

Extraction of Local Hydrogen Exchange Data from HDX CAD MS Measurements by Deconvolution of Isotopic Distributions of Fragment Ions

Rinat R. Abzalimov and Igor A. Kaltashov

Department of Chemistry, University of Massachusetts at Amherst, Amherst, Massachusetts, USA

Hydrogen/deuterium exchange (HDX) coupled to protein fragmentation either in solution (by means of proteolysis) or in the gas phase (using collisional activation of protein ions) and followed by mass spectral measurements of deuterium content of individual fragments has become one of the major experimental tools to probe protein structure and dynamics. One difficulty, which often arises in the course of interpretation of HDX MS data, is a need to separate deuterium contribution to the observed isotopic patterns from that of naturally occurring isotopes. Another frequently encountered problem, especially when HDX in solution is followed by protein ion fragmentation in the gas phase, is a need to determine the deuterium content of an internal protein segment based on the measured isotopic distributions of overlapping fragments. While several algorithms were developed in the past several years to address the first problem, the second one did not enjoy as much attention. Here we report a new algorithm based on a maximum entropy principle, which is capable of extracting local exchange data from the isotope distribution of overlapping fragments, as well as subtracting the background due to the presence of natural isotopes and residual deuterium in exchange buffers. The new method is tested with several proteins and appears to generate stable solutions even under unfavorable circumstances, e.g., when the resolving power of a mass analyzer is not sufficient to avoid signal interference or when the isotopic distributions of individual fragments are complex and cannot be approximated with simple binomial distributions. The latter feature makes the algorithm particularly useful when the exchange in solution is correlated or semicorrelated, paving the way to precise structural characterization of non-native protein states in solution. (J Am Soc Mass Spectrom 2006, 17, 1543–1551) © 2006 American Society for Mass Spectrometry

Hydrogen/deuterium exchange (HDX) in combination with mass spectrometric (MS) detection has become a powerful biophysical tool capable of probing protein structure and dynamics in solution under a variety of conditions [1, 2] and on different time scales [3]. MS offers several important advantages compared to other means of monitoring the progress of HDX reactions, such as NMR. First, the superior sensitivity of electrospray (ESI) and matrix-assisted laser desorption/ionization (MALDI) MS allows many HDX experiments to be carried out using only minute quantities of proteins. In many cases, this enables studies of protein behavior at, or even below, their endogenous levels without encountering artifacts associated with protein aggregation in solution. NMR, on the other hand, often requires that HDX experiments

be carried out at mM or high sub-mM protein concentration. Second, ESI MS greatly outperforms high field NMR in its ability to handle larger proteins and their complexes and, unlike high-field NMR, it is tolerant to paramagnetic ligands. Third, under certain conditions, MS detection allows protein characterization to be carried out in a conformer-specific fashion, providing a means of making clear distinction between various equilibrium intermediates of the same protein [4–7]. Finally, ESI MS is very successful in addressing a serious problem inherent to all “traditional” biophysical tools, namely the great difficulty associated with the analysis of protein structure and dynamics in multi-component systems. The ionic signals corresponding to different species in protein mixtures (e.g., solutions containing mixtures of several proteins and their complexes) do not generally overlap. Even the overlapping ion peaks can be easily resolved in many cases based on the difference of charge states [8] or exact masses [9] of the corresponding ionic species.

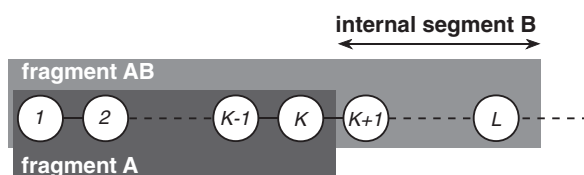
Local information on exchange kinetics within

Published online August 24, 2006

Address reprint requests to Dr. Igor A. Kaltashov, Department of Chemistry, University of Massachusetts-Amherst, 710 North Pleasant Street, Amherst, MA 01003, USA. E-mail: kaltashov@chem.umass.edu

polypeptide chains is obtained in HDX MS experiments by introducing a protein fragmentation step before MS detection. Protein fragmentation can be carried out in solution (using proteolysis under slow exchange conditions to minimize the exchange within unstructured proteolytic fragments [2] or in the gas phase (using fast-heating ion activation of ESI-generated protein ions to avoid hydrogen scrambling [10]). In the latter case, a series of terminal (*b*- or *y*-type) ions are typically produced, and the deuterium content of internal segments must be calculated based on measured isotopic distributions of the terminal fragments flanking each such segment (Scheme 1). Since HDX of most proteins under native conditions proceeds via the so-called EX2 regime [11] and is, therefore, uncorrelated [12], isotopic distributions of both intact protein ions and various fragment ions under these conditions are nearly symmetrical. Deuterium content of each fragment ion in this case can be adequately described with a single number (the average number of ^2H atoms in the segment). Therefore, deuterium content of various internal segments of the protein can be determined with a simple subtraction. However, correlated or semicorrelated exchange (which often occur under conditions favoring the so-called EX1 regime or an intermediate EXX regime [12], situations often encountered under mildly denaturing conditions) give rise to isotopic patterns that are asymmetrical and cannot even be represented by simple binomial distributions. In this case an average number of ^2H atoms retained by a fragment ion is no longer sufficient to describe its deuterium content, and a simple subtraction would no longer produce a meaningful estimate of the deuterium content of internal protein segments.

The aim of this work is to develop a deconvolution procedure capable of calculating deuterium distribution within internal polypeptide segments based on experimentally determined isotopic distributions of the flanking fragment ions (e.g., fragments **A** and **AB** in Scheme 1). The end result of such deconvolution is a distribution of ^2H atoms that are indicative of the HDX reactions, i.e., the corrections are made not only for the natural abundances of various isotopes, but also for the finite $^1\text{H}/^2\text{H}$ ratio in the exchange buffer. Various deconvolution procedures were used in the past as a means to subtract the contributions of natural isotopes from the measured isotopic distributions [13–15] to aid interpretation of both gas-phase and solution-phase HDX measurements. Most procedures ignore the fact that the $^2\text{H}/^1\text{H}$ ratio in fully exchanged peptides is



Scheme 1

always finite, since the exchange reactions are usually initiated by a rapid dilution of a concentrated stock solution of completely deuterated protein (prepared in $^2\text{H}_2\text{O}$ -based buffer) to an $^1\text{H}_2\text{O}$ -based buffer. Conversely, a stock solution of unlabeled protein in $^1\text{H}_2\text{O}$ can be diluted in $^2\text{H}_2\text{O}$. The typical dilution factors range from 10 to 100, meaning that the residual content of ^2H in the $^1\text{H}_2\text{O}$ -based exchange buffer (or ^1H in the $^2\text{H}_2\text{O}$ -based exchange buffer) could be as high as 10%.

