ISSN 1061-9348, Journal of Analytical Chemistry, 2014, Vol. 69, No. 13, pp. 1270–1277. © Pleiades Publishing, Ltd., 2014.
Original Russian Text © V.V. Raznikov, M.O. Raznikova, 2013, published in Mass-spektrometria, 2013, Vol. 10, No. 3, pp. 175–181.

ARTICLES

Decomposition of Multidimensional Charge State Distributions of Ions Produced by Electrospray Ionization of Bioorganic Compounds. Part 1: Basic Formalism and Implementation of the Method

V. V. Raznikov[a] and M. O. Raznikova[b, *]

[a] *The Branch of Talrose Institute for Energy Problems of Chemical Physics, Russian Academy of Sciences, Chernogolovka, Moscow oblast, 142432 Russia*

[b] *Institute of Problems of Chemical Physics, Russian Academy of Sciences, Chernogolovka, Moscow oblast, 142432 Russia*

**e-mail: mari@icp.ac.ru*

Received June 17, 2013; in final form, July 23, 2013

Abstract—Initially our method of the decomposition of charge state distributions of biopolymer ions was designed to estimate the probabilities of the retention of charges by biopolymer ionogenic groups without distinction of different charge carriers (proton, alkali metal ions, or other ions). The analysis of mass spectra of several peptides and proteins produced by electrospray ionization carried out on this basis made possible some conclusions about their ion structure and behavior inside the ion source. A generalization of the method for the case of multidimensional charge state distributions corresponding to the retention of various charge carriers by ions of biomolecules is described. Hopefully, this will make possible gaining more substantial information concerning the structure and behavior of biomolecule ions in solutions and under electrospray conditions, which can be important for various biological applications.

Keywords: electrospray, biomolecules, multiply charged ions, charge state distribution, decomposition of distributions, ionogenic groups, probabilities of charge retention

DOI: 10.1134/S1061934814130073

INTRODUCTION

Electrospray ionization (**ESI**) [1, 2] is one of the most important methods of mild ionization and widely used for structural and conformational studies of biomolecules. The mass spectra of rather big molecules usually contain series of peaks of multiply charged ions with the mass m close to the molecular mass of the molecule. The presence of series of peaks of multiply charged ions particularly in the analysis of biopolymer mixtures can cause difficulties in calculations of molecular masses of individual compounds. In some cases these difficulties are overcome using various algorithms of the so-called deconvolution of multiple charge state distributions, including the method of maximum entropy [3].

Attempts to relate the intensity distributions of multiply charged ions (or charge state distributions) with the structural peculiarities of biomolecules and processes occurring in the electrospray ion source were made in a series of works [4–6]. Among these were correlations between the maximum number of the observed charges in the mass spectrum and the number of ionogenic groups in a biomolecule, the presence and interconversions between the native and denaturated forms of the studied biopolymer in solution or in an ion source, etc. In some works the expansion of the experimental charge state distribution into Gaussian components was performed and, on this basis, a conclusion about the number and contributions of the structural forms of the studied biopolymer present in solution (or in ion source) was made [7, 8]. Felitsyn et al. in [9] approximated distribution by the number of sodium ions retained by ions of the studied biopolymer of a certain charge by a binomial distribution based on a certain estimate of the averaged probability of their retention. However, the difference between the observed and calculated distributions in this case appeared rather great. The question to which degree the results obtained corresponded to the properties of the studied biopolymers in solutions and what features are determined by the electrospray process and the transfer of ions to the gas phase remains quite unclear.

While considering the behavior of biomolecule ions in solutions analysts usually suppose that charges in these ions are retained by specific functional groups, in other words, by ionogenic sites or charged centers of the molecule which can exchange charge carriers with the environment. It seems natural to expand this

model to the processes of ion formation in spraying biopolymer solutions. However, an alternative assumption was also made about a more or less uniform distribution of charges on the external surface of a native biomolecule arising in the electrospray process [6]. Essentially, this is a consequence of the known residue charge model, dating back to Dole work [10]. In the work by Kebarle [11], this model was named Dole's Charge Residue Model, CRM, and cited work [10]; however, we did not succeed to find the comprehensive description of this model in [10]. In this model, as it was described in [11], it was assumed that the measured charge of a bioion in electrospraying is equal to the charge close to the Rayleigh limit of the stability of the droplet surface with the radius equal to that of the bioion. Hence, a combination of a certain version of this model with the model of ion evaporation [5], in which small ions of a solvent, for example, solvated H_3O^+ at the final stage of the evaporation of a droplet containing a bioion can leave this droplet, competing for proton with ionogenic sites of the bioion, seems to be more realistic. The possibility of the evaporation of small ions was also mentioned in [10]. It should be noted that, in this work, the electrospray of solutions of macromolecules leading to the transfer of ions of these macromolecules into the gas phase was performed for the first time and the term electrospray was introduced.

