# APPLICATION NOTE

# Semi-Automated Data Processing of Hydrogen Exchange Mass Spectra Using *HX-Express*

## David D. Weis and John R. Engen
Department of Chemistry, University of New Mexico, Albuquerque, New Mexico, USA

## Ignatius J. Kass
Waters Corporation, Beverly, Massachusetts, USA

A Microsoft Excel utility, *HX-Express*, that significantly accelerates the analysis of hydrogen exchange mass spectrometry data is described. *HX-Express* generates deuterium uptake and peak width plots from peaks in mass spectral data. Data analysis is intentionally semi-automated, requiring that the user find the peaks to be analyzed. The peaks are entered in the form of $x$, $y$ lists of $m/z$ versus intensity or can be directly imported from Waters MassLynx software. Analysis of data with *HX-Express* provides the same results as manual data processing but is substantially faster. In addition to speed, *HX-Express* provides and preserves visual and numeric displays of the analysis process for quality control. (J Am Soc Mass Spectrom 2006, 17, 1700–1703) © 2006 American Society for Mass Spectrometry

Hydrogen exchange mass spectrometry (HXMS) has become a powerful tool in the analysis of protein conformation and dynamics [1]. In this technique, a protein is exposed to $D_2O$ for various periods of time (labeling time-points) during which amide hydrogens are replaced by deuterium. The rate of deuterium uptake, a function of a number of factors including hydrogen bonding and protein dynamics, reveals details about protein conformation and dynamics. To localize the deuterium uptake, the deuterium-labeled protein can be digested with a protease (typically pepsin, which tolerates the low pH required to quench the exchange reaction). The peptides are then separated by reverse phase HPLC using a rapid gradient before mass analysis.

Once mass analysis has occurred, the experimental portion of the analysis is complete and all that remains is the data analysis, or data processing steps. For a protein of ~50 kDa, proteolytic digestion may produce 30 to 50 peptides. In a typical experiment of 8 to 12 labeling time-points, therefore, there may be as many as 600 peptide peaks for which the average mass (and hence deuterium uptake) must be determined. If HXMS is employed in a comparative mode in which two or more states (e.g., ligand-bound versus free, native versus mutant) of a protein are compared, there may be as many as 1200 individual spectra that must be analyzed.

In our laboratory, as well as many others, the data analysis process consists of a number of labor-intensive steps. First, the total ion chromatogram for each labeling time is inspected to locate a given peptide. Next, individual peptide spectra from the chromatogram that have sufficient signal-to-noise ratio are averaged. The centroid of the isotopic distribution for one of the charge states of the peptide is then determined. We copy the spectrum from the mass spectrometer data processing software and paste it into MagTran (Zhongqi Zhang, Amgen, Thousand Oaks, CA). MagTran allows interactive centroid measurement in which we choose the $m/z$ window over which the center of mass should be determined. The centroid value displayed on the computer screen is then typed into an Excel spreadsheet. The next peptide is identified and the entire process is repeated. Once all the centroid values for all the time points have been entered, a deuterium uptake plot is constructed. For an adept user, the entire process requires ~10 to 15 min for each peptide, depending on the number of labeling time-points.

Recently, we have begun to use peak width as a diagnostic for EX1 exchange [2]. Peak width analysis requires pasting each spectrum one at a time into an Excel worksheet that uses a Visual Basic for Applications (VBA) macro to determine the peak width. The resulting peak width data is also plotted versus time. Including peak-width processing in the analysis of HXMS data typically increases the analysis time to roughly 20 to 30 min per peptide. Needless to say, it is usually the case that with or without peak width analysis, analyzing the data consumes more time than acquiring the data.

To speed up the analysis of HXMS data, we have developed a Microsoft Excel-based application, named *HX-Express*, which automatically measures the centroid and peak width for all of the spectra of a given peptide. The application then produces a report consisting of deuterium uptake and peak width plots. Since most of our HXMS experiments involve comparing two or more protein states, the report displays data for all the states (bound, free, active, etc.) on the same plot to aid in comparison. While other HXMS data analysis software exists [3–5] *HX-Express* is available as freeware and runs within the ubiquitous Microsoft Excel environment.

## Methods

### HX-Express

*HX-Express* consists of a Microsoft Excel worksheet containing embedded Visual Basic for Applications (VBA) macros (Redmond, WA). The application requires Microsoft Excel version 97 or higher and has been tested on Windows 2000 and Windows XP operating systems. A minimum of 512 Mb of system memory is recommended. *HX-Express* is available from the authors at http://www.hxms.com/HXExpress.

