

See discussions, stats, and author profiles for this publication at:
<https://www.researchgate.net/publication/221997904>

Applying a new algorithm for obtaining site specific rate constants for H/D exchange of the gas phase proton-bound arginine dimer

ARTICLE *in* CHEMICAL PHYSICS LETTERS · OCTOBER 2003

Impact Factor: 1.9 · DOI: 10.1016/j.cplett.2003.08.102

CITATIONS

15

READS

13

6 AUTHORS, INCLUDING:



Ya'acov Ritov

Hebrew University of Jerusalem

135 PUBLICATIONS 4,083 CITATIONS

SEE PROFILE



Melinda A Mcfarland

U.S. Food and Drug Administration

24 PUBLICATIONS 632 CITATIONS

SEE PROFILE

Applying a new algorithm for obtaining site specific rate constants for H/D exchange of the gas phase proton-bound arginine dimer

Bryan G. Reuben ^a, Ya'acov Ritov ^b, Orit Geller ^c, Melinda A. McFarland ^d,
Alan G. Marshall ^d, Chava Lifshitz ^{c,*}

^a School of Applied Science, South Bank University, London, UK

^b Department of Statistics, The Hebrew University of Jerusalem, Jerusalem 91905, Israel

^c Department of Physical Chemistry and The Farkas Center for Light Induced Processes,
The Hebrew University of Jerusalem, Givat Ram, Jerusalem 91904, Israel

^d Ion Cyclotron Resonance Program, National High Magnetic Field Laboratory, Florida State University,
1800 East Paul Dirac Drive, Tallahassee, FL 32310-4005, USA

Received 10 July 2003; in final form 20 August 2003

Published online: 25 September 2003

Abstract

A new algorithm has been developed for extracting site-specific rate constants for hydrogen/deuterium (H/D) exchange in gas phase protonated amino acids, their clusters and peptides. The algorithm minimizes the mutual entropy or the Kullback–Leibler information divergence between the observed concentrations and the model. Electrospray ionization-mass spectrometry (ESI-MS) results from fast flow tube and Fourier transform ion cyclotron resonance (FT-ICR) experiments, respectively, were modeled. The results for protonated glycine were in excellent agreement with previous literature data. Four rate constants were found, three of them identical corresponding to the three equivalent hydrogen atoms of the protonated amine group and a fourth higher one corresponding to the single carboxyl hydrogen. New results for the proton-bound dimer of arginine demonstrated a single high site-specific rate constant and fourteen low ones. These results are in agreement with the ion–zwitterion structure of (arginine)₂H⁺ that has a single carboxyl hydrogen atom.

© 2003 Elsevier B.V. All rights reserved.

1. Introduction

Much research has been reported in recent years on gas phase hydrogen–deuterium exchange, where

protonated biomolecules in an electrospray ionization (ESI) – Fourier transform ion cyclotron resonance mass spectrometer (FT-ICR MS) or flow tube system react with deuterated small molecules such as ND₃, D₂O or CH₃OD. The deconvolution of the spectra obtained from such experiments leads to sets of profiles for consecutive deuterium exchanges as a function of the time or the flow rate

* Corresponding author. Fax: +972-2-652-2472.

E-mail address: chavalu@vms.huji.ac.il (C. Lifshitz).

of the deuterating agent. The flow rate, in the kinetic analysis of flow tube experiments, is a surrogate for time.

The aim of the kinetics is the extraction of site-specific rate constants for the replacement of labile hydrogen atoms differently located in the amino acid or peptide structure [1–3]. The extraction of these rate constants depends on the observation that different reaction sequences give different kinetics. This analysis has been done in the past by the solution of sets of independent differential equations, in which the rate constants were adjusted by an optimization program to give the best fit to the experimental data. In the case of the interaction of D₂O or CH₃OD with protonated monoglycine, four rate constants were found, three of them identical corresponding to the three equivalent labile hydrogen atoms of the protonated amine group and a fourth corresponding to the single labile carboxyl hydrogen [1,2].

