analytical chemistry

Subscriber access provided by Kaohsiung Medical University

Letter

Native proteomics in discovery mode using size exclusion chromatography-capillary zone electrophoresis-tandem mass spectrometry

Xiaojing Shen, Qiang Kou, Ruiqiong Guo, Zhichang Yang,
Daoyang Chen, Xiaowen Liu, Heedeok Hong, and Liangliang Sun

Anal. Chem., **Just Accepted Manuscript** • DOI: 10.1021/acs.analchem.8b02725 • Publication Date (Web): 07 Aug 2018

Downloaded from http://pubs.acs.org on August 10, 2018

Just Accepted

"Just Accepted" manuscripts have been peer-reviewed and accepted for publication. They are posted online prior to technical editing, formatting for publication and author proofing. The American Chemical Society provides "Just Accepted" as a service to the research community to expedite the dissemination of scientific material as soon as possible after acceptance. "Just Accepted" manuscripts appear in full in PDF format accompanied by an HTML abstract. "Just Accepted" manuscripts have been fully peer reviewed, but should not be considered the official version of record. They are citable by the Digital Object Identifier (DOI®). "Just Accepted" is an optional service offered to authors. Therefore, the "Just Accepted" Web site may not include all articles that will be published in the journal. After a manuscript is technically edited and formatted, it will be removed from the "Just Accepted" Web site and published as an ASAP article. Note that technical editing may introduce minor changes to the manuscript text and/or graphics which could affect content, and all legal disclaimers and ethical guidelines that apply to the journal pertain. ACS cannot be held responsible for errors or consequences arising from the use of information contained in these "Just Accepted" manuscripts.

ACS Publications

is published by the American Chemical Society. 1155 Sixteenth Street N.W., Washington, DC 20036

Published by American Chemical Society. Copyright © American Chemical Society. However, no copyright claim is made to original U.S. Government works, or works produced by employees of any Commonwealth realm Crown government in the course of their duties.

Native proteomics in discovery mode using size exclusion chromatography-capillary zone electrophoresis-tandem mass spectrometry

Xiaojing Shen,[a] Qiang Kou,[b] Ruiqiong Guo,[a] Zhichang Yang,[a] Daoyang Chen,[a] Xiaowen Liu,[b,c] Heedeok Hong,[a] and Liangliang Sun[a] *

[a] Department of Chemistry, Michigan State University, 578 S Shaw Ln, East Lansing, MI 48824 USA

[b] Department of BioHealth Informatics, Indiana University-Purdue University Indianapolis, 719 Indiana Avenue, Indianapolis, IN 46202 USA

[c] Center for Computational Biology and Bioinformatics, Indiana University School of Medicine, 410 W. 10th Street, Indianapolis, IN 46202 USA

* Corresponding author. E-mail: lsun@chemistry.msu.edu

Abstract

Native proteomics aims to characterize complex proteomes under native conditions and ultimately produces a full picture of endogenous protein complexes in cells. It requires novel analytical platforms for high-resolution and liquid-phase separation of protein complexes prior to native mass spectrometry (MS) and MS/MS. In this work, size exclusion chromatography (SEC)-capillary zone electrophoresis (CZE)-MS/MS was developed for native proteomics in discovery mode, resulting in the identification of 144 proteins, 672 proteoforms, and 23 protein complexes from the *Escherichia coli* proteome. The protein complexes include four protein homodimers, 16 protein-metal complexes, two protein-[2Fe-2S] complexes, and one protein-glutamine complex. Half of them have not been reported in the literature. This work represents the first example of online liquid-phase separation-MS/MS for characterization of a complex proteome under the native condition, offering the proteomics community an efficient and simple platform for native proteomics.