Spatial resolution afforded by HDX MS measurements that use proteolysis before MS analysis is limited by the size of peptide fragments and can range from just a few to tens of amino acid residues. The resolution can be enhanced (in some favorable cases down to the residue level) if proteolytic cleavages produce overlapping fragments [16]. The deuteration level of the “difference” region is currently calculated by simple subtraction of the average ^2H levels of the two overlapping peptide fragments. Obviously, such measurements can also greatly benefit from a deconvolution algorithm that is designed to calculate ^2H distributions of internal segments based on the measured isotopic patterns of overlapping fragments.

Experimental

Materials

Ubiquitin was purchased from Sigma-Aldrich Chemical Co. (St. Louis, MO) and used without further purification. A pseudo-wild type cellular retinoic acid binding protein I (wt*-CRABP I) was expressed using a plasmid generously provided by Professor Lila M. Gierasch (UMass-Amherst). Heavy water ($^2\text{H}_2\text{O}$), *d*-acetic acid ($\text{C}^2\text{H}_3\text{CO}_2^2\text{H}$) and *d*-ammonium hydroxide ($\text{N}^2\text{H}_4\text{O}^2\text{H}$) were purchased from Cambridge Isotope Laboratories, Inc. (Andover, MA). All other chemicals and solvents were of analytical grade or higher.

Hydrogen Exchange

All proteins were completely deuterated before HDX measurements by going through at least three consecutive cycles of incubation in $^2\text{H}_2\text{O}$ at 36 °C followed by lyophilization. A concentrated stock solution of each protein in 10 mM $\text{C}^2\text{H}_3\text{CO}_2\text{N}^2\text{H}_4/^2\text{H}_2\text{O}$ was then prepared and completeness of deuteration was verified by mass spectrometry. The exchange reactions were initiated by rapidly diluting and vortexing an aliquot of the protein stock solution in 10 mM $\text{C}^1\text{H}_3\text{CO}_2\text{N}^1\text{H}_4$ ($^1\text{H}_2\text{O}$ solvent), whose pH was adjusted to the desired level with $\text{C}^1\text{H}_3\text{CO}_2^1\text{H}$ or $\text{N}^1\text{H}_4\text{O}^1\text{H}$. The solution was then placed in a glass syringe and continuously infused in the ESI source of a mass spectrometer using a syringe pump. The end-points of HDX were obtained by incubating an aliquot of the exchanging protein solution at elevated temperature (36 °C) and pH (9–10) for 24 h.

Mass Spectrometry

HDX MS and HDX CAD MS measurements were carried out with a 4.7 T Apex III (Bruker Daltonics, Inc., Billerica, MA) Fourier transform ion cyclotron resonance (FT ICR) MS equipped with a standard ESI source and a QStar-XL (Applied Biosystems, Framingham, MA) hybrid quadrupole-time-of-flight (Qq-TOF) MS equipped with a Turbospray ESI source. Protein ion CAD was induced in the ESI interface region of FT ICR MS by increasing the ion kinetic energy (CapExit potential was raised from 120 to 200 V) without mass selection of the precursor ions. CAD of protein ions in Qq-TOF MS was carried out in the RF-only quadrupole (q) following selection of the precursor ions in the first quadrupole region (Q). Argon was used as a collision gas.

Results and Discussion

Formulation of the Problem

HDX CAD MS experiments provide isotopic distributions of terminal (*b*- or *y*-type) fragment ions, such as segments **A** and **AB** shown in Scheme 1. Since the segment **AB** comprises another terminal segment **A** and an internal segment **B**, its isotopic distribution is a convolution of the isotopic distributions of **A** and **B**:

$$\begin{aligned} P_{AB}(n) &= \sum_{m,k \leq n} P_A(m) \cdot P_B(k) \cdot \delta_{n,m+k} \\ &= \sum_{m \leq n} P_A(m) \cdot P_B(n-m), \end{aligned} \quad (1)$$

where $P_X(n)$ is a normalized isotopic distribution of segment **X** ($n = 0$ corresponds to the monoisotopic mass). Rewritten in a more compact form, eq 1 becomes

$$P_{AB}(n) = P_A(m) \otimes P_B(k). \quad (2)$$

Since both distributions $P_A(m)$ and $P_{AB}(n)$ can be measured, the isotopic distribution of the internal segment $P_B(k)$ can be found using a deconvolution procedure. The isotopic distributions $P_X(n)$ contain contributions from the population of labile ^2H atoms that become depleted in the course of exchange reactions, naturally occurring isotopes (^{13}C , ^{15}N , ^{18}O and, in the case of methionine- and cysteine-containing peptides, ^{33}S and ^{34}S), as well as ^2H atoms that are present in the exchange solvent due to the finite dilution factor of the protein stock solution in the exchange buffer. The latter two contributions combined give an end-point of the exchange reaction $P_X^*(n)$, which is an isotopic distribution of segment **X** following the completion of HDX (i.e., $P_X^*(n) = P_X(n)$ at $t = \infty$). If the dilution of the deuterated protein stock solution in the protiated exchange buffer were infinite, $P_X^*(n)$ would be a natural isotopic distribution of segment **X**. $P_A^*(n)$ and $P_{AB}^*(n)$ can be determined experimentally, and the end-point isotopic distribution of the internal segment $P_B^*(n)$ can then be found by deconvolution. Another part of $P_X(n)$

is a distribution $D_X(n)$ of labile ^2H atoms that become depleted in the course of exchange reactions. $D_B(n)$ provides information on the ^2H retention within the internal segment **B** unaffected by naturally occurring isotopes and residual ^2H atoms in the exchange solvent, and its calculation is the ultimate goal of this work. It is easy to see that $P_X(n)$ is a convolution of $P_X^*(n)$ and $D_X(n)$:

$$\begin{aligned} P_X(n) &= \sum_{q,r \leq n} D_X(q) \cdot P_X^*(r) \cdot \delta_{n,q+r} \\ &= \sum_{q \leq n} D_X(q) \cdot P_X^*(n-q), \end{aligned} \quad (3)$$

