



## RESEARCH ARTICLE

# ExMS: Data Analysis for HX-MS Experiments

Zhong-Yuan Kan,<sup>1</sup> Leland Mayne,<sup>1</sup> Palaniappan Sevugan Chetty,<sup>2</sup> S. Walter Englander<sup>1</sup><sup>1</sup>Johnson Research Foundation, Department of Biochemistry and Biophysics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA<sup>2</sup>Department of GI, Nutrition, and Hepatology, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA

## Abstract

A previous paper considered the problems that presently limit the hydrogen exchange - mass spectrometry (HX-MS) method for studying the biophysical and functional properties of proteins. Many of these problems can be overcome by obtaining and analyzing hundreds of sequentially overlapping peptide fragments that cover the protein many times over (Mayne et al. J. Am. Soc. Mass Spectrom. 2011: [10.1007/s13361-011-0235-4](https://doi.org/10.1007/s13361-011-0235-4)). This paper describes a computer program called ExMS that furthers this advance by making it possible to efficiently process crowded mass spectra and definitively identify and characterize these many peptide fragments. ExMS automatically scans through high resolution MS data to find the individual isotopic peaks and isotopic envelopes of a list of peptides previously identified by MS/MS. It performs a number of tests to ensure correct identification in spite of peptide overlap in both chromatographic and mass spectrometric dimensions and possible multi-modal envelopes due to static or dynamic structural heterogeneity or HX EX1 behavior. The program can automatically process data from many sequential HX time points with no operator intervention at the rate of ~2 sec per peptide per HX time point using desktop computer equipment, but it also provides for rapid manual checking and decision when ambiguity exists. Additional subroutines can provide a step by step report of performance at each test along the way and parameter adjustment, deconvolute isotopic envelopes, and plot the time course of single and multi-modal H-D exchange. The program will be available on an open source basis at: <http://HX2.med.upenn.edu/download.html>

**Key words:** Hydrogen exchange, Proteins, HDX

## Introduction

The proteolytic fragmentation-mass spectrometry method extends hydrogen exchange (HX) experimentation to protein systems that are out of the reach of HX-NMR methods. In the HX-MS method an experimental protein is exposed to H to D exchange (or the reverse), samples are quenched to a slow HX condition, the protein is proteolytically fragmented, the peptide fragments are separated by liquid chromatography and mass spectrometry, and the

resulting mass spectra are examined to identify each peptide fragment and determine the number of deuterons that it carries. The method intrinsically provides HX data only at the resolution of peptide fragments. However, the comparison of accurate HX-MS results for many sequentially overlapping peptide fragments can make it possible to determine protein HX behavior at high resolution, perhaps reaching ultimately to the individual amino acid level.

This goal will require the capability to obtain and analyze many sequentially overlapping peptide fragments of experimental proteins. A previous paper [1] describes methods that can provide very many fragments under conditions that are advantageous for H-D exchange experiments. In order to determine the number of D atoms accumulated in HX experiments by each of these many peptides, it is necessary to efficiently analyze large LC/MS data files, identify and

**Electronic supplementary material** The online version of this article (doi:10.1007/s13361-011-0236-3) contains supplementary material, which is available to authorized users.

Correspondence to: Zhong-Yuan Kan; e-mail: [kanz@mail.med.upenn.edu](mailto:kanz@mail.med.upenn.edu)

Received: 20 May 2011  
Revised: 11 August 2011  
Accepted: 12 August 2011  
Published online: 15 September 2011

characterize hundreds of peptides with varying degrees of deuteration, and measure the deuterium content of each one. This must be accomplished in spite of the presence of peptides that may be poorly separated chromatographically and in the MS spectra and that may have more than one level of D-labeling due to static or dynamic structural heterogeneity or to EX1 HX behavior. Available data analysis programs have been designed with emphasis on different aspects of the HX-MS analysis problem [2–13]. This paper describes a program called ExMS that can efficiently accomplish all of these tasks. ExMS is able to analyze high resolution HX-MS data for hundreds of peptide fragments and definitively identify and characterize them at the rate of  $\sim 2$  s per peptide per HX time point with desktop computer equipment and minimal operator intervention.

The ExMS program does not in its present form solve the ultimate problem of reaching amino acid resolution. However, experimental operations described before [1] together with ExMS analysis do appear to overcome what has been the major barrier to achieving high structural resolution. This is the need to obtain very large numbers of overlapping peptide fragments and to analyze, efficiently and accurately, the large quantity of complex HX information that they contain. Such results implicitly contain HX information at an amino acid-resolved level. One might assume that the ability to extract that information will be forthcoming.

Here, we first present an overview of the steps in the ExMS analysis, and then a more detailed description of the analytic operations and their logic. Examples taken from different kinds of HX experiments illustrate the tests used by ExMS to ensure correct peptide identification and the data analysis capabilities that result.

## Overview

In a typical day of HX experimentation, we run MS analyses of the following: (1) an undeuterated (*all-H*) protein sample in order to determine the HPLC retention time (*RT*) of each peptide for that day; (2) an initially *all-D* protein sample to calibrate the recovery of D-label for each peptide (back exchange); (3) up to 10 experimental H-D exchange time points. Common types of HX experiment range from the straightforward time-dependent HX behavior of the native state to the study of structure formation and kinetics during protein folding. These data are submitted to ExMS for analysis.

The workflow of the analysis is shown in Figure 1. For each experimental protein, we initially generate a working list of proteolytic peptides by MS/MS analysis of fragments produced in the same online analysis system to be used in H-D exchange experiments. We use SEQUEST (ThermoScientific Bioworks ver. 3.3.1) to identify the peptides in competition with an extensive decoy sequence database. Given this information ExMS establishes a list of accepted peptides based on a quality cut

off (Bioworks  $P_{\text{pep}}$  score) that was shown to reduce the false identification rate to 1/1000.

In HX experiments, ExMS searches through MS data to find and validate each listed peptide. It can be instructed to sequentially process an entire peptide list, or to search for any given peptide, or to focus on peptides from some specified protein region, or to resume an interrupted analysis. For each search peptide, ExMS systematically processes through the MS scans that fall within a specified chromatographic retention time (*RT*) range. It compares MS spectra with calculable mass spectrometric information (monoisotopic mass, charge state, isotopic peak positions, peak amplitude distribution), and with other local spectral features to validate or to question as ambiguous each potential peptide spectrum. MS scans that contain ambiguous and unambiguous peptide spectra are flagged as not matched and matched, respectively, with the goal of finding a run of matched scans for integration into a single enhanced *s/n* spectrum for a more definitive determination using a similar strategy.

