cm length, packed with 1.5-3.6 μm superficially porous and nonporous C8 and C18 particles) to separate the

tested cell lysate. The resultant proteomic SI-RPLC chromatograms consistently showed reduced complexity and greater peak tailing than corresponding WI-RPLC chromatograms; while replacing C18 with C8 in SI-RPLC only provided a limited improvement for top-down proteomic separation (Fig. 4*b–c* and Supporting Fig. 3). SI-RPLC may provide some selectivity for specific proteoforms through hydrophobic interactions, but this is less important for proteomics applications due to the complexity of the samples and unavoidable overlaps of peak profiles.

The surface physical structure of particles is currently one of major criterions for selection of stationary phases to perform top-down RPLC proteomics applications. According to Renkin's formula [26], increasing molecular size from 10 to 40 kDa (and diameter from 30 to 50 Å, assuming the molecules as spheres) reduces the molecular diffusion by ~20% in 300 Å pores. Such limited reduction does not explain the peak broadening (e.g., doubling the peak widths) that we observed for 20–40 kDa proteins compared to ~10 kDa proteins. Also, the 200 Å pore particle-packed column allowed elution of 70–110 kDa proteoforms with the similar performance (evaluated with MS detection of proteoforms) as the 400 Å pore particle-packed column (Fig. 6 and Supporting Fig. 5). These evidences show that under RPLC conditions (i.e., proteoform denaturing) the pore sizes in a range of 200–400 Å on the particle surface has a limited contribution for proteoform elution. In contrast, it is commonly observed that proteins having similar sizes (masses) can have significantly different peak widths on RPLC, while in size exclusion chromatography greatly different sizes of proteins can elute from 300 Å porous particle-packed columns with similar peak shapes (widths). Thus, we believe that some detrimental (e.g., irreversible and/or secondary) interactions between proteins and packing particles' surface should be the major factor responsible for broadening protein peaks during elution along RPLC columns. The structure of superficially and totally porous particles appears to affect recovery (especially for late eluted proteins, Supporting Figs. 8 and 9), but little peak shapes. Reduced sample capacity remains the major limitation for use of nonporous particles for top-down proteomics.

Long column WI-RPLC enables eluting large (e.g., >100 kDa) proteoforms (Fig. 6, Supporting Figs. 6 and 8), which is benefited by the better recovery from the weak hydrophobic stationary phases than strong interaction reversed phases (i.e., long alkyl chains) [27,28]. The proteoforms confidently identified in this work was in the <50 kDa range. We ascribe the identification absence of larger proteins to low sensitivity of the Orbitrap FT MS applied in this work, especially in the *m/z* range of 1500–3000. The low FT MS sensitivity mainly resulted from poor transfer of large ions from ion trap to Orbitrap, as 50–110 kDa proteins of the proteomic sample were detected by using a simple ion trap mass spectrometer. (The parameters related to such ion transfer for the instrument used in this work are not allowed to be optimized by users). Pre-enrichment of large proteoforms, e.g., through gel electrophoresis that has been applied to improve detection of a 100 kDa protein [6], may help for their identification, but this will sacrifice analysis sensitivity and throughput.

The peak capacities of ~400 achieved in this work for WI-RPLC of proteins are equivalent to performances typically obtained for small peptides in conventional bottom-up proteomics RPLC separations (e.g., 3.5 μm C18 packed 25 cm column) [29] that often lead to identification of dozens of thousands of peptides when coupled with advanced mass spectrometry [1]. However, with such separation power, we cannot achieve comparable proteome coverage from top-down approach. For instance, our initial results using a Thermo Elite mass spectrometer resulted in assignment of ~900 proteoforms at 1% FDR. We attribute this mainly to inadequate FT MS/MS performances. Advances in sensitive high-performance mass spectrometry, including effective detection of ions in a large *m/z* range (e.g., up to 3000), advanced software (e.g., enabling 'smart' MS/MS of multiple charge states), and more efficient dissociation methods, are needed to realize the full potential of high-resolution intact protein separations reported herein. We are currently testing a new version Orbitrap Fusion Lumos mass spectrometer and exploring higher-field (e.g., 21T) FTICR MS with implement of ultraviolet photodissociation (UVPD, previously shown to provide extensive and essentially charge state independent fragmentation of intact proteins [30]) for top-down proteomics, and results will be presented elsewhere.