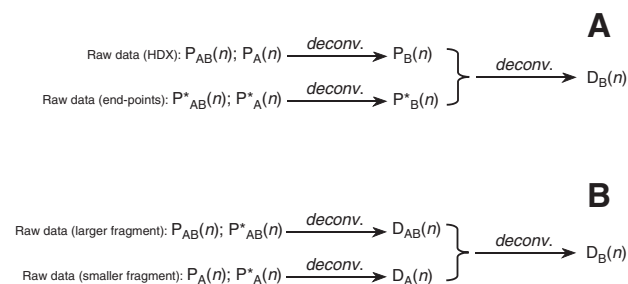
or simply

$$P_X(n) = D_X(q) \otimes P_X^*(r). \quad (4)$$

Two strategies can be used to find $D_B(n)$. For example, $D_B(n)$ can be obtained by deconvoluting $P_B^*(n)$ from $P_B(n)$. These two distributions can be obtained in turn by deconvoluting the end-point distributions ($P_A^*(n)$ from $P_{AB}^*(n)$ to obtain $P_B^*(n)$) and the isotopic distributions of fragments **A** and **AB** at a certain time point during HDX [$P_A(n)$ from $P_{AB}(n)$ to obtain $P_B(n)$]. This sequence is shown in Scheme 2A. Alternatively, one can obtain ^2H distributions for fragments **A** and **AB** by deconvolution and then use $D_{AB}(n)$ and $D_A(n)$ to calculate $D_B(n)$ (Scheme 2B). Both strategies calculate $D_B(n)$ using three deconvolution steps and should produce similar (although not necessarily identical) results.

Deconvolution Procedure

Deconvolution is generally an ill-conditioned problem that does not necessarily has a unique solution even in the absence of noise. The experimentally determined distributions $P_A(n)$ and $P_{AB}(n)$ are inevitably modulated by noise, which makes the inversion (deconvolution) even more problematic [17]. A variety of mathematical tools have been developed to deal with the deconvolution problem in general and processing of mass spectral data in particular. These include classical determinist regularization approaches (with the least-squares method being one of the simplest examples), information theory-based approaches, and probabilistic ap-



Scheme 2

proaches, such as the Bayesian estimation [17]. The two most popular approaches to deconvolution of isotopic distributions utilize least-squares (or, more generally, χ^2) methods or maximum entropy methods (MEM) [18]. Zhang et al. carried out a detailed analysis of both least-squares and MEM-based deconvolution as applied to isotopic distributions and concluded that MEM is better suited to handle larger polypeptides [13]. Since the protein ion fragments in HDX CAD MS experiments can be fairly large (several kDa), we chose MEM over the least-squares methods as a means to carry out deconvolution.

In this work we adopt a definition of entropy given by Zhang et al. [13]:

$$S = -\sum_q D_x(q) \cdot (\ln(D_x(q)) - 1). \quad (5)$$

Entropies for other distributions to be found by deconvolution are defined analogously. The search for maximum S is restricted by a constraint imposed on the deviation of the re-convoluted distribution from the original one. For example, if $D_x(q)$ is found by deconvolution of $P^*_x(q)$ from $P_x(q)$, a χ^2 factor is introduced:

$$\chi^2 = \sum_q G(q) \left(\sum_{m,k \leq q} D_x(m) \cdot P^*_x(k) \cdot \delta_{m+k,q} - P_x(q) \right)^2, \quad (6)$$

where $G(q)$ is a parameter designed to account for the signal modulation by the noise. Since in our case the spectral noise is independent of q , we assume that $G(q)$ is constant and ignore it in the following discussion. Finding maximum for S and minimum for χ^2 can be accomplished simultaneously by searching for a maximum of a modified entropy function Q defined as follows:

$$Q = S - \lambda \cdot \chi^2, \quad (7)$$

where λ is a user-defined Lagrange parameter. The program for the deconvolution procedure was written on the basis of MathCAD (Mathsoft Engineering and Education, Inc., Cambridge, MA) and is available from the authors upon request.

Deconvolution of Uncorrelated Exchange Patterns: EX2 Regime

The deconvolution strategies outlined in the preceding sections were evaluated by calculating the deuterium distribution of an internal segment of a small protein ubiquitin (*Ub*) undergoing HDX at pH 7.0. Under these conditions HDX of *Ub* proceeds through the EX2 mechanism [6] and is clearly uncorrelated, as is evident from the shape of the isotopic distributions of protein ions throughout the course of the exchange reaction (Figure 1). CAD of *Ub* ions in the ESI interface region of FT ICR MS produces a series of abundant *b*- and *y*-fragment ions with a fairly uniform coverage of the entire amino acid sequence. The isotopic distributions of two repre-

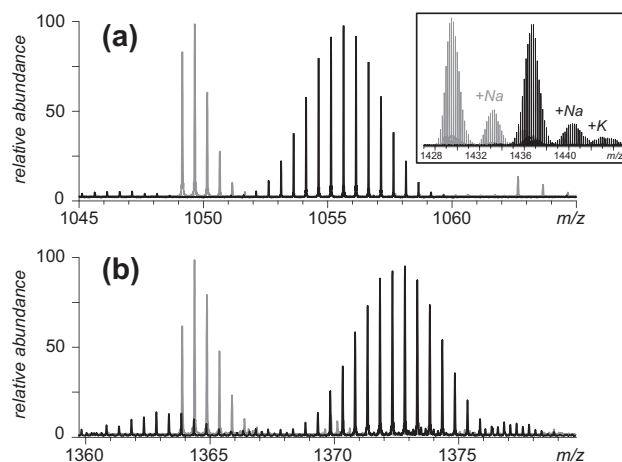


Figure 1. HDX CAD MS of *Ub*: isotopic profiles of y_{18}^{2+} (a) and y_{24}^{2+} (b) fragment ions of *Ub* following 10 min of protein exposure to the exchange buffer (FT ICR MS data). The gray traces represent the end-point of the exchange reaction. Isotopic distribution of the intact protein ion (shown in inset) indicates that the exchange is uncorrelated.

sentative fragment ions, y_{18}^{2+} and y_{24}^{2+} , are shown in Figure 1. Both isotopic distributions are largely interference-free and, as such, provide an ideal testing system for the deconvolution procedures.

Deconvolution of the end-point isotopic distributions (shown in gray color in Figure 1) from the isotopic distributions recorded at 10 min of the protein exposure to the exchange buffer yields distributions of labile ^2H atoms retained by each of the fragment. These distributions (gray bars on top and middle diagrams in Figure 2) do not contain contributions from the residual deuterium in the exchange buffer. The error bars on these diagrams illustrate stability of the deconvolution procedure with respect to the Lagrange parameter λ , which was varied in the range 7 to 215 for $D_{AB}(n)$ and 4 to 465 for $D_A(n)$. Deconvolution of the ^2H distribution of the y_{18}^{2+} fragment from that of y_{24}^{2+} [i.e., $D_A(n)$ from $D_{AB}(n)$] gives the isotopic distribution of the internal segment Gly (53) \rightarrow Asp (58) (shown as $D_B(n)$ at the bottom diagram of Figure 2). To evaluate the overall stability of the deconvolution procedure, optimal solutions were found for a range of $D_{AB}(n)$ and $D_A(n)$ (the $D_B(n)$ distribution shown in Figure 2 is actually the average distribution and the error bars represent the standard deviation of each peak within the entire tested range of the Lagrange parameters λ). The overall shape of the deconvoluted distribution $D_B(n)$ does not change significantly within the tested range of λ and the reconvoluted ^2H distribution for the y_{24}^{2+} fragment appears to closely match the original distribution (compare gray and black bars in the top diagram, Figure 2). Nevertheless, the standard deviation is rather significant, as it is comparable with the absolute intensity of each peak in the calculated distribution $D_B(n)$.

