Available online at www.sciencedirect.com
SCIENCE DIRECT
ELSEVIER
International Journal of Mass Spectrometry 254 (2006) 101–105
International Journal of Mass Spectrometry
www.elsevier.com/locate/ijms

Random protein association analyses of MALDI-TOF mass spectra of two- and three-component protein systems using binomial and multinomial distribution

Bao-Shiang Lee *, Sangeeth Krishnanchettiar, Syed Salman Lateef, Chun-Ting Li, Shalini Gupta

Protein Research Laboratory, Research Resources Center, University of Illinois, 835 S. Wolcott Avenue, Chicago, IL 60612, USA

Received 29 March 2006; received in revised form 31 May 2006; accepted 31 May 2006

Abstract

The phenomenon of protein randomly associating among themselves during MALDI-TOF mass spectrometric measurements is manifest on all the proteins tested here. The magnitude of this random association seems to be protein-dependent. However, a detailed mathematical analysis of this process has not been reported so far. Here, binomial and multinomial equations are used to analyze the relative populations of multimer ions formed by random protein association during MALDI-TOF mass spectrometric measurements. Hemoglobin A (which consists of two α-globins and two β-globins) and biotinylated insulin (which contains intact, singly biotinylated, and doubly biotinylated insulin) are used as the test cases for two- and three-component protein systems, respectively. MALDI-TOF spectra are acquired using standard MALDI-TOF techniques and equipment. The binomial distribution matches the relative populations of multimer ions of Hb A perfectly. For biotinylated insulin sample, taking lesser relative populations for doubly biotinylated insulin and intact insulin compared with singly biotinylated insulin into account, the relative populations of multimer ions of the biotinylated insulin confirms the prediction of multinomial equation. The pairs of unrelated proteins such as myoglobin, avidin, and lysozyme show lesser intensities for heteromultimers than for homomultimers, indicating weaker propensities to associate between different proteins during MALDI-TOF mass spectrometric measurements. Contrary to the suggestion that the multimer ions are formed in the solution phase prior to MALDI-TOF mass spectrometric measurement through multistage sequential reactions of the aggregation of protein molecules, we postulate that multimer ions of proteins are formed after the protein molecules have been vaporized into the gas phase through the assistance of the laser and matrix.

Keywords: Random protein association; MALDI-TOF MS; Binomial and multinomial distributions

1. Introduction

There has been a great interest in the mass spectrometric characterization of the intact non-covalent protein association. Most of the work has been conducted using the electrospray ionization mass spectrometry (ESI MS) [1–7], where, the optimum parameters (protein concentration, buffer pH, inlet capillary temperature, and skimmer offset voltage) for observing the protein complexes have to be determined empirically, and the efficacy of the technique is protein-dependent. Alternatively, the matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) [8] has been used to study the intact non-covalent protein association. Benefits of using MALDI-TOF MS compared with ESI MS include a simple sample preparation, high sensitivity (femtomole range), and short analysis times. In addition, the use of a time-of-flight mass analyzer allows the detection of higher molecular weight ions as high as ∼megadaltons [8]. Furthermore, routinely, only singly and doubly charged protein molecular ions are observed with MALDI-TOF MS, which facilitates the spectral interpretation. Using intensity-fading MALDI-TOF mass spectrometry, the intact non-covalent protein association of small protein inhibitors (serine proteases, cysteine proteases, and carboxypeptidases) and their corresponding partner enzymes have been detected [9]. Several publications on the intact non-covalent association of monomeric protein chain enhanced by the chem-

* Corresponding author. Tel.: +1 312 996 1411; fax: +1 312 996 1898.
E-mail address: boblee@uic.edu (B.-S. Lee).

1387-3806/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.ijms.2006.05.023

ical cross-linking, diverse matrix materials, various laser wavelengths, or one-shot techniques have been reported [10–12]. A complication in the measurement of intact non-covalent protein association by MALDI-TOF MS is that proteins can associate through random association during mass spectrometric measurement. These protein complexes are artifacts in that they are not the same form that exist in biological system. Information about the random protein association process is needed to deal with this complication. However, the detailed analysis of this random protein association has not been reported so far. Here we report analyses of the MALDI-TOF mass spectra of the random protein association of two- and three-component protein systems. Two-component protein system hemoglobin A (Hb A; which consists of two α-globins and two β-globins) and three-component protein system biotinylated bovine insulin (which contains intact insulin, singly biotinylated insulin, and doubly biotinylated insulin) are used as the test cases. Mass spectra of protein multimer complexes are detected by standard MALDI-TOF MS techniques and analyzed using binominal and multinomial equations [13]. The pairs of unrelated proteins such as myoglobin, avidin, and lysozyme have also been tested.

