Published in final edited form as:

Anal Chem. 2015 ; 87(10): 5363–5371. doi:10.1021/acs.analchem.5b00657.

Three Dimensional Liquid Chromatography Coupling IEC/HIC/RPC for Effective Protein Separation in Top-Down Proteomics

Santosh G. Valeja^{#a}, Lichen Xiu^{#b}, Zachery R. Gregorich^{a,c}, Huseyin Guner^{a,d}, Song Jin^b, and Ying Ge^{a,b,c,d,*}

^a Department of Cell and Regenerative Biology, University of Wisconsin-Madison, Madison, Wisconsin, USA

^b Department of Chemistry, University of Wisconsin-Madison, Madison, Wisconsin, USA

^c Molecular and Cellular Pharmacology Training Program, University of Wisconsin-Madison, Wisconsin, USA

^d Human Proteomics Program, School of Medicine and Public Health, University of Wisconsin-Madison, Madison, Wisconsin, USA

[#] These authors contributed equally to this work.

Abstract

To address the complexity of proteome in mass spectrometry (MS)-based top-down proteomics, multi-dimensional liquid chromatography (MDLC) strategies that can effectively separate proteins with high resolution and automation are highly desirable. Although various MDLC methods that can effectively separate peptides from protein digests exist, very few MDLC strategies, primarily consisting of 2DLC, are available for intact protein separation, which is insufficient to address the proteome complexity. We recently demonstrated that hydrophobic interaction chromatography (HIC) utilizing a MS-compatible salt can provide high resolution separation of intact proteins for top-down proteomics. Herein, we have developed a novel 3DLC strategy by coupling HIC with ion exchange chromatography (IEC) and reverse phase chromatography (RPC) for intact protein separation. We demonstrated that a 3D (IECHIC-RPC) approach greatly outperformed the conventional 2D IEC-RPC approach. For the same IEC fraction (out of 35 fractions) from a crude HEK 293 cell lysate, a total of 640 proteins were identified in the 3D approach (corresponding to 201 non-redundant proteins) as compared to 47 in the 2D approach, whereas simply prolonging the gradients in RPC in the 2D approach only led to minimal improvement in protein separation and identifications. Therefore this novel 3DLC method has great potential for effective separation of intact proteins to achieve deep proteome coverage in top-down proteomics.

*To whom correspondence may be addressed: Ying Ge, 1300 University Ave., SMI 130, Madison, Wisconsin 53706, USA. ge2@wisc.edu; Tel: 608-263-9212; Fax: 608-265-5512..

Supporting Information Available:

Additional information as noted in the text. This material is available free of charge via the Internet at <http://pubs.acs.org>.

INTRODUCTION

To better understand disease mechanisms and discover new biomarkers for clinical diagnostics, it is essential to perform deep proteome profiling, which includes the identification, characterization, and quantification of “proteoforms”¹ arising from genetic variations, alternative RNA splicing, and post-translational modifications.¹⁻⁸ The well-established bottom-up proteomics approach requires digestion of proteins into many peptides, which complicates the identification, quantification, and characterization of proteoforms (the ‘peptide to protein inference problem’).⁸ In contrast, top-down mass spectrometry (MS)⁹⁻¹¹-based proteomics analyzes intact proteins and has unique advantages for the comprehensive analysis of proteoforms.^{4-6,12-22} However, the extreme complexity of the proteome, which is comprised of thousands of proteins corresponding to millions of proteoforms, presents a significant challenge in top-down proteomics. Consequently, multi-dimensional separation strategies have been developed to decrease the complexity of the proteome prior to MS analysis.^{4,17,23-25} Among these strategies, multi-dimensional liquid chromatography (MDLC) strategies, which can be coupled to the mass spectrometer and are amenable to automation, are highly desired.^{17,25,26}

Following the pioneering work of multi-dimensional protein identification technology (MudPIT),²⁷ a plethora of MDLC methods have been developed to effectively separate peptides for bottom-up proteomics.^{25,26} In contrast, the separation of intact proteins remains challenging due to the diverse protein properties, MS-incompatibility of the buffers used for solubilizing proteins, and poor chromatographic resolution.¹⁷ Thus, very few MDLC approaches have been developed to separate intact proteins for use in top-down proteomic analyses. To the best of our knowledge, only two-dimensional (2D) LC strategies, usually by coupling either ion exchange chromatography (IEC), size exclusion chromatography (SEC), or more recently, hydrophilic interaction chromatography (HILIC), with reverse phase chromatography (RPC) have been employed for the separation of intact proteins.²⁸⁻³³ However, such 2DLC strategies are insufficient to address the proteome complexity and, thus, the use of additional dimensions of separation may hold promise for reducing the complexity of the proteome and increasing the depth of top-down proteomic analyses.

Hydrophobic interaction chromatography (HIC) is considered a high resolution chromatography technique for the separation of intact proteins under a non-denaturing mode,³⁴⁻³⁷ but the nonvolatile salts conventionally employed (e.g. ammonium sulfate) render HIC incompatible for direct MS analysis.³⁸⁻⁴⁰ Recently, we have identified ammonium tartrate $[(\text{NH}_4)_2\text{C}_4\text{H}_4\text{O}_6]$ as a new MS-compatible salt and, using HIC in combination with this salt, demonstrated high resolution protein separations comparable to that achieved with the commonly used ammonium sulfate salt.⁴¹ In this study, we have further developed a novel 3DLC strategy using IEC-HIC-RPC/MS by integrating this new MS-compatible HIC mode with the conventionally used MudPIT-like 2DLC (IEC-RPC/MS). Owing to the mutual orthogonality among these three chromatography modes, 3D IEC-HIC-RPC allowed for the separation of intact proteins with higher resolution and significantly enhanced protein identifications in comparison to the conventional 2D IEC-RPC/MS approach. From a single IEC fraction (out of 35 fractions) from a crude cell lysate, 640 total proteins (corresponding to 201 non-redundant proteins) were identified by this new

3DLC technique compared with 47 total proteins (corresponding to 47 non-redundant proteins) identified by the conventional 2D approach. To the best of our knowledge, this is the first time a 3DLC strategy has been used for top-down proteomics, which has high potential to achieve deep profiling of the complex proteome.