Subsequent work has extended these analyses to larger molecules with more replaceable hydrogens. In general, there is considerable interest in large gas phase biomolecules. Conformational changes in proteins are probed by hydrogen-exchange ESI-MS [4,5]. The algorithms for determination of the site specific rate constants, however, become very large. If one assumes initially that all protons are kinetically distinguishable, then the number of protonated and deuterated structures for a species containing two protons is only four: core-HD and core-DH together with the initial and final forms core-HH and core-DD. For three protons, this rises to eight: core-HHD, core-HDH and core-DHH with one proton replaced, and core-DDH, core-DHD and core-HDD with two protons replaced, together with starting and finishing species core-HHH and core-DDD. If there are N replaceable protons, the number of intermediate structures follows the binomial coefficients and the total number of possible structures is 2^N . The array of differential equations to be solved thus rises exponentially with their number. It requires enhanced effort to set-up the equations. Since the equations are integrated numerically, their solution becomes a time-consuming procedure and one for which the optimization program is liable to get stuck in numerous local minima. We have devised

a new algorithm to avoid these problems and to permit optimization of large numbers of site-specific rate constants. We present the algorithm and demonstrate its capabilities by use of the well-known example of protonated glycine cited earlier as well as some new data obtained for H/D exchange of the proton-bound dimer of arginine.

Consider a species (amino acid, peptide or cluster) with N replaceable hydrogens. Each molecule is represented by an N bit binary number, where 1 represents deuterium and 0 represents hydrogen. In the hydrogen–deuterium exchange system, it is assumed that each of these protons is replaced independently in a first-order process with rate constants $\theta_1, \theta_2, \dots, \theta_n, \dots, \theta_N$. After time t , the concentration x of hydrogens that have reacted (initial concentration is a) is given by the familiar equation

$$\theta_n t = \ln(a/(a - x)).$$

The probability $p_{n,t}$ that the n th hydrogen will have been replaced by deuterium after time t is given by rearrangement

$$p(D)_{n,t} = 1 - \exp(-\theta_n t).$$

The probability that the n th hydrogen will not have been replaced is then

$$p(H)_{n,t} = \exp(-\theta_n t).$$

The overall probability of any structure occurring is the product of the individual probabilities, $p(D)_{n,t}$ and $p(H)_{n,t}$ selected according to whether the digit in the corresponding binary number is 1 or 0.

For example if $N = 5$, the probability p_t of the structure core-HHDDH (binary equivalent 00110) is given by

$$\begin{aligned} P_t &= p(H)_{1,t} * p(H)_{2,t} * p(D)_{3,t} * p(D)_{4,t} * p(H)_{5,t} \\ &= \exp(-\theta_1 t) * \exp(-\theta_2 t) * (1 - \exp(-\theta_3 t)) \\ &\quad * (1 - \exp(-\theta_4 t)) * \exp(-\theta_5 t). \end{aligned}$$

Summation of the digits of the binary number gives the number of deuteriums in the intermediate, and the probabilities of structures containing the same numbers of deuteriums can be summed to give an overall probability for that number of replacements.

The experimental profiles for the abundance of species containing $0, 1, 2, \dots, n, \dots, N$ deuterium atoms after various reaction periods (for the FT-ICR experiment) or flow rates (for the flow-tube experiment) are then normalized and compared with the above calculation of the predicted probabilities at the same flow rates. Fitting the experimental curves by a theoretical model can be done by minimization of the overall mean squares deviation by variation of $\theta_1, \theta_2, \dots, \theta_n, \dots, \theta_N$ to give the site specific rate constants for the exchanges. Alternatively, one can minimize the distance between the theoretical and experimental profiles. The Kullback–Leibler (KL) divergence [6] has some good theoretical properties, and there is a simple algorithm that can be employed.

2. The algorithm

The algorithm minimizes the mutual entropy or the Kullback–Leibler [6] information divergence between the observed concentrations and the model. Alternatively, it maximizes the ‘likelihood’

$$L(\vartheta) = \sum_t \sum_{i=0}^N x_i(t) \log(p_i(t; \vartheta)),$$

in which t denotes reaction period, N is the number of sites, x_i is the observed concentration at reaction period t of ions with i exchanged hydrogen atoms. Finally, $\vartheta = (\vartheta_1, \dots, \vartheta_N)$ is the set of unknown parameters and $p_i(t; \vartheta)$ the prescribed probability of i exchanged hydrogen atoms at t , if the parameter is ϑ . Note, if for some integer $N, Nx_i(t)$ were all integers, then $L(\vartheta)$ would be the likelihood function of a multinomial model.