An initial search is done through the *all-H* MS data to find each listed peptide and accurately establish its chromatographic *RT*. This search is aided by the fact that the *m/z* values for the peptide's isotopic peaks and their relative intensities are accurately known in advance (unlike the deuterated case). For each peptide, ExMS searches the MS scans that fall within an *RT* window based on an earlier MS/MS run, and can afford a wide *RT* window to allow for possible chromatographic drift in the intervening weeks. To ensure the correctness of the potential peptide identification, ExMS performs a number of tests. Individual isotopic peaks and peak envelopes must exceed noise threshold cutoffs. The individually resolved peaks must match expected peak positions on the *m/z* axis ( $\Delta m/z$  tolerance). The isotopic envelope must match the theoretically calculated shape (*intensity distribution* test). There must be no extraneous peaks, either above or below or within the possible *m/z* range of the search peptide that might signal the presence a misassignment (*intensity out of range* test). Possible matches must pass all of these tests; else they are marked as not matched. Scans that contain matches and non-matches for each search peptide are flagged.

A *sliding window* procedure is then applied to identify a run of sequential scans that contain the search peptide. These scans are summed. ExMS subjects the summed spectrum to all of the same tests that were done before at the single scan level plus an additional *isobaric/overlapping* test. This is the *autocheck* stage. Peptide assignments that pass the computer *autocheck* have very high reliability and can be automatically accepted. Assignments that remain less than completely definitive can be presented to the user for examination and judgment in the following *manual check* stage. These operations establish the sequential MS scans that contain correct matches for each listed peptide and thus define an accurate *RT* range for that peptide for that day. These procedures are illustrated in Figure 2.

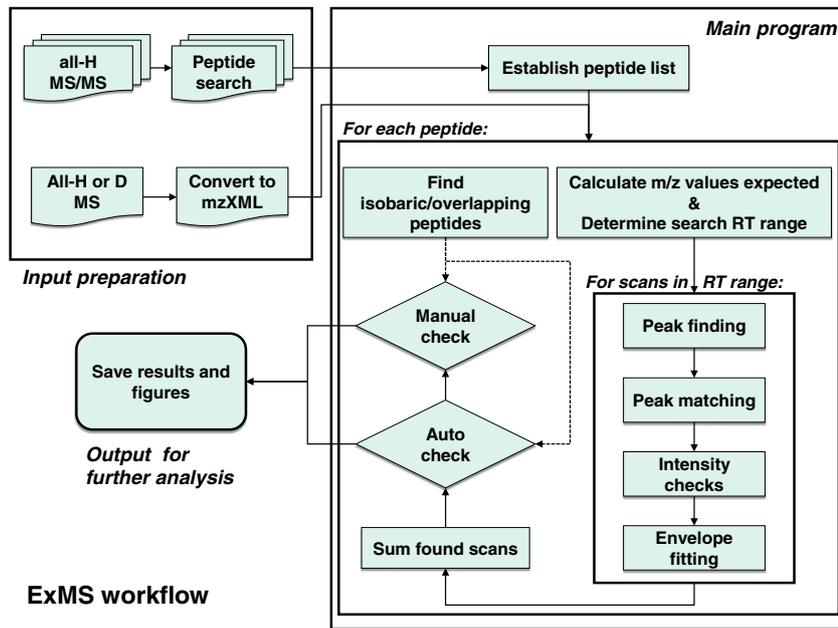


Figure 1. Workflow summarizing the operational steps performed by the ExMS program

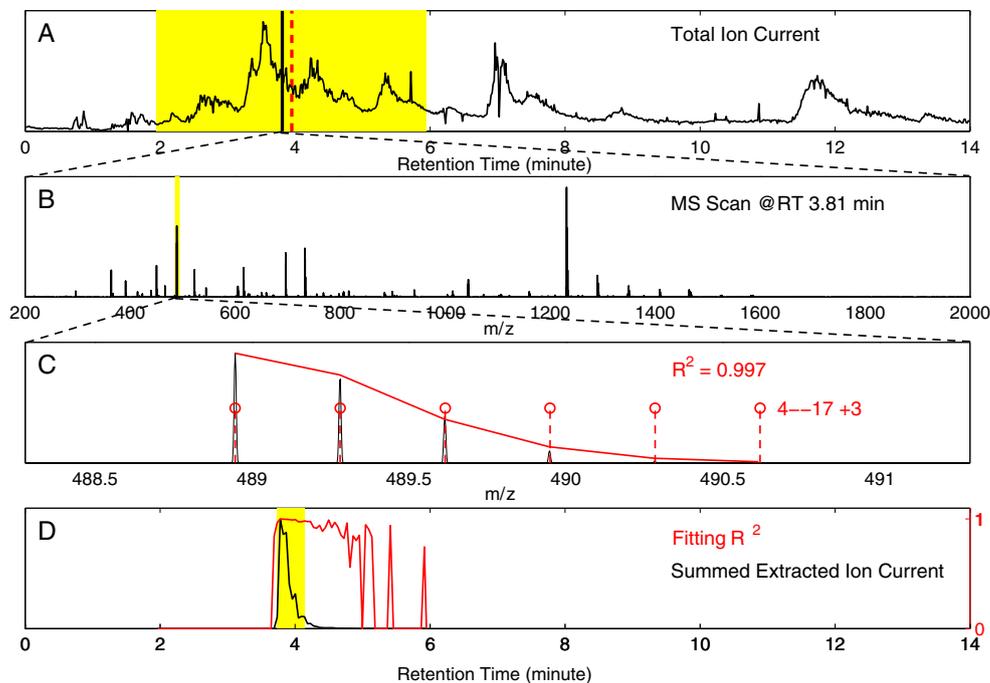


Figure 2. Illustration of ExMS processing for one peptide from an *all-H* sample. **(A)** TIC for the peptic peptides of an *all-H* sample. ExMS analyzes each scan within a settable window (yellow band) around the *RT* of each search peptide taken from a previous MS/MS analysis (dashed red line at 3.95 min for peptide 4–17 +3). **(B)** The *m/z* spectrum of one such scan (at *RT*=3.81 min). **(C)** Expanded section from **(B)**. ExMS finds peaks in each scan within the *RT* search range (9 point method, peaks shown). They are compared with the theoretical *m/z* of the search peptide peaks (dashed red lines). Potentially matching peaks (theoretical $\pm$ 10 ppm) and their envelope are tested to ensure a correct match. Result of the *intensity distribution* test is shown (isotopic envelope compared with theoretical intensity distribution). Scans that pass all tests are marked as matched. Matched spectra are summed and the summed spectrum is subjected to the same tests plus others (*autocheck* stage) to accept or question the assignment. **(D)** For each scan in the search range, a plot of the summed intensity of each matched envelope (l, black) and its score in the *intensity distribution* test ( $R^2$ , red). The  $l \times R^2$  parameter and other tests are used to specify the *all-H RT* range for that peptide for that day (yellow band). For subsequent analysis of deuterated samples, scans in this narrow *RT* range are analyzed similarly to identify each listed peptide, and measure its centroid mass, and therefore the number of carried deuterons