The majority of proteins in a cell function as protein complexes. Comprehensive characterization of complex proteomes under native conditions, termed “native proteomics”, will ultimately produce a full picture of endogenous protein complexes in a cell.[1] Native proteomics requires high-resolution and liquid-phase separation of a complex proteome prior to native electrospray ionization (nESI)-MS and MS/MS. nESI-MS has been widely used for characterization of purified protein complexes, antibodies and virus assemblies via direct infusion.[2-9] Some work has been done using liquid-phase separation-nESI-MS for characterization of standard protein complexes or samples with very low complexity.[10-16] Recently, Skinner *et al.* coupled off-line ion exchange chromatography or clear native gel-eluted liquid fraction entrapment electrophoresis[17] to direct infusion nESI-MS/MS for native proteomics of mouse hearts and four human cell lines, leading to the identification of 164 proteins and 125 protein complexes from 600 fractions.[1] This is the first example of native proteomics. However, the workflow is labor- and time-consuming. Coupling an online and high-resolution separation technique to nESI-MS and MS/MS is required to boost the throughput and scale of native proteomics. Capillary zone electrophoresis (CZE)-MS/MS has a great potential for native proteomics due to the high separation efficiency of CZE for intact proteins,[18-20] the mature CE-MS interfaces,[21-24] and its capability for high-resolution separation and high sensitive detection of protein complexes under native conditions.[13,16] However, there is still no report on evaluating CZE-MS/MS for native proteomics of complex proteomes.

In this work, we coupled size exclusion chromatography (SEC) prefractionation to online CZE-MS/MS for native proteomics in discovery mode, **Figure 1**. *Escherichia coli* (*E.coli*) cells were lysed in PBS buffer. The extracted proteins were fractionated with SEC into 8 fractions. The mobile phase was 100 mM ammonium acetate (NH_4Ac, pH 7.0). After simple protein concentration and buffer exchange with Microcon-30 kDa centrifugal filter units, the SEC fractions were analyzed by CZE-MS/MS. We evaluated the sample loss during the buffer exchange using SDS-PAGE, **Figure S1**. We did not observe significant differences in protein abundance before and after the buffer exchange. The protein abundance in the flow-through sample was ignorable compared with the original sample. For the CZE-MS/MS, the commercialized electro-kinetically pumped sheath

flow interface (CMP Scientific, Brooklyn, NY) was used to couple CZE to MS.[23,24] The background electrolyte (BGE) and the sheath buffer were 50 mM NH_4Ac (pH 6.9) and 25 mM NH_4Ac (pH 6.9), respectively. A Q-Exactive HF mass spectrometer (Thermo Fisher Scientific) was used. The SEC-CZE-MS/MS platform is straightforward. The CZE-MS/MS analyses of the 8 SEC fractions took 16 hours. One example electropherogram is shown in **Figure 1B**. The native CZE-MS/MS run obtained 15 major peaks and approached a 1-hour separation window. TopPIC (Top-down mass spectrometry based Proteoform Identification and Characterization) software was used for the database search of the acquired MS/MS spectra for proteoform identification.[25,26] We note that the Q-Exactive HF cannot isolate ions with the mass-to-charge ratio (*m/z*) higher than 2500 for fragmentation. In this proof-of-principle work, we focused on identification of protein complexes with mass lower than 30 kDa. The experimental details are described in Supporting Information I.

A total of 144 proteins and 672 proteoforms were identified from the *E.coli* lysate with a 1% spectrum-level false discovery rate (FDR) and a 5% proteoform-level FDR. The identified proteoforms are listed in Supporting Information II. The number of protein identifications from each SEC fraction and the protein-level overlap between adjacent SEC fractions are shown in **Figure S2**. The data indicate that SEC can reach a reasonable protein separation under the native condition. Most of the identified proteoforms have mass lower than 30 kDa, **Figure 1C**. 23 protein complexes from 17 proteins were identified, including four homodimers, 16 protein-metal complexes, two protein-[2Fe-2S] complexes, and one protein-glutamine complex. 14 out of the 23 protein complexes have not been reported before. The details of those protein complexes are listed in **Table S1**. The SEC-CZE-MS/MS performed native proteomics in discovery mode because the identities of protein complexes were unknown before analysis.

The protein complexes were identified through several steps. First, during CZE-MS/MS analysis, a protein or a whole protein complex was isolated in the quadrupole, followed by high energy collision dissociation (HCD). The MS and MS/MS spectra of the proteins and protein complexes were acquired. Second, proteoforms were identified through

