This is an open access article published under an ACS AuthorChoice License, which permits copying and redistribution of the article or any adaptations for non-commercial purposes.

analytical chemistry

Letter

pubs.acs.org/ac

# Sheathless Capillary Electrophoresis-Tandem Mass Spectrometry for Top-Down Characterization of *Pyrococcus furiosus* Proteins on a Proteome Scale

Xuemei Han,[†] Yueju Wang,[†] Aaron Aslanian,[†,‡] Marshall Bern,[∥,§] Mathieu Lavallée-Adam,[†] and John R. Yates, III*[,†]

[†]Department of Chemical Physiology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, United States

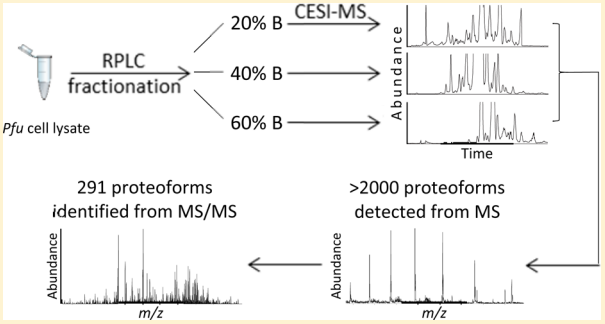
[‡]Molecular and Cell Biology Laboratory, Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, California 92037, United States

[∥]Protein Metrics, 1622 San Carlos Avenue, San Carlos, California 94070, United States

[§]Palo Alto Research Center, 3333 Coyote Hill Road, Palo Alto, California 94304, United States

Supporting Information

**ABSTRACT:** Intact protein analysis via top-down mass spectrometry (MS) provides the unique capability of fully characterizing protein isoforms and combinatorial post-translational modifications (PTMs) compared to the bottom-up MS approach. Front-end protein separation poses a challenge for analyzing complex mixtures of intact proteins on a proteomic scale. Here we applied capillary electrophoresis (CE) through a sheathless capillary electrophoresis-electrospray ionization (CESI) interface coupled to an Orbitrap Elite mass spectrometer to profile the proteome from *Pyrococcus furiosus*. CESI-top-down MS analysis of *Pyrococcus furiosus* cell lysate identified 134 proteins and 291 proteoforms with a total sample consumption of 270 ng in 120 min of total analysis time. Truncations and various PTMs were detected, including acetylation, disulfide bonds, oxidation, glycosylation, and hypusine. This is the largest scale analysis of intact proteins by CE-top-down MS to date.




Mass spectrometry (MS)-based proteomics has grown to become a major analytical tool for the study of complex biological processes. The high-throughput bottom-up MS approach, in which protein proteolytic digests are separated, analyzed, and used to infer protein identity, is dominant in the proteomics field due to the rapid development of peptide separation techniques, as well as the advances in MS instrumentation and bioinformatic tools tailored to peptide analysis. However, information pertaining to combinatorial post-translational modifications (PTMs) and protein splice variant isoforms is often lost after proteolysis.[1,2] Two dimensional electrophoresis (2DE) is a protein separation technique that has been used to investigate and identify splice variant isoforms and PTMs.[3−5] However, a major drawback of 2DE is sensitivity primarily due to the sample quantity requirements of protein visualization methods and losses associated with peptide/protein extraction from the gel. Also, analyzing large numbers of gel spots obtained by 2DE can be very time-consuming. Alternatively, the limitations of bottom-up MS analysis can be overcome by the top-down MS approach, in which the intact proteins are measured as a whole and fragmented directly in the mass spectrometer. This technique has primarily been applied to targeted analysis of single proteins or simple protein mixtures since its initial introduction 2 decades ago.[6−8] Very recently, with the development of high-resolution MS instrumentation,[9,10] front-end separations,[11−13] and top-down software,[14−17] top-down analyses of various organisms have been reported on a proteome scale.[18−22] For example, Tran et al.[21] identified 1043 proteins, including over 3000 proteoforms, from HeLa S3 cells with a total analysis time of approximately 45 h using a customized three-stage separation system. Ansong et al.[22] reported the identification of 563 proteins including 1665 proteoforms from *Salmonella typhimurium* with a single dimension ultrahigh-pressure liquid chromatography (LC) separation that enabled the analysis of 5 μg of sample in approximately 4 h. However, these studies still lag behind the capabilities of bottom-up analyses by an order of magnitude in regard to protein identification and require greater sample consumption.