Let Δ be the set of all N digit binary expansions: $\Delta = \{d_k = (d_{k1}, \dots, d_{kN}), d_{kj} \in \{0, 1\}\}$. Let Δ_i be the indexes of those members of Δ with exactly i 1's: $\Delta_i = [k : d_k \in \Delta, \sum_j d_{kj} = i]$. Then

$$p_i(t; \vartheta) = \sum_{k \in \Delta_i} \pi_k(t; \vartheta),$$

where

$$\pi_k = \prod_{j=1}^N e^{-\vartheta_j d_{kj}} (1 - e^{-\vartheta_j})^{d_{kj}}.$$

The algorithm is essentially an EM procedure [7], in which concentrations take the place of proba-

bilities, see [8]. It proceeds as follows. Let $\vartheta^{(n)}$ be the estimate of ϑ after the i th iteration. Then $\vartheta^{(n+1)}$ maximizes

$$\begin{aligned} \arg \max_{\vartheta} L^{(n)}(\vartheta) &= \arg \max_{\vartheta} \sum_t \sum_{i=0}^N x_i(t) \\ &\times \sum_{k \in \Delta_i} \frac{\pi_k(t; \vartheta^{(n)})}{p_i(t; \vartheta^{(n)})} \log(\pi_k(t; \vartheta^{(n)})). \end{aligned} \quad (1.1)$$

Note that by definition $L^{(n+1)}(\vartheta^{(n+1)}) \geq L^{(n)}(\vartheta^{(n)})$. Hence

$$\begin{aligned} 0 &\leq \sum_t \sum_{i=0}^N x_i(t) \sum_{k \in \Delta_i} \frac{\pi_k(t; \vartheta^{(n)})}{p_i(t; \vartheta^{(n)})} \log\left(\frac{\pi_k(t; \vartheta^{(n+1)})}{\pi_k(t; \vartheta^{(n)})}\right) \\ &\leq \sum_t \sum_{i=0}^N x_i(t) \sum_{k \in \Delta_i} \log\left(\frac{\pi_k(t; \vartheta^{(n)})}{p_i(t; \vartheta^{(n)})} \frac{\pi_k(t; \vartheta^{(n+1)})}{\pi_k(t; \vartheta^{(n)})}\right) \\ &= \sum_t \sum_{i=0}^N x_i(t) \sum_{k \in \Delta_i} \log\left(\frac{\pi_k(t; \vartheta^{(n+1)})}{p_i(t; \vartheta^{(n)})}\right) \\ &= L(\vartheta^{(n+1)}) - L(\vartheta^{(n)}). \end{aligned} \quad (1.2)$$

The second line follows since the log function is concave. But then $L(\vartheta)$ is increasing in each step of the algorithm.

However, the maximization of (1.1) is simple. Let, for $k \in \Delta_i$

$$y_k^{(n)} = x_i(t) \frac{\pi_k(t; \vartheta^{(n)})}{p_i(t; \vartheta^{(n)})},$$

be the ‘guessed’ concentration of the k th structure. Then

$$\begin{aligned} L^{(n)}(\vartheta) &= \sum_t \sum_{i=0}^N \sum_{k \in \Delta_i} y_k^{(n)}(t) \log(\pi_k(t; \vartheta)) \\ &= \sum_{j=1}^N \sum_t \left((-\vartheta_j t) \sum_{k: d_{kj}=1} y_k^{(n)}(t) \right. \\ &\quad \left. + \log(1 - e^{-\vartheta_j t}) \sum_{k: d_{kj}=0} y_k^{(n)}(t) \right). \end{aligned}$$

Hence, each ϑ_j is found separately by maximizing a simple concave function.

As an initial value we fit the observed mean number of sites occupied: $z(t) = \sum_{i=1}^N i x_i(t)$, $t =$

1, 2, ... to the theoretical value: $\sum_{s=1}^N e^{-\vartheta_s t}$ by non-linear least squares. The problem is convex minimization and is solved by a modified Newton–Raphson algorithm.