database search of the acquired MS and MS/MS spectra against a UniProt *E. coli* database with the TopPIC software.[25,26] Third, the mass shifts of the identified proteoforms were compared with the masses of known protein co-factors in the UniProt *E. coli* database manually. The co-factors are shown in **Table S2**. If they matched with each other within a 4-Da mass difference, we assumed that the mass shift corresponded to the specific cofactor. We obtained a list of proteoforms that were potential protein complexes with the cofactors. Fourth, we compared those proteoforms with that identified in our recent deep top-down proteomics work. We identified nearly 6000 proteoforms and 850 proteins from the *E. coli* proteome using denaturing top-down approach. [27] Because of the denaturing conditions, the non-covalently bound cofactors were lost during that experiment. If the potential protein complexes detected in this work matched well with some proteoforms in reference [27], the corresponding mass shifts should represent some covalent modifications. Those potential protein complexes were removed from the initial protein complex list. After this step, we obtained a list of identified protein complexes with bound cofactors. Finally, we searched the UniProt database to find information on those identified protein complexes in the literature. Protein complexes without literature information were considered as unreported protein complexes. The workflow using RNA polymerase-binding transcription factor DksA-zinc complex as an example was described in supporting information I. **Figure 2A** shows one deconvoluted spectrum of the DksA-zinc complex. **Figure S3** shows the sequence, observed fragmentation pattern, and detected mass shift of the DksA-zinc complex after the database search. Using this approach, we identified 17 protein complexes including 16 protein-zinc/copper complexes and one protein-[2Fe-2S] complex, **Table S1**. Besides the detected 2Fe-2S ferredoxin complex, we observed another form of the complex with additional 21-Da modification. Seven out of the identified protein complexes have been reported in the literature including YcaR-zinc complex,[28] DksA-zinc complex,[29] DsbA-zinc complex,[30] thioredoxin 1-copper complex,[31] RcnB-copper complex,[32] YobA-copper complex,[33] and ferredoxin-[2Fe-2S] complex.[34] The data clearly indicate that those non-covalent interactions can be preserved during the SEC-CZE-MS analysis. The 11 unreported protein complexes include the 2Fe-2S ferredoxin complex with additional 21-Da modification, a truncated

DsbA-zinc complex, and protein-metal complexes from nine novel zinc/copper-binding proteins.

We noted that the mass spectrometer used in this work limited our capability to localize the protein co-factors in the protein sequences. For example, as shown in **Figure S3**, the zinc ion is localized between the 50th and 72nd amino acids based on the database search result. However, the zinc ion should bind with the four cysteine amino acids at positions 114,117,135, and 138 based on the UniProt database. During the HCD fragmentation, the zinc ion and the DksA protein fell apart, leading to a challenge for accurately localizing the zinc ion in the protein sequence. During the database search with the TopPIC, the mass shift corresponding to the co-factor was assigned to a region that no fragment ion could cover. In order to improve the localization of protein co-factors, we will employ mass spectrometers with electron transfer dissociation (ETD) [35] or electron capture dissociation (ECD) [9] in our future work.

We identified four homodimers from two proteins, acid stress chaperone HdeA and phosphocarrier protein HPr, **Table S1**. The masses of these homodimers are 19 487 Da (HdeA), 19 466 Da (HdeA), 19 527 Da (HdeA), and 18 227 Da (HPr). For those two proteins, the mass shifts of some proteoforms are about 50% of the proteoform mass. For example, a mass shift of 9 731 Da was detected from one HdeA proteoform that had a mass of 19 466 Da. The proteoform is the homodimer of HdeA. The data agree well with the literature.[36] HdeA is homodimer at neutral pH and dissociates into monomer at pH 4. Using the same approach, we detected the homodimer of phosphocarrier protein HPr. Another two forms of HdeA homodimer were identified with additional 18-Da and 58-Da modifications. Those three protein complexes have not been reported previously. We identified one proteoform of glutamine-binding periplasmic protein (glnH) with a mass shift of 146 Da. The mass shift matches well with the mass of glutamine. The proteoform represents the glnH-glutamine complex. GlnH is involved in glutamine transport. One crystal structure of the glnH-glutamine complex has been reported,[37] **Figure 2B**. The data highlight the capability of our platform for identification of various protein complexes.

