David Hage Publications

Published Research - Department of Chemistry

6-2016

Optimizing Sequence Coverage for a Moderate Mass Protein in Nano-Electrospray Ionization Quadrupole Time-of-Flight Mass Spectrometry

Ryan E. Matsuda
University of Nebraska-Lincoln, rmatsuda@huskers.unl.edu

Venkata Kolli
University of Nebraska-Lincoln, venkata.kolli@huskers.unl.edu

Megan Woods
University of Nebraska-Lincoln

Eric D. Dodds
University of Nebraska-Lincoln, edodds2@unl.edu

David S. Hage
University of Nebraska - Lincoln, dhage1@unl.edu

Follow this and additional works at: http://digitalcommons.unl.edu/chemistryhage

Part of the Medicinal-Pharmaceutical Chemistry Commons

Accepted Manuscript

Optimizing Sequence Coverage for a Moderate Mass Protein in Nano-Electrospray Ionization Quadrupole Time-of-Flight Mass Spectrometry

Ryan Matsuda, Venkata Kolli, Megan Woods, Eric D. Dodds, David S. Hage

Optimizing Sequence Coverage for a Moderate Mass Protein in Nano-Electrospray Ionization Quadrupole Time-of-Flight Mass Spectrometry
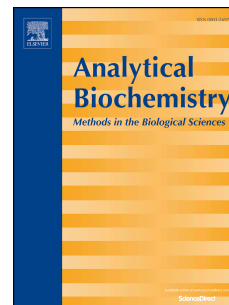
Ryan Matsuda, Venkata Kolli, Megan Woods, Eric D. Dodds, and David S. Hage*

Department of Chemistry

University of Nebraska-Lincoln

Lincoln, NE 68588-0304 (USA)

*Author for correspondence: 704 Hamilton Hall, Chemistry Department, University of Nebraska, Lincoln, NE 68588-0304 USA. Phone: 402-472-2744; FAX: 402-472-9402; Email: dhage1@unl.edu

Abstract

Sample pretreatment was optimized to obtain high sequence coverage for human serum albumin (HSA, 66.5 kDa) when using nano-electrospray ionization quadrupole time-of-flight mass spectrometry (nESI-Q-TOF-MS). Use of the final method with trypsin, Lys-C and Glu-C digests gave a combined coverage of 98.8%. The addition of peptide fractionation resulted in 99.7% coverage. These results were comparable to those obtained previously with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS). The sample pretreatment/nESI-Q-TOF-MS method was also used with collision-induced dissociation to analyze HSA digests and to identify peptides that could be employed as internal mass calibrants in future studies of modifications to HSA.

Mass spectrometry (MS) is an important tool for identifying proteins and providing information on the primary sequences of proteins and their post-translational modifications (PTMs) [1,2]. Several recent studies have used MS to examine human serum albumin (HSA, 66.5 kDa) and PTMs such as non-enzymatic glycation that occur on this protein [3-5]. The effects of such modifications have been of concern given the important role played by HSA in transporting many low mass substances in blood (e.g., hormones, fatty acids, and drugs) [6].

Several past studies have utilized matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS) to investigate the structure of HSA and glycated HSA [2,3,7-12]. Sequence coverages of up to 97.4% have been reached by optimizing the MALDI matrix and conditions that were used for sample pretreatment; however, obtaining this level of coverage required multiple enzyme digestions and a series of steps for peptide fractionation [2]. The purpose of this current study was to determine if more convenient pretreatment conditions could still obtain high sequence coverage for a moderately high mass protein such as HSA and when using nano-electrospray quadrupole time-of-flight mass spectrometry (nESI-Q-TOF-MS) for analysis of the digests.

Two pretreatment procedures were initially considered (see Supplementary Material for more details). Pretreatment Method 1 followed a scheme similar to the previous method used with MALDI-TOF-MS [2] but without any peptide fractionation. In this method, a sample of HSA was first treated by means of alkylation and reduction, followed by dialysis to remove any excess reagents. The protein sample was then digested separately with three proteolytic enzymes (i.e., trypsin, Lys-C and Glu-C) and analyzed by MS. In Pretreatment Method 2, all of the pretreatment and digestion processes were performed sequentially in a single vial, resulting in the ability to use lower concentrations of both the protein and pretreatment reagents. The

pretreatment conditions for alkylation and reduction were similar to those in Pretreatment Method 1, with some slight variations in the reaction times and temperatures. Also in Pretreatment Method 2, dialysis was no longer used to remove the excess pretreatment reagents, and each digest was fractionated by using ZipTip$_\mu$-C$_{18}$ pipette tips.