Figure 3 illustrates another approach to calculating the distribution of exchanging ^2H atoms within the

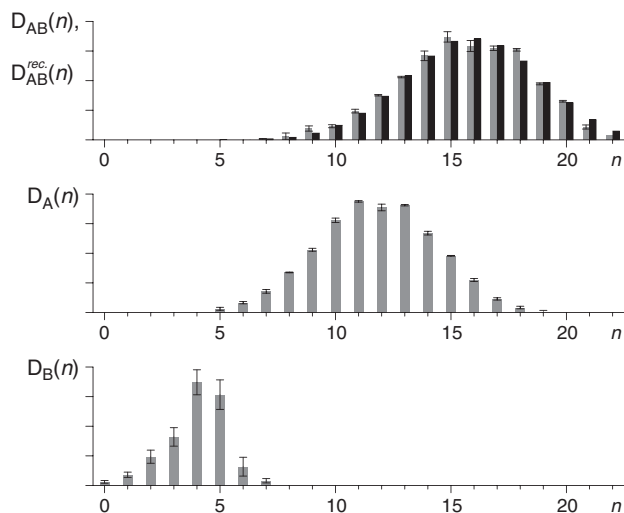


Figure 2. Distribution of retained ^2H atoms within a $[\text{Gly}^{53}\rightarrow\text{Asp}^{58}]$ segment of ubiquitin following 10 min HDX at pH 7.0 obtained from the isotopic distributions of y_{24}^{2+} and y_{18}^{2+} fragment ions (raw data shown in Figure 1) by calculating the ^2H distributions of the fragment ions first. The upper panels show the isotopic distribution of the two fragments following a correction for the end-point isotopic distributions (y_{24}^{2+} , top, and y_{18}^{2+} , middle). The error bars show the variance of optimal solutions with respect to the Lagrange parameter λ used in calculations (top diagram: $\lambda = 7\text{--}215$; middle diagram: $\lambda = 4\text{--}465$). The error bars of the ^2H distribution for the internal segment (bottom diagram) indicate the range of solutions $D_B(n)$ generated by varying the Lagrange parameter λ from 7 to 420 in addition to taking into account variations of both $D_{AB}(n)$ and $D_A(n)$. Reconvoluted distribution of ^2H atoms for the y_{24}^{2+} fragment ion is shown on the top diagram (black bars).

$[\text{Gly}^{53}\rightarrow\text{Asp}^{58}]$ segment. Here the isotopic distributions of the internal segment are calculated at 10 min HDX (by deconvolution isotopic distribution of y_{18}^{2+} fragment from that of y_{24}^{2+} following 10 min of exchange, $P_B(n)$ in the top diagram) and at the end-point of the exchange reaction ($P_B^*(n)$ in the middle diagram in Figure 3). Once again, the error bars represent the sensitivity of the calculated distributions $P_B(n)$ and $P_B^*(n)$ with respect to the Lagrange parameter λ . These two distributions are then used to obtain the distribution of exchanging ^2H atoms for the internal segment ($D_B(n)$ in the bottom diagram, Figure 3). While this distribution appears to be quite insensitive with respect to the Lagrange parameter λ used in the final deconvolution procedure (the variance of each peak did not exceed 1% of its intensity within the entire tested region of λ from 7 to 820), the uncertainty in calculating distribution $P_B(n)$ gives rise to a much more noticeable variation of optimal solutions $D_B(n)$ (as illustrated with error bars in the bottom diagram, Figure 3). Nonetheless, the overall uncertainty of this final solution $D_B(n)$ is still less significant than that of $D_B(n)$ calculated previously (Figure 2), suggesting that this deconvolution algorithm is more stable.

It is instructive to compare the deconvoluted distribution $D_B(n)$ shown in Figure 3 with the results of

calculating the average deuterium content of the $[\text{Gly}^{53}\rightarrow\text{Asp}^{58}]$ segment using common procedures. The average deuterium content of the y_{18}^{2+} fragment ion can be calculated as the difference between the centroids of two isotopic distributions, one representing the isotopic makeup of this fragment ion following 10 min of the exchange reaction in solution and another representing the natural isotope abundance. However, this calculation would lead to an overestimation of the backbone protection within the y_{18}^{2+} fragment ion, as it does not account for the increased $^2\text{H}/^1\text{H}$ ratio in the exchange solution (finite dilution factor). To make this correction, it is necessary to replace the natural isotopic distribution of y_{18}^{2+} with that at the exchange reaction endpoint (gray trace in Figure 1, top panel). This calculation allows the protection of the y_{18}^{2+} fragment ion after 10 min of protein exposure to the exchange buffer to be estimated as 11.8. Analogously, average protection of a larger polypeptide segment (represented by the y_{24}^{2+} fragment ion) can be estimated as 15.6 following 10 min of the exchange reaction. A difference between these two numbers gives backbone protection of the $[\text{Gly}^{53}\rightarrow\text{Asp}^{58}]$ segment, 3.8. While this number is the same as the average number of protected ^2H atoms deduced from the $D_B(n)$ distribution (3.8 with a standard deviation of 1.4 for the distribution shown in Figure 3), it does not provide as much information compared to $D_B(n)$. For example, the small bias of the distribution towards states with low degrees of protection, which is clearly revealed by the asymmetric character of $D_B(n)$,