In contrast to the method we described previously, i.e., the decomposition of one-dimensional charge state distributions leading to the estimation of the probabilities of proton retention by single sites of polyprotonated biomolecules [12], in this case we assume a possibility of the retention of one charge carrier of another type (for example, Na^+, K^+, Ca^{2+}, etc.) along with a proton by each ionogenic site. In considering the types of charge carriers, multidimensional charge state distributions over all combinations of numbers of retained carriers measured in this case, as was demonstrated in [11], can bear more substantial information about the structure of the bioion and ion formation in solution, for example, in the presence of sodium ions in solution, by estimating the number of acidic groups on the bioion surface. If the ion formation mechanism for acidic groups described in [11] is applicable, the decomposition of multidimensional charge state distributions with a change in the composition of solution can provide quantitative estimates of ion equilibrium constants for acidic groups on the bioion surface in solution, provided that these groups are virtually neutral in solution but not necessarily protonated, or where protons can be substituted by other charge carrier. In this case, the state of these groups can hardly have enough time to change during the electrospay process.

The aim of this work was the development of a multidimensional version of the decomposition of charge distributions of biomolecule ions obtained in electrospray ion source and the demonstration of some capabilities of this model.

THEORETICAL BACKGROUND OF THE METHOD

In considering several types of charge carriers, the retention of a certain charge carrier by one or another ionogenic group or the absence of such carriers should be considered as an event with the number of possible outcomes greater than two. For the convenience of the further analysis, let us consider that these outcomes are estimated for initially neutral basic groups and initially negatively charged (detached proton) acidic groups of a biomolecule. Thus, the set of observed multiply charged ions for a given biomolecule can be presented as a multidimensional distribution of the intensities of their peaks by the number of charge carriers retained by the given ion. In this case, the charge of such an ion will be equal to the sum of charges of the retained carriers minus the total number of the acidic ionogenic groups.

If we normalize such distribution for the unit sum of intensities of corresponding ion peaks, we can consider it as an estimation of a distribution of probabilities for the retention of the specified numbers of charge carriers of each type by the given biomolecule. A possible form of the presentation of such distribution can be a polynomial, in which the symbols of the corresponding charge carriers, for example, H, Na, K, etc., are used as formal variables. In this case, for example, notation $p_{000} + p_{100}\mathrm{H} + p_{200}\mathrm{H}^2 + p_{210}\mathrm{H}^2\mathrm{Na} + p_{032}\mathrm{Na}^3\mathrm{K}^2$ means that the biomolecule does not retain charge carriers with the probability p_{000}, it retains one proton with the probability p_{100}, two protons with the probability p_{200}, two protons and one sodium ion with the probability p_{210}, and three sodium ions and two potassium ions with the probability p_{032}. Thus, index of power at a formal variable indicates the number of retained carriers of a corresponding type. The absence of any combination of formal variables in such notation is equivalent to a zero coefficient at this combination and means that such ions do not present in the mass spectrum.

Convenience of the proposed presentation of the distributions under consideration is that, in the case when the events of charge retention by individual ionogenic sites are considered statistically independent, the notation of the resulting multidimensional charge state distribution for a biomolecule will be the product of polynomial notations of such distributions of these sites. For a single ionogenic site in the biomolecule, in this case the three-dimensional charge distribution for the studied charge carriers of its ions can be presented as ${}^1_0P + {}^1_\mathrm{H}P\mathrm{H} + {}^1_\mathrm{Na}P\mathrm{Na} + {}^1_\mathrm{K}P\mathrm{K}$. Here ${}^1_0P$ is the probability of the absence of charges in this site, ${}^1_\mathrm{H}P$ is the probability of retaining proton, ${}^1_\mathrm{Na}P$ is the probability

of retaining a sodium ion, and ${}^{1}_{K}P$ is the probability of retaining a potassium ion. The sum of these probabilities is equal to 1. In future let us refer to such distribution in the general case for each site as to the distribution of charge retention (**DCR**) by this site, and to a set of many such distributions for all ionogenic sites of a biomolecule as to the multiple distribution of charge retention (mDCR). When the second ionogenic site is present, its DCR can be similarly written in the case under consideration: ${}^{2}_{0}P + {}^{2}_{H}P\mathrm{H} + {}^{2}_{Na}P\mathrm{Na} + {}^{2}_{K}P\mathrm{K}$. A three-dimensional charge state distribution at the independent retention of no more than one of three charge carriers by each of two sites can be written as the product of their DCR notations considered as polynomials of the formal variables H, Na, and K:

