RAPID COMMUNICATIONS IN MASS SPECTROMETRY, VOL. 7, 58–62 (1993)

The Use of Hydrogen-Deuterium Exchange to Facilitate Peptide Sequencing by Electrospray Tandem Mass Spectrometry

Nikolai F. Sepetov, Olga L. Issakova and Michal Lebl
The Selectide Corporation, Tucson, AZ 85737, USA

Kristine Swiderek, Douglas C. Stahl and Terry D. Lee*
Division of Immunology, Beckman Research Institute of the City of Hope, 1450 E. Duarte Road, Duarte, CA 91010, USA

SPONSOR REFEREE: Dr Marshall M. Siegel, American Cyanamid Company, Pearl River, NY 10965, USA

The utility of hydrogen-deuterium exchange for sequencing peptides by mass spectrometry is demonstrated. The number of exchangeable hydrogens in a peptide is readily obtained by electrospray analysis of the peptide dissolved in deuterated solvents. This information can be used, in conjunction with published computer algorithms for interpreting peptide mass spectra, to reduce significantly the number of candidate sequences that fit the experimental data. This information, when combined with fragment-ion information in the mass spectrum, greatly increases the reliability of sequence determination.

Mass spectrometry has slowly developed into a powerful and sensitive tool for peptide sequencing which can be used both as an alternative and a complement to automated Edman degradation.[1–3] Generally, tandem mass spectrometry (often referred to as MS/MS) is used to generate fragment-ion spectra from collision-induced dissociation (CID) of intact peptide ions created by either fast-atom bombardment (FAB) or electrospray (ES) ionization techniques. While peptide spectra can be obtained in a matter of minutes, the time needed to interpret the spectrum is often an order of magnitude longer, and requires a high level of skill and training. A number of computer programs have been written with the goal of improving both the speed and accuracy of interpreting peptide mass spectral data.[4–13] The output from such programs is a list of possible sequences ranked by score. In those instances when the score of one sequence is significantly better than the others, there is some confidence in the result. However, all too often there are a number of possible sequences with similar scores and the result is much less certain.

Various derivatization procedures have been used to facilitate sample analyses and interpretation of the spectra. Before the advent of ionization methods compatible with polar, thermally labile molecules, acetylation of amine functionalities and methylation of carboxylic acid groups was necessary to obtain spectra.[14] By forming a mixture of deuterated and non-deuterated acetyl derivatives, it is possible to distinguishing N-terminal from C-terminal ion series. Derivatization of amino and carboxylic acid functionality is not necessary to obtain spectra, using techniques like FAB or ES, but is still done to improve sensitivity or resolve ambiguities in the interpretation.[15–18] The formation of 2-trimethylammonium acetyl N-terminal derivatives has been used to direct charge distributions that favor fragmentation processes that permit the differentiation of leucine and isoleucine in high energy CID spectra.[19] ^{18}O-labelling of carboxyl oxygens has been used to identify the C-terminal amino acid residue and probe mechanisms of peptide fragmentation in the mass spectrometer.[20,21]

In this paper, we describe a new approach to facilitate sequence analysis of peptides by mass spectrometry that utilizes information about the number of labile hydrogens (derived from —OH, —NH—, —NH_2, —SH and —COOH groups) in the peptide. Hydrogen-deuterium exchange is an established technique for the characterization of low molecular weight organic molecules by mass spectrometry.[22] Katta and Chait have used hydrogen-deuterium exchange and ES mass spectrometry to determine the number of labile protons in proteins.[23] To our knowledge, the use of hydrogen-deuterium exchange for peptide sequencing has not yet been reported.

EXPERIMENTAL

Peptide samples. The peptides used in this study were synthesized on a solid phase carrier utilizing Fmoc/But strategy on acid cleavable linkers. Samples were either purified by high-performance liquid chromatography (HPLC) or used unpurified.