We noted that NfuA-[4Fe-4S] complex and 50S ribosomal protein L31-zinc complex have been reported with bound cofactors.[38,39] However, we only identified proteoforms corresponding to those two proteins without the cofactors through the database search and did not identify the whole protein complexes. Cysteine(C)149 and C152 of NfuA are known to play central roles in binding the [4Fe-4S] cluster, and C16 of 50S ribosomal protein L31 is crucial for zinc ion binding. We detected mass shifts as -4 and -2 Da from the identified 50S ribosomal protein L31 and NfuA proteoforms, **Figure 2C** and **Figure S4**. Based on those mass shifts and their location in the protein sequences, we concluded that those mass shifts represented two disulfide bonds among the four cysteines (C16, C18, C37, and C40) in 50S ribosomal protein L31 and one disulfide bond (C149-C152) in NfuA. Therefore, 50S ribosomal protein L31 and NfuA were detected without the zinc ion and [4Fe-4S]. We further performed mass deconvolution on the averaged mass spectra across the peaks of the identified 50S ribosomal protein L31 and NfuA proteoforms without the cofactors, **Figures S5** and **S6**. It is clear that 50S ribosomal protein L31 with two disulfide bonds and NfuA with one disulfide bond dominate the spectra. There are not very strong protein peaks corresponding to the 50S ribosomal protein L31-zinc complex and NfuA-[4Fe-4S] complex in **Figures S5** and **S6**. The results demonstrate that a larger fraction of 50S ribosomal protein L31 and NfuA exist as apo forms lacking the cofactors in the *E.coli* cells used in the experiment, providing a new insight into the cofactor binding of these two protein complexes.

The SEC-CZE-MS/MS platform identified 16 protein-metal complexes. These metalloproteins are involved in metal ion binding, catalysis, enzyme regulation, transcription, and transmembrane transport, **Figure 2D**. This work agrees with the literature regarding the molecular function distribution of metalloproteins.[40] The platform enabled us to determine the metal binding stoichiometry of most of those metalloproteins, **Table S3** and **Figure 2E**. Protein YcaR and DksA bind zinc ion through sulfur from cysteine (C), [28,29] and others most likely bind metal ions through nitrogen from histidine (H) and/or oxygen from acidic amino acids (aspartic acid (D) and glutamic acid (E)). Zinc ion binding through sulfur and nitrogen is generally more stable than that through nitrogen and oxygen.[41] Our metal binding stoichiometry results agree well with the general concept from the literature. For YcaR and DksA, the abundance of the

metal-binding form is at least 8 times higher than the non-binding form. For other proteins, the metal-binding form has lower abundance than the non-binding form. Our results highlight the potential of the native SEC-CZE-MS/MS platform for high throughput characterization of metal ion binding on metalloproteins directly from complex proteomes.

We identified many post-translational modifications (PTMs) including N-terminal acetylation, phosphorylation, C-terminal thiocarboxylation, 4'-phosphopantetheine, biotinylation, and disulfide bond. An example of those PTMs is shown in **Figure S7**. Some of the PTMs and corresponding proteins are listed in **Table S4**. We identified 55 proteoforms with N-terminal acetylation. Disulfide bonds were identified on eight proteins, and four of them are reported for the first time. We detected unreported signal peptide cleavage and initial methionine excision on 25 proteins, **Table S5**. An example of unreported signal peptide cleavage is shown in **Figure S8**.

In summary, we developed a novel, efficient and high-throughput SEC-CZE-MS/MS platform for characterization of endogenous protein complexes in cells. The proof-of-principle study of the *E.coli* proteome identified 144 proteins, 672 proteoforms, and 23 protein complexes in discovery mode. The platform will be useful for the proteomics community for charactrization of complex proteomes under the native condition. The platform can be further improved through optimization of the SEC and CZE conditions for better separation of protein complexes, via use of multiple fragmentation techniques (*e.g.*, HCD,[42] ETD,[35] ECD[9], and ultraviolet photodissociation[43]) for more comprehensive fragmentation of intact protein complexes, and by employing a high-resolution mass spectrometer that is capable for detection of large protein complexes with very high *m/z* and isolation of those protein complexes for fragmentation.[44-46]

Supporting Information

Experimental procedures; The list of the identified protein complexes; The list of protein co-factors; The metal binding stoichiometry of some identified metalloproteins; The list of some PTMs detected in this work; The list of proteins with unreported signal peptide

cleavage and initial methionine excision; Image of the SDS-PAGE result; The number of protein IDs from each SEC fraction and protein overlaps between adjacent SEC fractions; The observed fragmentation pattern and the mass shift of the DksA; The observed fragmentation pattern and the mass shift of the Fe/S biogenesis protein NfuA; The deconvoluted spectrum of 50S ribosomal protein L31; The deconvoluted spectrum of the Fe/S biogenesis protein NfuA; The observed fragmentation pattern and the modifications of the 50S ribosomal protein L7/L12; The observed fragmentation pattern and the modifications of the 50S ribosomal protein L25. (PDF)

The identified proteoforms (XLSX)

Notes

The authors declare no competing financial interest.