Because of the existence of isotopes of other atoms, we do not observe directly $p_i(t; \vartheta)$ but its convolution $p_i^*(t; \vartheta) = \sum_{j=0}^i p_j(t; \vartheta) q_{i-j}$ in which q_i are the abundances of other isotopes. If we denote by $\hat{x}(t; \vartheta)$ the estimated concentration of the ion with i exchanged hydrogen atoms, then the above algorithm is modified to cope with the convolution by replacing the observed $x_i(t)$ by

$$\hat{x}_i(t; \vartheta) = \sum_{j=i} x_j(t) \frac{p_i(t; \vartheta) q_{j-i}}{\sum_{k=0}^j p_k(t; \vartheta) q_{j-k}}. \quad (1.3)$$

We summarize the steps to be taken:

1. Find initial values $\vartheta^{(0)} = (\vartheta_1^{(0)}, \dots, \vartheta_N^{(0)})$. Let $n = 0$.
2. Calculate the deconvoluted concentrations according to (1.3) with $\vartheta = \vartheta^{(n)}$.
3. Calculate

$$y_k^{(n)}(t) = \hat{x}_i(t; \vartheta^{(n)}) \frac{\pi_k(t; \vartheta^{(n)})}{p_i(t; \vartheta^{(n)})}.$$

4. Maximize for each $j = 1, \dots, N$

$$\sum_t \left(\left(-\vartheta_j^{(n+1)} t \right) \sum_{k: d_{kj}=1} y_k^{(n)}(t) + \log \left(1 - e^{-\vartheta_j^{(n+1)} t} \right) \sum_{k: d_{kj}=0} y_k^{(n)}(t) \right),$$

over $\vartheta_1^{(n+1)} \leq \vartheta_2^{(n+1)} + \varepsilon \leq \dots \leq \vartheta_N^{(n+1)} + n\varepsilon$ for some small ε .

5. Set $n = n + 1$ and return to step 2 unless convergence has been declared.

3. Protonated glycine

We have studied the H/D exchange of protonated glycine with CH_3OD with our electrospray ionization-flow tube apparatus and deduced its site specific rate constants [9]. The present data are the result of a repeat experiment and are presented in Fig. 1. Analysis of the data with the new algorithm (see Fig. 1) leads to one site-specific rate constant,

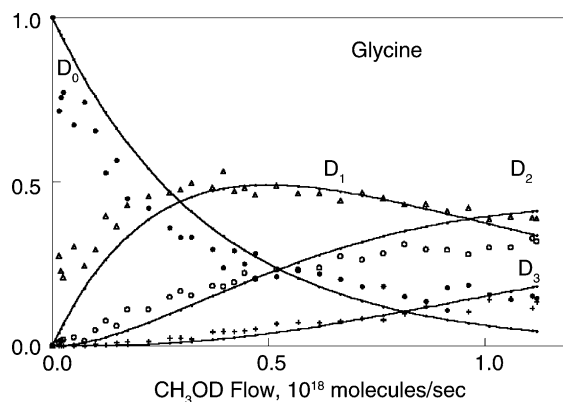


Fig. 1. Relative abundance vs. neutral flow rate in molecules/s for the various indicated cations with no exchange (D_0), one exchange (D_1), etc. in the reaction of protonated glycine with CH_3OD . The experiment was carried out on the ESI-flow tube. The points are experimental; the continuous curves are the results of the simulation based on the algorithm described in this paper that minimizes the Kullback–Leibler (KL) divergence (see text).

7.6×10^{-11} cc/molecule s and three equal site-specific rate constants, 1.5×10^{-11} cc/molecule s each. These results are in reasonable agreement with the previous results of Green and Lebrilla [2] according to which there is one fast-reacting site, $k_a = 6.6 \times 10^{-11}$ cc/molecule s and three equivalent slow sites, $k_b = 0.8 \times 10^{-11}$ cc/molecule s. Furthermore, these results support their proposed structure of glycineH^+ in which the proton is localized on the amino group, resulting in three equivalent amine hydrogens and one carboxyl hydrogen. The amino-protonated form is the consensus structure because the proton affinity of the NH_2 group is higher than that of the carboxyl group. Estimates based on model compounds suggest a difference in PA between the NH_2 and the carboxyl group of glycine of ~ 19.8 kcal/mol [10,11].