2. Experimental

2.1. Materials

All chemicals and proteins were purchased from Sigma (St. Louis, MO) unless otherwise indicated and used without further purification.

2.2. Preparations of Hb A, α-globin of Hb A, and β-globin of Hb A

Informed written consent from healthy subjects was obtained before collecting the blood samples. Blood specimens were collected in heparin (anticoagulant). Hemolysates were prepared by lysis of the washed erythrocytes with three to five volumes of cold water and gassed with carbon monoxide. The cell debris was removed by centrifugation for 20 min at $3000 \times g$ before storage of a clear hemolysate at −85 °C. The pure hemoglobin solution was prepared from hemolysate by diethylaminoethyl Sephadex anion exchanger chromatography and/or carboxymethyl Sephadex cation exchanger chromatography and was extensively dialyzed against cold deionized water. Preparation of pure α-globin and β-globin of the Hb A was done by using carboxymethyl Sephadex cation exchanger chromatography in 8 M urea [14].

2.3. MALDI-TOF mass spectrometric analysis

Protein solution for MS analysis was prepared in deionized water, and sinapinic acid (SA) was used as the matrix. Aliquots (1.3 μl) of the matrix solution [3–10 mg SA in 1 ml aqueous solution of 50% (v/v) acetonitrile containing 0.1% (v/v) TFA] were mixed with the protein solution and spotted onto a MALDI-TOF target. A Voyager-DE PRO Mass Spectrometer (Applied Biosystems, Forst City, CA) equipped with a 337 nm pulsed nitrogen laser was used to analyze the samples. Protein mass was measured using the positive-ion linear mode. Delayed extraction (DE) was operated with an accelerating voltage of 25 kV, a grid voltage of 92%, a guide wire of 0.15%, and a delay time of 300 ns. Mass spectra were obtained by averaging 100 laser shots. External mass calibration was performed using the peaks of a mixture of insulin (bovine) at *m/z* 5735, cytochrome *c* (equine) at *m/z* 12,362, apomyoglobin (equine) at *m/z* 16,952, adolase (rabbit muscle) at *m/z* 39,212, and albumin (bovine serum) at *m/z* 66,430. Curve deconvolution and integrated area of the computer program PeakFit (Systat Software, Richmond, CA) were used to determine intensities of the molecular ions of proteins.

3. Results and discussion

Hb A consists of two α-globin (*m/z* 15,127) and two β-globin (*m/z* 15,867) with a tetramer structure ($\alpha_2\beta_2$). In the gas phase during MALDI-TOF mass spectrometric measurement, α-globin (α) and β-globin (β) randomly collide and associate with each other. Assuming that the association abilities in the gas phase during MALDI-TOF mass spectrometric measurement of αα, αβ, and ββ are equal and the relative population of either α or β is 1/2, the relative population distribution of all the possible multimers can be calculated using the binomial equation.

$$(x+y)^n = x^n + \binom{n}{1} x^{n-1}y + \binom{n}{2} x^{n-2}y^2 + \binom{n}{3} x^{n-3}y^3 + \cdots + \binom{n}{n} y^n \quad (1)$$

where x and y stand for the probabilities (relative populations) of α and β, respectively; n is the total number of protein molecules in the multimers; the coefficients, called binomial coefficients, are given by

$$\binom{n}{k} = \frac{n(n-1)(n-2)\cdots(n-k+1)}{k!} = \frac{n!}{k!(n-k)!} = \binom{n}{n-k}$$

where $k! = 1 \times 2 \times 3 \times \cdots \times k$ and $0! = 1$.