$$\begin{aligned}
&({}^{1}_{0}P + {}^{1}_{H}P\mathrm{H} + {}^{1}_{Na}P\mathrm{Na} + {}^{1}_{K}P\mathrm{K})\\
&\times ({}^{2}_{0}P + {}^{2}_{H}P\mathrm{H} + {}^{2}_{Na}P\mathrm{Na} + {}^{2}_{K}P\mathrm{K})\\
&{}^{1}_{0}P\,{}^{2}_{0}P + ({}^{1}_{H}P\,{}^{2}_{0}P + {}^{2}_{H}P\,{}^{1}_{0}P)\mathrm{H} + ({}^{1}_{Na}P\,{}^{2}_{0}P + {}^{2}_{Na}P\,{}^{1}_{0}P)\mathrm{N}\\
&+ ({}^{1}_{K}P\,{}^{2}_{0}P + {}^{2}_{K}P\,{}^{1}_{0}P)\mathrm{K} + {}^{1}_{H}P\,{}^{2}_{H}P\mathrm{H}^2 \qquad (1)\\
&+ ({}^{1}_{H}P\,{}^{2}_{Na}P + {}^{2}_{H}P\,{}^{1}_{Na}P)\mathrm{HNa}\\
&+ ({}^{1}_{H}P\,{}^{2}_{K}P + {}^{2}_{H}P\,{}^{1}_{K}P)\mathrm{HK} + {}^{1}_{Na}P\,{}^{2}_{Na}P\mathrm{Na}^2\\
&+ ({}^{1}_{Na}P\,{}^{2}_{K}P + {}^{1}_{K}P\,{}^{2}_{Na}P)\mathrm{NaK} + {}^{1}_{K}P\,{}^{2}_{K}P\mathrm{K}^2.
\end{aligned}$$

In this case, we take into account all versions of the retention of charge carriers, multiplication of probabilities at the independent execution of events, and their addition when the events are incompatible. Such calculations demonstrate that the proposed method of getting notations of charge distributions will be also valid for three sites and also for any number of sites independently retaining charge carriers in a bioion. To perform an inverse operation in certain approximation, i.e., to factorize the polynomial representing a registered multidimensional charge distribution into factors linear with respect to formal variables means to decompose this distribution. As a result, we will obtain estimates of the distributions of retention of charges for all ionogenic sites of the biomolecule (mDCR).

A strong evidence for such a correlation between the assumed distributions of charge retention for individual sites (mDCR) and the observed multidimensional charge state distribution for a biomolecule in a general case for any number of sites and any number of types of charge carriers can be obtained using the mathematical induction method. This property of multidimensional charge state distributions is the generalization of the formalism of generating functions [13], which we used to decompose one-dimensional charge distributions of polyprotonated biopolymer ions [12]. A similar approach involving calculations of the product of polynomials of symbols of isotopic atoms was used to calculate versions of isotopic combinations for calculating isotopic distributions of peaks in mass spectra of complex molecules [14]. In this case, no coefficients at symbols of atoms were used initially.

Of course, the assumption of the complete independence of events of the retention of charge carriers by various sites of the biomolecule can be true only with a certain degree of approximation. For one-dimensional charge state distributions, this problem in sufficient detail was discussed previously [12]. The presented considerations and the results of analysis of experimental mass spectra of multiply charged ions on this basis demonstrate that the situation is not as bad as it could seem first. Thus, for example, it appeared that if such dependence could be reduced to pairwise interactions of charge carriers on closely lying sites, the formalism of independent charge retentions works, though the mutual influence of closely lying charged sites should be taken into account in the interpretation of the results of calculations.

Notation (1) can be considered as a generating function (of three variables) for two ionogenic sites capable of retaining proton or sodium or potassium ions. For the general case of n types of charge carriers corresponding to x_i variables and N charge-retaining sites in the biomolecule, the generating function in notations slightly different from those used in (1) will take the form:

$$\begin{aligned}
&\sum_{k_1, \ldots, k_n = 0}^{k_1 + \ldots + k_n = N} J_{k_1 \ldots k_n} x_1^{k_1} \ldots x_n^{k_n}\\
&\approx \sum_{k_1, \ldots, k_n = 0}^{k_1 + \ldots + k_n = N} p_{k_1, \ldots, k_n} x_1^{k_1} \ldots x_n^{k_n} \qquad (2)\\
&= \prod_{i=1}^{N} \left(1 + \sum_{s=1}^{n} {}^{i}P_s (x_s - 1) \right).
\end{aligned}$$

Here the left side is an *approximate* expression for the generating function via normalized to unit sum the measured intensities of mass-spectral peaks $J_{k_1, \ldots, k_n}$, corresponding to the retention k_1 carriers of the first type, k_2 carriers of the second type, …, and k_n carriers of the type n by the ion. Expression on the right side is the presentation of the generating function via the product for composition of distributions of charge retentions by all ionogenic sites of the biomolecule. We took into account that the sum of probabilities of the retention of all possible charge carriers by a certain site and the probabilitiy of the absence of charge is equal to unity. Therefore, the notation of this last probability ${}^{i}P_0$ is absent in the expression and the value