The same analysis is then applied to the more difficult search through MS spectra for the deuterated samples. This search is aided by the fact that an accurate *RT* range is now known for each peptide, so that the search can be limited to scans within a narrow *RT* window. However, the overall *m/z* range that must be considered is larger because the number of carried deuterons is not known. Also, a larger *m/z* tolerance must be used in the peak search because of the difference between the  $^2\text{H} - ^1\text{H}$  and the  $^{13}\text{C} - ^{12}\text{C}$  mass defect ( $\Delta\Delta \sim 0.003$  u). Also, the peak amplitude distribution cannot be prespecified since the number and distribution of carried deuterons is unknown. Thus the *intensity distribution* test must use different trial fitting functions, as detailed below. The other tests are the same as for the *all-H* data. Again, the scans are appropriately flagged and the sliding window procedure is applied. Scans that pass are summed to produce a single summed spectrum for each peptide. The summed spectrum is subjected to the *autocheck* tests, namely all of the same tests as before plus two others (*isobaric/overlapping* peptide test and *RT range overlap* test), which either validates or questions the peptide identification.

ExMS can optionally present to the operator for visual examination and acceptance or rejection (*manual check*) the assignments that pass the *autocheck* and more importantly those that remain ambiguous. We find that most identified peptides successfully pass all the way through the *autocheck*. Those that fail most often do so for a reason that becomes apparent to the user in the *manual check* where they can be individually viewed and judged in seconds. ExMS also allows for manual adjustment of *RT* range, manual cleaning (flagging noise peaks), and then operator decision.

In the analysis of time-dependent HX-MS data, these operations are used to search MS spectra for the large lists of peptides obtained as described in [1]. ExMS then calculates the centroid mass of each peptide from the *all-H* and the deuterated samples, and thus the average number of deuterons carried by each peptide. It will display the time course of HX labeling through sequential samples for both single and bimodal labeling. Figure S1 shows a straightforward example; other figures in the main text show results from other HX experiments. Also, ExMS records and can present the pass/fail record of each search peptide at each step in the analysis. An example is in Table S2. This feature aids the informed choice of settable parameters, and more generally will facilitate modification and improvement of the program.

## Detailed Operations

The ExMS program is implemented in MATLAB using routines from the Bioinformatics, Curve Fitting, and Optimization Toolboxes. Once the MS/MS peptide list is established, ExMS is able to automatically process MS spectra for sequential deuterated data sets (HX time

points, etc.), identifying and analyzing peptides at the rate of  $\sim 2$  sec/peptide on a typical personal computer ( $\sim 1000$  scans in a 15 min chromatogram with  $\sim 80$  scans searched per peptide). While the program has only been tested with LTQ Orbitrap data, input uses mzXML format files so it should work with other MS data that resolves individual isotopic peaks, although some parameter adjustments will be necessary. ExMS was developed and tested in extensive work with the proteins described in [1] and others. Details of the analysis are described below.

## The Peptide List

For each proteolysis condition used, MS/MS runs are done with *all-H* protein samples to assemble a list of peptides that will be available for HX analysis. We use SEQUEST to identify the sequence of each peptide, record its chromatographic retention time (*RT*), and specify an identification quality score (Bioworks  $P_{\text{pep}}$  is used). Given this data, ExMS constructs an initial list of potentially useful peptides and their properties. By using an empirically calibrated  $P_{\text{pep}}$  quality test, it culls the peptide list to reduce incorrect identifications to  $< 1/1000$  [1]. If SEQUEST assigns a peptide to a given sequence but it appears in the list with more than one retention time (e.g., due to residual peptide intensity in the chromatogram after expiration of its dynamic exclusion listing), ExMS selects the one with the best  $P_{\text{pep}}$  score. If a peptide sequence can be placed at more than one location in the protein (usually only with very short peptides), ExMS discards it.

For MS spectral searching, ExMS calculates the exact monoisotopic *m/z* ( $m_{\text{mono}/z}$ ) of each listed peptide, the expected observable number of peaks (*maxH*) above  $m_{\text{mono}/z}$  in its heavy atom natural abundance isotopic envelope ( $^{13}\text{C}$ ,  $^{15}\text{N}$ ,  $^{18}\text{O}$ ,  $^{34}\text{S}$ ), the *m/z* of each peak, their expected intensity distribution, and the maximum possible number of exchangeable D atoms that can be added (*maxD*). *MaxH* refers to the heaviest peak in the *all-H* isotopic distribution expected to be present with appreciable intensity ( $> 0.1\%$  of the highest amplitude peak). *MaxD* is taken as the number of amino acids minus proline residues minus 2 because the first two residues are expected to lose their D label due to back exchange during the analysis [14]. (For simplicity we write  $m_{\text{mono}}$  to include the mass of the additional protons that produce the ionic charge *z*.)

## Peak Searching and Matching

ExMS searches for each listed peptide in turn through the MS scans that fall within that peptide's retention time search window. For *all-H* samples, the *RT* range is taken from the MS/MS run in which that peptide was identified  $\pm$  a user settable window which can be set very wide, with no appreciable disadvantage (e.g.,  $\pm 2$  min as in Figure 2A), in order to allow the use of MS/MS results obtained long

before, changing HPLC columns, etc. For deuterated samples, the more accurate *RT* range found from that day's *all-H* sample is used (Figure 2D), typically  $\pm 0.5$  min.

To find experimental peaks in each MS scan, the default is a 9-point algorithm in which a peak is picked if the central data point is greater than the 4 points on either side. The *m/z* location of the peak is taken as the centroid of the 9 points and the peak intensity as their sum. This algorithm works well for Orbitrap data with resolution from 100,000 to 7500. An optional Gaussian fit through the 9 points and the 'mspeaks' function in the MATLAB Bioinformatics Toolbox give peak locations and intensities in good agreement with the simpler 9 point method but they execute more slowly.