**Received:** September 12, 2014
**Accepted:** October 26, 2014
**Published:** October 26, 2014

 © 2014 American Chemical Society 

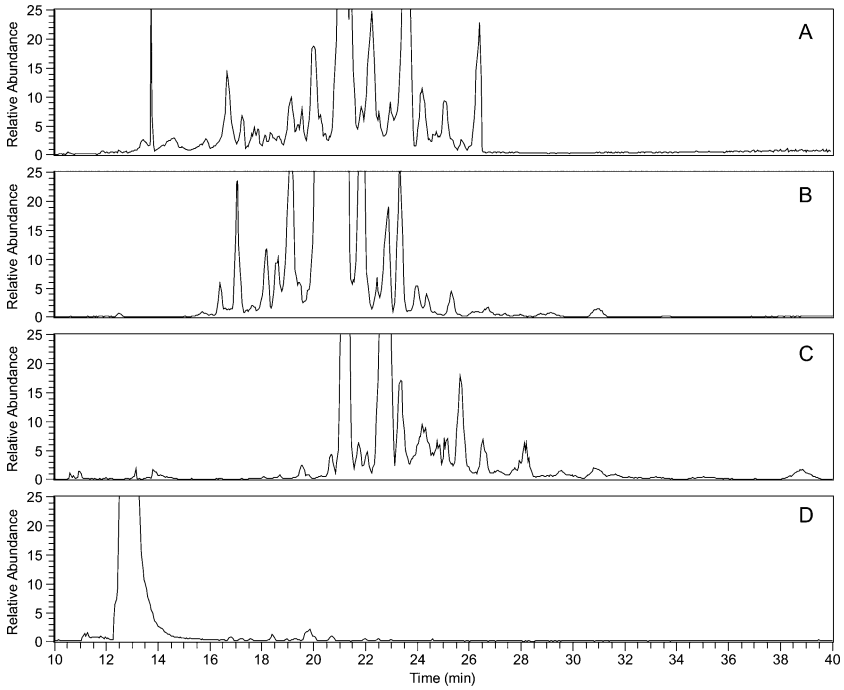



**Figure 1.** Electropherograms of the *Pyrococcus furiosus* intact proteins analyzed by CE-MS. 90 ng of RPLC step eluted fractions with (A) 20% buffer B, (B) 40% buffer B, (C) 60% buffer B, and (D) 100% buffer B were injected. CE separation conditions: PEI coated capillary; voltage −30 kV; BGE 0.1% acetic acid, 20% IPA. Taking into account the ∼30 min separation window, the average peak capacities in parts A−C were calculated to be ∼120 (using the average peak width at half peak height).

New technological developments are needed to improve the sensitivity, sequence coverage, throughput, and robustness of top-down MS characterization of complex protein mixtures. One of the key challenges for top-down proteomics is the lack of a high-efficiency and high-resolution intact protein separation technique. Reverse phase liquid chromatography (RPLC) is the method of choice for separations but has drawbacks when applied to intact proteins, such as irreversible protein adsorption to the stationary phase. Although recent advances[11,12,23] with solution-based separation techniques have facilitated the further development of top-down MS technologies, the development of alternative techniques is still a necessary step for the maturation of this discipline.

Capillary electrophoresis (CE) has been recognized as a powerful method for separating intact proteins, possessing advantages such as high separation efficiency, short separation time, as well as low sample consumption.[24−27] Smith and coworkers have pioneered CE top-down proteomics by coupling capillary isoelectric focusing (CIEF) to Fourier transform ion cyclotron resonance (FTICR) mass spectrometry for *Escherichia coli* proteome characterization.[28] CIEF separates proteins according to the differences in p*I* and normally needs the addition of ampholytes for the generation of a pH gradient, which can cause ion suppression and contamination in the MS signal. In contrast, capillary zone electrophoresis (CZE) is another attractive CE mode which will generate much lower background noise, due to the lack of ampholytes, and is much easier in system automation. CZE separates analytes based on their mass to charge ratios and its separation efficiency can be very high for proteins since they have relatively low diffusion coefficients, thereby restricting band broadening. One of the main reasons for the unpopularity of CE-MS relative to LC-MS is the difficulty of online interfacing CE to ESI-MS. The sheath liquid interface[29] is most widely used but it is always associated with analyte dilution. Recently, two low-flow sheath liquid sprayers have been designed independently by the Chen and Dovichi laboratories to minimize band broadening and analyte dilution.[27,30−32] The Dovichi group developed an electrokinetically pumped low sheath-flow electrospray CE-MS interface and has shown its application not only in bottom-up proteomics but also top-down protein analysis. Other than model protein separation and Orbitrap Velos detection for top-down demonstration,[27] they recently also applied this platform for top-down analysis of more complex biological samples,[33] in which the *Mycobacterium marinum* secretome was separated by CZE and analyzed by Q-Exactive mass spectrometry via the electrokinetic sheath-flow interface. They were able to identify 22 proteins in a single 1-h analysis with a total sample consumption of 500 ng. The Kelleher group also reported the use of the electrokinetically pumped electrospray interface for *Pseudomonas aeruginosa* top-down analysis.[34] They applied GELFrEE as sample prefractionation followed by CZE-MS/MS analysis and were able to identify 30 proteins in the mass range of 30−80 kDa.

Other efforts have focused on the development of sheathless CESI interfaces, which completely eliminate the sheath liquid to maximize electrospray sensitivity. One of the most attractive subsets is the porous tip sheathless electrospray interface.[35,36] This prototype interface has been recently evaluated by several research groups and demonstrated highly sensitive proteomic analysis.[37−40] Using this sheathless interface, the Lindner group[39] demonstrated that CE-MS/MS is 10-fold more sensitive than nLC-MS/MS for the analysis of histone H1 peptides. The Mayboroda group[41] has applied transient ITP (tITP)-CZE-MS/MS for sensitive glycopeptide characterization and found a 40-fold increase in sensitivity for IgG1 Fc glycopeptide analysis when compared to a conventional strategy. In addition to peptide level analysis, this sheathless


dx.doi.org/10.1021/ac503439n | *Anal. Chem.* 2014, 86, 11006−11012

CESI interface has also been applied to the characterization of intact proteins. Haselberg et al.[42] showed that the detection limits of four model proteins improved by a factor of 50−140 compared to the sheath-liquid interface; more recently they also demonstrated the effectiveness and high sensitivity in glycoform profiling of intact pharmaceutical proteins by CE-TOF MS analysis.[26]