$1 - \sum_{s=1}^{n} {}^{i}P_s$ was used instead of it. The problem is to estimate the values ${}^{i}P_s$ in the multipliers on the right side of expression (2) or, in other words, to factorize approximately the polynomial on the left side expression (2) into linear terms. As the order of multipliers on the right side of expression (2) is of no importance, this indicates that the set of ionogenic sites of the biomolecule with their probabilities of retaining charge carriers is an unordered set. Calculation of the product on the right side of expression (2) gives the sum in the middle part of Eq. (2) and transforms these into an ordered $p_{k_1, \ldots, k_n}$. The values ${}^{i}P_s$ in the multipliers of the right side of Eq. (2) can be found, for example by minimizing the sum of squared deviations of coefficients at $x_1^{k_1} \ldots x_n^{k_n}$ on the left and right sides of Eq. (2):

$$\mathrm{coef}_{k_1 \ldots k_n} \sum_{l_1, \ldots, l_n = 0}^{l_1 + \ldots + l_n = N} J_{l_1 \ldots l_n} x_1^{l_1} \ldots x_n^{l_n} = J_{k_1 \ldots k_n}$$
$$= \frac{\partial^{k_1 + \ldots + k_n}}{k_1! \ldots k_n! \partial x_1^{k_1} \ldots \partial x_n^{k_n}}$$
$$\times \sum_{l_1, \ldots, l_n = 0}^{l_1 + \ldots + l_n = N} J_{l_1 \ldots l_n} x_1^{l_1} \ldots x_n^{l_n} \Big|_{x_1 = 0, \ldots, x_n = 0},$$

$$\mathrm{coef}_{k_1 \ldots k_n} \prod_{i=1}^{N} \left(1 + \sum_{s=1}^{n} {}^{i}P_s (x_s - 1)\right) = p_{k_1, \ldots, k_n}$$
$$= \frac{\partial^{k_1 + \ldots + k_n}}{k_1! \ldots k_n! \partial x_1^{k_1} \ldots \partial x_n^{k_n}}$$
$$\times \prod_{i=1}^{N} \left(1 + \sum_{s=1}^{n} {}^{i}P_s (x_s - 1)\right) \Big|_{x_1 = 0, \ldots, x_n = 0}.$$

The condition that the first-order derivatives of ${}^{j}P_s$ with respect to the unknown quantities for a certain site (numbered j) are equal to zero at the known values of ${}^{k}P_s$ for all other sites ($k \neq j$) leads to a linear problem. The system of equations for determining the probabilities ${}^{j}P_s$ of retaining corresponding charge carriers by site j at the given such probabilities for the other sites can be written as follows:

$$\begin{pmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \ldots & a_{nn} \end{pmatrix} \begin{pmatrix} {}^{j}P_1 \\ \vdots \\ {}^{j}P_n \end{pmatrix} = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}, \quad (3)$$

where

$$a_{pq} = \sum_{k_1, \ldots, k_n = 0}^{k_1 + \ldots + k_n = N} \left[\mathrm{coef}_{k_1 \ldots k_n} (x_p - 1) \times \prod_{i=1; i \neq j}^{N} \left(1 + \sum_{s=1}^{n} {}^{i}P_s (x_s - 1)\right) \right]$$
$$\times \left[\mathrm{coef}_{k_1 \ldots k_n} (x_q - 1) \prod_{i=1; i \neq j}^{N} \left(1 + \sum_{s=1}^{n} {}^{i}P_s (x_s - 1)\right) \right],$$

$$c_p = \sum_{k_1, \ldots, k_n = 0}^{k_1 + \ldots + k_n = N} \left[J_{k_1 \ldots k_n} - \mathrm{coef}_{k_1 \ldots k_n} \times \prod_{i=1; i \neq j}^{N} \left(1 + \sum_{s=1}^{n} {}^{i}P_s (x_s - 1)\right) \right]$$
$$\times \mathrm{coef}_{k_1 \ldots k_n} (x_p - 1) \prod_{i=1; i \neq j}^{N} \left(1 + \sum_{s=1}^{n} {}^{i}P_s (x_s - 1)\right).$$

Thus, the complete procedure of determining the probabilities of retaining charges by all ionogenic sites of the biomolecule will be iterative, starting from a certain initial set of values of these probabilities preset, for example, using a random number generator. As the target function, which is the sum of squared deviations of the experimental and calculated charge state distributions, at each iteration can only decrease and is limited from below (≥ 0), the iterative procedure converges and its result will be determined by the initial values of the target quantities. In the one-dimensional case implemented earlier, the scattering of the obtained estimates of probabilities of retaining proton at a random selection of the initial values of the descent could be quite significant (at a level of 0.1 and greater), so that an acceptable accuracy of the results could be obtained only using averaging. In this case, the lack of ordering of ionogenic sites was overcome by the transposition of the found probabilities of protonation for each initial set of the required probabilities in the descending order. For the multidimensional case, such ordering is impossible. However, in this case the importance of averaging the values obtained was not obvious. For polyprotonated molecules, the number of determined parameters actually coincides with the number of measured peak intensities. For multidimensional charge state distributions, the number of measured components can be many times greater than the number of target quantities. Thus, for example, for eight sites and two types of charge carriers, the number of determined parameters is 16, while the number of measured quantities is almost 3 times greater: $(9 \times 9 + 9)/2 - 1 = 44$. The calculations are illustrated in Fig. 1. For polyprotonation at the independent retention of protons by ionogenic sites, the