Figure 1 and the Supplementary Material summarize the results that were obtained by nESI-Q-TOF-MS for digests of HSA when using these pretreatment methods. The use of Pretreatment Method 1 with a tryptic digest gave 42 identified peptides and a sequence coverage of 36.9% for HSA, while the Glu-C or Lys-C digests gave sequence coverages of 39.1% or 32.1% and 43 or 21 identified peptides, respectively. For Pretreatment Method 2, the sequence coverages were 86.0% for the tryptic digest, 84.1% for Lys-C digest and 78.2% for Glu-C digest, which corresponded to 104, 52, or 54 identified peptides in these digests. Because there was both improved sequence coverage and an increase in the number of identified peptides with Pretreatment Method 2, this method was selected for use in the remainder of this study.

The results obtained by Pretreatment Method 2, nESI-Q-TOF-MS, and using a single elution step for peptide fractionation were compared to the results from the previous pretreatment method that had been used with MALDI-TOF-MS to examine HSA (see Supplementary Material) [2]. The previous MALDI-TOF-MS method gave sequence coverages of 82.9%, 77.6% or 64.6% and 56, 41, or 28 identified peptides when HSA was digested with trypsin, Lys-C or Glu-C, respectively, and these digests were fractionated on ZipTip$_\mu$-C$_{18}$ tips through a series of several elution steps [2]. The approach based on Pretreatment Method 2 and nESI-Q-TOF-MS, which used a simpler alkylation/reduction protocol and only a single elution step for peptide fractionation, gave 3.1-13.6% higher sequence coverages for these digests and a greater number of identifiable peaks (i.e., 11-48 more peptides per digest). The use of multiple

4

elution steps for peptide fractionation, as described in the Supplementary Material, gave a small increase in the sequence coverages for Pretreatment Method 2 and nESI-Q-TOF-MS to levels of 83.2% and 91.5% for tryptic and Glu-C digests and a slight decrease to 72.6% for the Lys-C digest.

When a single elution step was used for peptide fractionation in Pretreatment Method 2, the total sequence coverage obtained by nESI-Q-TOF-MS for HSA was 98.8%. This coverage included 578 out of the 585 amino acids in HSA, with the only excluded region being residues 535-541. This coverage included 56 out of 59 lysines in HSA and all of the arginine residues, which were of interest because lysine and arginine are both potential sites for glycation-related modifications (Note: the three lysines in residues 535-541 that were not included in this coverage are not major sites for such modifications) [3]. When multiple elution steps were used for peptide fractionation, the combined digests gave an overall sequence coverage was 99.7%, in which a total of 583 out of 585 residues were now covered. The only missing residues in this second case came from residues 312-313; however, previous reports have shown that K313 is highly susceptible to modification by glycation [3].

Both of these situations were an improvement over the previous pretreatment method used with MALDI-TOF-MS, in which a total sequence coverage of 97.4% was obtained for HSA when data were combined for the three enzyme digests [2]. In this prior work, 570 out of the 585 amino acids were included, as well as 54 of the 59 lysines and all of the arginines. In the missed regions for this previous method, K4, K536, K538, and K541 are not major sites for glycation-related modifications, but K317 can be an important location for such PTMs [3]. Neither the previous MALDI-TOF-MS method nor the use of Pretreatment Method 2 with nESI-Q-TOF-MS and a single elution step for fractionation covered the region 535-541 of HSA.

However, if the data were combined for nESI-Q-TOF-MS and Pretreatment Method 2 when using both a single step and multiple steps for peptide fractionation, then all 585 amino acids of HSA included and effectively 100% coverage was achieved.

In the previous work with HSA and MALDI-TOF-MS [2], 0.5 µL portions of the fractionated protein digests (protein concentrations: ~0.822 µg/µL in tryptic digest, ~0.870 µg/µL in Lys-C digest, and ~ 1.25 µg/µL in Glu-C digest) were individually mixed with the MALDI matrix and spotted onto a plate for MS analysis. These multiple fractions increased the overall amount of sample that was needed for analysis. Although a 20 µL sample of the purified protein digest (protein concentration: ~ 1 µg/µL in each digest) was prepared for use in Pretreatment Method 2 and nESI-Q-TOF-MS, the sample could be placed directly into the nano-ESI emitter and sprayed in nanoliter volumes to obtain a mass spectrum. In addition, the remaining sample from the nano-ESI emitter could be recovered and stored for future experiments.