MATERIALS AND METHODS

Chemicals and reagents

All reagents were purchased from Sigma-Aldrich Inc. (St. Louis, MO) unless noted otherwise. HPLC grade water and acetonitrile were purchased from Fisher Scientific (Fair Lawn, NJ) and cell lysis buffer was purchased from Thermo Fisher Scientific (Rockford, IL). 0.5 mL centrifugal filters with 10 kDa molecular weight cut-off were purchased from Merck Millipore Ltd. (Bedford, MA).

Sample preparation

The following standard protein samples: Apr, aprotinin from bovine lung; Cyt, cytochrome C from equine heart; RiA, ribonuclease A from bovine pancreas; Myo, myoglobin from equine heart; RiB, ribonuclease B from bovine pancreas; ChA, α -chymotrypsinogen A from bovine pancreas; Chy, α -chymotrypsin from bovine pancreas; Oval, ovalbumin from chicken egg white; BSA, albumin from bovine serum; Con, conalbumin from chicken egg white; Thg, thyroglobulin from bovine thyroid; were used without further purification. For IEC and HIC, all standard protein samples were first prepared in 10 mg/mL with HPLC-grade water, and subsequently diluted to 0.1 - 1.5 mg/mL. The 4-protein mixture (BSA, RiB, RiA, and Chy) and 5-protein mixture (Oval, Thg, Myo, Chy and ChA) were prepared to assess the orthogonality between IEC and HIC.

Human embryonic kidney (HEK) 293 cells (~80 million) grown in-house were lysed in 450 μ L of cell lysis buffer containing protease and phosphatase inhibitor cocktails (Roche, Indianapolis, IN), as well as phenylmethanesulfonyl fluoride (PMSF) (100 mM), by briefly vortexing and then shaking for one hour at 4 °C. The resulting lysate was centrifuged at 4 °C for 30 min at 16,000 g. The supernatant was utilized for further chromatographic separations and the pellet was discarded.

Ion exchange chromatography (IEC)

IEC was performed on a Shimadzu HPLC system (Shimadzu Scientific Instruments Inc., Columbia, MD) equipped with a mixed-bed PolyCATWAX ATM column (200 mm $\times$ 4.6 mm i.d., 5 μ m, 1000 Å, PolyLC Inc., Columbia, MD). Mobile phase A (MPA) contained 10 mM ammonium tartrate and mobile phase B (MPB) contained either 0.5 M (optimized for standard proteins) or 1.0 M (for HEK 293 cell lysate sample) ammonium tartrate, respectively. All solutions were adjusted to pH 7.0 with 10% ammonium hydroxide (NH₄OH) solution. A 30-min linear gradient (from 100% MPA to 100% MPB) was utilized to elute proteins followed by 5-min of isocratic flow at 100% MPB to ensure elution, both at a constant flow rate of 1 mL/min. All samples were diluted with equal volume of MPA (i.e., 1:1 v/v) to avoid injection viscosity differences. The injection volume was 50 μ L for standard proteins and mixtures thereof, and 100 μ L for HEK 293 cell lysate samples.

Baseline subtraction was performed for all IEC chromatograms. The collected IEC fractions from HEK 293 cell lysate samples were quickly concentrated with 10 kDa ultracentrifugal filters before separation in the next dimension.

Hydrophobic interaction chromatography (HIC)

HIC was conducted on a Shimadzu HPLC system (Shimadzu Scientific Instruments Inc., Columbia, MD) equipped with a PolyPROPYL A™ column (100 mm × 4.6 mm i.d., 3 μm, 1500 Å, PolyLC Inc., Columbia, MD), similar to what was described previously.⁴¹ Here, 1.8 M and 20 mM ammonium tartrate solutions, adjusted with 10% ammonium hydroxide (NH₄OH) solution to pH 7.0, were utilized as MPA and MPB, respectively, for HIC separation. A 30-min linear gradient (from 100% MPA to 100% MPB) was used to elute proteins followed by isocratic flow at 100% MPB for 5 min to ensure elution, both at a flow rate of 1 mL/min. For standard protein samples, the gradient profile was slightly optimized to achieve better separation: two isocratic regions from 12 min to 14.5 min (at the proportion of 48.3% MPB) and 15 min to 19 min (63.3% MPB) were interjected. All standard protein samples were diluted 1:1 (v/v) with MPA to avoid injection viscosity differences, and the sample injection volume was 50 μL. For IEC fractions, approximately 40 μL solution was obtained after centrifugal filtration, and the HIC MPA was added to make the total volume to 105 μL for a 100 μL HIC injection. Baseline subtraction was performed for all HIC chromatograms. Other chromatographic conditions are given in the figure legends. The collected HIC fractions from HEK 293 cell lysate samples were concentrated or desalted with 10 kDa ultracentrifugal filters before separation in the next dimension.