5. Conclusions

Proteoforms (including intact proteins and their truncations) having masses up to ~50 kDa can be efficiently separated with peak capacity of >400 under the denaturing conditions applied in RPLC. This quality of separations and the complexity of chromatograms presented in this work are previously unattainable by any single-dimension separation technique such as LC and capillary electrophoresis. The RPLC resolving power for proteins was affected by column dimensions and physicochemical properties of the packing stationary phases, and column length was found to be the most important factor for separating proteins with masses up to 50 kDa. Shortening column degraded both RPLC resolution and subsequent MS detection sensitivity regardless of size and surface structure of the particles packed in columns. Long separations (> 10 h) with slow gradients along long columns (> 1 m) were particularly favorable when coupled with FTMS. For a microbial *S. oneidensis* lysate, ~900 proteoforms can be confidently (1%FDR) identified from a single RPLC-FT MS/MS analysis of low microgram sizes of proteomic samples using in-house developed search engine MSPathFinder. Large proteoforms (50–110 kDa) were chromatographed on long RPLC columns, while the pore sizes (200–400 Å) and surface structure (porous and superficially porous) of particle packings had minor effect on RPLC separation. In contrast, the length of alkyl chains bonded on particles greatly influenced RPLC resolution, even for moderate sizes of proteoforms (e.g., 20–30 kDa). Long alkyl chain (C8 and C18) bonded-phases greatly reduced RPLC separation resolution (and thus the chromatogram complexity) and lowered MS sensitivity for detection of proteoforms, and such strong interaction RPLC should be primarily utilized for separation of <10 kDa proteoforms with moderate hydrophobicity.

Long column RPLC will become a mainstream approach for top-down proteomics analysis, at least for proteoforms with masses below 50 kDa. The existing commercial LC instruments have an operation limit of ~100 MPa (15 K psi, Waters UPLC or Thermo UltiMate) can support use of 1 m long columns packed particles down to 3 μm, while use of smaller particles (e.g., <2 μm) packed long columns requires development of practical higher pressure LC instruments (e.g., >200 MPa or >30 K psi). The RPLC separation quality attained in this work is close to that typically obtained in bottom-up proteomics and advancements in MS detection and fragmentation efficiencies, especially for >20 kDa proteoforms are critically needed to further extend the utility of top-down proteomics.

Acknowledgments

Portions of this research were supported by Department of Energy Office of Biological and Environmental Research Genomic Sciences Program under the Pan-omics project. We acknowledge support by the U.S. Department of Energy (DOE) Office of Science, Office of Biological and Environmental Research (BER). This research was performed using EMSL, a DOE Office of Science User

Facility sponsored by BER and located at Pacific Northwest National Laboratory (PNNL). PNNL is operated by Battelle for DOE.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.chroma.2017.01.008.