In this work, we coupled CZE with an Orbitrap Elite mass spectrometer through the prototype sheathless CESI interface for the top-down characterization of the *Pyrococcus furiosus* proteome. *Pyrococcus furiosus* is a hyperthermophilic archaeon that grows optimally at 100 °C. It has 1908 kilobases of DNA sequence and 2065 open reading frames. The average molecular weight of all proteins in the *Pyrococcus furiosus* Uniprot database is 32 kDa. *Pyrococcus furiosus* has been established as a proteomics standard[43] due to its moderate complexity and very little duplication within the genome. We applied the aforementioned CE-MS platform to analyze these *Pyrococcus furiosus* proteins. After a failed initial attempt of separating the whole lysate with CE, we decided to reduce the sample complexity by step elution from a reverse phase separation. The resulting four fractions were dried, redissolved in water, and then injected individually as described in the Supporting Information, Experimental Section. We were able to solubilize these fractions in water, without needing to add 70% acetic acid as reported by the Dovichi group. We optimized the BGE buffer composition for effective separation of *Pyrococcus furiosus* proteins and found that 0.1% acetic acid with 20% IPA gave the best protein separation and the most sensitive MS detection. As shown in Figure 1, about 90 ng injection of each *Pyrococcus furiosus* fraction yielded a unique electropherogram with a good signal-to-noise ratio. All protein species came out within a 30 min separation window with peak widths as short as 10 s. For this CE analysis, a 180 s injection was used to increase the sample loading, which corresponds to a total of 13 cm or 14% of the injection plug; this large sample plug will introduce peak broadening in regular CE separation. In order to minimize the peak broadening and increase the separation resolution, tITP was used for sample stacking and preconcentration.

To conservatively estimate the number of detected protein species in each fraction, deconvoluted mass lists generated by Xtract were further grouped with an in-house built program (Supporting Information, Experimental Section) to merge the mass over different charge states and scans as well as to correct isotope shifted assignments. Such analyses of the full scan FTMS spectra revealed hundreds of unique protein masses in each fraction (Table 1) ranging from 2 kDa to 30 kDa with an average of 4.7 kDa. Interestingly the mass distribution is biased toward low mass species (Figure 2). This could be due to the decreased sensitivity for larger proteins with Orbitrap detection as the numbers of charge states and isotopic peaks increase as MW increases, and this could result in poorly resolved isotopic envelopes to be deconvoluted readily with any decharge, deisotope software. Moreover, possible endogenous proteolysis during sample preparation and/or storage may also contribute to the observed low molecular weights.

**Table 1. Number of Detected Species in MS1 and Identified *Pyrococcus furiosus* Proteoforms in MS2 for Each of Four RPLC Fractions by CE Top-Down MS Analysis**

| RPLC fraction | detected protein species in MS | identified proteoforms in MS/MS | identified proteins in MS/MS | proteoform level FDR (%) |
|---|---|---|---|---|
| 20% B | 1346 | 144 | 71 | 1.4 |
| 40% B | 834 | 126 | 67 | 2.4 |
| 60% B | 385 | 63 | 47 | 4.8 |
| 100% B | 289 | 0 | 0 | N/A |
| total | 2585 | 291 | 134 | 2.7 |

Sensitive and efficient CE separation coupled with top-down tandem mass spectrometry on the Orbitrap Elite mass spectrometer was used to profile the *Pyrococcus furiosus* proteome. Among four step-eluted *Pyrococcus furiosus* fractions, CE-top-down MS/MS identified 71, 67, and 47 proteins and 144, 126, and 63 proteoforms in the 20% buffer B, 40% buffer B, and 60% buffer B fraction, respectively. These numbers represent a small portion of the detected protein species in each fraction (Table 1), and the identified proteins are the most abundant detected protein species based on a frequency calculation in the MS1 detected species (data not shown). A total of 291 proteoforms and 134 proteins were identified at a 2.7% proteoform level false-discovery rate (FDR) and a 6.0% protein level FDR with total sample consumption of 270 ng in 120 min of total analysis time. This accounts for 9% of 1517 proteins identified through the combined efforts of several bottom-up proteomics studies for this archaeon.[43] Additionally, the current top-down study also identified six proteins that were not found by the previous bottom-up studies.[43,44] Among those some are very small proteins being missed for some reason in the bottom-up measurements, and some have uncharacterized modifications that could be missed by routine bottom-up analysis. Such proteome coverage (approximately 10% of expressed proteins) is comparable to the largest top-down bacterial data sets[16,18,22] but with improved sensitivity.

Even though *Pyrococcus furiosus* is one of the most extensively studied hyperthermophilic archaea, very little experimentally verified PTM information is available. Bottom-up proteomics studies have identified 73% of the total ORFs, but none of them discussed PTMs in the proteome. There is only limited annotation in the Uniprot database for *Pyrococcus furiosus*, and most of the annotated PTM entries are based on homology as opposed to direct observation. Top-down interrogation of this proteome could provide more information about the occurrence of PTMs. The 134 unique *Pyrococcus furiosus* proteins identified in this study are represented by 291 proteoforms, 48% of which bear diverse PTMs or unexplainable modifications (Supplementary Table S1 in the Supporting Information). We found 43 proteins with N-terminal methionine excision, representing ∼32% of the proteins identified (Supplementary Table S1 in the Supporting Information). We also found six proteins with an N-terminal acetylation modification, representing ∼4% of the proteins identified. These numbers, which are similar to data from *Salmonella* and *Escherichia coli*, are different than what is found for eukaryotes in which the majority of proteins undergo N-terminal acetylation. A total of 12 proteins were found to contain disulfide bonds, and interestingly, proteins with an even number of cysteines all form disulfide bonds except for those proteins bearing unexplained modifications. This modification may make these proteins more thermostable or could be an outcome of growing at a temperature of 80 °C or higher. Another interesting finding is that many of proteins were present in oxidized proteoforms. Even though electrospray ionization can produce oxidation, given that a similar analysis of the Dam1 complex did not show any obvious oxidation, the oxidized *Pyrococcus furiosus* proteoforms may represent an