Reverse phase chromatography (RPC)

RPC was carried out on a Thermo EASY nano-LC 1000 system (Thermo Fisher Scientific) equipped with a PicoFrit™ PLRP-S column (100 mm x 100 μm i.d., 5 μm, 1000 Å, New Objective, Inc., Woburn, MA) as described previously.⁴¹ Buffer A consisted of water with 0.25% formic acid and Buffer B consisted of acetonitrile with 0.25% formic acid. The nano-LC was operated at a constant flow rate of 500 nL/min and 3 μL of sample was injected with an autosampler post-equilibration of the capillary column. For the separation of complex HEK 293cell lysate proteins from different IEC and HIC fractions, an 80-min optimized RPC gradient was utilized consisting of the following concentrations of buffer B: 5% for 15 min, 25% at 25 min, 60% at 70 min, 95% at 75 min, and then back to 5% at 80 min. The collected IEC and HIC fractions from HEK 293 cell lysate samples were concentrated or desalted with 10 kDa ultracentrifugal filters before injection onto the RPC column.

Top-down MS analysis

Samples were electrosprayed with a “nanoflex” ionization source into a Q Exactive™ benchtop Orbitrap mass spectrometer (ThermoFisher Scientific, Bremen, Germany) for online data-dependent nano-LC/MS/MS experiments, similar to what was described previously.⁴¹ The spray voltage, heated metal capillary temperature, and s-lens voltage were experimentally optimized to 2,500 V, 300 °C, and 50 V, respectively, for improved ion desolvation and transmission.^{42,43} LC/MS and LC/MS/MS data was acquired with 6 micro scans at a high mass resolving power of 140,000 (theoretical maximum resolving power

setting on QE: m/z $m_{50\%} = 140,000$ at m/z 200, in which $m_{50\%}$ denotes mass spectral peak full width at half-maximum peak height) and 8 micro scans at a resolving power of 70,000, respectively, with automatic gain control set to 5 E5 ions. A 10 V offset in the source was used for all the experiments. In top 2 data-dependent MS/MS scans, the intact protein ions were injected into the collision cell for higher-energy collision dissociation (HCD)^{44,45} at a previously optimized setting of 25 V. The dynamic exclusion is 10 sec. Charge exclusion was applied for unassigned and lower charge states from 1 to 8 for data-dependent HCD. Data was collected with Xcalibur™ 2.2 software (ThermoFisher Scientific) and the total RPC/MS data acquisition period was adjusted to fit the RPC gradient (80 min, 3 h, or 7 h). The MASH-suite software⁴⁶ was employed to calculate the isotopic distributions on the basis of observed molecular weights and the averaging model for observed intact proteins.^{46,47}

Protein identification

MS-Deconv, a combinatorial algorithm⁴⁸ was first utilized for isotopic distribution deconvolution and charge state assignment for all the observed ions in the raw LC/MS/MS data as well as to generate MSAlign files.⁴⁸ These MSAlign files, which contain the monoisotopic mass, intensity, and charge of all the detected ions, were then searched with the NCBI Human database (Uniprot-Swissprot database, released January, 2013, containing 20,232 protein sequences) with alignment-based MS-Align+ algorithm for top-down intact protein identification based on protein-spectrum matches.⁴⁹ A fragment mass tolerance of 15 ppm was used for the assignment of *b* and *y* ions. Protein identification results with statistically significant lower P and E value (<0.01) and a satisfactorily higher fragment number (at least 10 fragments assigned) were manually validated. Furthermore, a reversed database was separately employed to search for false discovery rates (FDRs) with same MSAlign+ algorithm. Though cut-off for FDR was set at 0.5%, nearly all of the identified proteins had significantly lower FDRs (0.001%-0.1%). Furthermore, all identified proteins were manually validated.

RESULTS AND DISCUSSION

Evaluation of the orthogonality between IEC and HIC

In this work, we incorporated HIC, a new MS-compatible protein separation chromatography mode,⁴¹ into a MDLC separation platform to address the increasing demand for improved intact protein separation in top-down proteomics. Two popular chromatographic methods, IEC and RPC, were chosen to couple with HIC. IEC separates molecules based on ionic interactions and can be subdivided into cation exchange chromatography and anion exchange chromatography. Here, we selected a mixed-bed IEC column containing equal amounts of weak cation and weak anion materials for intact protein separation since it results in improved separation relative to a single mode IEC column.⁵⁰ We first sought to determine whether these three chromatographic methods (HIC, IEC, and RPC) are orthogonal. IEC and RPC have been reported previously to be orthogonal.^{51,52} Additionally, we have recently shown that HIC and RPC have different selectivity for intact protein separation despite the fact that both methods separate proteins based on

hydrophobicity.⁴¹ Therefore, here we focused on demonstrating the orthogonality between IEC and HIC.

We utilized standard proteins to demonstrate the orthogonality between IEC and HIC. UV chromatograms from individual injections of eleven standard proteins are overlaid in Figure 1a, b, revealing satisfactory peak shapes for the majority of the intact proteins for both IEC and HIC. The standard proteins were selected to cover a wide range of molecular weights (6 – 330 kDa), pI (4.94 – 9.32), and polarity. Note that the order of elution and the retention times of the proteins are drastically different in IEC and HIC, suggesting that the selectivity of these two chromatographic methods is different. Furthermore, we selected two different sets of protein mixtures, namely “4-mix” (a mixture of four standard proteins containing BSA, ChA, RiA, and RiB) and “5-mix” (a mixture of five standard proteins with ChA, Chy, Myo, Oval, and Thg) to perform the separation on both IEC and HIC, respectively. According to retention time in reference to the individual protein injections (Figure 1a, b), all peaks were identified in the chromatograms of 4-mix and 5-mix samples. Notably, 4-mix was well-resolved by IEC (Figure 1c), whereas HIC was unable to separate the same 4-protein mixture (Figure 1d), with RiA and RiB as well as BSA and Chy co-eluting. On the other hand, 5-mix could not be separated using IEC (Figure 1e) but was well-resolved by HIC (Figure 1f). These results further demonstrate the orthogonality between HIC and IEC.