Acknowledgements

We thank Profs. Daniel Jones and Jian Hu at the Department of Biochemistry and Molecular Biology at Michigan State University for kindly providing help on the project. We thank the support from the National Institute of General Medical Sciences, National Institutes of Health (NIH), through Grant R01GM118470 (X. Liu), R01GM125991 (L. Sun and X. Liu) and R01GM118685 (H. Hong).

References

[1] Skinner, O. S.; Haverland, N. A.; Fornelli, L.; Melani, R. D.; Do Vale, L. H. F.; Seckler, H. S.; Doubleday, P. F.; Schachner, L. F.; Srzentić, K.; Kelleher, N. L.; Compton, P. D. *Nat. Chem. Biol.* **2018**, *14*, 36-41.

[2] Benesch, J. L.; Ruotolo, B. T.; Simmons, D. A.; Robinson, C. V. *Chem. Rev.* **2007**, *107*, 3544-3567.

[3] Lu, Y.; Zhang, H.; Niedzwiedzki, D. M.; Jiang, J.; Blankenship, R. E.; Gross, M. L. *Anal. Chem.* **2016**, *88*, 8827-8834.

[4] Rosati, S.; Rose, R. J.; Thompson, N. J.; van Duijn, E.; Damoc, E.; Denisov, E.; Makarov, A.; Heck, A. J. *Angew. Chem. Int. Ed.* **2012**, *51*, 12992-12996.

[5] Snijder, J.; Rose, R. J.; Veesler, D.; Johnson, J. E.; Heck, A. J. *Angew. Chem. Int. Ed.* **2013**, *52*, 4020-4023.

[6] Quintyn, R. S.; Zhou, M.; Yan, J.; Wysocki, V. H. *Anal. Chem.* **2015**, *87*, 11879-11886.

[7] Susa, A. C.; Xia, Z.; Williams, E. R. *Angew. Chem. Int. Ed.* **2017**, *56*, 7912-7915.

[8] Marty, M. T.; Hoi, K. K.; Gault, J.; Robinson, C. V. *Angew. Chem. Int. Ed.* **2016**, *55*, 550-554.

[9] Li, H.; Nguyen, H. H.; Ogorzalek Loo, R. R.; Campuzano, I. D. G.; Loo, J. A. *Nat. Chem.* **2018**, *10*, 139-148.

[10] Chen, B.; Peng, Y.; Valeja, S. G.; Xiu, L.; Alpert, A. J.; Ge, Y. *Anal. Chem.* **2016**, *88*, 1885-1891.

[11] Muneeruddin, K.; Nazzaro, M.; Kaltashov, I. A. *Anal. Chem.* **2015**, *87*, 10138-10145.

[12] Muneeruddin, K.; Thomas, J. J.; Salinas, P. A.; Kaltashov, I. A. *Anal. Chem.* **2014**, *86*, 10692-10699.

[13] Nguyen, A.; Moini, M. *Anal. Chem.* **2008**, *80*, 7169-7173.

[14] Marie, A. L.; Dominguez-Vega, E.; Saller, F.; Plantier, J. L.; Urbain, R.; Borgel, D.; Tran, N. T.; Somsen, G. W.; Taverna, M. *Anal. Chim. Acta.* **2016**, *947*, 58-65.

[15] Said, N.; Gahoual, R.; Kuhn, L.; Beck, A.; François, Y. N.; Leize-Wagner, E. *Anal. Chim. Acta.* **2016**, *918*, 50-59.

[16] Belov, A. M.; Viner, R.; Santos, M. R.; Horn, D. M.; Bern, M.; Karger, B. L.; Ivanov, A. R. *J. Am. Soc. Mass. Spectrom.* **2017***, 28,* 2614-2634.

[17] Melani, R. D.; Seckler, H. S.; Skinner, O. S.; Do Vale, L. H.; Catherman, A. D.; Havugimana, P. C.; Valle de Sousa, M.; Domont, G. B.; Kelleher, N. L.; Compton, P. D. *J. Vis. Exp.* **2016**, *108*, e53597.