Also, Pascal's triangle (Fig. 1) can be used to predict the relative population distribution of Eq. (1). Fig. 2 shows the MALDI-TOF mass spectra of Hb A; the experimental population distribution of Hb A association matches the prediction of the binomial distribution perfectly. Contrary to the suggestion that the multimer is formed in the solution phase through multistage sequential reactions of the aggregation of protein molecules prior to MALDI-TOF mass spectrometric measurement as suggested in the previous publication on toxin [12], the fact that the population distribution of Hb A associations match the prediction of the binomial distribution demonstrates that the multimer is formed in the gas phase. In addition, no change in the experimental data upon changing protein's concentration has been observed in this study (data not shown), which does not

$(x+y)^n$

$n=1$						1		1						
$n=2$					1		2		1					
$n=3$				1		3		3		1				
$n=4$			1		4		6		4		1			
$n=5$		1		5		10		10		5		1		
$n=6$	1		6		15		20		15		6		1	
$n=7$	1	7		21		35		35		21		7		1

Fig. 1. Pascal's triangle. *n* is the total number of molecules in the protein complex. Numerical values represent the relative populations. X and Y represent two different proteins. Notice that the borders of the array consist of 1's, and any number in the array not on the border is the sum of the two closest numbers in the preceding row.

support the protein aggregation in solution. In fact, more protein multimers are detected by increasing the laser power (which generates more protein molecular ions in the gas phase). We postulated that multimer ions of Hb A are formed after the protein molecules have been vaporized into the gas phase through the assistance of the laser and matrix.

Fig. 3 shows MALDI-TOF mass spectra of the random protein association of protein pairs of pure α (1 μg/μl) of the Hb A and pure β (1 μg/μl) of the Hb A, avidin (2 μg/μl) and Mb (0.02 μg/μl), lysozyme (0.2 μg/μl) and avidin (2 μg/μl), and lysozyme (0.2 μg/μl) and Mb (0.02 μg/μl). The concentration of each protein is chosen such that the relative intensities are at similar magnitude. The data shows that the relative populations of the dimer ions αα, αβ, and ββ are ∼1:1:1 (Fig. 3I). This data indicates that although the relative population of the αβ is twice that of either αα or ββ, the association ability between purified α and purified β is only half that of between purified α and

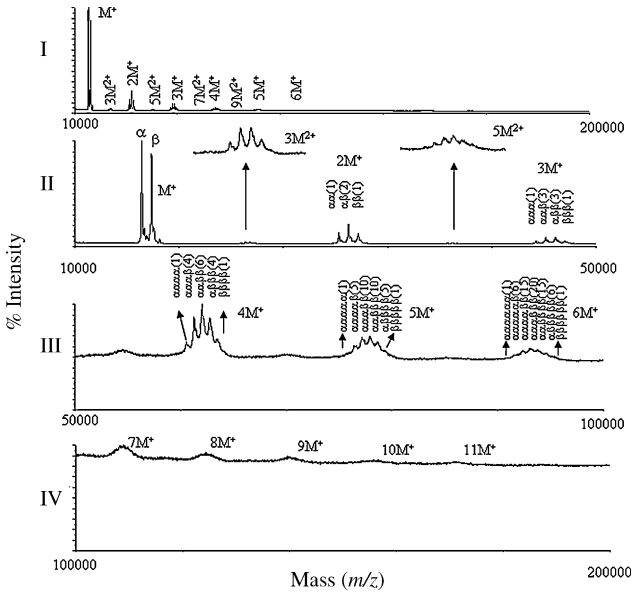

Fig. 2. MALDI-TOF mass spectra of Hb A (1 μg/μl): (I) mass range from 10,000 to 200,000 *m/z*; (II) mass range from 10,000 to 50,000 *m/z*; (III) mass range from 50,000 to 100,000 *m/z*; (IV) mass range from 100,000 to 200,000 *m/z*. Value in parenthesis indicates the relative population. α stands for α-chain of the Hb A; β stands for β-chain of the Hb A.