For each search peptide, ExMS calculates the expected *m/z* locations of its isotopic peaks. For *all-H* samples, peak *m/z* values based on natural isotopic abundance are well defined. They take into account the  $^{13}\text{C} - ^{12}\text{C}$  mass difference ( $\Delta m\text{C}=1.003355$ ), and are calculated from  $m_{\text{mono}}/z$  out to  $(m_{\text{mono}} + \text{maxH})/z$ . For deuterated samples, expected peak locations must also take into account the D - H difference ( $\Delta m\text{H}=1.006277$ ). We adopt the following compromise. For the first two peaks above  $m_{\text{mono}}/z$ , the  $\Delta m/z$  added for each heavier peak is taken as  $\Delta m\text{C}/z$ . For the third and fourth peaks, the  $\Delta m/z$  added is the average of  $\Delta m\text{C}$  and  $\Delta m\text{H}$  ( $[\Delta m\text{C} + \Delta m\text{H}]/2z$ ). For all heavier peaks  $\Delta m\text{H}/z$  is added for each new peak. These positions are calculated out to  $(m_{\text{mono}} + \text{maxH} + \text{maxD})/z$ .

An experimental peak with intensity above a specified noise threshold ( $3e3$ , adjustable) is considered to match an expected peak only if it is within  $\pm 10$  ppm (adjustable) in *m/z* for *all-H* samples, or  $\pm 20$  ppm (adjustable) for deuterated samples (larger tolerance due to the mass uncertainty just described). Given sufficiently resolved mass spectra, this procedure is able to make the correct distinctions, as illustrated in Figure 3. If there is more than one experimental peak within an expected peak *m/z* window, ExMS selects one based on peak intensity and nearness to the expected *m/z*, using an intensity/nearness parameter  $[\text{intensity}/\exp(200\Delta m/z)]$  designed to appropriately weight the discrepancy in *m/z*. These decisions act to exclude noise and overlapping peaks and to select the correct peaks, which are used in further testing and fitting operations.

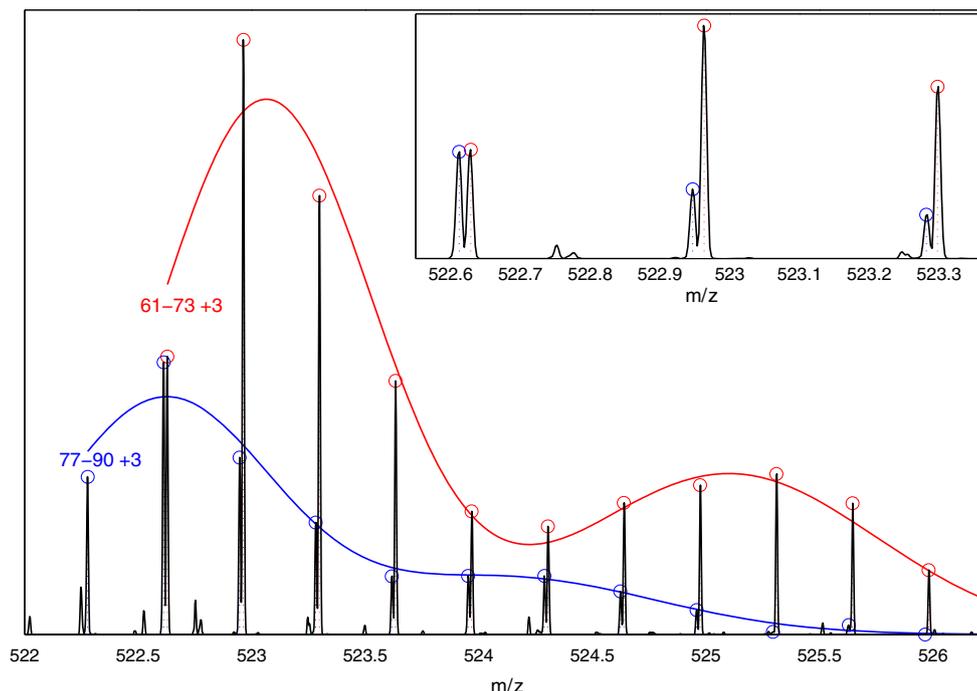
### Isotopic Peak and Envelope Validation

Given peaks within an MS scan that appear to match the search peptide, tests are done to ascertain whether they do reliably represent the expected peaks and envelope. An initial *noise threshold* test requires that the summed intensity of the apparently matched peaks is greater than  $1e4$  (adjustable).

*Intensity out of range* tests check whether the assigned peaks or their quantitation are made questionable by the presence of any other peaks that might indicate a

misassignment or the presence of another interfering peptide. Scans are flagged as not matched in the following cases. (1) There is significant intensity ( $>5\%$  of the highest isotopic peak; adjustable) at  $(m_{\text{mono}} - 1)/z$  and also a peak at  $m_{\text{mono}}/z$  above the noise intensity threshold ( $3e3$ , adjustable). This would indicate that the assignment is incorrect because the correct peptide cannot have intensity below its monoisotopic mass. The test for a peak at  $m_{\text{mono}}/z$  for deuterated samples, where all of the intensity is moved above the monoisotopic mass, avoids incorrect rejection due to random noise at the  $-1$  position. (2) For *all-H* samples, there is significant intensity at  $(m_{\text{mono}} + \text{maxH} + 1)/z$ , and also a peak at  $(m_{\text{mono}} + \text{maxH})/z$  above the noise threshold. For deuterated samples, there is significant intensity at  $(m_{\text{mono}} + \text{maxH} + \text{maxD} + 1)/z$  and also a peak at  $(m_{\text{mono}} + \text{maxH} + \text{maxD})/z$  above the noise threshold. This would indicate misassignment to or overlap with a peptide that is heavier than the search peptide. (3) There are peaks at *m/z* positions between the expected isotopic peaks, commensurate with a higher charge state ion, and the summed intensity of the wrong charge state peaks is above a noise threshold ( $>5\%$ ; adjustable) of the summed intensity of the supposedly correct expected peaks. The various threshold tests are intended to avoid rejecting correct spectra due to low intensity noise. Figure 4 shows examples of the *intensity out of range* tests.