[18] Jorgenson, J. W.; Lukacs, K. D. *Science* **1983**, *222*, 266-272.

[19] Haselberg, R.; de Jong, G. J.; Somsen, G. W. *Anal. Chem.* **2013**, *85*, 2289-2296.

[20] Han, X.; Wang, Y.; Aslanian, A.; Fonslow, B.; Graczyk, B.; Davis, T. N.; Yates, J. R. 3rd, *J. Proteome Res.* **2014**, *13*, 6078-6086.

[21] Smith, R. D.; Barinaga, C. J.; Udseth, H. R. *Anal. Chem.* **1988**, *60*, 1948-1952.

[22] Moini, M. *Anal. Chem.* **2007**, *79*, 4241-4246.

[23] Wojcik, R.; Dada, O. O.; Sadilek, M.; Dovichi, N. J. *Rapid Commun. Mass Spectrom.* **2010**, *24*, 2554-2560.

[24] Sun, L.; Zhu, G.; Zhao, Y.; Yan, X.; Mou, S.; Dovichi, N. J. *Angew. Chem. Int. Ed.* **2013**, *52*, 13661-13664.

[25] Kou, Q.; Xun, L.; Liu, X. *Bioinformatics* **2016**, *32*, 3495-3497.

[26] Liu, X.; Inbar, Y.; Dorrestein, P. C.; Wynne, C.; Edwards, N.; Souda, P.; Whitelegge, J. P.; Bafna, V.; Pevzner, P. A. *Mol. Cell. Proteomics* **2010**, *9*, 2772-2782.

[27] McCool, E. N.; Lubeckyj, R. A.; Shen, X.; Chen, D.; Kou, Q.; Liu, X.; Sun, L. *Anal. Chem.* **2018**, *90*, 5529-5533.

[28] Bourgeois, G.; Létoquart, J.; van Tran, N.; Graille, M. *Biomolecules* **2017**, *7*, pii: E7.

[29] Perederina, A.; Svetlov, V.; Vassylyeva, M. N.; Tahirov, T. H.; Yokoyama, S.; Artsimovitch, I.; Vassylyev, D. G. *Cell* **2004**, *118*, 297-309.

[30] Inaba, K.; Murakami, S.; Suzuki, M.; Nakagawa, A.; Yamashita, E.; Okada, K.; Ito, K. *Cell* **2006**, *127*, 789-801.

[31] Rudresh; Jain, R.; Dani, V.; Mitra, A.; Srivastava, S.; Sarma, S. P.; Varadarajan, R.; Ramakumar, S. *Protein Eng.* **2002**, *15*, 627-633.

[32] Blériot, C.; Gault, M.; Gueguen, E.; Arnoux, P.; Pignol, D.; Mandrand-Berthelot, M. A.; Rodrigue, A. *Metallomics* **2014**, *6*, 1400-1409.

[33] Rensing, C.; Grass, G. *FEMS Microbiol. Rev.* **2003**, *27*, 197-213.

[34] Knoell, H. E.; Knappe, J. *Eur. J. Biochem.* **1974**, *50*, 245-252.

[35] Syka, J. E.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. Proc. Natl. Acad. Sci. U.S.A. 2004, 101, 9528-9533.

[36] Gajiwala, K. S.; Burley, S. K. *J. Mol. Biol.* **2000**, *295*, 605-612.

[37] Sun, Y. J.; Rose, J.; Wang, B. C.; Hsiao, C. D. *J. Mol. Biol.* **1998**, *278*, 219-229.

[38] Hensley, M. P.; Gunasekera, T. S.; Easton, J. A.; Sigdel, T. K.; Sugarbaker, S. A.; Klingbeil, L.; Breece, R. M.; Tierney, D. L.; Crowder, M. W. *J. Inorg. Biochem.* **2012**, *111*, 164-172.

[39] Angelini, S.; Gerez, C.; Ollagnier-de Choudens, S.; Sanakis, Y.; Fontecave, M.; Barras, F.; Py, B. *J. Biol. Chem.* **2008**, *283*, 14084-14091.

[40] Andreini, C.; Bertini, I.; Rosato, A. *Acc. Chem. Res.* **2009**, *42*, 1471-1479.

[41] Pearson, R. G. *J. Am. Chem. Soc.* **1963**, *85*, 3533-3539.