Rationale for the order of IEC-HIC-RPC in the 3DLC separation strategy

With the mutual orthogonality among IEC, HIC, and RPC demonstrated, we next determined the rational order of the 3D separation to be IEC-HIC-RPC (Scheme 1). RPC uses MS-friendly organic solvents and can be directly coupled to the mass spectrometer for online LC/MS analysis so it was chosen as the last dimension. Since both IEC and HIC methods utilize salt gradients (and here we use ammonium tartrate, a new MS-compatible salt for HIC we recently reported⁴¹) for protein separation, it is important to note the differences in the gradients employed in these methods. In IEC, proteins are eluted with an increasing concentration of salt in the mobile phase whereas in HIC, the starting salt concentration is extremely high and declines during the LC run. Thus, we decided to utilize IEC as the first dimension of separation followed by HIC in the second dimension. Such IEC-HIC order takes advantage of the starting low salt concentrations in IEC fractions (10 mM) relative to the starting high salt concentration of HIC mobile phase (conventionally, 1.5-1.8 M), thereby avoiding the need to desalt the IEC fractions; whereas if HIC were to precede IEC, the HIC fractions would need to be desalted to be compatible with the low starting salt concentration utilized in IEC (10 mM).

Next, we evaluated the impact of the additional HIC dimension in the separation of proteins after IEC fractionation (Scheme 1). We first collected a total of 35 IEC fractions from the HEK 293 cell lysate. After a quick concentration step, each IEC fraction was further separated by HIC. The UV chromatograms of HIC for all 35 IEC fractions from HEK 293 cell lysate samples were overlaid (Figure 2a, b). Moreover, a heat map indicating the intensity of the respective UV chromatograms of HIC from IEC fractions revealed additional fractionation of IEC-separated proteins by HIC (Figure 2c). As shown in Figure

2d, e, IEC fraction #3 could be further separated into a range of peaks using HIC. Thus, HIC not only works orthogonally to IEC, but also leads to additional protein separation after IEC.

Comparison of intact protein separation between 3D and 2D strategies

Next, we compared the separation capability of the new 3DLC (IEC-HIC-RPC) strategy with a conventional MudPIT-like 2DLC (IEC-RPC) platform (Scheme 1). We employ identical IEC and RPC/MS settings, thus the addition of the HIC step is the sole variable between the 2DLC and 3DLC approaches. We selected IEC fraction #3, which has relatively abundant and complex protein components (Figure 2d), for further analysis by RPC/MS or HIC-RPC/MS for a direct comparison of the 2DLC and 3DLC approaches. The total sample loading amounts were the same in 2D and 3D experiments.

In the 2D approach, the total ion current (TIC) chromatogram for IEC fraction #3 (Figure 3a) revealed a variety of intact proteins eluting out between ~25 to 60 min. Figure 3b, c show representative high resolution mass spectra of intact proteins with accurate mass measurements. Note that some intact proteins co-eluted in RPC TIC chromatograms (such as 23.2 kDa and 16.8 kDa; 19.5 kDa and 14.4 kDa), indicating that the complexity of IEC fractions remained after a 2DLC separation. Such co-elution problem, commonly seen in IEC-RPC approach, will likely compromise protein identification by top-down MS especially for low abundant intact proteins, necessitating additional dimension(s) of chromatographic separation.

In comparison, in the 3D approach, the IEC fraction #3 was first separated by HIC into 35 fractions. We then used a typical HIC fraction #20 collected from IEC fraction #3 for online RPC/MS analysis as an example (Figure 4). The RPC TIC chromatogram in the 3D approach (Figure 4a) revealed peaks with much better separation than those seen in the 2D approach, especially in the retention time ranging from 40 to 55 min. Figure 4b, c show representative MS spectra of intact proteins detected in the HIC fraction #20. The intact protein with a mass of 23.2 kDa did not co-elute with another protein after 3D separation as it did following 2D separation (Figure 4b), suggesting that the co-elution problem may be comparatively reduced by the addition of a third dimension of chromatographic separation. Moreover, some of the proteins that were detected by top-down MS after 3DLC (e.g. 13.4 kDa, 14.3 kDa and 11.8 kDa) were not observed following 2DLC separation.

Comparison of protein identifications between 3D and 2D by top-down tandem MS (MS/MS)

Since accurate mass measurement of intact proteins alone does not allow for unequivocal protein identification, we employed MS/MS with higher-energy collision dissociation (HCD) to fragment and subsequently identify the proteins in the fractions generated using the 2DLC and 3DLC approaches. Similar to collisionally activated dissociation, HCD predominantly produces *b* and *y* ions by cleaving the amide bonds in the peptide backbone. Overall, a total of 640 intact proteins were identified by top-down RPC/MS/MS analysis of the 35 HIC fractions generated from IEC fraction #3, which correspond to 201 non-redundant proteins since some intact proteins were identified in multiple HIC fractions (see Table S1 in the Supporting Information). As an example, 31 non-redundant proteins were

identified in a HIC fraction #20 alone (see Table S2 in the Supporting Information), and Figure 5 shows the single scan HCD mass spectra for three of the proteins identified in that fraction (the protein masses shown in Figure 4b, c, 13.4 kDa, 14.3 kDa and 11.8 kDa). These proteins were identified as human transcription elongation factor B polypeptide 2, histidine triad nucleotide-binding protein 2 mitochondrial, and u6 snRNA-associated Sm-like protein LSm7 protein based on a total of 44, 63, and 52 unique *b/y* fragment ions, respectively. It is estimated that the limits of detection for intact proteins (MW < 30 kDa) is 0.1–0.5 µg/µL using the ESI-Q Exactive MS with the parameters described in the experimental section.