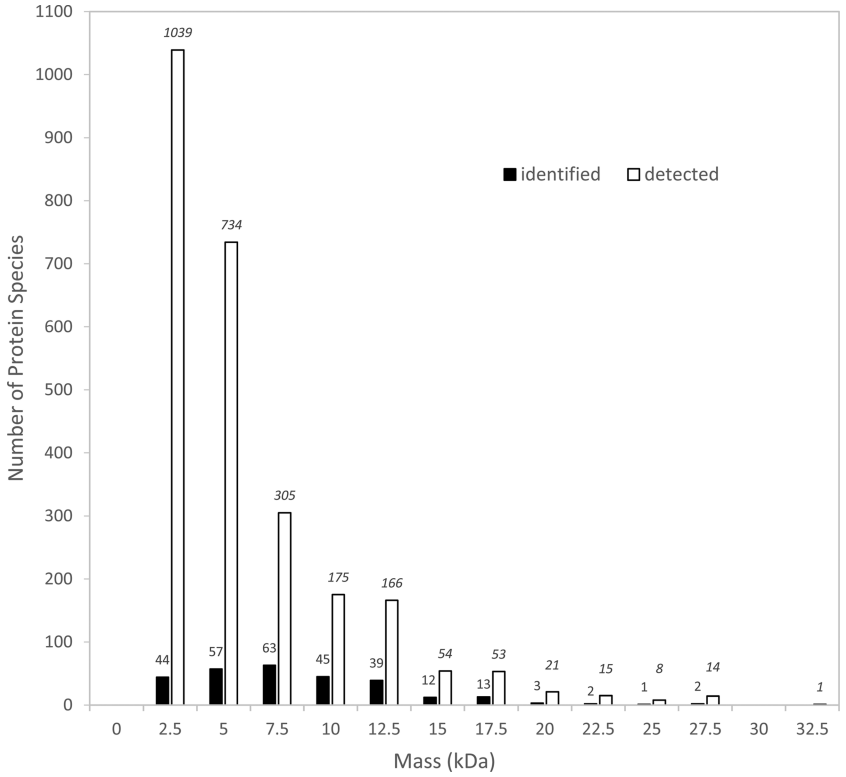



**Figure 2.** Molecular weight distribution for the detected (□) species in MS1 (Table 1) and unique *Pyrococcus furiosus* protein identifications (■) obtained through MS2 (Supplementary Table S1 in the Supporting Information) in CE-MS/MS experiments, combined from 20% buffer B, 40% buffer B, and 60% buffer B fractions.

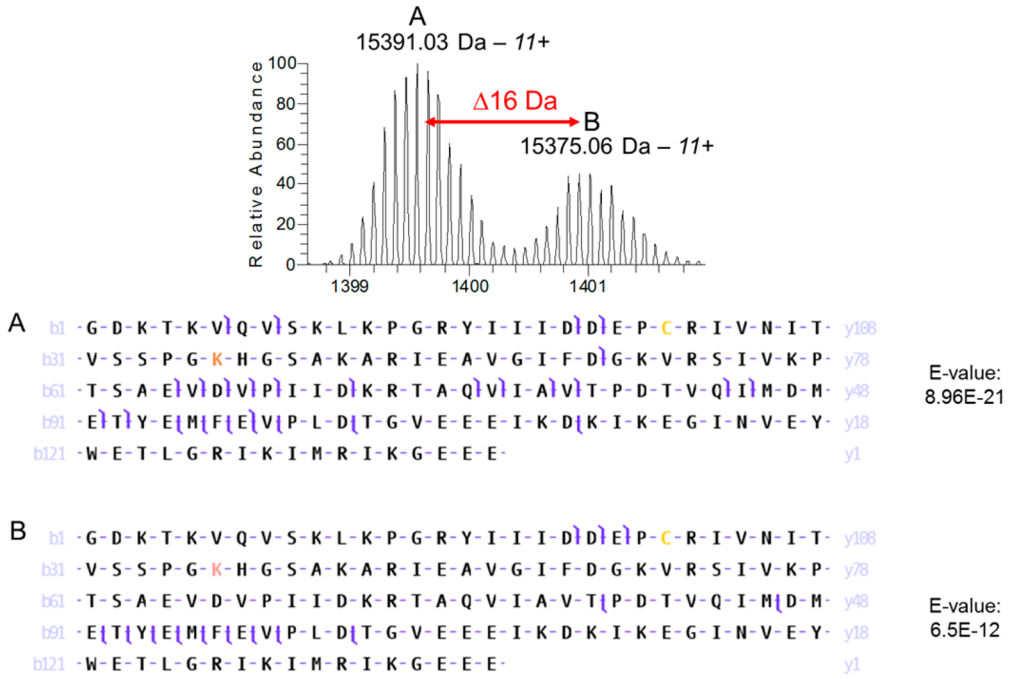



**Figure 3.** Characterization of hypusine/deoxyhypusine containing protein, translation initiation factor 5A (IF5A), in *Pyrococcus furiosus*. Graphical fragmentation maps generated with ProSightPC are shown for (A) deoxyhypusinated IF5A and (B) hypusinated IF5A.

endogenous modification. The observed oxidations occur not only on methionine but also on tryptophan and tyrosine. This may be an outcome of oxidative stress *in vivo* in the high temperature environment.