[42] Olsen, J. V.; Macek, B.; Lange, O.; Makarov, A.; Horning, S.; Mann, M. *Nat. Methods* **2007**, *4*, 709-712.

[43] Shaw, J. B.; Li, W.; Holden, D. D.; Zhang, Y.; Griep-Raming, J.; Fellers, R. T.; Early, B. P.; Thomas, P. M.; Kelleher, N. L.; Brodbelt, J. S. *J. Am. Chem. Soc.* **2013**, *135*, 12646-12651.

[44] van de Waterbeemd, M.; Fort, K. L.; Boll, D.; Reinhardt-Szyba, M.; Routh, A.; Makarov, A.; Heck, A. J. *Nat. Methods* **2017**, *14*, 283-286.

[45] Tamara, S.; Dyachenko, A.; Fort, K. L.; Makarov, A. A.; Scheltema, R. A.; Heck, A. J. *J. Am. Chem. Soc.* **2016**, *138*, 10860-10868.

[46] Dyachenko, A.; Wang, G.; Belov, M.; Makarov, A.; de Jong, R. N.; van den Bremer, E. T.; Parren, P. W.; Heck, A. J. *Anal. Chem.* **2015**, *87*, 6095-6102.

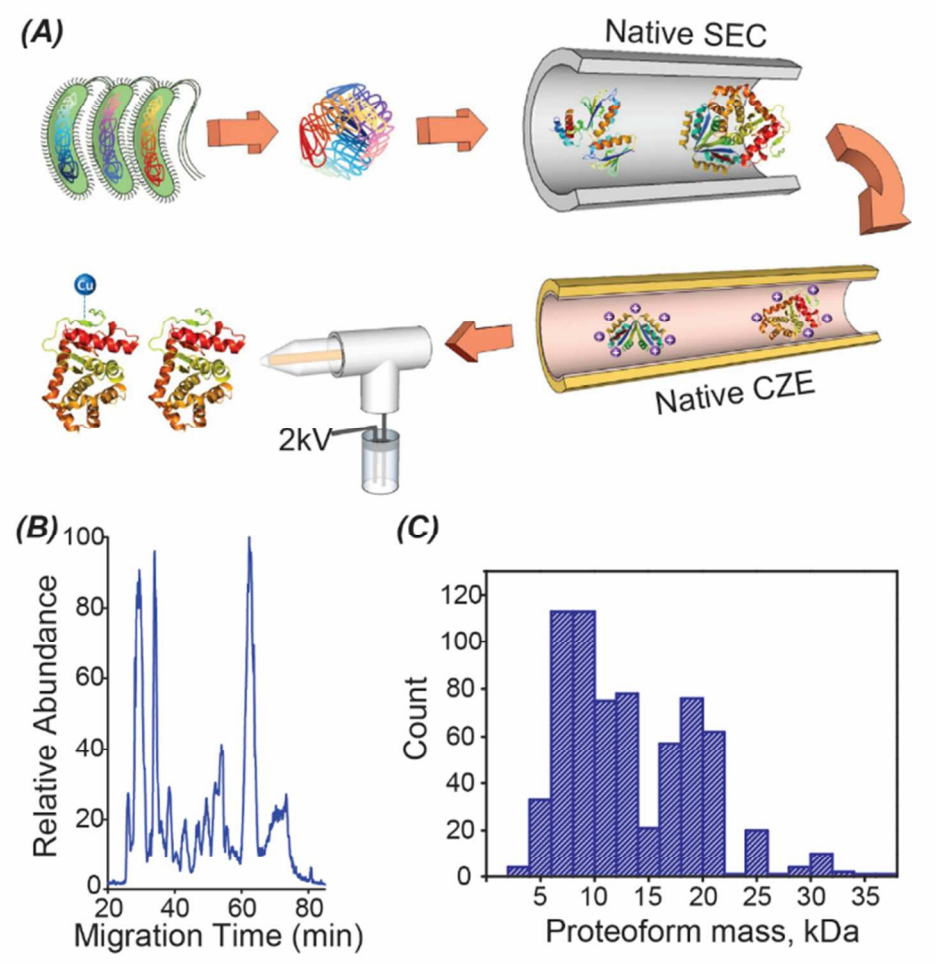


Figure 1. (A) The SEC-CZE-ESI-MS/MS platform for native proteomics. (B) An example base peak electropherogram of an SEC fraction of the *E. coli* lysate after CZE-MS/MS analysis. (C) The mass distribution of the identified proteoforms from the *E. coli* proteome.