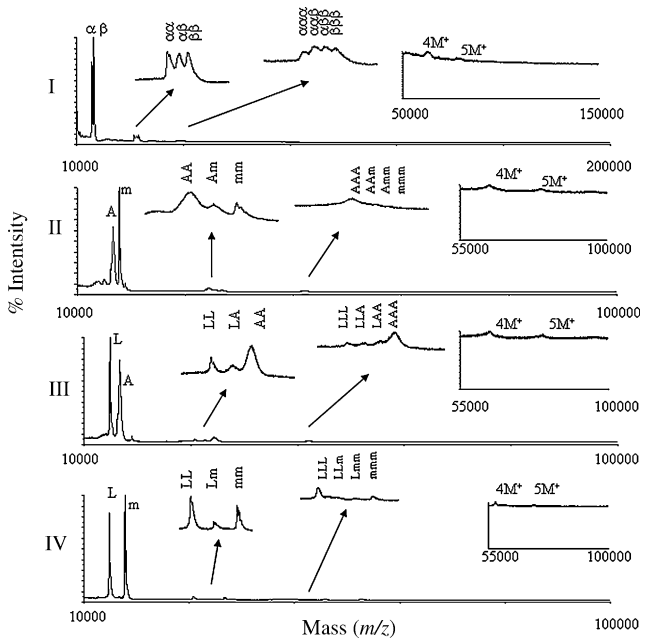

Fig. 3. MALDI-TOF mass spectra of: (I) pure α-globin (1 μg/μl) of the Hb A + pure β-globin (1 μg/μl) of the Hb A; (II) avidin (2 μg/μl) + Mb (0.02 μg/μl); (III) lysozyme (0.2 μg/μl) + avidin (2 μg/μl); (IV) lysozyme (0.2 μg/μl) + Mb (0.02 μg/μl). Number in parenthesis indicates the relative population. α stands for purified α-globin of the Hb A; β stands for purified β-globin of the Hb A. L stands for lysozyme; A stands for avidin; m stands for myoglobin.

purified α or between purified β and purified β. However, α and β from the intact Hb A show equal association abilities among intact αα, intact αβ, and intact ββ as manifest with the relative populations of ∼1:2:1 for dimer ions αα, αβ, and ββ (Fig. 2). This observation is not surprising since 8 M urea was used during the preparation of pure α-globin and pure β-globin of Hb A, which may have altered the globin's association ability during MALDI-TOF mass spectrometric measurements. We speculate that the structure of the protein, which affects the orientation during the collision may be responsible for the protein-dependent association ability that causes deviation from the bionomial distribution predictions. It is noted that the involvement of the heme cofactor on the association ability can be ruled out because all the multimer ions observed here agree with molecular ions without the heme cofactors. In addition, due to this reduced association ability between purified α and purified β, the relative populations of the trimer ions ααα, ααβ, αββ, and βββ are ∼1:1:1:1 (Fig. 3I) while it is 1:3:3:1 for the intact α and intact β of intact Hb A (Fig. 2II). Data shown in Fig. 3II–IV of the multimer ions of the protein pairs of avidin, myoglobin, and lysozyme further demonstrates the reduced association ability during MALDI-TOF mass spectrometric measurements. The population of heterodimer pair between avidin and Mb is only about 1/3 of their corresponding homodimers (Fig. 3II). The populations of heterodimer pairs between lysozyme and avidin (Fig. 3III) and between lysozyme and Mb (Fig. 3IV) are only about 1/4 of their corresponding homodimers.

Fig. 4 shows MALDI-TOF mass spectra of biotinylated bovine insulin. The sample obtained from Sigma Company is reported to contain intact insulin (a; ∼30.6%), singly biotiny-