The *intensity distribution* of the matching isotopic peaks is tested to ensure that potentially matched peaks are all from the target peptide. For *all-H* samples the intensity pattern of the trial isotopic envelope is tested against the expected natural abundance distribution obtained by direct calculation (criterion is  $R^2 > 0.7$ ). For deuterated samples, the envelope shape cannot be predicted. Instead, the goodness of fit to a smooth curve through the data is measured. Because of the statistical nature of isotopic distributions, any correctly identified peptide must have a smooth intensity distribution, whereas chemical noise or other kinds of mismatch will not. For deuterated samples, multimodal mass distributions are also possible due to EX1 HX behavior or to the presence of heterogeneous protein conformations during H-D exchange labeling. These will require (at least) a bimodal fitting function for the smoothness test. ExMS allows the operator to choose either a bi-Gaussian or a polynomial fit constrained to decay to zero intensity at  $(m_{\text{mono}} + \text{maxH} + \text{maxD} + 1)/z$ . The polynomial fit is faster but requires a more lax  $R^2$  threshold to avoid incorrect rejection. We use a goodness-of-fit  $R^2$  for the intensity distribution test  $\geq 0.7$  (adjustable) for a bi-Gaussian and  $\geq 0.5$  (adjustable) for a polynomial. If  $R^2 < 0.7$  for the bi-Gaussian fit, a second curve fit is tried as the sum of the calculated *all-H* distribution plus a single Gaussian. This option accommodates smaller peptides with low deuteration for which the  $^{13}\text{C}$ -dependent amplitude distribution is far from Gaussian. The better goodness-of-fit  $R^2$  value



**Figure 3.** Peak matching example shown for two closely overlapping peptides. Comparison of experimental and theoretical peak positions and envelopes for peptides 61–73 +3 and 77–90 +3. The ExMS analysis finds each peptide in its turn, automatically, in the *autocheck* stage. In this case, both pass all tests. This figure illustrates the ability of ExMS to distinguish spectrally overlapping envelopes from different peptides, even when the isotopic peaks are nearly isobaric (inset) and the envelopes are bimodal. These data, taken from an H-D exchange pulse labeling experiment during kinetic protein folding, exhibit a transient folding intermediate and some already folded native protein. Additional data taken as a function of pre-folding time (not shown) show the rate for formation and loss of the intermediate and for formation of the native protein

is recorded. A two binomial fit, with speed between the bi-Gaussian and polynomial, can also be used. Figure 5 illustrates the fit to an *all-H* peptide by its theoretical distribution (panel a) and by a bi-Gaussian for fully deuterated (panel b) and partially deuterated (panel c) samples (see also Figure 3).

### Defining the RT Range (Sliding Window Procedure)

To be flagged as matched, each analyzed scan must pass all of the above tests. In addition, the integrated intensity weighted by the intensity distribution goodness of fit ( $I \times R^2$ ) in each scan must be higher than 5% (adjustable) of the maximum value of the quantity  $I \times R^2$  in any scan for that peptide. A sliding window procedure is then applied. A window of adjacent scans, for example 5, is checked. If 3 of the 5 scans in any window are matches, then all 5 are counted as matched. This is done for each 5 scan window, sliding over by one scan at a time. In order to speed the analysis of the many scans within the RT range to be searched for each peptide, ExMS has an option to initially examine only every  $n^{\text{th}}$  scan (typically every third scan). In this case the sliding window procedure would use a window of 3 examined scans and a 2 out of 3 match test. The range of matching windows defines the RT range. All scans within

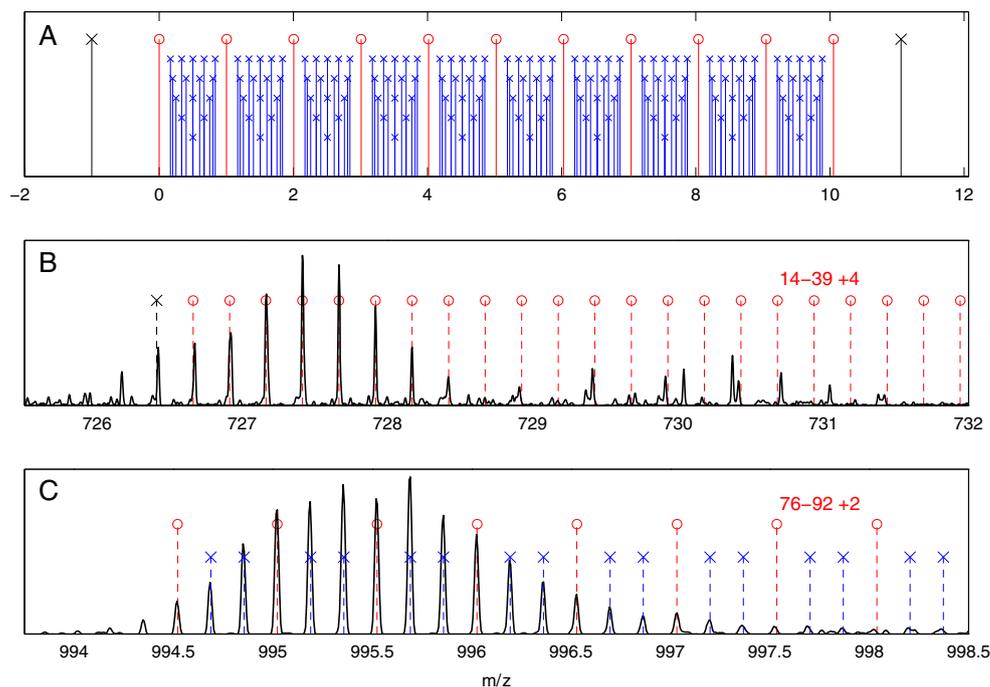
the RT range are summed, including intervening individual scans not specifically examined and those not matched. This allows the analysis to tolerate occasional bad scans, for example due to an unstable spray.

This procedure yields an RT range and a single summed envelope with enhanced s/n for each peptide to be used for subsequent operations (*autocheck* test, *manual check*, centroid evaluation, isotopic deconvolution, time plots, etc.). If more than one merged RT range, separated by non-matched scans, is generated by this procedure, all are saved for the *autocheck* and the final *manual check*.

### Autocheck

In the *autocheck* stage, the above tests are applied to the summed MS data. In order to pass *autocheck*, the following requirements must be met. The summed peak intensity divided by the number of summed scans must exceed a *noise threshold cutoff* ( $1e4$ ; adjustable). The spectrum must pass the *intensity out of range* tests. The *intensity distribution* test must achieve a fit  $R^2 > 0.95$ . For a deuterated sample, the RT range found must overlap the narrow RT range found for the *all-H* sample.

In an additional *isobaric/overlapping* test, the initial MS/MS peptide list is searched for any other peptide that could mimic the search peptide, given any level of deuteration.