Mass Spectrometry. All mass spectra were obtained using a TSQ-700 triple sector quadruople mass spectrometer (Finnigan MAT, San Jose, CA, USA), equipped with a standard electrospray ion source. The various multiply charged ions for horse-heart apomyoglobin were used to calibrate the mass scale. Peptide samples were dissolved in 50% aqueous methanol at a concentration of 20 pmol/μL. A Harvard model 22 syringe pump (Harvard Apparatus, South Natik, MA, USA) was used to infuse the sample into the electrospray source at a rate of 2 μL/min. The instrument was operated in a programmed mode for data collection. Profile spectra were collected over the mass range *m/z* 200–1500 with a 3 s scan time. A centroid spectrum was derived from the average of four profile spectra. From the centroid spectrum, the data system selected

* Author to whom correspondence should be addressed.

0951–4198/93/010058–05 $07.50

Received 6 November 1992
Accepted (revised) 6 November 1992

Table 1. Results of the computer analysis of the CID spectrum of an α-chymotryptic peptide reported by Johnson and Biemann along with the calculated number of exchangeable hydrogens (EH). Those entries with the same number of EH as the correct sequence (underlined) are indicated with boldface type

Entry no.	Score	Sequence	EH
1	**0.898**	**VNSQIQPGQVVVF**	**23**
2	0.870	GGVSQIQPGQVVVF	21
3	0.870	NVSQIQPGQVVVF	23
4	0.867	VSGGQIQPGQVVVF	22
5	0.864	VNSXQQPGQVVF	22
6	0.864	GRSQIQPGQVVVF	25
7	0.862	GGADQIQPGQVVVF	22
8	0.862	XADQIQPGQVVVF	21
9	0.862	NADQIQPGQVVVF	23
10	0.857	GQDQIQPGQVVVF	23
11	0.852	SPDQIQPGQVVVF	21
12	0.850	SVGGQIQPGQVVVF	22
13	0.845	GGEGQIQPGQVVVF	22
14	0.845	NEGQIQPGQVVVF	23
15	0.842	SRGQIQPGQVVVF	25
16	0.837	VNSQIQPGQVTPF	23
17	0.836	GGVSXQQPGQVVF	21
18	0.836	NVSXQQPGQVVVF	23
19	0.835	DQGQIQPGQVVVF	23
20	0.835	QDGQIQPGQVVVF	23

all ions above 10% of the base peak for CID MS/MS analysis. Under data system control, the instrument was set to pass a selected ion through the first quadrupole mass analyzer. Argon collision gas was introduced into the octapole collision cell, and the second quadrupole analyzer was used to scan the spectrum over the mass range 1–1500 u with a 3 s scan time. Various offset voltages for each quadrupole analyzer and the collision cell were set according to a formula that varies values according to the mass of the precursor ion. Generally, higher collision energies are needed for larger peptides. Eight profile scans were averaged and converted to a centroid spectrum for computer analysis. Deuterium-exchanged samples were prepared by dissolving the peptide samples in 50% D_2O (Aldrich (Milwawkee, WI, USA), 99.9% d), methanol-d_4 (Cambridge Isotope Labs (Woburn, MA, USA), 99.8% d). No acid was used to enhance the exchange rate.

Computer interpretation. A preliminary version of the program PepSeq[13] supplied by Finnigan MAT was used to interpret the CID spectra. The program was set for fully automated ES mode, using average mass values, a mass error of 0.7 u, and a threshold value of 0.1% of the base peak. Sequences were extended from the C-terminus, and the C-terminal amide group was specified. An intermediate count of 2000 was used to limit the number of subsequences carried on to the next analysis cycle. The observed mass and charge state (Table 1) were entered for each peptide spectrum.