In contrast, a total of 47 intact proteins (all of them non-redundant protein IDs) were identified in IEC fraction #3 of the HEK 293 cell lysate by RPC/MS/MS analysis (without HIC separation) using an 80-min RPC gradient (see Table S3 in the Supporting Information). Figure S1 shows single scan HCD mass spectra for two of the proteins from IEC fraction #3 of the HEK 293 cell lysate (the protein masses shown in Figure 3b, c). A database search using the MSAlign+ algorithm identified the 14.9 kDa protein with the RPC retention time of ~37 min as human-profilin 1 protein (a total of 94 *b* and *y* ions observed, Figure S1-top) and the 19.5 kDa protein with the ~50 min RPC retention time as human adenine phosphoribosyltransferase protein (a total of 66 *b* and *y* ions observed, Figure S1-bottom).

By further examining the 2DLC and 3DLC results together, we tabulate in Tables S4 a list of the 39 intact proteins that were identified by both the 2DLC and 3DLC approaches. Besides the 39 proteins that were identified in both the 2D and 3D approaches, 8 and 162 proteins were uniquely identified in the 2DLC and 3DLC approaches, respectively (Figure 6a). Note that the total sample loading amounts were the same in 2D and 3D experiments. These results highlight the significant enhancement in intact protein identification achieved using the IEC-HIC-RPC/MS approach in comparison to the IEC-RPC/MS.

Limited improvements in protein identifications using longer RPC gradients in the 2D IEC-RPC strategy

We also investigated whether extended RPC gradient profiles may yield enhanced separation and increased protein identification in 2DLC top-down MS strategy. Aliquots of the same IEC fraction #3 used for the analyses described above were separated using longer RPC gradients of 3 h and 7 h, respectively. The intact proteins identified in these longer RPC gradient profiles are listed in Table S5 and Table S6, respectively. A total of 63 and 67 non-redundant intact proteins were identified in the 3-h and 7-h RPC gradients in comparison to 47 total protein identifications in the 80-min gradient (Figure 6b and Table S7). Of the proteins identified using the different RPC gradients, 33 proteins (70%) were commonly identified by RPC/MS/MS. Therefore, very limited improvement is observed in the protein identifications with longer RPC gradients.

The potential of IEC-HIC-RPC 3DLC strategy for deep top-down proteomics profiling

Unlike peptides, proteins have a much more diverse range of physiochemical properties (i.e. charge, hydrophobicity, and size).¹⁷ Here, utilizing a 3DLC approach employing three mutually orthogonal chromatography modes, we have shown that MDLC has great potential

for reducing the complexity of the proteome to aid in top-down proteomics analyses.² As demonstrated above, our new 3DLC strategy coupling IEC-HIC-RPC provides considerably enhanced chromatographic separation and yields a significantly higher number of unique intact protein identifications than the conventional 2D IEC-RPC approach. For a single IEC fraction (out of 35 fractions), 640 total protein were identified corresponding to 201 non-redundant proteins by 3DLC method, which is in sharp contrast to the 47, 63, and 67 total protein identifications (corresponding to the same number of non-redundant proteins in the 2DLC strategy with 80-min, 3-h and 7-h RPC gradients). These results clearly show the promise of this novel 3DLC approach for top-down proteomics. Nevertheless, further improvements are still needed to fully realize its potential.

First, the proteins identified in this study have low MWs (<30 kDa), even though larger proteins (30-250 kDa) were clearly present in the HIC fractions as shown in the sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) analysis of selected HIC fractions for IEC fraction #3 of the HEK 293 cell lysate (Figure S3). Presumably the lack of detection of larger intact proteins was due to the exponential decay in S/N with increasing MW and/or the rapid decay in the time-domain signal for higher kinetic energies of ions in the Orbitrap mass spectrometer.^{53,54} Therefore, further advancements in mass spectrometer instrumentation could improve the detection of high MW proteins with higher sensitivity.^{2,55-58} Additionally, the separation of high MW proteins from low MW proteins using size-exclusion chromatography⁵⁹ or gel-eluted liquid fraction entrapment electrophoresis (GELFrEE)^{4,60} could also improve the detection of large proteins.

Second, the loss of protein sample during offline fractionation could prevent the detection of low abundance proteins, which might be lost during the concentration step. The current IEC and HIC separations were performed on an offline HPLC with a concentration step required for each fractionation. Paša-Tolić and co-workers demonstrated previously that an online 2DLC system offers significant improvement in sensitivity, with several orders of magnitude reduction in sample requirement as well as a reduction in the overall analysis time compared with the offline 2DLC approach.³² We envision that a truly on-line MDLC will offer significant improvement in sensitivity.

Third, for any MDLC method, there is typically a tradeoff between throughput and depth of information. Conceivably, it would be time consuming to run all 1225 (35×35) fractions. So it could be beneficial to reduce the number of fractions to a more manageable level for improved throughput (e.g. select 15×15 most abundant fractions in each dimension or combine similar fractions based on the UV absorption heat map). On the other hand, this could be at the cost of decreased chromatographic resolution or increased complexity in each fraction that most likely will result in fewer protein identifications or reduced proteome coverage. Likely, further improvements in chromatography and MS systems in terms of speed, sensitivity, and resolution will make the MDLC approach feasible for routine proteomic analyses.