**Figure 4.** *Intensity out of range tests.* **(A)** Expected  $m/z$  positions of isotopic peaks for the search peptide, relative to  $m_{\text{mono}}/z$  taken as zero, are marked in red. Locations marked in black are examined to check for a possible misassignment due to a co-eluting and overlapping peptide (check at  $(m_{\text{mono}} - 1)/z$  and at  $(m_{\text{mono}} + \text{maxH} + \text{maxD} + 1)/z$ ). Locations marked in blue are examined to check for a possible misassignment due to a peptide with higher charge state indicated by intervening peaks. ExMS checks up to  $z=8$ ;  $z=2$  to 6 are shown; actual positions will depend on the charge state of the search peptide. **(B)** In the search for peptide 14–39 +4, ExMS finds the possible match shown but this identification is questionable due to the presence of a significant *out of range* peak at  $(m_{\text{mono}} - 1)/z$  (black x). The scan is therefore marked not matched. **(C)** In the search for peptide 76–92 +2, ExMS finds the possible match shown, but this peptide fails the *out of range* test due to intervening peaks (blue). The scan is therefore marked not matched

(Examples are shown in Table S1.) This test is similar to the *intensity out of range* tests which check for peaks from interfering peptides outside of the possible range of the search peptide. The *isobaric/overlapping* test extends that examination to cases where a known peptide may have isobaric peaks that can overlap and potentially interfere with the search peptide. In this test ExMS examines whether the peptide list has any peptide with possible peak  $m/z$  values that can match possible  $m/z$  values of the test peptide, and also the same charge state and a nearby retention time. The  $m/z$  window is  $\pm 10$  ppm (adjustable) for *all-H* samples or  $\pm 20$  ppm (adjustable) for deuterated samples. The  $RT$  window is set very wide because of the possibility of chromatographic tailing by an abundant peptide, namely between 3 min earlier than the target peptide and 1.5 min after (adjustable). This is the cautious strategy. If these conditions exist, the identification of the search peptide will fail the *autocheck*, and will be submitted to *manual check* for operator examination.

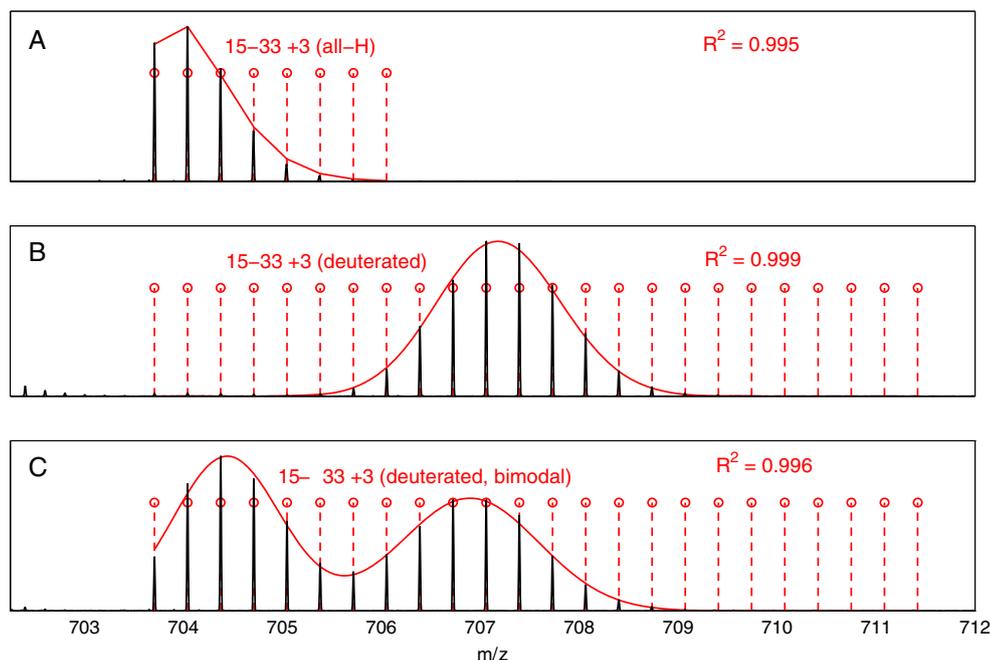
In cases where more than one separated  $RT$  range passes the *autocheck*, ExMS will choose the correct  $RT$  range of the search peptide based on summed peak intensity and nearness to its expected  $RT$  value. The comparison parameter used is  $(\text{intensity} \times \text{fitting } R^2 / \exp(\Delta RT))$ , designed to appropriately

weight discrimination by  $\Delta RT$ . For *all-H* samples  $\Delta RT$  is taken relative to the MS/MS  $RT$ . For deuterated samples  $\Delta RT$  is measured relative to the *all-H* range. This situation may arise when one of the identifications is incorrect. Alternatively, both may represent the same correct peptide when an abundant peptide has a long chromatographic tail that is interrupted due to local noise or an intervening peptide or a bad spray, which can insert no matched scans in the sliding window procedure. In general, signals from chromatographic tailing will be eliminated by the  $I \times R^2$  parameter shown in Figure 2D.

In our experience, close to 100% of the *all-H* peptides and 50% to 75% of peptides from partially deuterated samples successfully pass the *autocheck* (see Table S2 and Table S1 in [1]). Potential peptide assignments that do not automatically pass all the way through the *autocheck* are saved and optionally presented to the user for assessment or adjustment in a final *manual check* stage.

### *Manual Check*

In the *manual check* stage, ExMS presents the peptide spectra that pass and that do not pass the *autocheck*. Most summed scans that fail *autocheck* do contain the correct



**Figure 5.** Intensity distribution test. Theoretical isotopic peak positions are in red. **(A)** The match between theoretical and experimental peak intensities for an *all-H* sample. **(B)** Bi-Gaussian fit to a fully deuterated sample. **(C)** Bi-Gaussian fit to a partially deuterated sample. All three potential identifications pass the *intensity distribution* test since  $R^2$  is  $>0.7$ . (These data, from a time-dependent H-D exchange experiment on  $\alpha$ -synuclein amyloid (Parkinson's disease), demonstrate static structural heterogeneity near the molecular N-terminus, define its position, and measure the population fractions. Time-dependent HX data (not shown) measures the segment-resolved stability of each)

search peptide but also some noise or other suspicious feature, or ExMS finds another peptide in the initial MS/MS peptide list with isotopic peaks that could overlap peaks of the search peptide. These cases become immediately apparent in the *manual check* presentation and can be reviewed and judged in a few seconds per peptide.

Figure 6 illustrates the *manual check* presentation for two peptides that failed *autocheck*. Chromatographic data for each search peptide along with nearby *RT* data are presented in the upper window, and MS data including the expected  $m/z$  locations of the search peptide are presented in the lower window. The operator can tell at a glance that the peptide search failed the *autocheck* because ExMS found an isobaric peptide in the known peptide list. In both cases, the tentative assignment suggested in the upper panel is supported by the good chromatographic quality of the search peptide, by the good fit to its isotopic envelope, and by the absence of other ambiguities. In general, the user may choose to accept or reject individual peptide assignments, may request a re-analysis of any peptide, or may choose to make some adjustments, for example to manually set the *RT* range or to flag certain peaks as interfering and not to be included in the analysis.