RESULTS AND DISCUSSION

The various amino acid residues contain different numbers of exchangeable hydrogens (G, A, V, L, I, M, F—one labile hydrogen; S, T, D, E, Y, H, C—two; N, K, Q—three; R—five). Peptides with the same molecular weight but different compositions may have d ent numbers of labile hydrogens. The numbe exchangeable hydrogens can be used to filter the l candidate sequences provided by a computer inter tation program. For example, computer analysis of CID mass spectrum of a 12-residue peptide by Joh and Biemann[10] gave a list of 20 sequences with scores over a narrow range (Table 1). The aut analyzed each of these possibilities in detail to certain that the sequence with the highest score wa correct sequence. Knowing that the number of exch geable hydrogens is 23 reduces the list of possibiliti nine. In the same work, the authors use the spectru a 9-residue peptide KKGQKVGFF as an example one difficult to analyze. Computer analysis of the s trum resulted in a list of 20 possible sequences similar scores (Table 2). The correct sequence ranked number 11. However, using information on number of exchangable hydrogens reduces the lis possibilities to three and the correct one has the hig score. Unfortunately, ambiguities with respect to ly (K) and glutamine (Q) cannot be resolved because h have the same mass and the same number of excha able hydrogens. Likewise, the number of exchange hydrogens cannot be used to distinguish between l cine (L) and isoleucine (I). In these and subseq examples, Q is used to designate both Lys and Gln, I is used to designate both Ile and Leu.

The number of exchangeable hydrogens in a pep is readily determined using electrospray m spectrometry. When dissolved in deuterated wate methanol solutions at the concentrations suitable electrospray analysis (<50 pmol μL), there is excha of all hydrogens attached to oxygen, nitrogen sulfur. Results for a test set of peptides having from to 24 exchangeable hydrogens (EH) are given in Ta 3. Spectra collected were obtained at low resolu

Table 2. Results of the computer analysis of the CID spectr of the peptide KKGQKVGEF reported by Joh and Biemann along with the calculated numbe exchangeable hydrogens (EH). Those entries the same number of EH as the correct seque (underlined) are indicated with boldface type

Entry No.	Score	Sequence
1	0.850	GAQGAGQVGEF
2	0.850	GAQGQAGVGEF
3	0.850	GAQGQQVGEF
4	0.850	GAQGQQGVEF
5	0.848	AANGQQVGEF
6	0.840	AGQGAQVGEF
7	0.840	AGQGQAGVGEF
8	0.840	AGQGQQVGEF
9	0.840	QQGQAGVGEF
10	0.840	QQGQAGVGEF
11	**0.834**	**QQGQQVGEF**
12	0.834	NAAGQQVGEF
13	0.832	AGGAGQQVGEF
14	0.832	GAGAGQQVGEF
15	0.832	GAGAGQQGVEF
16	0.832	QGAGQQVGEF
17	0.827	AGQGQQGVEF
18	0.827	QQGQQGVEF
19	0.826	AANGQQVWF
20	0.826	QQGQANVEF

Table 3. Calculated and observed average mass values for peptide molecular ion species with and without deuterium exchange. For each entry, the total number of possible compositions for peptides of that length and having that mass as well as the number of those compositions consistent with the number of exchangeable hydrogens (EH) are given

Entry no.	Sequence	Number of Charges	Before exchange Calculated	Before exchange Observed	EH[a]	After exchange Calculated	After exchange Observed	Total possible composition	Total consistent with EH
1	LAYWKa	1	679.8	679.9	12	692.8	692.7	73	15
		2	340.4	340.2		347.4	347.3		
2	WNYFKa	1	756.9	756.7	14	771.9	771.7	15	3
		2	379.0	378.8		387.0	386.8		
3	TYTAGGa	1	568.6	568.5	12	581.6	581.5	56	11
		2	284.8	—		291.8	—		
4	KGSGAVASa	1	675.8	675.6	15	691.8	691.7	61	39
		2	338.4	338.3		346.9	346.9		
5	DGSRYRTSa	1	941.0	940.8	24	966.0	965.8	2643	28
		2	471.0	470.9		484.0	483.9		
6	WVFDYa	1	728.8	728.7	11	740.8	740.7	34	7
		2	364.9	—		371.4	—		
7	YWKLKAa	1	679.8	679.7	12	692.8	692.6	73	15
		2	340.4	340.3		347.4	347.4		
8	KGSRHTASa	1	842.9	842.7	21	864.9	864.8	1353	28
		2	421.9	421.8		433.4	433.3		
		3	281.6	281.5		289.3	289.6		
9	KFWKTa	1	708.9	708.7	14	723.9	723.7	47	8
		2	354.9	354.8		362.9	362.8		

[a] Number given is for the neutral molecule. For an ion with n charges, n is added to the number of exchangeable hydrogens.