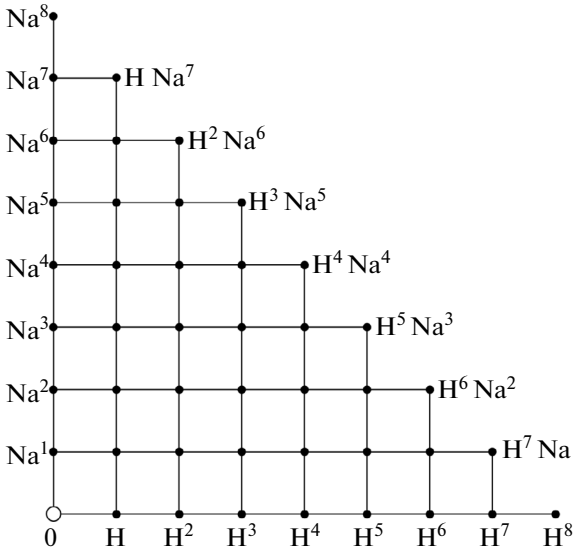

Fig. 1. Illustration of the calculation of the number of members of charge state distributions for eight ionogenic sites in a virtual biomolecule retaining protons or sodium ions. Black points correspond to the measured members of charge state distribution.

conditions of the central limit theorem can be fulfilled [13]. Then, with an increase in the number of such sites, the charge state distributions should be approximated by normal distributions with improving accuracy; the latters are determined by two parameters: the mean value and standard deviation. Thus, with an increase in the number of ionogenic sites, the accuracy of estimating retention probabilities should get worse for this reason as well. For multidimensional distributions, the situation also seems more favorable.

IMPLEMENTATION OF THE METHOD, DISCUSSION ON POSSIBLE VERSIONS AND LIMITATIONS

Tests of the presented scheme of the decomposition of multidimensional charge state distributions demonstrated the necessity of averaging the results obtained, because in many cases significant scattering of the obtained estimates of probabilities of charges retention for different initial values of their search was obtained; it was comparable with similar scatter for one-dimensional case. At the same time, in estimating such probabilities for a virtual bioion with completely identical sites, simple averaging resulted in a very high accuracy. However, we succeeded to obtain such results only after the introduction of randomization of the iteration process of the search for a minimum, in contrast to the earlier used one-dimensional version of the decomposition of charge state distributions. In this version, the numbers of optimized sites at every iteration step were set deterministically and the direct averaging for identical sites even without ordering of the results obtained led to noticeable errors of the result. The random selection of optimized site numbers in the new version of decomposition solved the problem both for one-dimensional and multidimensional cases.

To average unordered sets of probabilities of charge retention in a general case, one can use an analogue of this procedure, accepted in statistics of nonnumeric data [15]. The starting point is the property of regular mean values, which involves the minimization of the sum of squared distances from the averaged data y_k to the average value $\bar{y}$:

$$\bar{y} = \frac{1}{n}\sum_{k=1}^{n} y_k \Rightarrow \min_{\alpha}\sum_{k=1}^{n}(y_k - \alpha)^2 \Rightarrow \alpha = \bar{y}. \qquad (4)$$

Thus, if the concept of the distance between the sets of data is defined or, in other words, a data space metric is introduced, one can try to find the point with the least mean-square deviation from the set of the data obtained and to accept this point as the mean value of these data. The right side of expression (2) converts the unordered set of distributions of charge retentions (mDCR) ${}^{j}P_s$ into an ordered sequence of values $p_{k_1,\ldots,k_n}$, which naturally gives a possibility of making the space (mDCR) a metric one with the squared distance between the distributions ${}^{j}_{1}P_s \to {}_1p_{k_1}, \ldots, k_1$ and ${}^{j}_{2}P_s \to {}_2p_{k_1}, \ldots, k_1$, equal to the sum of squared deviations ${}_1p_{k_1}, \ldots, k_1$ from ${}_2p_{k_1}, \ldots, k_1$:

$$\rho^2({}^{j}_{1}P_s, {}^{j}_{2}P_s) = \sum_{k_1,\ldots,k_n=0}^{k_1+\ldots+k_n=N} ({}_1p_{k_1,\ldots,k_n} - {}_2p_{k_1,\ldots,k_n})^2. \qquad (5)$$

If we obtain a particular set of L values ${}^{j}_{l}P_s \to {}_lp_{k_1,\ldots,k_n}$, we can calculate the average charge state distribution corresponding to this set:

$$\bar{p}_{k_1,\ldots,k_n} = \frac{1}{L}\sum_{l=1}^{L} {}_lp_{k_1,\ldots,k_n},$$

and among ${}^{j}_{l}P_s$ find ${}^{j}_{l_0}P_s = {}^{j}P_s^{\min 1}$, which has the minimum deviation from this average value:

$$\sum_{k_1,\ldots,k_n=0}^{k_1+\ldots+k_n=N} ({}_{l_0}p_{k_1,\ldots,k_n} - \bar{p}_{k_1,\ldots,k_n})^2 = \min_{l}\sum_{k_1,\ldots,k_n=0}^{k_1+\ldots+k_n=N} ({}_lp_{k_1,\ldots,k_n} - \bar{p}_{k_1,\ldots,k_n})^2. \qquad (6)$$