CONCLUSION

In summary, we have developed an effective 3D (IEC-HIC-RPC) chromatography strategy for protein separation for the first time that is highly promising for overcoming the separation bottleneck that has hampered progress in top-down proteomics. We demonstrated the separation power of this 3DLC strategy by coupling HIC with IEC and RPC for top-down proteomics of complex cell lysate and compared it with the conventional 2D (IEC-RPC) MudPIT-like method. Overall, 201 non-redundant intact proteins were identified by this new 3D IEC-HIC-RPC approach relative to 47 proteins identified by the 2D IEC-RPC method for a single IEC fraction. With the significant improvement in chromatographic separation and far greater number of protein identifications than the conventional 2DLC methods, this novel 3DLC strategy presents great potential for deep proteome profiling in top-down proteomics.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENT

This work is supported by NIH R01HL096971 and R01HL109810 (to Y.G.) and R21EB013847 (to S.J. and Y.G.). We thank Dr. Serife Ayaz-Guner for assistance with growing the HEK 293 cells and Dr. Xiaowen Liu for helpful discussions on the MS data analysis with the MSAlign+ algorithm. We are thankful to Dr. Andrew Alpert of PolyLC, Inc. for providing the PolyCATWAX A™ and PolyPROPYL A™ columns and for helpful discussion. We also would like to thank Ms. Amanda Berg and Dr. Gary Valaskovic of New Objective Inc., for providing the PicoFrit™ PLRP-S column.

REFERENCES

1. Smith LM, Kelleher NL. Nat. Methods. 2013; 10:186–187. [PubMed: 23443629]
2. Gregorich ZR, Ge Y. Proteomics. 2014; 14:1195–1210. [PubMed: 24723472]
3. Siuti N, Kelleher NL. Nat. Methods. 2007; 4:817–821. [PubMed: 17901871]
4. Tran JC, Zamborg L, Ahlf DR, Lee JE, Catherman AD, Durbin KR, Tipton JD, Vellaichamy A, Kellie JF, Li M, Wu C, Sweet SM, Early BP, Siuti N, LeDuc RD, Compton PD, Thomas PM, Kelleher NL. Nature. 2011; 480:254–258. [PubMed: 22037311]
5. Zhang H, Ge Y. Circ.: Cardiovasc. Genet. 2011; 4:711. [PubMed: 22187450]
6. Ryan CM, Souda P, Bassilian S, Ujwal R, Zhang J, Abramson J, Ping P, Durazo A, Bowie JU, Hasan SS, Baniulis D, Cramer WA, Faull KF, Whitelegge JP. Mol. Cell. Proteomics. 2010; 9:791–803. [PubMed: 20093275]
7. Tipton JD, Tran JC, Catherman AD, Ahlf DR, Durbin KR, Lee JE, Kellie JF, Kelleher NL, Hendrickson CL, Marshall AG. Anal. Chem. 2012; 84:2111–2117. [PubMed: 22356091]
8. Kelleher NL, Thomas PM, Ntai I, Compton PD, LeDuc RD. Expert Rev. Proteomic. 2014; 11:649–651.
9. Kelleher NL, Lin HY, Valaskovic GA, Aaserud DJ, Fridriksson EK, McLafferty FW. J. Am. Chem. Soc. 1999; 121:806–812.
10. Ge Y, Lawhorn BG, ElNaggar M, Strauss E, Park JH, Begley TP, McLafferty FW. J. Am. Chem. Soc. 2002; 124:672–678. [PubMed: 11804498]
11. Han X, Jin M, Breuker K, McLafferty FW. Science. 2006; 314:109–112. [PubMed: 17023655]
12. Ansong C, Wu S, Meng D, Liu XW, Brewer HM, Kaiser BLD, Nakayasu ES, Cort JR, Pevzner P, Smith RD, Heffron F, Adkins JN, Pasa-Tolic L. Proc. Natl. Acad. Sci. U.S.A. 2013; 110:10153–10158. [PubMed: 23720318]