the isotopic cluster was not resolved, and values reported are for average mass. The value of the number of exchangeable hydrogens (EH) in the Table is for the neutral molecule. In ES, peptide ions are generally formed by proton attachment. Thus, the number of EH is increased by one for each charge on the ion. For each peptide in the test set, the observed difference in m/z value between samples run without deuterated solvents and those run in deuterated solvents was within 0.1 u of the calculated difference. Previously, Katta and Chait have reported that back-exchange can occur during the electrospray process.[23] They demonstrated that the exclusion of water vapour from the spray region eliminates back-exchange. The atmospheric pressure region of the Finnigan MAT TSQ-700 source is constantly flushed with hot, dry nitrogen to promote solvent evaporation which eliminates any possibility for back-exchange.

Also included in Table 3 are the number of possible amino acid compositions for amidated peptides that correspond to the observed mass and the number compositions consistent with the observed number of EH. This is a good measure of the degree of simplification possible when information about the EH is used in sequence analysis of peptide mass spectra.

The program (PepSeq) described by Yeates and coworkers[15] was used to interpret the CID spectra collected for each peptide in the test set. Results of the analysis are summarized in Table 4. In these preliminary studies, we purposely avoided optimizing instrument and computer interpretation parameters for each individual sample. All spectra were acquired under identical conditions. As a consequence, some fragment-ion spectra were of low intensity and not well

Table 4. Summary of computer analysis (PepSeq) of CID fragment-ion spectra for the set of test peptides. The rank of the correct sequence and the total number in the range of scores indicated is given before and after filtering the list using the experimentally determined number of exchangeable hydrogens (EH)

Entry no.	Sequence	Charges	Before EH filtering Rank	Before EH filtering Total number	Range of scores	After EH filtering Rank	After EH filtering Total number
1	LALYWKa	1	1	20	1.6–4.1	1	10
		2	1	16	1.4–3.8	1	6
2	WNYFKa	1	1	18	3.2–6.9	1	6
		2	2	18	1.3–3.9	1	3
3	TYTAGGa	1	1	15	8.0–16.1	1	8
4	KGSGAVASa	1	—	30	2.6–4.1	—	2
		2	—	30	1.2–2.6	—	8
5	DGSRYRTSa	2	—	—	—	—	—
6	WVFDYa	1	1	17	3.0–4.7	1	4
7	YWKLAa	1	1	18	1.2–2.7	1	8
8	KGSRHTASa	2	—	30	0.3–0.6	—	0
9	KFWKTa	1	—	30	1.0–2.2	—	8
		2	5	17	0.2–0.4	2	5

Table 5. Result of PepSeq analysis of CID spectra of both the singly and doubly protonated molecule for the peptide WNYFKa.

1+ Charge state			2+ Charge state		
Sequence	Score	EH	Sequence	Score	EH
WNYFQ	7.33	14	WIYFQ	3.86	12
QGDYFQ	6.44	15	WNYFQ	3.53	14
YHYFQ	6.06	13	WMHRQ	2.30	16
YEYPGE	5.91	12	WISHSQ	2.30	15
HYYFQ	5.59	13	WISNHT	2.16	15
EYYPGE	5.50	12	WISSHQ	2.15	15
ANDYFQ	4.29	15	WIYQF	1.94	12
NADYFQ	4.29	15	WIFYQ	1.89	12
QGDYQF	4.29	15	WNYQF	1.75	14
QGDYTSS	3.94	17	WISHNT	1.58	15
NWYFQ	3.63	14	WNFYQ	1.50	14
YEYPW	4.50	11	MFVSSW	1.48	12
GAGDYFQ	3.19	14	QYSSSW	1.44	16
QGDYFAG	3.19	14	YHYFQ	1.37	13
AGGDYFQ	3.19	14	YQSTMQ	1.36	16
QGDYFGA	3.19	14	WIFQY	1.34	12
WNYFGA	3.13	13	YQSTRC	1.33	19
WNYFAG	3.13	13	YQSECQ	1.31	17