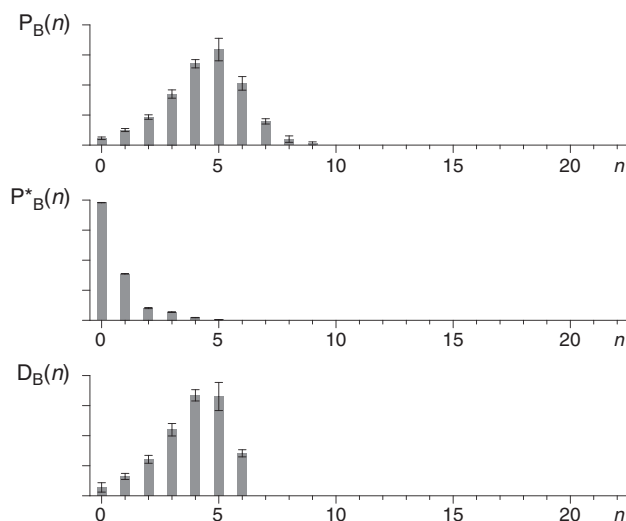


Figure 3. Distribution of retained ^2H within a $[\text{Gly}^{53}\rightarrow\text{Asp}^{58}]$ segment of ubiquitin following 10 min HDX at pH 7.0 calculated based on the isotopic distributions of y_{24}^{2+} and y_{18}^{2+} fragment ions (raw data shown in Figure 1). The upper panels show the isotopic distribution of the $[\text{Gly}^{53}\rightarrow\text{Asp}^{58}]$ segment following 10 min of HDX (top) and at the end-point of the exchange reaction. The error bars show the variance of optimal solutions with respect to the Lagrange parameter λ used in calculations (top diagram: $\lambda = 6\text{--}57$; middle diagram: $\lambda = 13\text{--}820$). The error bars of the $D(n)$ distribution (bottom diagram) indicate the range of solutions generated by varying the Lagrange parameter λ from 7 to 820 in addition to taking into account variations of both $P_B(n)$ and $P_B^*(n)$.

cannot be deduced from the results of calculations carried out using common methodology.

Deconvolution of Uncorrelated Exchange Patterns in the Presence of Signal Interference

The two distributions $D_B(n)$ obtained in the previous section by implementing different sequences of deconvolution steps are qualitatively very similar (compare bottom diagrams in Figures 2 and 3). However, the set of final solutions shown in Figure 3 displays significantly reduced variation and clearly indicates that the corresponding procedure is more stable and is better suited to handle simple interference-free isotopic distributions. To assess the stability of this procedure in the presence of signal interference, it was applied to calculate the distribution of exchanging ^2H atoms within a $[\text{Pro}^{19} \rightarrow \text{Glu}^{34}]$ segment of *Ub*. The isotopic distribution of one of the fragment ions used in calculations (y_{42}^{4+}) is largely interference free, while another (y_{58}^{5+}) partially overlaps with isotopic distributions of other fragment ions (Figure 4). Although the resolving power of FT ICR MS is sufficient to allow a clear distinction to be made between the isotopic peaks of y_{58}^{5+} and one of the interfering ions (y_{46}^{4+}), the peaks of an unidentified singly charged ion could not be resolved from the isotopic peaks of y_{58}^{5+} .

Since the signal interference in this case is clearly above the noise level, an adjustment was made to the experimentally obtained isotopic distribution of y_{58}^{5+} ($P_{AB}(n)$ in Figure 5, inset in the top diagram). Intensities of the two isotopic peaks containing contributions from a singly charged ionic species ($n = 35$ and 40) were assigned as half-sums of their neighbors. The isotopic distribution of the internal fragment after 10 min of

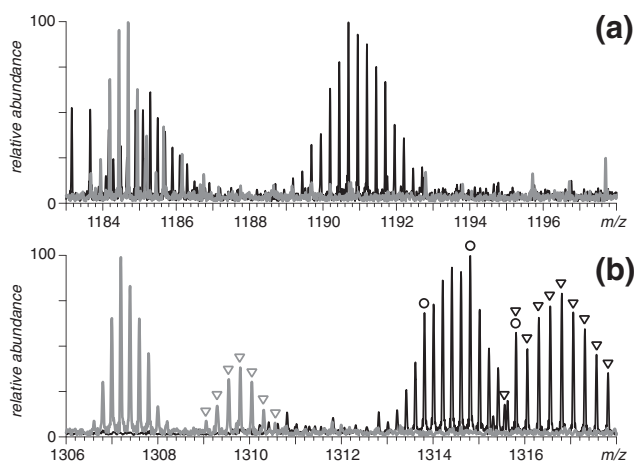


Figure 4. HDX CAD MS of *Ub*: isotopic profiles of y_{42}^{4+} (A) and y_{46}^{4+} and y_{58}^{5+} (B) fragment ions of *Ub* following 10 min of protein exposure to the exchange buffer (FT ICR MS data). Isotopic peaks of y_{58}^{5+} fragment labeled with circles contain contributions from an unidentified singly charged ionic species. Ion peaks in the isotopic cluster of y_{46}^{4+} fragment (labeled with triangles) do not interfere with the ionic signal of y_{58}^{5+} . The gray traces on both panels represent the end-point of the exchange reaction.

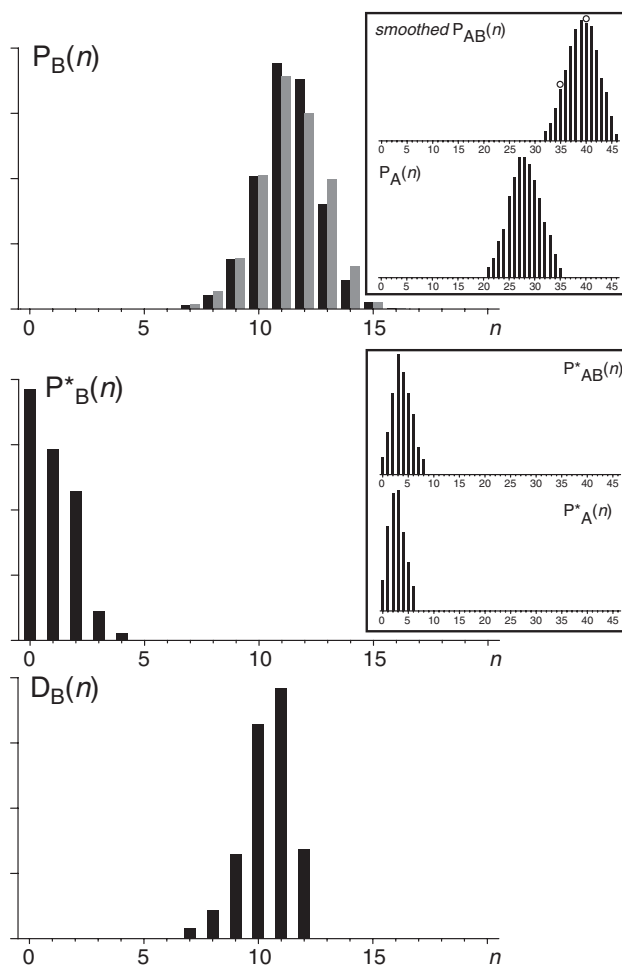


Figure 5. Distribution of retained ^2H within a $[\text{Pro}^{19} \rightarrow \text{Glu}^{34}]$ segment of ubiquitin following 10 min HDX at pH 7.0 calculated based on the isotopic distributions of y_{58}^{5+} and y_{42}^{4+} fragment ions (raw data shown in Figure 4). The upper panels show the isotopic distribution of the $[\text{Pro}^{19} \rightarrow \text{Glu}^{34}]$ segment following 10 min of HDX (top) and at the end-point of the exchange reaction. The calculations were carried out after smoothing the isotopic distribution of y_{58}^{5+} fragment ion: the isotopic peaks whose intensity is distorted by an interfering singly charged species (labeled with circles in Figure 4) were calculated as half-sums of their neighbors (the isotopic distributions used for deconvolution are shown in insets). The gray bars in the top diagram represent the isotopic distribution of the $[\text{Pro}^{19} \rightarrow \text{Glu}^{34}]$ segment calculated without smoothing the original isotopic distribution of the y_{58}^{5+} fragment ion (not used for further calculations).