Another advantage of the nESI-Q-TOF-MS method over MALDI-TOF-MS was its ability to perform MS/MS experiments on the digests, such as by using CID to create peptide fragmentation [13]. This type of experiment was performed on the various digests of HSA to confirm the identity of the peptides, as demonstrated in Figure 2(a). Some of these peptides were also tested for use as internal *m/z* calibrants to improve mass accuracy when searching for PTMs [14,15] like glycation-related modifications [2,3,7-10], as illustrated in Figure 2(b). Around 6-7 unique peptides were selected through these experiments from each type of enzyme digest of HSA, with each peptide having a relatively high signal intensity (see Supplementary Material). One of these peptides is shown in Figure 2(a) and corresponded to residues 337-348 and RHPDYSVVLLLR; this peptide was selected because it contained no lysine residues and would

6

not be subject to PTMs at such sites. Other peptides were chosen because they appeared in a variety of regions in the mass spectra.

These peptides were next used as internal calibrants in the analysis of a separate set of HSA digests that were prepared by using trypsin, Lys-C, or Glu-C. A calibrated spectrum of a Glu-C digest of HSA resulted in a root mean square mass error of 19.9 ppm, with similar results being obtained with the trypsin and Lys-C digests. This low mass measurement error made it possible to lower the mass tolerance that was used for sequence analysis (i.e., a previous value of 50 ppm); this change had the advantage of reducing the possibility of false positive matches while also improving confidence in the assignment of putative peptides, including those harboring PTMs. The effect of this change on the sequence coverage was evaluated by using the internally calibrated mass spectra for the tryptic, Lys-C and Glu-C digests along with a 25 ppm mass tolerance. Under these conditions, the individual digests had sequence coverages of 79.8%, 86.5%, and 78.9% for tryptic, Lys-C and Glu-C digests, with a 96.1% total coverage for the combined results. These results were comparable to the previous values acquired by MALDI-TOF-MS [2] and were only slightly lower than the coverages seen earlier with nESI-Q-TOF-MS when using a 50 ppm mass tolerance and no internal calibration.

In summary, a sample pretreatment was developed for use with nESI-Q-TOF-MS to obtain high sequence coverage for HSA. The combined use of nESI-Q-TOF-MS and the final pretreatment method with single or multiple elution steps for peptide fractionation gave sequence coverages of 98.8-99.7% for HSA, with a combined coverage of essentially 100%. It was also shown how this combined approach could be used with CID to confirm the structures of the peptides and to select some of these peptides for use as internal calibrants in MS/MS experiments. Although this study focused on the analysis of HSA by nESI-Q-TOF-MS, the same

methods could be applied to other proteins with moderately high masses.

References

1. R. Aebersold, M. Mann, Mass spectrometry-based proteomics, Nature 422 (2003) 198-207.

2. C. Wa, R. Cerny, D.S. Hage, Obtaining high sequence coverage in matrix-assisted laser desorption time-of-flight mass spectrometry for studies of protein modification: analysis of human serum albumin as a model. Anal. Biochem. 349 (2006) 229-241.

3. J. Anguizola, R. Matsuda, O.S. Barnaby, K.S. Joseph, C. Wa, E. Debolt, M. Koke, D.S. Hage, Review: glycation of human serum albumin. Clin. Chim. Acta 425 (2013) 64–76.

4. A. Lapolla, D. Fedele, R. Seraglia, P. Traldi, The role of mass spectrometry in the study of non-enzymatic protein glycation in diabetes: an update, Mass Spectrom. Rev. 25 (2006) 775-797.

5. Q. Zhang, J.M. Ames, R.D. Smith, J.W. Baynes, T.O. Metz, A perspective on the Maillard reaction and the analysis of protein glycation by mass spectrometry: probing the pathogenesis of chronic disease, J. Proteome Res. 8 (2009) 754-769.