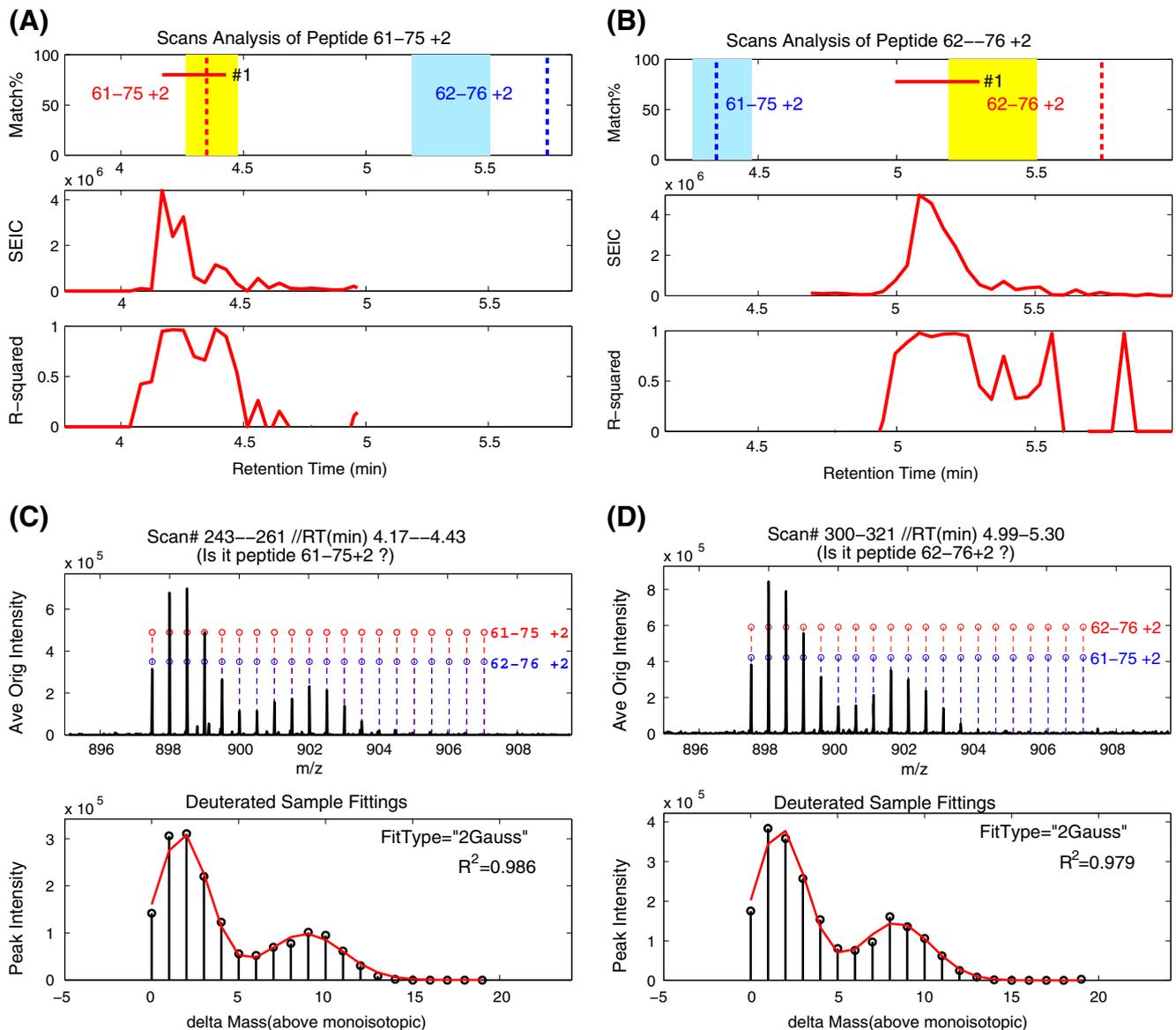
### The Pass/Fail Record

To assess performance, ExMS keeps track of the pass/fail record for each test that it applies at both the *autocheck* and

*manual check* stages. Table S2 lists detailed pass/fail statistics for *all-H* and 50% deuterated samples for  $\alpha$ -synuclein and Hsp104, the smallest and largest proteins that we studied (140 and 908 residues, respectively). For *all-H* samples, almost all peptides pass the *autocheck*, with no operator intervention, while a majority of the 50% deuterated peptides do so. Essentially all spectra that are correct but failed the *autocheck* could be rescued at *manual check*. The first three tests listed (1e4 cutoff, *out of range*, *intervening position*) fail very few candidates at the summed scan stage because they had been applied at the previous individual scan level. The failure rate for goodness of envelope fit (test #4) is about the same for *all-H* and deuterated samples suggesting that the fitting functions used for deuterated samples function as well as for the definitely known *all-H* sample. In the *isobaric/overlapping* test, deuterated samples fail more often due to the wider  $m/z$  window, as might be expected, and peptides from the larger protein fail more often due to the greater probability for spectral overlap. Table S2 also lists statistics for the small number of peptides for which ExMS did not find a correct *RT* range.

### Final Data Output

For each confirmed peptide, ExMS calculates the experimental  $m/z$  centroid. A correction for back exchange during the analysis, measured using the *all-D* sample, can be



**Figure 6.** Sample ExMS *manual check* presentation. Two different *manual check* presentations are shown. The search for 61-75 +2 (left) in a deuterated sample does not pass the *autocheck* because ExMS finds an isobaric peptide in the peptide list (62-76 +2), and the converse in the right windows. The upper window shows results on the retention time axis including (top panel) the RT value for the search peptide (from MS/MS; red dashed line), the RT range determined by ExMS for this peptide in the *all-H* sample (MS; yellow band, as in Figure 2), and the RT range for the two suggested peptide envelopes in the deuterated sample (MS; horizontal red lines). The next two panels in this window assess the quality of the possible match. As in Figure 2D, they show plots of the summed intensity of the isotopic peaks of the peptide spectrum in each individual scan (SEIC is summed extracted ion current), and the intensity distribution test R<sup>2</sup> value of each scan. The lower window (C) and (D) shows MS data for the putative search peptide, compared to expected *m/z* locations of the search peptide (red) and the confusing isobaric peptide (blue). The red and blue *m/z* positions are identical because the peptides in question are isobaric. The lower panel shows the fitted spectrum putatively found for the search peptide (*intensity distribution* test for the summed scans; bi-Gaussian fit to the MS data), and also demonstrates the absence of other ambiguous peaks (*intensity out of range* test). Given this display, the operator can quickly judge that the results definitively remove the initial ambiguity and clearly identify the MS spectra of the two isobaric peptides. Data are from the same H-D exchange pulse labeling experiments in Figure 3 showing a folding intermediate and some already folded protein. Here the peptides shown are absolutely isobaric as well as spectrally overlapping, but they can be individually identified in the manual check stage because they separate chromatographically (top panel)

applied. This leads to the average number of deuterons carried by each peptide. Depending on the nature of the H-D exchange experiment, ExMS can implement other modules,

including deconvolution of natural isotopic abundance to display the D distribution pattern alone, and fitting of multimodal data to two or more populations. The time course of

centroid or population data can be plotted to display the time dependence of H-D labeling. An example is shown in Figure S1.