exchange (calculated by deconvolution of $P_A(n)$ from $P_{AB}(n)$) is shown in Figure 5 (top diagram) with black bars. Other ways of correcting the isotopic distribution of y_{58}^{5+} for the presence of interfering signal (e.g., Gaussian smoothing) resulted in very similar distributions (data not shown). However, ignoring the interference and carrying out the deconvolution using the original (uncorrected) isotopic distribution of y_{58}^{5+} resulted in noticeable distortion of the calculated distribution $P_B(n)$. This isotopic distribution of the internal fragment (shown with gray bars in Figure 5, top diagram) is biased towards higher n , although the most abundant peak in the cluster remains unchanged ($n = 11$).

Further calculations were carried out using $P_B(n)$ derived from the interference-corrected isotopic distribution $P_{AB}(n)$ and the end-result is shown in the bottom diagram of Figure 5. Despite a relatively large size of the segment (16 amides), the distribution of ^2H atoms retained after 10 min exposure to the protiated exchange buffer is very narrow. The distribution indicates that 10.5 ± 1.5 amides remain protected at this time, providing a precision that cannot be attained by simply comparing the centroids of the isotopic distributions of y_{58}^{5+} and y_{42}^{4+} fragment ions (Figure 4).

Interestingly, the alternative deconvolution strategy (calculating $D_{AB}(n)$ and $D_A(n)$ first, followed by calculation of $D_B(n)$) was very sensitive to the choice of the procedure used to correct for the presence of the interference in $P_{AB}(n)$ (data not shown). Since this procedure was already shown to be inferior as far as solution variance even in the absence of signal interference (vide supra), a decision was made to exclude it from further consideration.

Deconvolution of Semicorrelated Exchange Patterns

The examples considered in the preceding sections clearly suggest that the deconvolution-based calculation of distributions of retained ^2H atoms is quite useful even in a situation when the exchange is uncorrelated. Indeed, even though the average protection of internal segments can be often estimated by simple subtraction of the deuterium levels of two overlapping fragments, implementation of the deconvolution-based strategy provides much higher precision. The use of deconvolution becomes essential in a situation when the exchange is correlated or semicorrelated. Correlated exchange patterns are observed when HDX reactions follow the EX1 regime, while semicorrelated patterns can either signal the presence of multiple transitions leading to a loss of backbone amide protection [6] or else indicate the occurrence of HDX in the intermediate EXX regime [12]. The bimodal appearance of the isotopic distributions of protein ions undergoing semicorrelated exchange does provide a means to distinguish a native (highly protected) state from those whose protection is compromised [6]. However, making a clear distinction among various (partially) unprotected states is often much more problematic, as the isotopic distributions of the corresponding ionic species are usually poorly resolved or not resolved at all [6].

A typical example of a semicorrelated exchange patterns is presented in the inset of Figure 6, which shows a peak shape of a molecular ion (+14 charge state) of a pseudo-wild type cellular retinoic acid binding protein I (wt*-CRABP I) following 10 min of exposure to a protiated exchange buffer at pH 3.5. The protein is significantly destabilized at this pH (as suggested by the appearance of the ionic charge state distribution in ESI mass spectra), and a significant

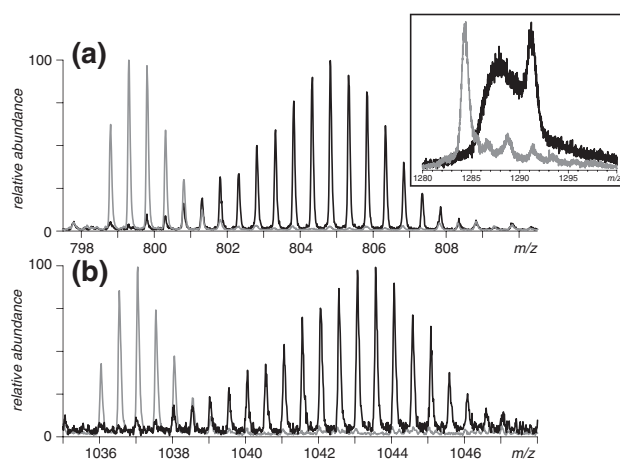


Figure 6. HDX CAD MS/MS of wt*-CRABP I under acidic conditions: isotopic profiles of y_{14}^{2+} (a) and y_{18}^{2+} (b) fragment ions following 10 min of protein exposure to the exchange buffer at pH 3.5 (Qq-TOF MS data). The gray traces represent the end-point of the exchange reaction. Isotopic distributions of the intact protein ions (charge state +14) used as precursor ions in MS/MS experiments indicate that the exchange is semicorrelated (inset).

fraction of the protein molecules populate an unfolded (fully unprotected) state in solution under these conditions. The absence of fully exchanged protein ions following 10 min of exchange is because the intrinsic exchange rates of backbone amides are very slow at this pH. This results in incomplete (although significant) labeling of unprotected segments during each unfolding event, a situation previously described in a two-state protein system as an intermediate EXX exchange regime [12]. The additional complication for multi-state protein systems (such as wt*-CRABP I) arises due to the possibility of existence of partially structured non-native protein states alongside the fully unstructured (but not completely exchanged) unfolded state.

Should the lower- m/z portion of the isotopic cluster of the intact protein ion represent only the incompletely exchanged unfolded state, a uniform loss of ^2H is expected across the entire protein sequence. If, however, contributions of the partially structured non-native state(s) make up a significant proportion of the ionic signal, distribution of ^2H across the polypeptide sequence should not be uniform but, instead, be indicative of the structure of such intermediate state(s). We expect the latter to be the case, since the results of the analysis of charge state distributions of wt*-CRABP I ions in ESI MS at pH 3.5 strongly suggest that the intermediate states are present in solution alongside the native and fully unstructured states.