4. The proton-bound arginine dimer

Amino acids are known to exist as zwitterions in solution. However, in the gas phase even arginine, the most basic amino acid, exists in the non-ionic neutral configuration [12]. Calculations for the monomer show [13] that the neutral conformer

is energetically slightly more stable than the zwitterionic form. At the same time DFT calculations on the proton-bound dimer of arginine [14] have demonstrated a reversed stability pattern with the salt-bridge or ion–zwitterion form found to be more stable than the simple protonated or ion–molecule form.

We have applied recently [15] the ESI-fast flow technique to study the gas phase H/D exchange reactions of protonated monomers and dimers of L-arginine with ND_3 and CH_3OD . The structural interpretation of the H/D exchange data even in systems as small as protonated arginine and its dimer turned out to be complex and the exchange mechanism has been discussed in some detail [15]. Whereas, CH_3OD has the advantage over ND_3 of providing only one exchangeable deuterium per collision, it is a rather inefficient exchange reagent and leads to the exchange of only a small fraction of the labile hydrogens in the protonated arginine monomer and dimer.

The protonated arginine dimer, $(\text{arginine})_2\text{H}^+$ has 15 labile hydrogens. To extract site-specific rate constants, we adopted an algorithm [15] that solves a set of independent simultaneous differential equations for a suggested reaction mechanism by the Runge–Kutta method. This worked for the monomer but failed for the dimer.

The H/D exchange experiment for $(\text{arginine})_2\text{H}^+$ with ND_3 was repeated with the 9.4 Tesla Q-FT-ICR instrument at The National High Magnetic Field Laboratory in Tallahassee, FL. The experimental results and the simulated fits based on the new algorithm are presented in Fig. 2. Fifteen site-specific pseudo-first-order rate constants were extracted as follows (in units of s^{-1}): 1.9×10^{-2} (1), 4.4×10^{-2} (1), $(8.2\text{--}8.6) \times 10^{-2}$ (12), and 2.8×10^{-1} (1). The ND_3 pressure has been estimated to be less than 10^{-6} Torr in the FT-ICR experiment but is not known accurately due to inherent difficulties in acquiring accurate pressure measurements in an ICR cell [16], and second-order absolute rate constants could not be directly deduced. The experiment was repeated on the electrospray ionization-flow tube apparatus (Fig. 3) and the 15 corresponding second-order rate constants found are (in units of $\text{cc}/\text{molecule s}$): 1.8×10^{-12} (1), 2.2×10^{-12} (1), 4.1×10^{-12} (1),