### Recommended Choices

For optimum performance in terms of processing speed and the avoidance of false identifications, the following options are suggested. For *all-H* samples, use an  $R^2$  cutoff of 0.7 for the *intensity distribution* test. Search every third scan in the *RT* search window, determined as  $\pm 2$  min from the MS/MS-defined *RT*. Use a 2 out of 3 *sliding window* test to find the *RT* range. For *autocheck*, apply an  $R^2$  threshold of 0.95 for the *intensity distribution* test. For deuterated samples, search every third scan in the *RT* search window given by the *all-H RT range*  $\pm 0.5$  min. Apply the *intensity distribution* test to each examined scan using a polynomial function with  $R^2$  cutoff of 0.5. Use a 2 out of 3 *sliding window* test to define the *RT* range. For *autocheck*, apply all of the tests using a bi-Gaussian or a two binomial function for the *intensity distribution* test with  $R^2$  cutoff of 0.95. Best performance is obtained by using the faster polynomial fitting at the individual scan stage and the slower Gaussian fitting in *autocheck* where only one summed spectrum need be examined for each peptide.

### Conclusions

The ExMS analysis is able to accurately and efficiently identify and analyze, with minimal operator intervention, the hundreds of peptide fragments that can now be obtained for use in H-D exchange experimentation. These capabilities together are able to achieve high structural resolution and represent a step in moving toward ultimate resolution at the amino acid level.

### Acknowledgements

We thank the ExMS test users – Alec Ricciuti, Ben Walters, Tanya Panchenko, Sam Getchell – and also David Stranz (Sierra Analytics) for many valuable suggestions. We thank John Engen, Yoshi Hamuro, and Virgil Woods Jr. for a critical reading of the manuscript and exceptionally helpful suggestions. This work was supported by research grants

from the NIH (GM031847), NSF (MCB1020649), and the Mathers Foundation.

### References

1. Mayne, L.; Kan, Z.; Chetty, SP.; Ricciuti, A.; Walters, BT.; Englander, S. W. Many overlapping peptides for protein hydrogen exchange experiments by the fragment separation/mass spectrometry method. *J. Am. Soc. Mass Spectrom.* (2011). doi:10.1007/s13361-011-0235-4
2. Englander, J.J., Del Mar, C., Li, W., Englander, S.W., Kim, J.S., Stranz, D.D., Hamuro, Y., Woods, V.L.: Protein structure change studied by hydrogen-deuterium exchange, functional labeling, and mass spectrometry. *Proc. Natl. Acad. Sci. USA* **100**, 7057–7062 (2003)
3. Slys, G.W., Baker, C.A., Bozsa, B.M., Dang, A., Percy, A.J., Bennett, M., Schriemer, D.C.: Hydra: Software for tailored processing of H/D exchange data from MS or tandem MS analyses. *Bmc Bioinformatics* **10**, 162 (2009)
4. Pascal, B.D., Chalmers, M.J., Busby, S.A., Griffin, P.R.: HD desktop: An integrated platform for the analysis and visualization of H/D exchange data. *J. Am. Soc. Mass Spectrom.* **20**, 601–610 (2009)
5. Pascal, B. D.; Chalmers, M. J.; Busby, S. A.; Mader, C. C.; Southern, M. R.; Tsinoremas, N. F.; Griffin, P. R. The deuterator: Software for the determination of backbone amide deuterium levels from H/D exchange MS data. *Bmc Bioinformatics* **8**, 156 (2007).
6. Nikamanon, P.; Pun, E.; Chou, W.; Koter, M. D.; Gershon, P. D. Tof2H. A precision toolbox for rapid, high density/high coverage hydrogen-deuterium exchange mass spectrometry via an LC-MALDI approach, covering the data pipeline from spectral acquisition to HDX rate analysis. *Bmc Bioinformatics* **9**, 387 (2008).
7. Lou, X.H., Kirchner, M., Renard, B.Y., Kothe, U., Boppel, S., Graf, C., Lee, C.T., Steen, J.A.J., Steen, H., Mayer, M.P., Hamprecht, F.A.: Deuteration distribution estimation with improved sequence coverage for HX/MS experiments. *Bioinformatics* **26**, 1535–1541 (2010)
8. Weis, D.D., Wales, T.E., Engen, J.R., Hotchko, M., Ten Eyck, L.F.: Identification and characterization of EX1 kinetics in H/D exchange mass spectrometry by peak width analysis. *J. Am. Soc. Mass Spectrom.* **17**, 1498–1509 (2006)
9. Hamuro, Y., Zawadzki, K.M., Kim, J.S., Stranz, D.D., Taylor, S.S., Woods Jr., V.L.: Dynamics of cAPK type II- $\beta$  activation revealed by enhanced amide  $^1\text{H}/^2\text{H}$  exchange mass spectrometry (DXMS). *J. Mol. Biol.* **327**, 1065–1076 (2003)
10. Hotchko, M., Anand, G.S., Komives, E.A., Ten Eyck, L.F.: Automated extraction of backbone deuteration levels from amide  $^1\text{H}/^2\text{H}$  mass spectrometry experiments. *Protein Sci* **15**, 583–601 (2006)
11. Palmblad, M., Buijs, J., Hakansson, P.: Automatic analysis of hydrogen/deuterium exchange mass spectra of peptides and proteins using calculations of isotopic distributions. *J. Am. Soc. Mass Spectrom.* **12**, 1153–1162 (2001)
12. Chik, J.K., Graaf, J.L.V., Schriemer, D.C.: Quantitating the statistical distribution of deuterium incorporation to extend the utility of H/D exchange MS data. *Anal. Chem.* **78**, 207–214 (2006)
13. Kreshuk, A., Stankiewicz, M., Lou, X., Kirchner, M., Hamprecht, F.A., Mayer, M.P.: Automated detection and analysis of bimodal isotope peak distributions in H/D exchange mass spectrometry using HEXICON. *Int. J. Mass Spectrom.* **302**, 125–131 (2011)
14. Bai, Y., Milne, J.S., Mayne, L., Englander, S.W.: Primary structure effects on peptide group hydrogen exchange. *Proteins* **17**, 75–86 (1993)