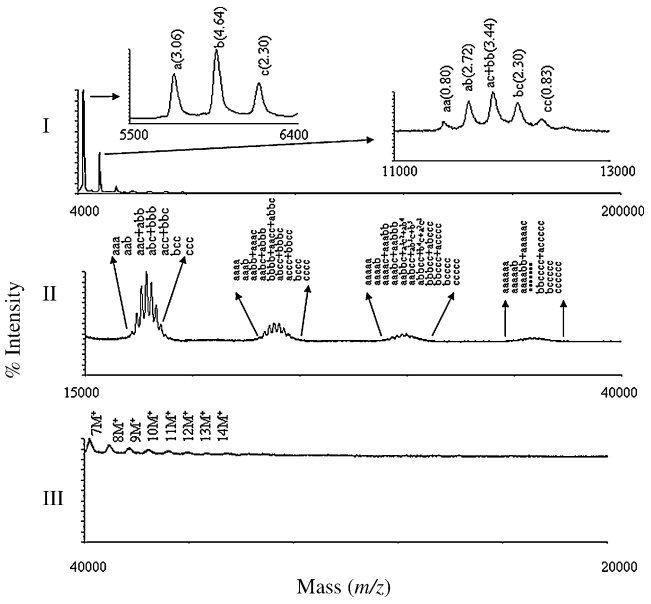

Fig. 4. MALDI-TOF mass spectra of biotinylated bovine insulin (10 μg/μl): (I) mass range from 4000 to 200,000 *m/z*; (II) mass range from 15,000 to 40,000 *m/z*; (III) mass range from 40,000 to 200,000 *m/z*. Value in parenthesis indicates the relative population. a stands for intact bovine insulin; b stands for singly biotinylated bovine insulin; c stands for doubly biotinylated bovine insulin. Masses of molecular ions ac and bb are equal.

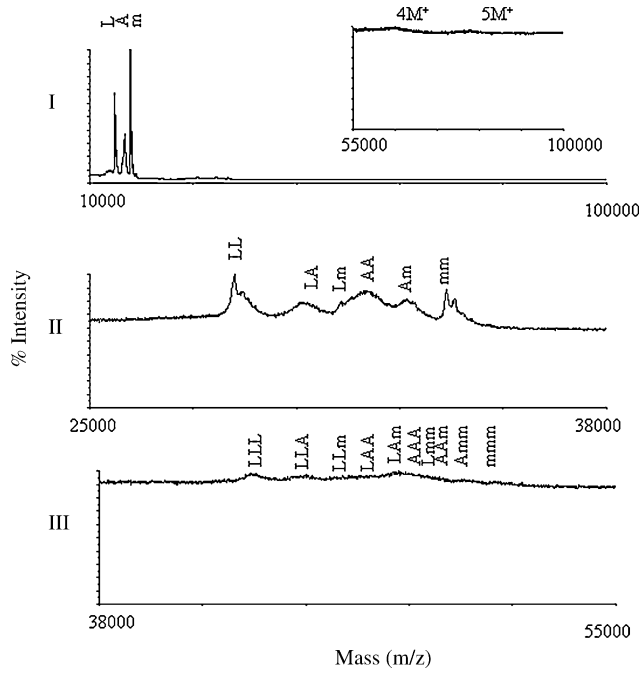

Fig. 5. MALDI-TOF mass spectra of lysozyme (0.2 μg/μl) + avidin (2 μg/μl) + Mb (0.02 μg/μl): (I) mass range from 10,000 to 100,000 *m/z*; (II) mass range from 25,000 to 38,000 *m/z*; (III) mass range from 38,000 to 55,000 *m/z*. L stands for lysozyme; A stands for avidin; m stands for myoglobin.

lated insulin (b; ∼46.4%), and doubly biotinylated insulin (c; ∼23.0%). In the gas phase during MALDI-TOF mass spectrometric measurement, these three molecular ions randomly collide and associate with each other. Assuming that the association abilities of these three molecular ions are the same and that the relative populations of a, b, and c are 0.306, 0.464, and 0.230 (Fig. 3I), respectively, the relative population distribution of all the possible multimers can be calculated using the multinomial equation.

$$(x_1 + x_2 + \cdots + x_p)^n = \sum \frac{n!}{n_1!n_2!\cdots n_p!} x_1^{n_1} x_2^{n_2} \ldots x_p^{n_p} \quad (2)$$

where x_1, x_2, and $x_{3=p}$ stand for probabilities of intact insulin, singly biotinylated insulin, and doubly biotinylated insulin, respectively; n is the total number of protein molecules in the multimers; the sum, denoted by Σ, is taken over all non-negative integers $n_1, n_2, \ldots, n_p$ for which $n_1 + n_2 + \cdots + n_p = n$.

According to Eq. (2), assuming that probabilities (relative populations) of x_1, x_2, and x_3 are the same (∼1/3), the relative populations of the molecular ions aa, ab, ac, bb, bc, and cc should be 1:2:2:1:2:1. Since masses of ac and bb are equal, the relative populations of the molecular ions aa, ab, ac + bb, bc, and cc should now be 1:2:3:2:1. Taking different relative populations of intact insulin (∼30.6%), singly biotinylated insulin (∼46.4%), and doubly biotinylated insulin (∼23%) into consideration, the predicted relative populations of the molecular ions aa, ab, ac + bb, bc, and cc become ∼0.94:2.84:3.55:2.13:0.53. This prediction of the relative populations of dimer matches the experimental values (0.80:2.72:3.44:2.30:0.83) nicely (Fig. 4I). The minor sixth peak with $m/z >$ the cc dimer in Fig. 4I originates from the minor component of the triply biotinylated insulin, which does not affect the analyses. The prediction is easily extended to trimer, tetramer, pentamer, and so on and is verified (Fig. 4II).