References

[1] B.R. Zhang, B. Shan, M.-C. Baek, J.R. Yates III, Yates Protein analysis by shotgun/bottom-up proteomics, Chem. Rev. 113 (2013) 2343–2394.
[2] M. Gingras, B. Raught, R. Aebersold, Analysis of protein complexes using mass spectrometry, Nat. Rev. Mol. Cell Biol. 8 (2007) 645–654.
[3] Y. Shen, T. Liu, N. Tolić, B.O. Petritis, R. Zhao, R.J. Moore, S.O. Purvine, D.G. Camp II, R.D. Smith, Strategy for degradomic-peptidomic analysis of the human blood plasma, J. Proteome Res. 9 (2010) 2339–2346.
[4] Y. Shen, N. Tolić, R. Xie, A.A. Purvine, G.A. Moore, R.D. Smith, Effectiveness of CID, HCD, and ETD with FT MS/MS for degradomic-peptidomic analysis: comparison of peptide identification methods, J. Proteome Res. 10 (2011) 3929–3943.
[5] Y. Shen, N. Tolić, T. Liu, R. Zhao, B.O. Petritis, M.A. Gritsenko, D.G. Camp, R.J. Moore, S.O. Purvine, F.J. Esteva, R.D. Smith, Blood peptidome-degradome profiles of breast cancer, PLoS One (2010), http://dx.doi.org/10.1371/journal.pone.0013133.
[6] J.C. Tran, L. Zamdborg, D.R. Ahlf, J.E. Lee, A.D. Catherman, K.R. Durbin, J.D. Tipton, A. Vellaichamy, J.F. Kellie, M. Li, C. Wu, S.M. Sweet, B.P. Early, N. Siuti, R.D. LeDuc, P.D. Compton, P.M. Thomas, N.L. Kelleher, Mapping intact protein isoforms in discovery mode using top-down proteomics, Nature 480 (2011) 254–258.
[7] W.C. Pearson, M.A. Hermodson, F.E. Regnier, Reversed-phase supports for the resolution of large denatured protein fragments, J. Chromatogr. 207 (1981) 325–332.
[8] Grace Vydac FAQs: What effect does column length have on my RP-HPLC column? http://www.seaviewsci.com/vydac/ques6.htm.
[9] Y. Shen, M.L. Lee, General expression of peak capacity for column chromatography, Anal. Chem. 70 (1998) 3853–3856.
[10] A.L. Capriotti, C. Cavaliere, P. Foglia, R. Samperi, A. Lagana, Intact protein separation by chromatographic and/or electrophoretic techniques for top-down proteomics, J. Chromatogr. A 1218 (2011) 8760–8776.
[11] R.E. Rogers, Z. Wu, M.J. Wirth, RPLC of intact proteins using sub-0. 5 μm particles and commercial instrumentation, Anal. Chem. 85 (2013) 6820–6825.
[12] Z. Wu, B. Wei, X. Zhang, M.J. Wirth, Efficient separations of intact proteins using slip-flow with nano-liquid chromatography-mass spectrometry, Anal. Chem. 86 (2014) 1592–1598.
[13] M.J. Roth, D.A. Plymire, A.N. Chang, J. Kim, E.M. Maresh, S.E. Larson, S.M. Patrie, Sensitive and reproducible intact mass analysis of complex protein mixtures with superficially porous capillary reversed-phase liquid chromatography mass spectrometry, Anal. Chem. 83 (2011) 9586–9592.
[14] J. Zhang, M.J. Roth, A.N. Chang, D.A. Plymire, J.R. Corbett, S. Greenberg, Top-down mass spectrometry on tissue extracts and biofluids with isoelectric focusing and superficially porous silica liquid chromatography, Anal. Chem. 85 (2013) 10377–10384.
[15] R. Mohr, M. Samonig, G. Böhm, C.G. Huber, High-efficiency nano- and micro-HPLC – high-resolution Orbitrap-MS platform for top-down proteomics, Proteomics 10 (2010) 3598–3609.
[16] B. Eeltink, M. Desmet, G.D. Blinco, A. Treumann, High-resolution separations of protein isoforms with liquid chromatography time-of-light mass spectrometry using polymer monolithic capillary columns, J. Chromatogr. A 1218 (2011) 5504–5511.
[17] K.A. Tweeten, T.N. Tweeten, Reversed-phase chromatography of proteins on resin-based wide-pore packings, J. Chromatogr. A 359 (1980) 111–119.
[18] C.S. Young, R.J. Weigand, An efficient approach to column selection in HPLC method development, LCGC North Am. 20 (2002) 464–473.
[19] Y. Shen, R. Zhang, R.J. Moore, J. Kim, T.O. Metz, K.K. Hixson, R. Zhao, E.A. Livesay, H.R. Udseth, R.D. Smith, Automated 20 Kpsi MS and MS/MS with chromatographic peak capacity of 1000–1500 for proteomics and metabolomics, Anal. Chem. 77 (2005) 3090–3100.
[20] R. Shen, G.A. Berger, N. Rodriguez, R.D. Smith, High-efficiency nanoscale liquid chromatography coupled on-line with mass spectrometry through nanoelectrospray Ionization, Anal. Chem. (2002) 4235–4249.
[21] T.S. Romine, L.A. Norbeck, M.S. Lipton, Identification of mobile elements and pseudogenes in the Shewanella oneidensis MR-1 genome, Appl. Environ. Microbiol. 74 (2008) 3257–3265.
[22] K. Jeong, S. Kim, N. Bandeira, False discovery rates in spectral identification, BMC Bioinf. 13 (2012) S2.
[23] Y. Shen, N. Tolić, S.O. Hixson, L. Paša-Tolić, W.-J. Qian, R.J. Adkins, R.D. Smith, Proteome-wide identification of proteins and their modifications with decreased ambiguities and improved false discovery rates, Anal. Chem. 80 (2008) 1871–1882.
[24] Y. Shen, N. Tolić, S.O. Hixson, G.A. Anderson, R.D. Smith, De novo sequencing of the unique sequence tags for discovery of posttranslational modification of proteins, Anal. Chem. 80 (2008) 7742–7754.
[25] N. Itoh, N. Kinoshita, M. Nomura, Fast protein separation by reversed-phase high-performance liquid chromatography on octadecylsilyl-bonded nonporous silica gel. II. Improvement in recovery of hydrophobic proteins, Anal. Biochem. 199 (1991) 7–10.
[26] E.M. Renkin, Filtration, diffusion, and molecular sieving through porous cellulose membranes, J. Gen. Physiol. 38 (1954) 225–243.
[27] M.W. Nice, N. Cooke, M.J. O'hare, Comparison of short and ultrashort-chain alkylsilane-bonded silicas for the high-performance liquid chromatography of proteins by hydrophobic interaction methods, J. Chromatogr. 218 (1981) 569–580.
[28] M.W. O'Hare, N.H.C. Nice, B.G. Archer, Factors influencing chromatography of proteins on short alkylsilane-bonded large pore-size silicas, Anal. Biochem. 126 (1982) 17–28.
[29] A.E. Gilar, U.D. Kele, J.C. Gebler, Implications of column peak capacity on the separation of complex peptide mixtures in single- and two-dimensional high-performance liquid chromatography, J. Chromatogr. A 1061 (2004) 183–192.
[30] J.B. Shaw, W. Li, D.D. Holden, Y. Zhang, J. Griep-Raming, R.T. Fellers, B.P. Early, P.M. Thomas, N.L. Kelleher, J.S. Brodbelt, J. Am. Chem. Soc. 135 (2013) 12646–12651.