CAD of the entire population of protein ions carrying 14 positive charges in the RF-only collision cell of a hybrid Qq-TOF MS generated over two dozens of abundant fragment ions of b - and y -type, two of which (y_{14}^{2+} and y_{18}^{2+}) are shown in Figure 6. The isotopic distributions of both fragment ions have convoluted shapes and it is rather difficult to make a confident assessment of the protection of internal segment

[Ile¹¹⁹→Phe¹²²] using either centroids or maxima of the two distributions shown in Figure 6. However, using a deconvolution procedure as a means to calculate the distribution of remaining ²H atoms in this internal segment allows the backbone protection to be easily determined. The deuterium distribution of the internal segment obtained by first calculating the isotopic distribution of the internal segment $P_B(n)$ following 10 min of HDX, and then correcting it for the isotopic distribution of the same segment at the end-point of the exchange reaction $P^*_B(n)$ is clearly bimodal and reveals the presence of two major conformational isomers (Figure 7). It is important to note that the variance of the calculated deuterium distribution $D_B(n)$ shown in Figure 7 is fairly low, strongly suggesting that the bimodal character of the calculated distribution is not an artifact. Furthermore, calculation of $D_B(n)$ using the alternative deconvolution strategy results in a very similar distribution, albeit with somewhat lower stability with respect to variation of the Lagrange parameter (Figure 8). Once again, lower variance of the results of deconvolution of HDX CAD MS data carried out by finding the isotopic distribution of the internal fragment first, followed by correction for its isotopic distribution at the end-point of HDX suggests that this a preferred method of carrying out deconvolution.

Not surprisingly, one of the two conformational isomers detected in this experiment is a highly protected species, which is likely to be the natively folded protein, while another isomer lacks any protection

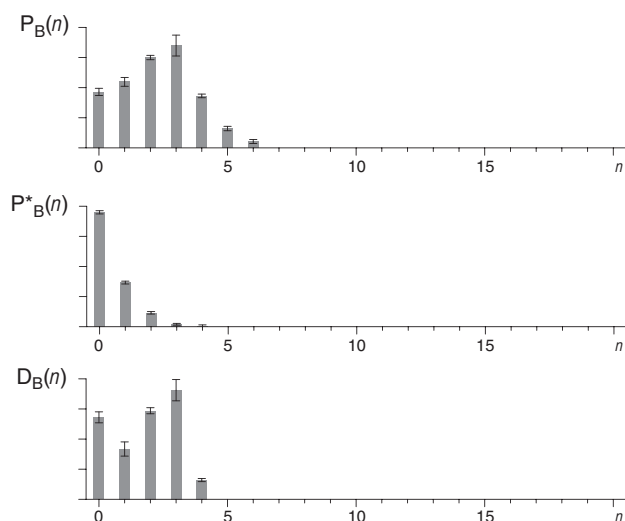


Figure 7. Distribution of retained ²H within a [Ile¹¹⁹→Phe¹²²] segment of wt*-CRABP I (β -strand 9) following 10 min HDX at pH 3.5 calculated based on the isotopic distributions of y_{14}^{2+} and y_{18}^{2+} fragment ions (raw data shown in Figure 6). The upper panels show the isotopic distribution of the [Ile¹¹⁹→Phe¹²²] segment following 10 min of HDX (top) and at the end-point of the exchange reaction (middle). The error bars show the variance of optimal solutions with respect to the Lagrange parameter λ used in calculations ($\lambda = 17$ –51). The error bars of the $D(n)$ distribution (bottom diagram) indicate the range of solutions generated by varying the Lagrange parameter λ from 1 to 13,100 in addition to taking into account variations of both $P_B(n)$ and $P^*_B(n)$.

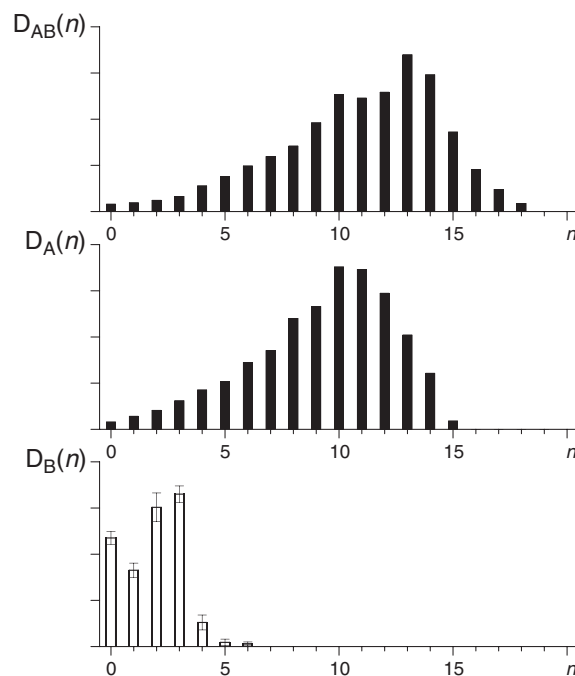


Figure 8. Distribution of retained ²H within a [Ile¹¹⁹→Phe¹²²] segment of wt*-CRABP I (β -strand 9) following 10 min HDX at pH 3.5 calculated based on the isotopic distributions of y_{14}^{2+} and y_{18}^{2+} fragment ions (raw data shown in Figure 6) by calculating the ²H distributions of the fragment ions first. The upper panels show the isotopic distribution of the two fragments following a correction for the end-point isotopic distributions (y_{18}^{2+} , top, and y_{14}^{2+} , middle). The error bars of the ²H distribution for the internal segment (bottom diagram) indicate the range of solutions generated by varying the Lagrange parameter λ from 1 to 1000.

within the [Ile¹¹⁹→Phe¹²²] segment. This is a very important observation as it clearly indicates that the residual protection exhibited by the lower- m/z portion of the isotopic cluster of the intact protein ion is not distributed uniformly across the entire protein sequence. Instead, some protein segments (such as [Ile¹¹⁹→Phe¹²²]) lack any protection within these non-native states, while others exhibit noticeable protection (data not shown). The lack of protection in the [Ile¹¹⁹→Phe¹²²] segment of the partially folded wt*-CRABP I is not surprising, as it belongs to a strand β_9 , whose flexibility is higher than that of the neighboring structural segments even under native conditions according to earlier HDX MS [19] and NMR studies (Eyles, unpublished data). Obviously, the calculated distribution of ²H atoms retained within the [Ile¹¹⁹→Phe¹²²] segment is much more informative than the numerical value of its deuterium content (1.8 with a standard deviation of 1.3), which is averaged across both conformational populations of the protein and does not provide adequate characterization of the segment's protection in either conformation. Clearly, subjecting the HDX CAD MS data to a MEM-based deconvolution allows the conformational analysis of the protein to be carried out at a level of detail unattainable with traditional methods of HDX MS data processing.