13. Chamot-Rooke J, Mikaty G, Malosse C, Soyer M, Dumont A, Gault J, Imhaus AF, Martin P, Trellet M, Clary G, Chafey P, Camoin L, Nilges M, Nassif X, Dumenil G. *Science*. 2011; 331:778–782. [PubMed: 21311024]
14. Dong XT, Sumandea CA, Chen YC, Garcia-Cazarin ML, Zhang J, Balke CW, Sumandea MP, Ge YJ. *Biol. Chem.* 2012; 287:848–857.
15. Mazur MT, Cardasis HL, Spellman DS, Liaw A, Yates NA, Hendrickson RC. *Proc. Natl. Acad. Sci. USA*. 2010; 107:7728–7733. [PubMed: 20388904]
16. Zhang J, Guy MJ, Norman HS, Chen YC, Xu QG, Dong XT, Guner H, Wang SJ, Kohmoto T, Young KH, Moss RL, Ge YJ. *Proteome Res.* 2011; 10:4054–4065.
17. Doucette AA, Tran JC, Wall MJ, Fitzsimmons S. *Expert Rev. Proteomics*. 2011; 8:787–800. [PubMed: 22087661]
18. Ge Y, Rybakova IN, Xu Q, Moss RL. *Proc. Natl. Acad. Sci. USA*. 2009; 106:12658–12663. [PubMed: 19541641]
19. Peng Y, Chen X, Zhang H, Xu Q, Hacker TA, Ge YJ. *Proteome Res.* 2013; 12:187–198.
20. Peng Y, Gregorich ZR, Valeja SG, Zhang H, Cai W, Chen Y-C, Guner H, Chen AJ, Schwahn DJ, Hacker TA, Liu X, Ge Y. *Mol. Cell. Proteomics*. 2014; 13:2752–2764. [PubMed: 24969035]
21. Thangaraj B, Ryan CM, Souda P, Krause K, Faull KF, Weber APM, Fromme P, Whitelegge JP. *Proteomics*. 2010; 10:3644–3656. [PubMed: 20845333]
22. Dang X, Scotcher J, Wu S, Chu RK, Tolic N, Ntai I, Thomas PM, Fellers RT, Early BP, Zheng Y, Durbin KR, LeDuc RD, Wolff JJ, Thompson CJ, Pan J, Han J, Shaw JB, Salisbury JP, Easterling M, Borchers CH, Brodbelt JS, Agar JN, Pasa-Tolic L, Kelleher NL, Young NL. *Proteomics*. 2014; 14:1130–1140. [PubMed: 24644084]
23. Wang H, Hanash S. *Mass Spectrom. Rev.* 2005; 24:413–426. [PubMed: 15389852]
24. Sheng S, Chen D, Van Eyk JE. *Mol. Cell. Proteomics*. 2006; 5:26–34. [PubMed: 16188874]
25. Zhang Z, Wu S, Stenoien DL, Pasa-Tolic L. *Annu. Rev. Anal. Chem. (Palo Alto, Calif.)*. 2014; 7:427–454. [PubMed: 25014346]
26. Zhang X, Fang A, Riley CP, Wang M, Regnier FE, Buck C. *Anal. Chim. Acta*. 2010; 664:101–113. [PubMed: 20363391]
27. Washburn MP, Wolters D, Yates JR 3rd. *Nat. Biotechnol.* 2001; 19:242–247. [PubMed: 11231557]
28. Pesavento JJ, Bullock CR, LeDuc RD, Mizzen CA, Kelleher NL. *J. Biol. Chem.* 2008; 283:14927–14937. [PubMed: 18381279]
29. Sharma S, Simpson DC, Tolic N, Jaitly N, Mayampurath AM, Smith RD, Pasa-Tolic L. *J. of Proteome Res.* 2007; 6:602–610. [PubMed: 17269717]
30. Simpson DC, Ahn S, Pasa-Tolic L, Bogdanov B, Mottaz HM, Vilkov AN, Anderson GA, Lipton MS, Smith RD. *Electrophoresis*. 2006; 27:2722–2733. [PubMed: 16732621]
31. Stobaugh JT, Fague KM, Jorgenson JW. *J. Proteome Res.* 2013; 12:626–636. [PubMed: 23205614]
32. Tian Z, Zhao R, Tolic N, Moore RJ, Stenoien DL, Robinson EW, Smith RD, Pasa-Tolic L. *Proteomics*. 2010; 10:3610–3620. [PubMed: 20879039]
33. Young NL, DiMaggio PA, Plazas-Mayorca MD, Baliban RC, Floudas CA, Garcia BA. *Mol. Cell. Proteomics*. 2009; 8:2266–2284. [PubMed: 19654425]
34. Determann H, Lampert K. *J. Chromatogr.* 1972; 69:123–128.
35. Hjerten S. *J. Chromatogr.* 1973; 87:325–331.
36. Hjerten S, Rosengren J, Pahlman S. *J. Chromatogr.* 1974; 101:281–288.
37. Regnier FE. *Methods Enzymol.* 1983; 91:137–190. [PubMed: 6343757]
38. Queiroz JA, Tomaz CT, Cabral JM. *S. J. Biotechnol.* 2001; 87:143–159.
39. Washabaugh MW, Collins KD. *J. Biol. Chem.* 1986; 261:2477–2485.
40. Collins KD. *Methods*. 2004; 34:300–311. [PubMed: 15325648]
41. Xiu L, Valeja SG, Alpert AJ, Jin S, Ge Y. *Anal. Chem.* 2014; 86:7899–7906. [PubMed: 24968279]
42. Valeja SG, Kaiser NK, Xian F, Hendrickson CL, Rouse JC, Marshall AG. *Anal. Chem.* 2011; 83:8391–8395. [PubMed: 22011246]
43. Flick TG, Cassou CA, Chang TM, Williams ER. *Anal. Chem.* 2012; 84:7511–7517. [PubMed: 22881839]