Our top-down data also confirmed the existence of a hypusine-containing protein translation initiation factor 5A, which has been inferred from homology. Hypusine is a nonstandard amino acid residue, found in all Eucarya, in a single protein—eukaryotic translation initiation factor 5A (eIF-5A). The hypusinated form of eIF5A is considered to be the active form, and to date most known functions of eIF5A (such as translational elongation, RNA binding, and protein—protein


dx.doi.org/10.1021/ac503439n | *Anal. Chem.* 2014, 86, 11006−11012

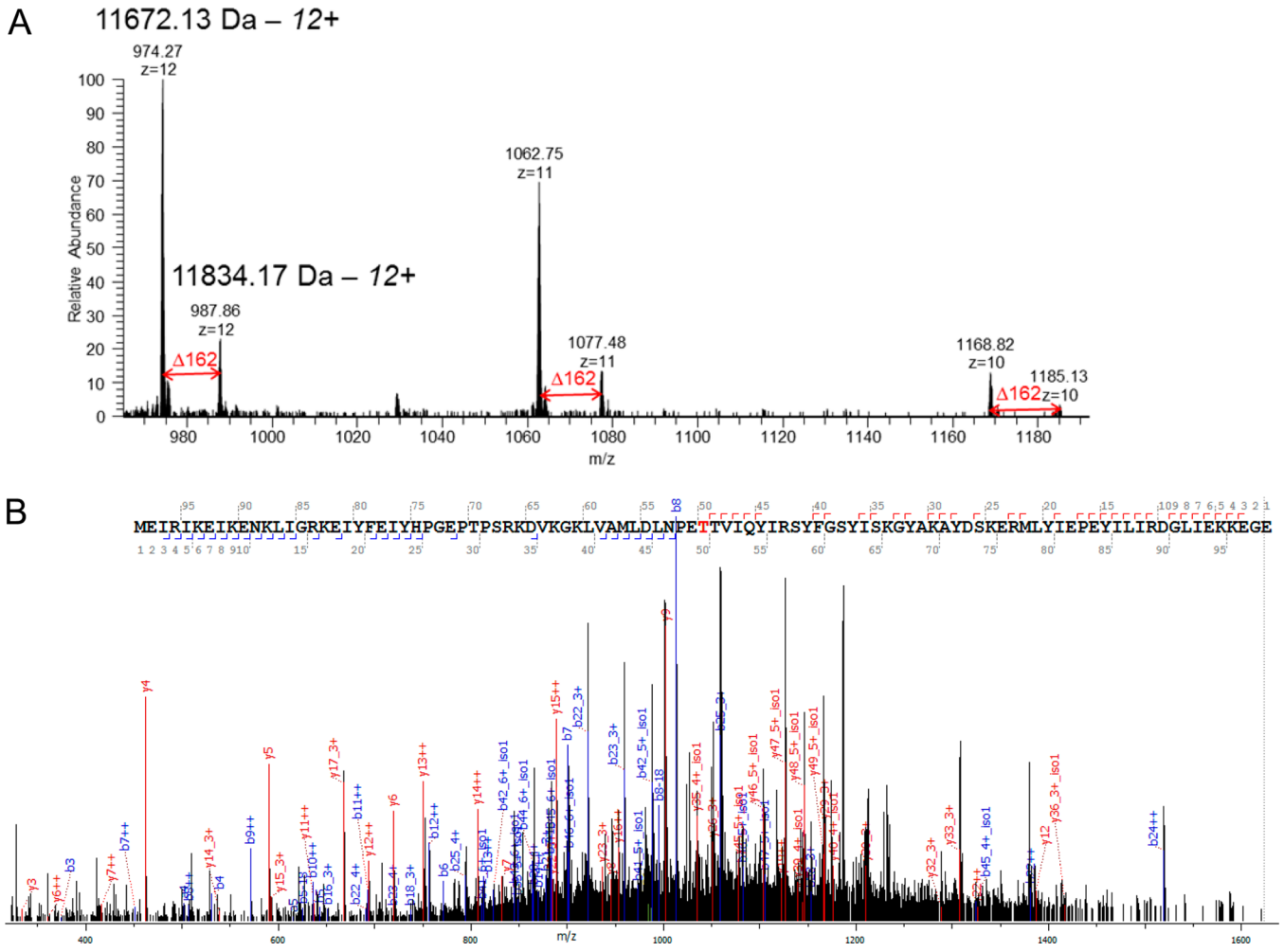



**Figure 4.** Characterization of hexose/maltose modified protein, Q8U442 (30S ribosomal protein S24e). (A) Orbitrap mass spectrum of unmodified and modified forms of Q8U442. (B) HCD fragmentation spectrum for precursor $987.86^{12+}$ and graphical fragmentation map generated with Byonic.

interactions) are wholly or partially dependent upon hypusination. Hypusine has also been shown to be present in Archaea (but not reported in *Pyrococcus furiosus*), where it is similarly found exclusively in aIF-5A, the archaeal homologue of eIF-5A. We also observed aIF-5A in the deoxyhypusinated form (+71 Da), its mature form, which was more prevalent than the hypusinated form in *Pyrococcus furiosus* (Figure 3). Although the modified site cannot be localized due to the limited fragmentation, K37 is assumed to be the modified residue based on similarity.