Fig. 5 shows MALDI-TOF mass spectra of a protein mixture, containing pure lysozyme (0.2 μg/μl), pure avidin (2 μg/μl), and pure Mb (0.02 μg/μl). The concentration of each protein is chosen such that the relative intensities are at similar magnitude. Due to the reduced association abilities during the MALDI-TOF mass spectrometric measurements between lysozyme and Mb, between lysozyme and avidin, and between avidin and Mb, the relative populations of the heterodimer molecular ions are much less than the prediction using Eq. (2).

4. Conclusion

We report here the analyses of the MALDI-TOF mass spectra of the random protein association of two- and three-component protein systems. Binomial and multinomial distributions are successfully used to predict the relative populations of the multimer ions of Hb A and biotinylated bovine insulin. Data indicates that the association abilities in the gas phase during MALDI-TOF mass spectrometric measurement between α and α, between α and β, and between β and β of the intact Hb A are equal. Similarly, association abilities among native insulin, singly biotinylated insulin, and doubly biotinylated insulin are equal. Data shows that multimer is formed by a random collision of proteins in the gas phase and matches mathematical predictions of randomness. Lesser relative intensities for heteromultimers than the prediction (assuming equal association abilities among proteins in gas phase) observed here for unrelated proteins such as myoglobin, avidin, and lysozyme indicates reduced association

abilities during the MALDI-TOF mass spectrometric measurement between different proteins in the gas phase.

Acknowledgement

We thank the support of the Research Resources Center at University of Illinois at Chicago.

References

[1] J.A. Loo, B. Berhane, C.S. Kaddis, K.M. Wooding, Y. Xie, J. Am. Soc. Mass Spectrom. 16 (2005) 998.
[2] L.L. Ilag, H. Videler, A.R. Mckay, F. Sobott, P. Fucini, K.H. Nierhaus, C.V. Robinson, Proc. Natl. Acad. Sci. U.S.A. 102 (2005) 8192.
[3] J.A. Aquilina, J.L.P. Benesch, L. Ding, O. Yaron, J. Horwitz, C.V. Robinson, J. Biol. Chem. 280 (2005) 14485.
[4] F. Sobott, C.V. Robinson, Curr. Opin. Struct. Biol. 12 (2002) 729.
[5] J.A. Loo, V. Thanabal, H. Wei, in: A.L. Burlingame, S.A. Carr, M.A. Baldwin (Eds.), Biology & Medicine, Humana Press, New Jersey, 2000, p. 73.
[6] J.A. Loo, Mass Spectrom. Rev. 16 (1997) 1.
[7] K.J. Light-Wahl, B.L. Schwartz, R.D. Smith, J. Am. Chem. Soc. 116 (1994) 5271.
[8] A. Aitken, in: J.M. Walker (Ed.), The Proteomics Protocols Handbook, Humana Press, New Jersey, 2005, p. 319.
[9] J. Villanueva, O. Yanes, E. Querol, L. Serrano, F.X. Aviles, Anal. Chem. 75 (2003) 3385.
[10] T.B. Farmer, R.M. Caprioli, J. Mass Spectrom. 33 (1998) 397.
[11] M. Sun, L. Ding, Y. Ji, D. Zhao, S. Liu, J. Ni, Rapid Commun. Mass Spectrom. 13 (1999) 150.
[12] L.R.H. Cohen, K. Strupat, F. Hillenkamp, J. Am. Soc. Mass Spectrom. 8 (1997) 1046.
[13] R.E. Walpole, R.H. Myers, Probability and Statistics for Engineers and Scientists, The Southeast Book Company, New York, 1972.
[14] J.B. Clegg, M.A. Naughton, D.J. Weatherall, J. Mol. Biol. 19 (1966) 91.