6. T. Peters, Jr., All About Albumin: Biochemistry, Genetics, and Medical Applications. Academic Press, San Diego, 1996.

7. O.S. Barnaby, R.L. Cerny, W. Clarke, D.S. Hage, Comparison of modification sites

formed on human serum albumin at various stages of glycation. Clin. Chim. Acta 412 (2011) 277-285.

8. O.S. Barnaby, R.L. Cerny, W. Clarke, D.S. Hage, Quantitative analysis of glycation patterns in human serum albumin using $^{16}O/^{18}O$-labeling and MALDI-TOF MS. Clin. Chim. Acta 412 (2011) 1606-1615.

9. O.S. Barnaby, C. Wa, R.L. Cerny, W. Clarke, D.S. Hage, Quantitative analysis of glycation sites on human serum albumin using $^{16}O/^{18}O$-labeling and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry, Clin. Chim. Acta 411 (2010) 1102-1110.

10. C. Wa, R. Cerny, W. Clarke, D.S. Hage, Characterization of glycation adducts on human serum albumin by matrix-assisted laser desorption ionization time-of-flight mass spectrometry, Clin. Chim. Acta 385 (2007) 48-60.

11. F.L. Brancia, J.Z. Bereszczak, A. Lapolla, D. Fedele, L. Baccarin, R. Seraglia, P. Traldi, Comprehensive analysis of glycated human serum albumin tryptic peptides by off-line liquid chromatography followed by MALDI analysis on a time-of-flight/curved field reflectron tandem mass spectrometer, J. Mass Spectrom. 41 (2006) 1179–1185.

12. B.S. Lee, S. Krishnanchettiar, S.S. Lateef, S. Gupta, Analyses of the in vitro nonenzymatic glycation of peptides/proteins by matrix-assisted laser desorption/ionization mass spectrometry, Int. J. Mass Spectrom. 260 (2007) 67–74.

13. J.H. Gross, Mass Spectrometry, A Textbook, 2nd ed., Springer, New York, 2004.

14. K. Hjerno, P Hojrup, in R. Mattiesen (Ed.), Mass Spectrometry Data Analysis in Proteomics, Humana Press, Totowa, 2007, Chap. 3.

15. B. Xiang, M. Prado, An accurate and clean calibration method for MALDI-MS, J.
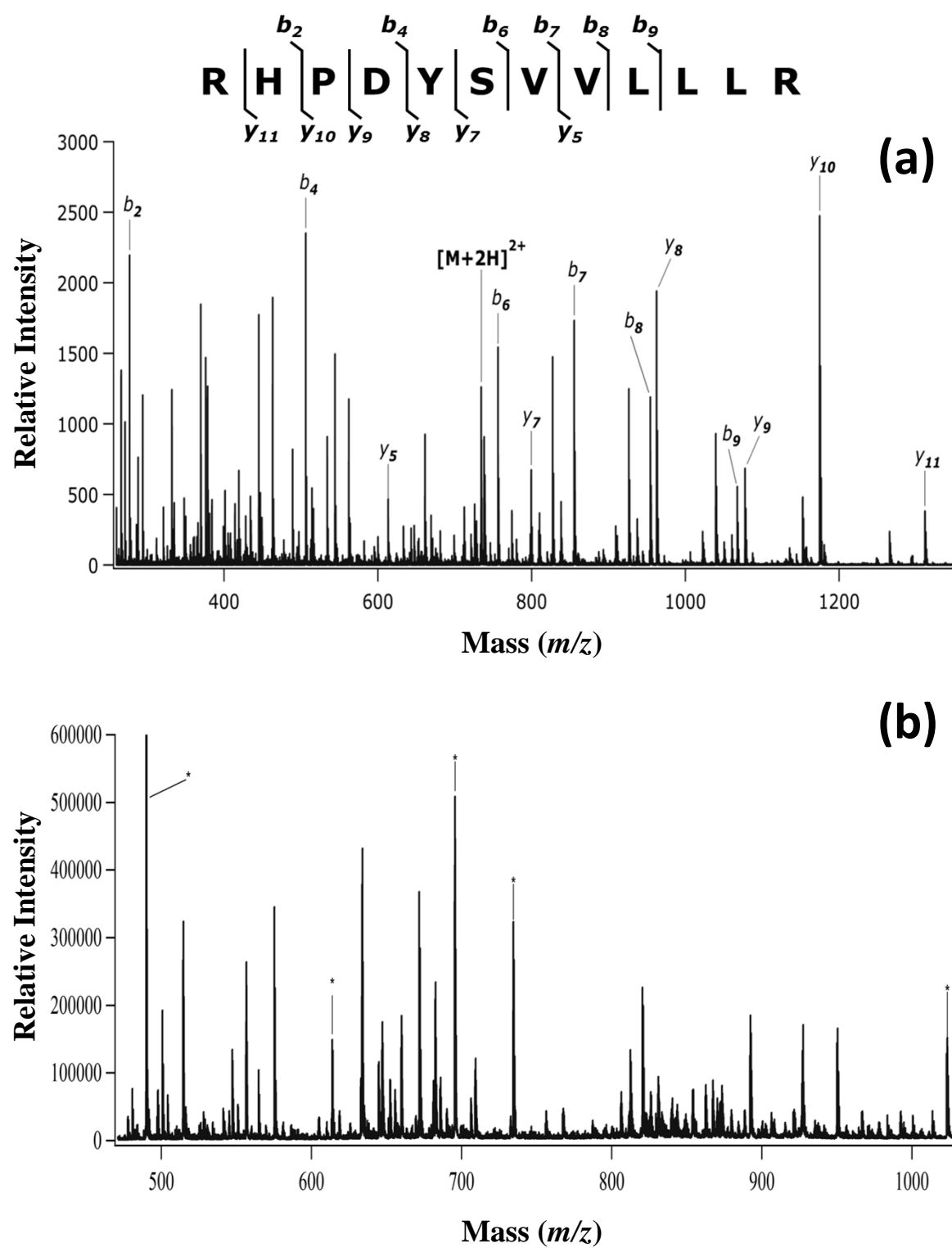
Biomol. Tech. 21 (2010) 116-119.