As was shown by tests on model data, the error of this estimate is slightly lower or comparable to that with the averaging of ordered data in one-dimensional case. However, it demonstrates much greater errors for completely identical sites compared to simple averaging of the found mDCRs without preliminary ordering of individual DCRs. In a general case, for the conve-

nience of the comparison of results, we ordered these DCRs by the increase in the probability of the absence of charge at a given site. As the benchmark for comparison we select mDCR ${}^{j}P_s^{\min 1}$ implementing the minimum of Eq. (6). It might be expected that some of these partially ordered mDCRs can be averaged together with ${}^{j}P_s^{\min 1}$, and for averaging other ones, one should transpose some of their sites. One can propose the following criterion of the acceptability of mDCR for averaging ${}^{j}_{l}P_s$ along with ${}^{j}P_s^{\min 1}$. Using a random number generator, we form a distorted mDCR ${}^{j}\tilde{P}_s = {}^{j}P_s^{\min 1} + {}^{j}\delta_s$, where ${}^{j}\delta_s$ are independent random variables uniformly distributed over the selected range of permissible errors $[-\delta, \delta]$ and satisfying the conditions

$$\sum_{s=1}^{n} {}^{j}P_s^{\min 1} + {}^{j}\delta_s \le 1 \text{ and } 0 \le {}^{j}P_s^{\min 1} + {}^{j}\delta_s \le 1.$$

Using the left side of Eq. (2) for ${}^{j}\tilde{P}_s$, let us calculate $\rho^2({}^{j}P_s^{\min 1}, {}^{j}\tilde{P}_s) = T_{\text{av}}$ keeping in mind Eq. (5) as a threshold of the acceptability of averaging ${}^{j}_{l}P_s$ along with ${}^{j}P_s^{\min 1}$.

$$\rho^2({}^{j}P_s^{\min 1}, ({}^{j}P_s^{\min 1} + {}^{j}_{l}P_s)/2) \le T_{\text{av}}. \quad (7)$$

Thus, the meaning of this criterion is in cutting off those found mDCRs that unacceptably far deviate from it in direct averaging with ${}^{j}P_s^{\min 1}$.

The criterion ${}^{j}_{l}P_s$ introduced for one-dimensional mDCRs sorted in ascending order of probabilities of the absence of charge appeared to be quite efficient. However, even in the presence of two possible charge carriers, the necessity of the transposition of some sites in the distribution ${}^{j}_{l}P_s$, to comply with selection criterion (7) arose quite often. As the exhaustive search of all site transpositions ($N!$) is unreal even for not very large number of sites N, we used a simplified approach to select an acceptable transposition. First a pair of sites in ${}^{j}_{l}P_s$ was found, ${}^{k}P_s^{\min 1}$, which differed from each other by the sum of squared deviations of the corresponding probabilities in DCR (j_1, k_1) to a lesser degree; then the second pair differing from (j_1, k_1) with the same property (j_2, k_2) was searched for, etc., until the last pair (j_n, k_n) not found previously was left. Then we transposes the sites in mDCR from place j_1 to k_1, from place j_2 to k_2, …, and from j_n to k_n, and checked the feasibility of criterion (7). If the result was successful, we included mDCR ${}^{j}_{l}P_s$ (after such transposition of sites) into averaging to obtain ${}^{j}P_s^{\text{av}1}$; in the opposite case, we proceeded to the next distribution.

This approach in tests on model data demonstrated in an average higher accuracy compared to the minimization method (6) described above. However, in most cases we did not succeed to obtain comparable accuracy in the direct averaging of distributions for identical sites. More progress was achieved in the organization of the second cycle of optimization, when two randomly selected sites for the initial mDCR, starting from ${}^{j}P_s^{\min 1}$ randomly approached each other up to their coincident instead of randomly specifying the initial descent point.

Then, the mDCR found as a result of iteration procedure (3) was tested for satisfaction of averaging criterion (7) including, if necessary, the procedure of site transposition described above. If criterion (7) is satisfied, the found mDCR is included into the procedure of new averaging to obtain ${}^{j}P_s^{\text{av}2}$ and becomes the initial approximation for the continuation of optimization. Otherwise, the initial distribution remained the same. In order that the whole second optimization procedure could converge in any case, the total number of attempts was limited by a certain predetermined number. If the desired accuracy to which criterion (7) must be met appeared to be too severe (δ is small) and, therefore, the number of averaged distributions seems to be insufficient, this accuracy is reduced (δ is increased) and recalculations are performed until an acceptable number of averaged distributions is obtained.