suited for computer interpretation. The program failed to interpret (correct sequence not listed) five of the spectra. The results for entry number 2 (sequence WNYFKa) are presented in detail in Table 5 for both the singly protonated (Fig. 1) and doubly protonated molecule (spectrum not shown). Analysis of the singly charged ions yields a list of possible sequences with the correct sequence ranked first. After EH filtering (selecting only those sequences with 14 exchangeable protons), the original list of 18 is reduced to 6. There are only 2 compositions represented, and the score for the correct sequence is much greater than the others. The analysis of the doubly protonated ions yields similar results. The correct sequence is ranked second with a score close to that of the first entry that has a substitution of Ile for Asn. The 0.7 mass error setting used in PepSeq allows a 1 mass unit variation in the molecular weight for doubly charged ions. After EH filtering, the list of possibilities is reduced to three, all with the same composition. Once again, the correct sequence has a significantly better score than the other possibilities.

The number of exchangeable protons can also be determined for each fragment in the CID spectrum. The CID spectra for the singly charged ions of the peptide WNYKFa with and without deuterium exchange are compared in Fig. 1. The fragmentation is not significantly affected by the exchange and the expected mass shifts for the fragment ions are observed. This information could be quite valuable in those instances when the assignment of an ion is ambiguous. Also, it would be easy to modify existing sequence interpretation programs to accommodate deuterium exchange. The masses of the amino acid residues and terminal groups would have to be changed to reflect the number of hydrogens that would be replaced by deuterium, and the protons transferred to generate fragment ion series such as Y″ would have to be changed to account for deuterium exchange.

The deuterium oxide used in these studies was contaminated with low levels of cations such as sodium, potassium and calcium. As a consequence, peptides

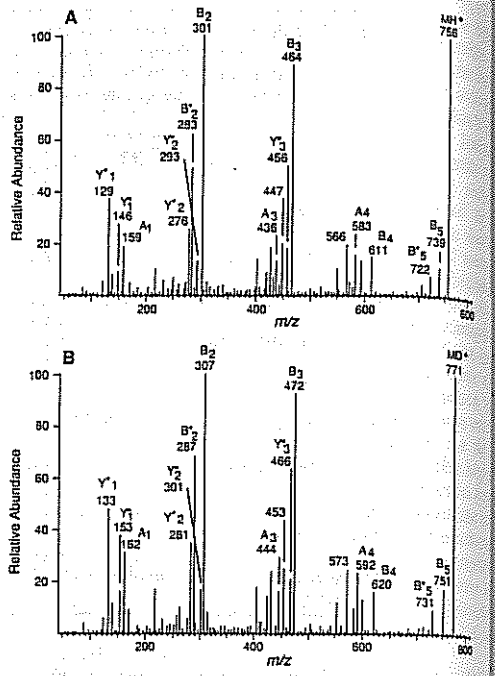

Figure 1. Collision-induced dissociation fragment-ion spectra of the single protonated peptide WNYFKa. Spectrum A is for the ion at *m/z* 756 formed by electrospray from a 20 pmol/μL solution in 50% aqueous methanol. Spectrum B is for the hydrogen-deuterium exchanged ion at *m/z* 771 obtained by electrospray using deuterated water and deuterated methanol. Ion series members are designated using the nomenclature of Roepstorff and Fohlman.[24] Ion series that result from loss of H_2O or NH_3 (commonly observed in low-energy CID spectra) in addition to the cleavage of the peptide backbone are designated with an asterisk.

with a high affinity for alkali cations (generally those with carboxylic acid functionality) exhibit a number of different ions during electrospray analysis. This does not affect the determination of the total number of exchangeable hydrogens because the various species are readily assigned and the number of EH can be determined from any one of them. However, the presence of such salts has a strong influence on the quality of CID fragment ion spectra obtained after deuterium exchange. Species formed by alkali cation attachment did not yield useful CID fragment-ion spectra.