We found four proteins (P61882, Q8TZV1, Q8U3S9, and Q8U442, Supplementary Table S1 in the Supporting Information) harboring glycosylation, most likely O-linked glycosylation. The modification site could not be localized possibly due to its labile property, which renders it unstable during collision-induced dissociation (CID) and higher energy collisional dissociation (HCD) activation. Further investigation on an electron-transfer dissociation (ETD)-capable mass spectrometer might be able to localize this type of PTM. The observed +162 and +324 might indicate a single hexose and a double hexose or disaccharide such as maltose. This is the first time that glycosylation has been detected in proteomic analysis of *Pyrococcus furiosus*. Among these four proteins, two of them are ribosomal proteins, Q8U3S9 and Q8U442 (Figure 4). The glycosylation of ribosomal proteins could contribute, through the formation and the stabilization of the ribosomes, to activation of the translational machinery. We also found glycosylation on Histone A (P16882) and DNA/RNA-binding protein Alba (Q8TZV1). The identification of these glycosylated proteins is consistent with the fact that this organism can utilize a range of both simple and complex carbohydrates as its primary carbon source.

We also found another nonstandard amino acid residue $\alpha$-aminoadipic acid (identified as Lys+15) in Archaeal Histone A (P61882, Supplementary Table S1 in the Supporting Information). The $\alpha$-aminoadipic acid is an intermediate in the $\alpha$-aminoadipic acid (AAA) pathway (present in yeast and some thermophilic bacteria) for the synthesis of the amino acid L-lysine. The identification of this amino acid in *Pyrococcus furiosus* is direct evidence that lysine biosynthesis via the AAA pathway is also present in this Archaeon as suggested through gene cluster analysis on hyperthermophilic archaea *P. horikoshii* and *P. abyssi*.[45,46]

Our top-down analysis found that many proteins are present as truncated forms. For example, among 32 ribosomal proteins identified, 24 of them were present as full-length forms (11 of them also truncated forms), while 8 were present as only truncated forms. This might suggest metabolic stability of these ribosomal proteins. Other truncated proteins might be the result of signal peptide processing.


dx.doi.org/10.1021/ac503439n | *Anal. Chem.* 2014, 86, 11006−11012

Identified proteoforms from MS/MS represent only 10% of the detected protein species in MS for *Pyrococcus furiosus* proteome analysis (Table 1). This is partially due to the long duty cycle of the top-down experiments, given that all scans were collected in FT mode and multiple microscans were applied to increase the S/N. On top of that, the fragmentation efficiency for intact proteins is not always good enough to permit identification especially for proteins above 15 kDa. Additional fragmentation methods such as electron capture dissociation (ECD)/ETD[47,48] and UVPD[49] could improve the fragmentation efficiency.

The biggest challenge in top-down database searching lies in assigning combinatorial PTMs especially when the delta mass cannot be explained by expected modifications. Several software programs are designed to overcome this problem such as MS-Align+[50] through spectral alignment and ProSightPC using ΔM mode. However, the ΔM mode considers the delta mass as a single entity to the modified sequence and thus is less effective in localizing the modified sites even though the protein can be identified regardless. Since *Pyrococcus furiosus* PTMs are poorly annotated in the Uniprot database, ProSightPC failed to assign many of the PTMs in our data set. To further characterize proteins in this category, we applied the Byonic top-down search to our data. Byonic matches theoretical ions to observed peaks *in situ* without decharging or deisotoping of the tandem mass spectra. Even though it identified fewer proteins than ProSightPC, Byonic's identifications were easier to validate manually. Byonic's wildcard search (blind modification search) provided complementary information in assigning unexpected modifications and localizing these modifications (Supplementary Table S1 in the Supporting Information).

The limited peak capacity of CE separation of intact proteins could be improved through further prefractionation to increase the proteome coverage. In bottom-up proteomics, a one-dimensional peptide separation platform coupled to MS is not sufficient for comprehensive analysis of complex biological samples; therefore, a variety of multidimensional peptide separation platforms have been employed by different research groups. Similar to this, intact protein level separation could also benefit from the development of a multidimensional separation platform. In this study, RPLC prefractionation has been applied to first reduce the sample complexity followed by the high-efficiency CE separation for further MS detection. RPLC separates proteins according to the hydrophobicity while CE separation is based on mass and charge; therefore, these two separation mechanisms are orthogonal. As a proof of concept, only 4 RPLC fractions were collected for CE-MS analysis; however, with even more first dimension prefractions the overall peak capacity could be easily improved (number of prefractions × CE peak capacity).

In conclusion, the sheathless CESI-top down MS provided an ultrasensitive, fast approach to tackle the challenge of the front-end separation for top-down proteomics. It is very suitable to analyze proteomes that are moderately complex with very low sample consumption. Combined with other orthogonal separation techniques such as RPLC, this could be applied to more complex biological samples to be analyzed through top-down interrogation.

## ■ ASSOCIATED CONTENT

### Supporting Information

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

## ■ AUTHOR INFORMATION


### Corresponding Author

*E-mail: jyates@scripps.edu. Phone: 858-784-8862. Fax: 858-784-8883.


### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS


This work was funded by grants from National Heart, Lung, and Blood Institute Proteomics Center Award HHSN268201000035C, R01 MH067880, and 8P41GM103533-17 to J.R.Y.,III, and NIH Grant R43 GM103352 to M.B. M.L.-A. holds a postdoctoral fellowship from FRQNT. We would also like to thank all lab members for thoughtful suggestions and comments.


## ■ REFERENCES


(1) Chait, B. T. *Science* **2006**, *314*, 65−66.
(2) Kelleher, N. L. *Anal. Chem.* **2004**, *76*, 197A−203A.
(3) Giometti, C. S.; Tollaksen, S. L.; Mukund, S.; Zhi Hao, Z.; Keren, M.; Xuhong, M.; Adams, M. W. W. *J. Chromatogr., A* **1995**, *698*, 341−349.
(4) Lim, H.; Eng, J.; Yates, J. R., 3rd; Tollaksen, S. L.; Giometti, C. S.; Holden, J. F.; Adams, M. W.; Reich, C. I.; Olsen, G. J.; Hays, L. G. *J. Am. Soc. Mass Spectrom.* **2003**, *14*, 957−970.
(5) Roepstorff, P. *Protein Cell* **2012**, *3*, 641−647.
(6) Kelleher, N. L.; Lin, H. Y.; Valaskovic, G. A.; Aaserud, D. J.; Fridriksson, E. K.; McLafferty, F. W. *J. Am. Chem. Soc.* **1999**, *121*, 806−812.
(7) Ge, Y.; Lawhorn, B. G.; ElNaggar, M.; Strauss, E.; Park, J.-H.; Begley, T. P.; McLafferty, F. W. *J. Am. Chem. Soc.* **2002**, *124*, 672−678.
(8) Han, X.; Jin, M.; Breuker, K.; McLafferty, F. W. *Science* **2006**, *314*, 109−112.
(9) Ahlf, D. R.; Compton, P. D.; Tran, J. C.; Early, B. P.; Thomas, P. M.; Kelleher, N. L. *J. Proteome Res.* **2012**, *11*, 4308−4314.
(10) Hu, Q.; Noll, R. J.; Li, H.; Makarov, A.; Hardman, M.; Graham Cooks, R. *J. Mass Spectrom.* **2005**, *40*, 430−443.
(11) Vellaichamy, A.; Tran, J. C.; Catherman, A. D.; Lee, J. E.; Kellie, J. F.; Sweet, S. M.; Zamdborg, L.; Thomas, P. M.; Ahlf, D. R.; Durbin, K. R.; Valaskovic, G. A.; Kelleher, N. L. *Anal. Chem.* **2010**, *82*, 1234−1244.
(12) Lee, J. E.; Kellie, J. F.; Tran, J. C.; Tipton, J. D.; Catherman, A. D.; Thomas, H. M.; Ahlf, D. R.; Durbin, K. R.; Vellaichamy, A.; Ntai, I.; Marshall, A. G.; Kelleher, N. L. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 2183−2191.
(13) Meng, F.; Cargile, B. J.; Patrie, S. M.; Johnson, J. R.; McLoughlin, S. M.; Kelleher, N. L. *Anal. Chem.* **2002**, *74*, 2923−2929.
(14) LeDuc, R. D.; Taylor, G. K.; Kim, Y.-B.; Januszyk, T. E.; Bynum, L. H.; Sola, J. V.; Garavelli, J. S.; Kelleher, N. L. *Nucleic Acids Res.* **2004**, *32*, W340−W345.
(15) Frank, A. M.; Pesavento, J. J.; Mizzen, C. A.; Kelleher, N. L.; Pevzner, P. A. *Anal. Chem.* **2008**, *80*, 2499−2505.
(16) Tsai, Y.; Scherl, A.; Shaw, J.; MacKay, C. L.; Shaffer, S.; Langridge-Smith, P. R.; Goodlett, D. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 2154−2166.
(17) Zamdborg, L.; LeDuc, R. D.; Glowacz, K. J.; Kim, Y. B.; Viswanathan, V.; Spaulding, I. T.; Early, B. P.; Bluhm, E. J.; Babai, S.; Kelleher, N. L. *Nucleic Acids Res.* **2007**, *35*, W701−706.
(18) Bunger, M. K.; Cargile, B. J.; Ngunjiri, A.; Bundy, J. L.; Stephenson, J. L., Jr. *Anal. Chem.* **2008**, *80*, 1459−1467.
(19) Kellie, J. F.; Catherman, A. D.; Durbin, K. R.; Tran, J. C.; Tipton, J. D.; Norris, J. L.; Witkowski, C. E., 2nd; Thomas, P. M.; Kelleher, N. L. *Anal. Chem.* **2012**, *84*, 209−215.
(20) Roth, M. J.; Parks, B. A.; Ferguson, J. T.; Boyne, M. T., 2nd; Kelleher, N. L. *Anal. Chem.* **2008**, *80*, 2857−2866.
(21) Tran, J. C.; Zamdborg, L.; Ahlf, D. R.; Lee, J. E.; Catherman, A. D.; Durbin, K. R.; Tipton, J. D.; Vellaichamy, A.; Kellie, J. F.; Li, M.;