In selecting distributions for averaging, a quite sufficient amount of these distributions can be rejected and not used in the construction of an averaged distribution. Therefore, it seems reasonable to use all mDCRs generated in the second optimization cycle to find a distribution minimally deviated from the average charge state distribution obtained in the first optimization cycle, similarly to Eq. (6), ${}^{j}P_s^{\min 2}$.

As in some cases the problem of the decomposition of charge state distributions can be an ill-defined problem or the available data may be insufficient for obtaining a stable result, and one of possible approaches for overcoming of such difficulties is the maximum entropy method mentioned above, the search for mDCRs from both optimization cycles with the maximum entropy ${}^{j}P_s^{\text{mEnt}}$ was also included into the general procedure of multidimensional decomposition:

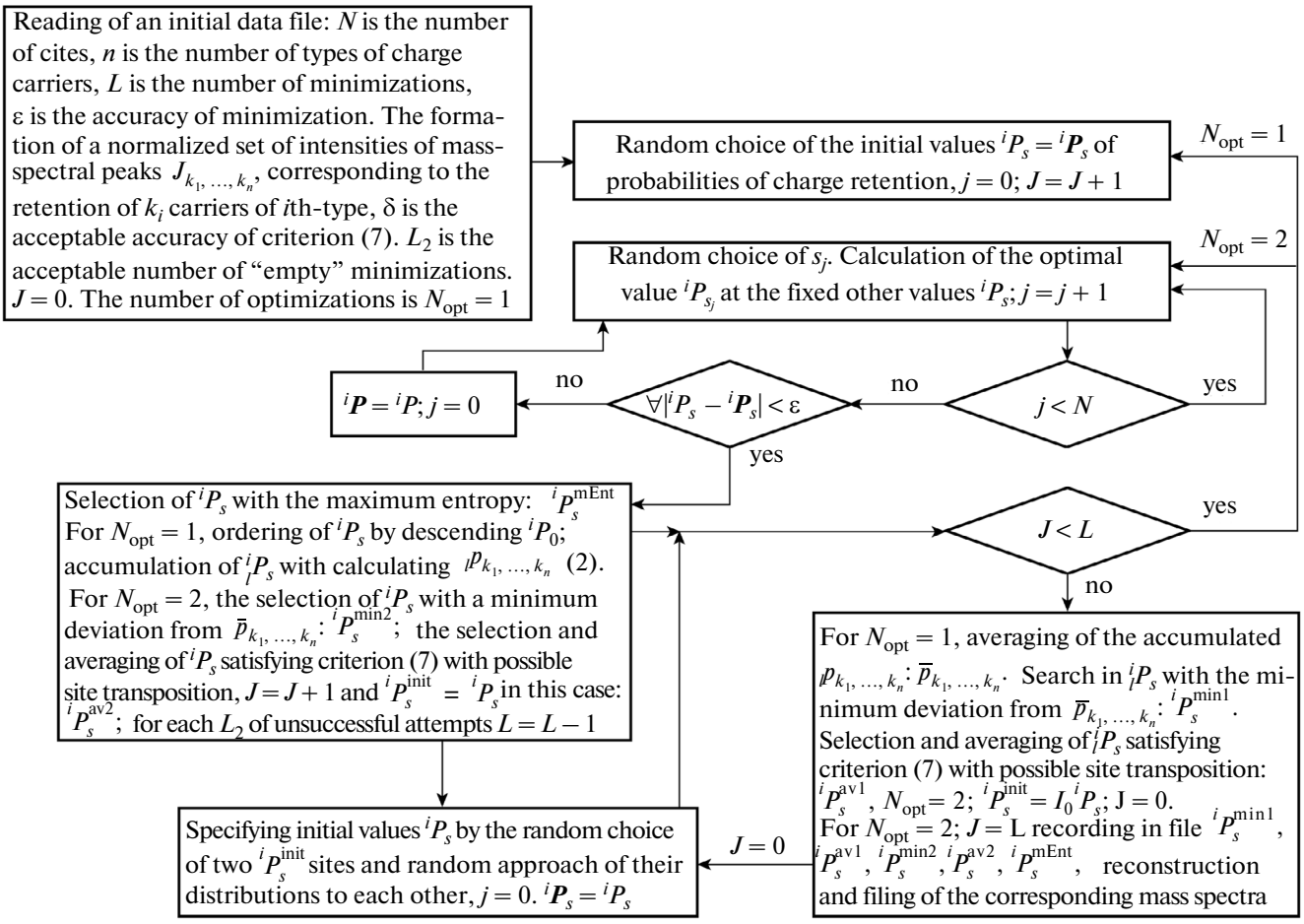

Fig. 2. Simplified block diagram of the procedure for the decomposition of multidimensional charge state distributions.

$$\max_k \sum_{s=1}^{N}\sum_{j=0}^{n} -{}^j_kP_s \ln {}^j_kP_s = \sum_{s=1}^{N}\sum_{j=0}^{n} -{}^jP_s^{mEnt} \ln {}^jP_s^{mEnt} = S({}^jP_s^{mEnt}). \quad (8)$$

The selection of mDCR with the maximum entropy actually means that the preferable mDCR is most close to the equally probable one and consistent with the experimental data.