Figure Legends

Figure 1. Primary sequence of human serum albumin (HSA) [2] and total sequence coverage of HSA that was obtained with Pretreatment Method 2 and nESI-Q-TOF-MS when combining the sequence coverages for tryptic (shown in underline), Lys-C (in bold), and Glu-C (in italics) digests. These results were obtained using a tolerance level that was equal to a maximum 50.0 ppm mass difference for each positive match.

Figure 2. (a) MS/MS spectrum for the fragmentation of a peptide with a mass-to-charge ratio of 734 *m/z*; this peptide corresponded to the 373-348 region of HSA, which has a sequence of RHPDYSVVLLLR. (b) Mass spectrum of a trypsin-digested HSA sample, in which peaks that were selected for internal calibration are indicated by an asterisk (*).

11

Residue No.	Sequence			
1-40	DAHKSEVAHR	FKDLGEENFK	ALVLIAFAQY	LQQCPFEDHV
41-80	KLVNEVTEFA	KTCVADESAE	NCDKSLHTLF	GDKLCTVATL
81-120	RETYGEMADC	CAKQEPERNE	CFLQHKDDNP	NLPRLVRPEV
121-160	DVMCTAFHDN	EETFLKKYLY	EIARRHPYFY	APELLFFAKR
161-200	YKAAFTECCQ	AADKAACLLP	KLDELRDEGK	ASSAKQRLKC
201-240	ASLQKFGERA	FKAWAVARLS	QRFPKAEFAE	VSKLVTDLTK
241-280	VHTECCHGDL	LECADDRADL	AKYICENQDS	ISSKLKECCE
281-320	KPLLEKSHCI	AEVENDEMPA	DLPSLAADFV	ESKDVCKNYA
321-360	EAKDVFLGMF	LYEYARRHPD	YSVVLLLRLA	KTYETTLEKC
361-400	CAAADPHECY	AKVFDEFKPL	VEEPQNLIKQ	NCELFEQLGE
401-440	YKFQNALLVR	YTKKVPQVST	PTLVEVSRNL	GKVGSKCCKH
441-480	PEAKRMPCAE	DYLSVVLNQL	CVLHEKTPVS	DRVTKCCTES
481-520	LVNRRPCFSA	LEVDETYVPK	EFNAETFTFH	ADICTLSEKE
521-560	RQIKKQTALV	ELVKHKPKAT	KEQLKAVMDD	FAAFVEKCCK
561-585	ADDKETCFAE	EGKKLVASSQ	AALGL	

Figure 1

Figure 2