Conclusions

A MEM-based deconvolution procedure developed and implemented in this work provides an easy way to process HDX CAD MS data and yields distributions of retained ^2H atoms within internal protein segments based on experimentally determined isotopic distributions of b - and y -fragment ions at various time points of the exchange reaction. The procedure corrects not only for the presence of naturally occurring isotopes, such as ^{13}C and ^{15}N , but also removes a contribution of residual ^2H in exchange buffers. The procedure is stable and produces meaningful solutions in the presence of unresolved signal interferences following minimal processing of raw data. Application of this procedure to HDX CAD MS data acquired under conditions favoring uncorrelated exchange in solution significantly increases the precision of assessing the average protection levels of internal protein segments. In situations favoring semicorrelated exchange, the new deconvolution procedure yields meaningful distributions of retained ^2H atoms that reveal contributions of various protein conformers to the exchange process. Although presently applied to processing of HDX CAD MS data, the new procedure can also be used to increase spatial resolution of HDX MS experiments carried out by fragmenting protein proteolytically under slow exchange conditions before MS analysis.

Acknowledgments

The authors acknowledge Dr. Hui Xiao and Joshua K. Hoerner for providing ubiquitin HDX CAD MS data, and Dr. Stephen J. Eyles for helpful discussions. This work was supported by a grant from the National Institutes of Health (R01 GM061666).

References

- Kaltashov, I. A.; Eyles, S. J. Studies of Biomolecular Conformations and Conformational Dynamics by Mass Spectrometry. *Mass Spectrom. Rev.* **2002**, *21*, 37–71.
- Wales, T. E.; Engen, J. R. Hydrogen Exchange Mass Spectrometry for the Analysis of Protein Dynamics. *Mass Spectrom. Rev.* **2006**, *25*, 158–170.
- Konermann, L.; Simmons, D. A. Protein Folding Kinetics and Mechanisms Studied by Pulse-Labeling and Mass Spectrometry. *Mass Spectrom. Rev.* **2003**, *22*, 1–26.
- Simmons, D. A.; Konermann, L. Characterization of Transient Protein Folding Intermediates During Myoglobin Reconstitution by Time-Resolved Electrospray Mass Spectrometry with On-Line Isotopic Pulse Labeling. *Biochemistry* **2002**, *41*, 1906–1914.
- Zhang, Y. H.; Yan, X.; Maier, C. S.; Schimerlik, M. I.; Deinzer, M. L. Conformational Analysis of Intermediates Involved in the *In Vitro* Folding Pathways of Recombinant Human Macrophage Colony Stimulating Factor β by Sulfhydryl Group Trapping and Hydrogen/Deuterium Pulsed Labeling. *Biochemistry* **2002**, *41*, 15495–15504.
- Hoerner, J. K.; Xiao, H.; Kaltashov, I. A. Structural and Dynamic Characteristics of a Partially Folded State of Ubiquitin Revealed by Hydrogen Exchange Mass Spectrometry. *Biochemistry* **2005**, *44*, 11286–11294.
- Pan, J.; Wilson, D. J.; Konermann, L. Pulsed Hydrogen Exchange and Electrospray Charge-State Distribution as Complementary Probes of Protein Structure in Kinetic Experiments: Implications for Ubiquitin Folding. *Biochemistry* **2005**, *44*, 8627–8633.
- Eyles, S. J.; Kaltashov, I. A. Methods to Study Protein Dynamics and Folding by Mass Spectrometry. *Methods* **2004**, *34*, 88–99.
- He, F.; Emmett, M. R.; Hakansson, K.; Hendrickson, C. L.; Marshall, A. G. Theoretical and Experimental Prospects for Protein Identification Based Solely on Accurate Mass Measurement. *J. Proteome Res.* **2004**, *3*, 61–67.
- Hoerner, J. K.; Xiao, H.; Dobo, A.; Kaltashov, I. A. Is There Hydrogen Scrambling in the Gas Phase? Energetic and Structural Determinants of Proton Mobility Within Protein Ions. *J. Am. Chem. Soc.* **2004**, *126*, 7709–7717.
- Krishna, M. M. G.; Hoang, L.; Lin, Y.; Englander, S. W. Hydrogen Exchange Methods to Study Protein Folding. *Methods* **2004**, *34*, 51–64.
- Xiao, H.; Hoerner, J. K.; Eyles, S. J.; Dobo, A.; Voigtman, E.; Mel'cuk, A. I.; Kaltashov, I. A. Mapping Protein Energy Landscapes with Amide Hydrogen Exchange and Mass Spectrometry. I. A Generalized Model for a Two-State Protein and Comparison with Experiment. *Protein Sci.* **2005**, *14*, 543–557.
- Zhang, Z. Q.; Guan, S. H.; Marshall, A. G. Enhancement of the Effective Resolution of Mass Spectra of High-Mass Biomolecules by Maximum Entropy-Based Deconvolution to Eliminate the Isotopic Natural Abundance Distribution. *J. Am. Soc. Mass Spectrom.* **1997**, *8*, 659–670.
- Hotchko, M.; Anand, G. S.; Komives, E. A.; Ten Eyck, L. F. Automated Extraction of Backbone Deuteration Levels from Amide H/ ^2H Mass Spectrometry Experiments. *Protein Sci.* **2006**, *15*, 583–601.
- Chik, J. K.; Van de Graaf, J. L.; Schriemer, D. C. Quantitating the Statistical Distribution of Deuterium Incorporation to Extend the Utility of H/D Exchange MS Data. *Anal. Chem.* **2006**, *78*, 207–214.
- Del Mar, C.; Greenbaum, E. A.; Mayne, L.; Englander, S. W.; Woods, V. L., Jr. Structure and Properties of α -Synuclein and Other Amyloids Determined at the Amino Acid Level. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 15477–15482.
- Mohammad-Djafari, A.; Giovannelli J.-F.; Demoment, G.; Idier, J. Regularization, Maximum Entropy and Probabilistic Methods in Mass Spectrometry Data Processing Problems. *Int. J. Mass Spectrom.* **2002**, *215*, 175–193.
- Buck B.; Macaulay V. A. Maximum Entropy in Action: A Collection of Expository Essays; Oxford University Press: New York, NY, 1991.
- Xiao, H.; Kaltashov, I. A. Transient Structural Disorder as a Facilitator of Protein-Ligand Binding: Native H/D Exchange-Mass Spectrometry Study of Cellular Retinoic Acid Binding Protein I. *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 869–879.