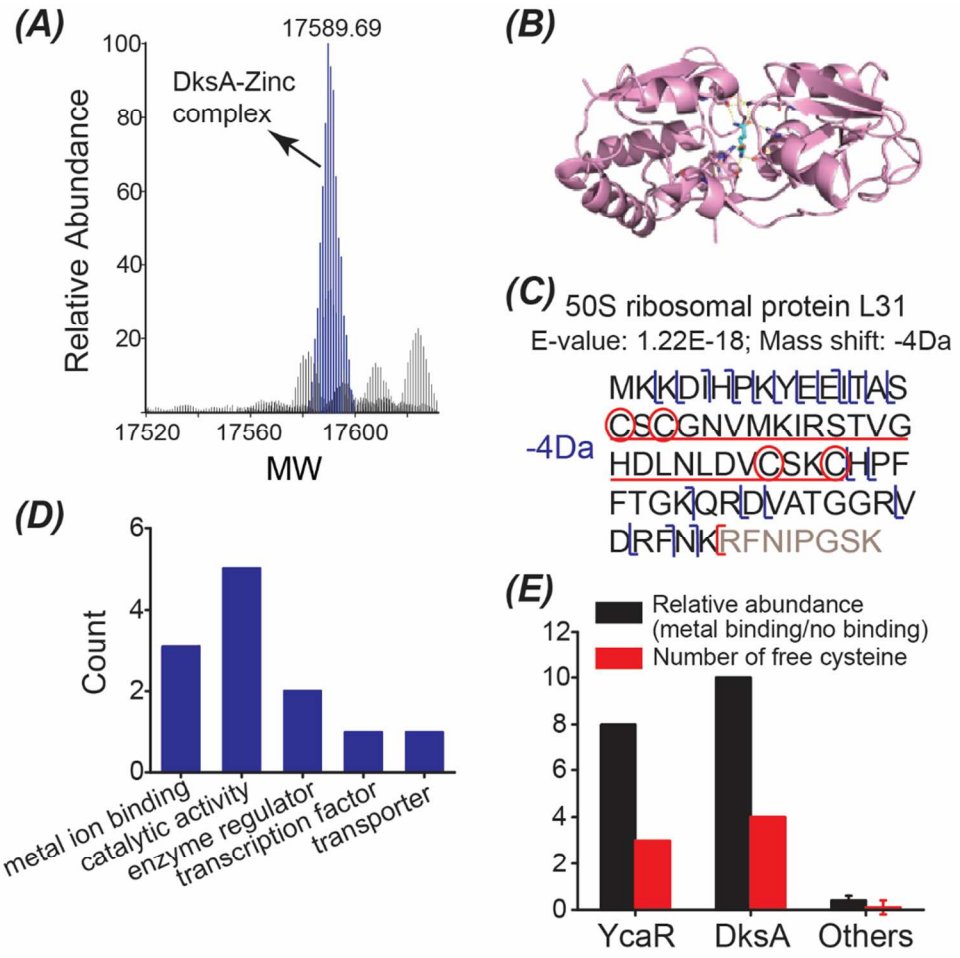


Figure 2. (A) One deconvoluted spectrum of the identified RNA polymerase-binding transcription factor DksA-zinc complex. The averaged mass spectrum across the peak of the complex was used for the mass deconvolution with the Xtract software (Thermo Fisher Scientific) using the default settings. The x-axis is molecular weight (MW). (B) The crystal structure of glutamine-binding periplasmic protein bound with a glutamine molecule. The image of the crystal structure was obtained from the Protein Data Bank in Europe (https://www.ebi.ac.uk/pdbe/). (C) The sequence, observed fragmentation pattern, and detected mass shift of the 50S ribosomal protein L31 through the database search. The location of the mass shift and the cysteine amino acids are highlighted. (D) The molecular function distribution of the identified metalloproteins. The Retrieve/ID mapping tool on the UniProt website (http://www.uniprot.org/uploadlists/) was used to obtain the molecular function information. (E) The metal binding stoichiometry of some identified metalloproteins. The detailed information is shown in Table S3. The error bars for “Others” represent the standard deviations of relative abundance and cysteine count from 13 metalloproteins.

Table of Contents (TOC)