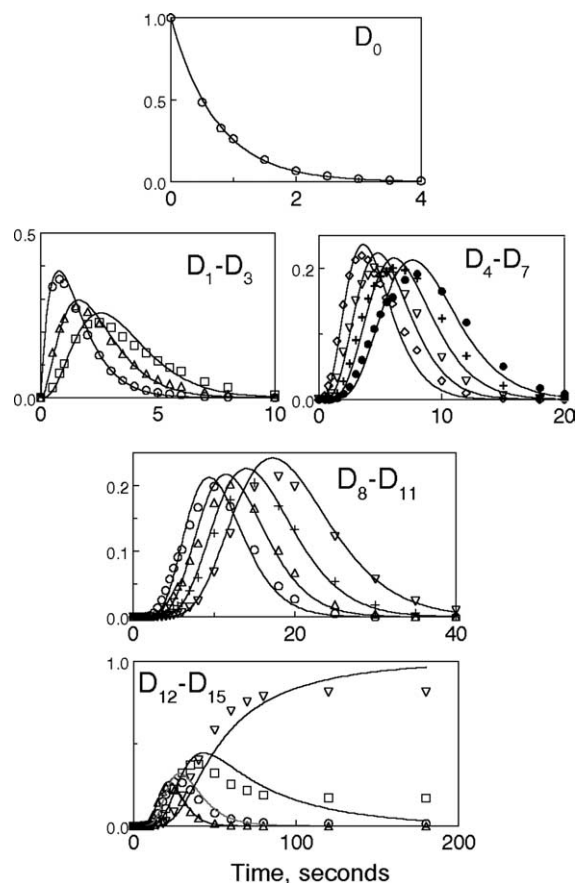


Fig. 2. Relative abundance vs. time (in s) for the 15 consecutive exchanges in the reaction of the protonated arginine dimer, $(\text{arginine})_2\text{H}^+$, with ND_3 . The experiment was carried out on the 9.4-T FT-ICR mass spectrometer at The National High Magnetic Field Laboratory in Tallahassee, FL. Monoisotopic ions were isolated for H/D exchange by stored waveform inverse Fourier transform (SWIFT) excitation. The points are experimental; the continuous curves are the results of the simulation based on the algorithm described in this paper that minimizes the KL divergence (see text). The normalization factor between the FT-ICR and flow-tube rates (see text) indicates an ND_3 pressure of 4.65×10^{-7} Torr in the FT-ICR experiment.

$(5.8\text{--}6.6) \times 10^{-12}$ (11), and 1.4×10^{-11} (1). The sum of these 15 site-specific rate constants, $9.0 \times 10^{-11} \text{ cc}/\text{molecule s}$, is in agreement, within experimental error, with the overall rate constant deduced from a semilog plot of the ESI-flow tube data [15], $(1.25 \pm 0.4) \times 10^{-10} \text{ cc}/\text{molecule s}$. By normalizing the sum of the pseudo-first-order rate

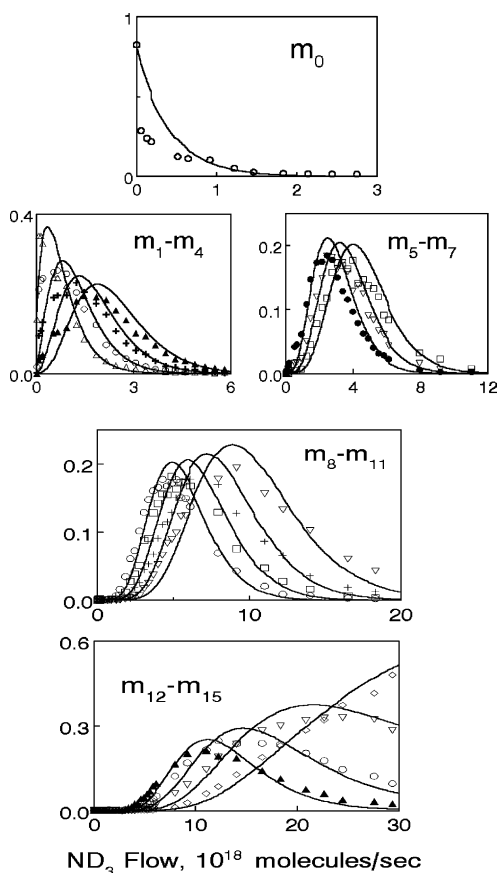


Fig. 3. Relative abundance vs. neutral flow rate in molecules/s for the various indicated cations for the 15 consecutive exchanges in the reaction of the protonated arginine dimer, (arginine)₂ H⁺, with ND₃. The experiment was carried out on the ESI-flow tube. The reaction time is 12.24 ms; helium carrier gas flow: 4.9 l/min; flow tube pressure: 0.286 Torr. The points are experimental; the continuous curves are the results of the simulation based on the algorithm described in this paper that minimizes the KL divergence (see text). The results are presented as m_0 , m_1 , etc. because the ions contain contributions from ¹³C, ¹⁵N and other natural isotopes in addition to the H and D isotopes. Convolutions with the natural isotopic abundance and the experimental peak shape were carried out for comparison of the simulated data with experiments.

constants from the FT-ICR experiment (1.36 s^{-1}) to the sum of the second-order flow tube results ($9 \times 10^{-11} \text{ cc/molecule s}$) we were able to deduce the following 15 site-specific rate constants from the FT-ICR experiment (in units of cc/molecule s): 1.3×10^{-12} (1), 2.9×10^{-12} (1), $(5.4 - 5.7) \times 10^{-12}$ (12), and 1.9×10^{-11} (1).