Wu, C.; Sweet, S. M.; Early, B. P.; Siuti, N.; LeDuc, R. D.; Compton, P. D.; Thomas, P. M.; Kelleher, N. L. *Nature* **2011**, *480*, 254−258.
(22) Ansong, C.; Wu, S.; Meng, D.; Liu, X.; Brewer, H. M.; Deatherage Kaiser, B. L.; Nakayasu, E. S.; Cort, J. R.; Pevzner, P.; Smith, R. D.; Heffron, F.; Adkins, J. N.; Pasa-Tolic, L. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 10153−10158.
(23) Chen, X.; Ge, Y. *Proteomics* **2013**, *13*, 2563−2566.
(24) Haselberg, R.; de Jong, G. J.; Somsen, G. W. *Electrophoresis* **2011**, *32*, 66−82.
(25) Haselberg, R.; de Jong, G. J.; Somsen, G. W. *Electrophoresis* **2013**, *34*, 99−112.
(26) Haselberg, R.; de Jong, G. J.; Somsen, G. W. *Anal. Chem.* **2013**, *85*, 2289−2296.
(27) Sun, L.; Knierman, M. D.; Zhu, G.; Dovichi, N. J. *Anal. Chem.* **2013**, *85*, 5989−5995.
(28) Jensen, P. K.; Paša-Tolić, L.; Anderson, G. A.; Horner, J. A.; Lipton, M. S.; Bruce, J. E.; Smith, R. D. *Anal. Chem.* **1999**, *71*, 2076−2084.
(29) Smith, R. D.; Barinaga, C. J.; Udseth, H. R. *Anal. Chem.* **1988**, *60*, 1948−1952.
(30) Maxwell, E. J.; Zhong, X.; Chen, D. D. Y. *Anal. Chem.* **2010**, *82*, 8377−8381.
(31) Wojcik, R.; Dada, O. O.; Sadilek, M.; Dovichi, N. J. *Rapid Commun. Mass Spectrom.* **2010**, *24*, 2554−2560.
(32) Zhong, X.; Maxwell, E. J.; Chen, D. D. Y. *Anal. Chem.* **2011**, *83*, 4916−4923.
(33) Zhao, Y.; Sun, L.; Champion, M. M.; Knierman, M. D.; Dovichi, N. J. *Anal. Chem.* **2014**, *86*, 4873−4878.
(34) Li, Y.; Compton, P. D.; Tran, J. C.; Ntai, I.; Kelleher, N. L. *Proteomics* **2014**, *14*, 1158−1164.
(35) Moini, M. *Anal. Chem.* **2007**, *79*, 4241−4246.
(36) Moini, M.; Huang, H. *Electrophoresis* **2004**, *25*, 1981−1987.
(37) Busnel, J. M.; Schoenmaker, B.; Ramautar, R.; Carrasco-Pancorbo, A.; Ratnayake, C.; Feitelson, J. S.; Chapman, J. D.; Deelder, A. M.; Mayboroda, O. A. *Anal. Chem.* **2010**, *82*, 9476−9483.
(38) Ramautar, R.; Busnel, J. M.; Deelder, A. M.; Mayboroda, O. A. *Anal. Chem.* **2012**, *84*, 885−892.
(39) Faserl, K.; Sarg, B.; Kremser, L.; Lindner, H. *Anal. Chem.* **2011**, *83*, 7297−7305.
(40) Wang, Y.; Fonslow, B. R.; Wong, C. C.; Nakorchevsky, A.; Yates, J. R., 3rd. *Anal. Chem.* **2012**, *84*, 8505−8513.
(41) Heemskerk, A. A.; Wuhrer, M.; Busnel, J. M.; Koeleman, C. A.; Selman, M. H.; Vidarsson, G.; Kapur, R.; Schoenmaker, B.; Derks, R. J.; Deelder, A. M.; Mayboroda, O. A. *Electrophoresis* **2013**, *34*, 383−387.
(42) Haselberg, R.; de Jong, G. J.; Somsen, G. W. *Anal. Chim. Acta* **2010**, *678*, 128−134.
(43) Wong, C. C.; Cociorva, D.; Miller, C. A.; Schmidt, A.; Monell, C.; Aebersold, R.; Yates, J. R., 3rd. *J. Proteome Res.* **2013**, *12*, 763−770.
(44) Lee, A. M.; Sevinsky, J. R.; Bundy, J. L.; Grunden, A. M.; Stephenson, J. L. *J. Proteome Res.* **2009**, *8*, 3844−3851.
(45) Nishida, H.; Nishiyama, M.; Kobashi, N.; Kosuge, T.; Hoshino, T.; Yamane, H. *Genome Res.* **1999**, *9*, 1175−1183.
(46) Cohen, G. N.; Barbe, V.; Flament, D.; Galperin, M.; Heilig, R.; Lecompte, O.; Poch, O.; Prieur, D.; Querellou, J.; Ripp, R.; Thierry, J. C.; Van der Oost, J.; Weissenbach, J.; Zivanovic, Y.; Forterre, P. *Mol. Microbiol.* **2003**, *47*, 1495−1512.
(47) Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W. *J. Am. Chem. Soc.* **1998**, *120*, 3265−3266.
(48) 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.
(49) 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.
(50) Liu, X.; Sirotkin, Y.; Shen, Y.; Anderson, G.; Tsai, Y. S.; Ting, Y. S.; Goodlett, D. R.; Smith, R. D.; Bafna, V.; Pevzner, P. A. *Mol. Cell. Proteomics* **2012**, *11*, M111.008524.