Figure 2 presents a simplified block diagram of the procedure for the decomposition of multidimensional charge state distributions described above. In this diagram the main procedures and sequence of their execution are schematically presented. Some auxiliary functions are not shown. In particular, modules for modeling of the initial data by the given set of probabilities of charge retention by the specified set of virtual ionogenic sites are not presented. Calculations of error estimates for the obtained values of probabilities of the retention of charge carriers are also omitted. The designations of the estimated values were given above, as well as the details of the calculation procedures.

CONCLUSIONS

An expansion of the previously developed approach to the decomposition of charge state distributions of polyprotonated biomolecules to the case of probable attachment of other possible charge carriers to these biomolecules was made. The problem was reduced to an iteration process with minimizing the sum of squared deviations between the measured intensities of mass-spectral peaks, corresponding to the retention of all possible sets of charge carriers by a given biomolecule and predicted by the model of independent retentions of one of these carriers by each ionogenic site of the molecule. To improve the accuracy of the result, four versions of averaging of the obtained distributions of charge retention by each site with varying the initial values for the search of minimum were performed. The selection of these distributions maximizing their total entropy was also included.

For the efficient implementation of the method in systems for the registration and analysis of mass spectrometric information, the development of a procedure for the determination of the contributions of peaks forming a multidimensional charge state distribution to be decomposed into the recorded mass spectrum is necessary. For time-of-flight mass spectra in this procedure one can use the approach described in our work [16].

ACKNOWLEDGMENTS

We are grateful to A.V. Raznikov and I.V. Sulimenkov for fruitful discussions. This work was partially supported by Program no. 9 for Basic Research of the Presidium of the Russian Academy of Sciences.

REFERENCES

1. Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F., and Whitehouse, C.M., *Science*, 1989, vol. 246, no. 4929, p. 464.
2. Aleksandrov, M.L., Gall', L.N., Krasnov, N.V., Nikolaev, V.I., Pavlenko, V.A., and Shkurov, V.A, *Dokl. Akad. Nauk SSSR,* 1984, vol. 277, no. 2, p. 379.
3. Ferrige, A.G., Seddon, M.J., Green, B.N., Jarvis, S.A., and Skilling, J., *Rapid Commun. Mass Spectrom.*, 1992, vol. 6, no. 7, p. 707.
4. Abzalimov, R.R., Fringpong, A.K., and Kaltashov, I.A., *Int. J. Mass Spectrom.*, 2006, vol. 253, no. 3, p. 207.
5. Hogan, C.J., Carroll, J.A., Rohrs, H.W., Biswas, P., and Gross, M.L., *J. Am. Chem. Soc.*, 2008, vol. 130, no. 12, p. 6926.
6. Kaltashov, I.A. and Mohimen, A., *Anal. Chem.*, 2005, vol. 77, no. 16, p. 5370.
7. Mohimen, A., Dobo, A., Hoerner, J.K., and Kaltashov, I.A., *Anal. Chem.*, 2003, vol. 75, no. 16, p. 4139.
8. Raznikov, V.V. and Raznikova, M.O., Which Information Can Be Acquired from the Distribution of Multicharged Ions Produced by Electrospray-Induced Fragmentation of Bioorganic Compounds?, *4 Vseross. konf. Mass-spektrometriya i ee prikladnye problemi* (4th All-Russia Conf. Mass-Spectrometry and Applied Problems), Moscow, 2011.
9. Felitsyn, N., Peschke, M., and Kebarle, P., *Int. J. Mass Spectrom.*, 2002, vol. 219, no. 1, p. 39.
10. Dole, M., Mack, L.L., Hines, R.L., Mobley, R.S., Feruson, L.D., and Alice, M.B., *J. Chem. Phys.*, 1968, vol. 49, no. 5, p. 2240.
11. Verkerk, U.H. and Kebarle, P., *J. Am. Soc. Mass Spectrom.*, 2005, vol. 16, no. 8, p. 1325.
12. Raznikov, V.V. and Raznikova, M.O., *Eur. J. Mass Spectrom.*, 2009, vol. 15, no. 3, p. 367.
13. Korolyuk, V.S., Portenko, N.I., Skorokhod, A.V., and Turbin, A.F., *Spravochnik po teorii veroyatnostei i matematicheskoi statistike* (Handbook on the Probability Theory and Mathematical Statistics), Moscow: Nauka, 1985.
14. Yergey, J.A., *Int. J. Mass Spectrom. Ion Phys.*, 1983, vol. 52, nos. 2–3, p. 337.
15. Orlov, A.I., *Prikladnaya statistika* (Applied Statistics), Moscow: Ekzamen, 2004.
16. Raznikov, V.V., Pikhtelev, A.R., and Raznikova, M.O., *Mass-Spektrometria*, 2006, vol. 3, no. 2, p. 113.

Translated by I. Duchovni