A mechanism that enables the exchange of all the 15 labile hydrogens of the protonated dimer albeit with relatively low rate constants is the ‘relay’ exchange mechanism [15]. Explicit assumptions are that exchange occurs independently at each site and that isotope effects in the relay mechanism are neglected. The H/D exchange data have demonstrated that a very pronounced change in structure occurs on going from the monomer to the dimer – the change from a protonated neutral arginine for the monomer to a charged zwitterionic species stabilized by a salt-bridge for the dimer. In the monomer, protons on the guanidyl group remain unexchanged. However, for the zwitterionic dimer there is a carboxylate group onto which the proton can be transferred. After this step one of the guanidyl groups becomes deprotonated, so it can be an acceptor for the next step, and in this way those protons can be exchanged as well. The fact that there is just a single high site-specific rate constant is in agreement with the ion–zwitterion structure calculated for (arginine)₂H⁺ [14] and may be ascribed to the single carboxyl hydrogen present. This result is an important outcome of the present analysis that was missing in our previous interpretation of the H/D exchange data [15].

The new algorithm described here will be used in the future on other protonated peptides and proteins to deduce large numbers of site-specific rate constants that could not be deduced to date with other existing procedures.

Acknowledgements

This work was supported by the United States-Israel Binational Science Foundation under Grant No. 2000026, by the NSF National High Field FT-ICR Facility (CHE-99-09502), Florida State University, and the National High Magnetic Field Laboratory in Tallahassee, FL. Prof. Reuben thanks the Royal Society for a study grant. We thank Prof. Assa Lifshitz and Dr. Julia Laskin for helpful discussions. The Farkas Research Center is supported by the Minerva Gesellschaft für die Forschung GmbH München.

References

- [1] F. He, A.G. Marshall, *J. Phys. Chem. A* 104 (2000) 562.
- [2] M.K. Green, C.B. Lebrilla, *Mass Spectrom. Rev.* 16 (1997) 53.
- [3] E. Levy-Seri, G. Koster, A. Kogan, K. Gutman, B.G. Reuben, C. Lifshitz, *J. Phys. Chem. A* 105 (2001) 5552.
- [4] D. Suckau, Y. Shi, S.C. Beu, M.W. Senko, J.P. Quinn, F.M. Wampler III, F.W. McLafferty, *Proc. Natl. Acad. Sci. USA* 90 (1993) 790.
- [5] S.J. Valentine, D.E. Clemmer, *J. Am. Soc. Mass Spectrom.* 13 (2002) 506.
- [6] P.J. Bickel, K. Doksum, in: *Mathematical Statistics*, vol. I, Prentice-Hall, Upper Saddle River, 2001, p. 116.
- [7] A.P. Dempster, N.M. Laird, D.B. Rubin, *J. R. Statist. Soc. B* 39 (1977) 1.
- [8] Y. Vardi, D. Lee, *J. R. Statist. Soc. B* 55 (1993) 569.
- [9] P. Ustyuzhanin, J. Ustyuzhanin, C. Lifshitz, *Int. J. Mass Spectrom.* 223/224 (2003) 491.
- [10] E. Gard, M.K. Green, J. Bregar, C.B. Lebrilla, *J. Am. Soc. Mass Spectrom.* 5 (1994) 623.
- [11] D.H. Aue, M.T. Bowers, in: M.T. Bowers (Ed.), *Gas Phase Ion Chemistry*, vol. 2, Academic Press, New York, 1979, p. 1.
- [12] C.J. Chapo, J.B. Paul, R.A. Provencal, K. Roth, R.J. Saykally, *J. Am. Chem. Soc.* 120 (1998) 12956.
- [13] Z.B. Maksić, B. Kovačević, *J. Chem. Soc. Perkin Trans. 2* (1999) 2623.
- [14] E.F. Strittmatter, E.R. Williams, *J. Phys. Chem. A* 104 (2000) 6069.
- [15] O. Geller, C. Lifshitz, *J. Phys. Chem. A* 107 (2003) 5654.
- [16] F. He, A.G. Marshall, M.A. Freitas, *J. Phys. Chem. B* 105 (2001) 2244.