44. Olsen JV, Macek B, Lange O, Makarov A, Horning S, Mann M. *Nat. Methods*. 2007; 4:709–712. [PubMed: 17721543]
45. Nagaraj N, D'Souza RC, Cox J, Olsen JV, Mann M. *J. Proteome Res.* 2010; 9:6786–6794. [PubMed: 20873877]
46. Guner H, Close PL, Cai W, Zhang H, Peng Y, Gregorich ZR, Ge Y. *J. Am. Soc. Mass Spectrom.* 2014; 25:464–470. [PubMed: 24385400]
47. Senko MW, Beu SC, McLafferty FW. *J. Am. Soc. Mass Spectrom.* 1995; 6:52–56. [PubMed: 24222060]
48. Liu X, Inbar Y, Dorrestein PC, Wynne C, Edwards N, Souda P, Whitelegge JP, Bafna V, Pevzner PA. *Mol. Cell. Proteomics.* 2010; 9:2772–2782. [PubMed: 20855543]
49. Liu X, Sirotkin Y, Shen Y, Anderson G, Tsai YS, Ting YS, Goodlett DR, Smith RD, Bafna V, Pevzner PA. *Mol. Cell. Proteomics.* 2012; 11:M111.008524.
50. Zhang L, Yao L, Zhang Y, Xue T, Dai G, Chen K, Hu X, Xu LX. *J. Chromatogr. B.* 2012; 905:96–104.
51. Pepaj M, Wilson SR, Novotna K, Lundanes E, Greibrokk T. *J. Chromatogr. A.* 2006; 1120:132–141. [PubMed: 16516903]
52. Farine S, Villard C, Moulin A, Marchis Mouren G, Puigserver A. *Int. J. Biol. Macromolec.* 1997; 21:109–114.
53. Compton PD, Zamborg L, Thomas PM, Kelleher NL. *Anal. Chem.* 2011; 83:6868–6874. [PubMed: 21744800]
54. Makarov A, Denisov E. *J. Am. Soc. Mass Spectrom.* 2009; 20:1486–1495. [PubMed: 19427230]
55. Kostyukevich YI, Vladimirov GN, Nikolaev EN. *J. Am. Soc. Mass Spectrom.* 2012; 23:2198–2207. [PubMed: 22993044]
56. Rose RJ, Damoc E, Denisov E, Makarov A, Heck AJR. *Nat. Methods.* 2012; 9:1084–1086. [PubMed: 23064518]
57. Michalski A, Damoc E, Lange O, Denisov E, Nolting D, Mueller M, Viner R, Schwartz J, Remes P, Belford M, Dunyach J-J, Cox J, Horning S, Mann M, Makarov A. *Mol. Cell. Proteomics.* 2012; 11:O111.013698. [PubMed: 22159718]
58. Li H, Wolff JJ, Van Orden SL, Loo JA. *Anal. Chem.* 2014; 86:317–320. [PubMed: 24313806]
59. Chen X, Ge Y. *Proteomics.* 2013; 13:2563–2566. [PubMed: 23794208]
60. Tran JC, Doucette AA. *Anal. Chem.* 2008; 80:1568–1573. [PubMed: 18229945]

Figure 1. The orthogonality between IEC and HIC

(a, b) overlay of the UV chromatograms obtained for individual standard proteins using IEC and HIC, respectively. (c-f) comparison of IEC and HIC separation of standard protein mixtures, 4-mix (BSA, Chy, RiA, RiB) in (c, d), and 5-mix (ChA, Chy, Myo, Oval, Thg) in (e, f), respectively. The UV detector was set to 280 nm for both IEC and HIC. Apr, aprotinin; BSA, bovine serum albumin; ChA, α -chymotrypsinogen A; Chy, α -chymotrypsin; Con, conalbumin; Cyt, cytochrome C; Myo, myoglobin; Oval, ovalbumin; RiA, ribonuclease A; RiB, ribonuclease B; Thg, thyroglobulin.

Scheme 1. Schematic work flow of 3DLC (IEC-HIC-RPC) in comparison with 2DLC (IEC-RPC) Cell lysate samples were fractionated with IEC in the first dimension (1D), and then the same IEC fraction was aliquoted and the aliquots (with the same loading amount) were used for both the 2DLC and 3DLC experiments to allow for a direct comparison between these approaches. In the 2D strategy, several RPC gradient profiles were utilized to investigate the impact of RPC gradient length on top-down protein identification by RPC/MS/MS. In the 3D approach, HIC was employed as a second dimension of separation (following IEC) prior to top-down RPC/MS/MS analysis.

Figure 2. Evaluation of HIC as the second dimension separation strategy following IEC fractionation of a crude cell lysate

(a) Overlay of UV chromatograms of 35 equal-volume IEC fractionations from the HEK 293 cell lysate. (b) Zoom-in UV chromatograms of IEC fractions. (c) Heat map of IEC-HIC separation of the HEK 293 cell lysate. (d) A representative IEC chromatogram of the HEK 293 cell lysate. (e) A representative HIC chromatogram of IEC fraction #3 of the HEK 293 cell lysate.

Figure 3. Online RPC/MS analysis of intact proteins following IEC fractionation (2DLC approach)

(a) A representative RPC/MS TIC from IEC fraction #3 from the HEK 293 cell lysate. (b) Representative mass spectra of the corresponding RPC peaks. (c) Zoom-in mass spectra with unit mass isotopic in chromatographic time-scale. A n 80 min RPC gradient were used.

Figure 4. Online RPC/MS analysis of intact proteins following IEC-HIC fractionations (3DLC approach)

Representative RPC/MS results of 3DLC (IEC-HIC-RPC) are depicted. (a) A UV chromatogram of a representative HIC fraction (fraction #20) from IEC fraction #3 from HEK 293 cell lysate samples. (b) Representative mass spectra for the corresponding RPC peaks. (c) Zoom-in mass spectra with unit mass isotopic resolution in chromatographic time-scale. An 80 min RPC gradient were used.

Figure 5. Online RPC/MS/MS protein identifications by HCD for HIC fraction #20 from IEC fraction #3

Representative MS/MS spectra and sequence maps of identified intact proteins with *b/y* ions and *P* values for identification. The insets highlight the isotopic resolution for typical fragments at *m/z* 1018, 1154, and 1569, respectively. The proteins are identified as (a) human transcription elongation factor B polypeptide 2, (b) histidine triad nucleotide-binding protein 2 mitochondrial, and (c) u6 snRNA-associated Sm-like protein LSM7 protein.

Figure 6. Comparison of protein identifications by top-down proteomics using the 2DLC (IEC-RPC) and 3DLC (IEC-HIC-RPC) separation strategies

A single IEC fraction (fraction #3) was used for both 2D and 3D approaches. (a) Venn diagram of protein identifications via 2D and 3D approaches, suggesting that the 3D approach significantly improves protein identifications. (b) Venn diagram of protein identifications using different RPC gradients (80-min, 3-h, and 7-h) via the 2D approach, showing limited improvement was achieved by longer gradients. * all protein identifications in the 3D approach, including redundant intact protein identifications in the 35 HIC fractions. ** non-redundant protein identifications via the 3D approach.