In conclusion, hydrogen-deuterium exchange is a facile method of derivatizing peptides for sequence analysis. The mass shift between deuterated and nondeuterated forms is a direct measure of the number of exchangeable hydrogens. This information places significant restraints on the number of different amino acid compositions corresponding to the determined molecular weight. Hydrogen-deuterium exchange also affects the mass of fragment ions generated by collision-induced dissociation. This methodology has the potential to greatly increase the efficiency of peptide sequence determination by mass spectrometry particularly when used in conjunction with computer algorithms for spectral interpretation. Work is in progress on new sequencing programs that more fully utilize exchangeable hydrogen information.

Acknowledgement

This work was partially funded by NIH Grant RR06217 (TDL) and NIH Cancer Center Core Grant CA33572.

REFERENCES

1. K. Biemann, in *Mass Spectrometry of Biological Materials*, ed. by C. N. McEwen and B. S. Larsen, Marcel Dekker, Inc., New York, 3 (1990).
2. D. F. Hunt, R. A. Henderson, J. Shabanowitz, K. Sakaguchi, H. Michel, N. Sevilir, A. L. Cox, E. Appella and V. H. Engelhard, *Science* **255**, 1261 (1192).
3. D. F. Hunt, J. Shabanowitz, J. R. Yates, P. R. Griffin and N. Z. Zhu, in *Mass Spectrometry of Biological Materials*, ed. by C. N. McEwen and B. S. Larsen, Marcel Dekker, Inc., New York, 169 (1990).
4. K. Biemann, *Biomed. Environ. Mass Spectrom.*, **16**, 99 (1988).
5. T. Sakurai, T. Matsuo, H. Matsuda and I. Katakuse, *Biomed. Mass Spectrom.* **11**, 396 (1984).
6. C. W. Hamm, W. E. Wilson and D. J. Harvan, *Computer Appl. Biosci.*, **2**, 115 (1986).
7. K. Ishikawa and Y. Niwa, *Biomed. Mass Spectrom.* **13**, 373 (1986).
8. M. M. Siegel and N. Bauman, *Biomed. Environ. Mass Spectrom.* **15**, 333 (1988).
9. M. M. Siegel, *Proceedings of the 40th Conference on Mass Spectrometry and Allied Topics*, p. 1785. Washington DC, ASMS, Sante Fe (1992).
10. R. S. Johnson and K. Biemann, *Biomed. Environ. Mass Spectrom.* **18**, 945 (1988).
11. H. Scoble, J. E. Biller and K. Biemann, *Frens. Z. Anal. Chem.*, **327**, 239, (1987).
12. C. Bartels, *Biomed. Environ. Mass Spectrom.* **19**, 365 (1990).
13. J. R. Yates, P. R. Griffin, L. E. Hood and J. X. Zhou, in *Techniques in Protein Chemistry*, ed. by J. J. Villafranca, p. 477. Academic Press, San Diego, CA (1991).
14. H. R. Morris, R. J. Dickensen and D. H. Williams, *Biochem. Biophys. Res. Commun.* **51**, 247 (1973).
15. A. M. Falick and D. A. Maltby, *Anal. Biochem.* **182**, 165 (1989).
16. D. F. Hunt, J. R. Yates, J. Shabanowitz and S. Winston, *Proc. Natl. Acad. Sci. USA* **83**, 6233 (1986).
17. D. F. Hunt, W. M. Bone, J. Shabanowitz, J. Rhodes and J. M. Ballard, *Anal. Chem.* **53**, 1704 (1981).
18. K. Biemann, in *Methods in Enzymology* Vol. 193, ed. by J. A. McCloskey, p. 455. Academic Press, San Diego, CA (1990).
19. J. E. Vath and K. Biemann, *Int. J. Mass Spectrom. Ion Processes* **100**, 287 (1990).
20. G. C. Thorne, K. D. Ballard and S. J. Gaskell, *J. Am. Soc. Mass Spectrom.* **1**, 249 (1990).
21. K. D. Ballard and S. J. Gaskell, *J. Am. Chem. Soc.* **114**, 64 (1992).
22. J. S. McCloskey, in *Methods in Enzymology* Vol. 193, ed. by J. A. McCloskey, p. 329. Academic Press, San Diego, CA (1990).
23. V. Katta and B. Chait, *Rapid Commun. Mass Spectrom.* **5**, 214 (1991).