To use *HX-Express*, the mass spectra must first be transferred into the *HX-Express* worksheet in two-column *x*, *y* format (*m/z*, intensity) using either a copy/
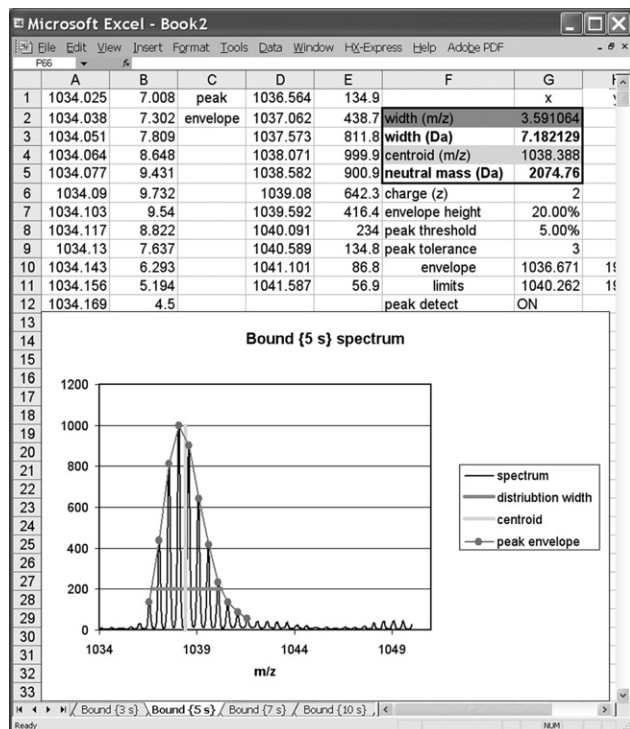


**Figure 2.** Screen capture of the summary page generated by *HX-Express*. The data are from a peptide in a ligand-bound and ligand-free protein. Deuterium uptake plots are shown on the left and peak width plots are shown on the right. Tabs on the bottom of the workbook lead to each of the processed spectra.

paste or a data import procedure. The transfer method will depend on the mass spectrometer software that was used to acquire the data. Automatic import of spectra is possible for data acquired with MassLynx software (Waters Corporation, Milford, MA). Data acquired from other instruments using software platforms can be copy/pasted, making *HX-Express* useful for any type of hydrogen exchange data, regardless of the instrument it was acquired on. Once the data are in the worksheet, the top row of the worksheet is used to indicate the deuterium exchange time for each spectrum. *HX-Express* includes a feature to assist with the exchange time annotation. Each series of data (e.g., free versus ligand bound, etc.) is placed in its own worksheet within the workbook. The worksheet tabs are used to designate the name of each series. To analyze the data, the user provides the charge state of the peptide, the percentage of base peak intensity (BPI) to use for peak width and centroid determination, an *m/z* tolerance for isotopic peak identification, and a noise threshold (as a percentage of base peak intensity) for isotopic peak detection. For spectra in which the isotopic peaks are unresolved, isotopic peak detection can be disabled. *HX-Express* creates a new workbook, distributes each *x*, *y* spectrum into its own individual named worksheet in the new workbook, and applies the analysis algorithms described below. A screen capture of a processed spectrum is shown in Figure 1. Finally, *HX-Express* generates a summary page containing the centroids, deuterium levels, and peak width values and plots of these values versus time. A screen capture of the summary page is shown in Figure 2. The deuterium uptake curves can be plotted in terms of peptide mass, relative deuterium level, or percent deuteration where



**Figure 1.** Screen capture of a spectrum processed by *HX-Express*. The circles and line connecting them shows the empirical peak-to-peak envelope. The horizontal bar shows the width at 20% of peak maximum and the vertical bar shows the centroid of the portion of the spectrum. Each spectrum in a dataset will have a tab like this in the Excel workbook.
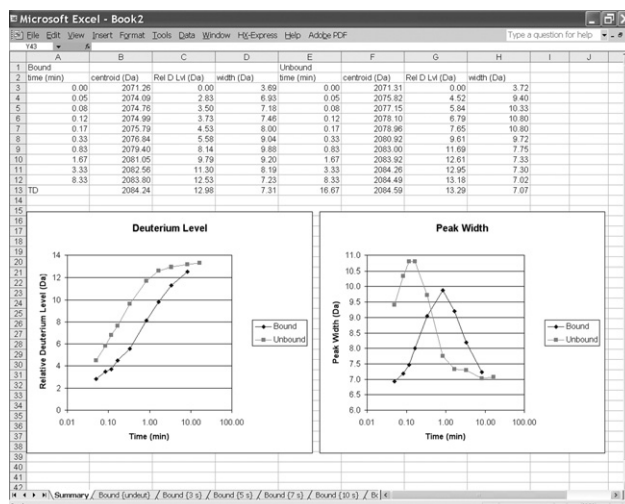
$$\text{relative deuterium level} = m - m_{0\%} \qquad (1)$$

and

$$\% \text{ deuteration} = \frac{m - m_{0\%}}{m_{100\%} - m_{0\%}} \times 100\% \qquad (2)$$

where $m_{0\%}$ and $m_{100\%}$ are the masses of an undeuterated and a fully deuterated control, respectively. It should be noted that unlike the DEX software package [4], there is no removal of the natural isotopic abundance by *HX-Express*. The amount of deuterium uptake determined by both *HX-Express* and DEX are the same, however, because in both cases the mass of the undeuterated species is subtracted from the deuterated masses in the final calculation of deuterium uptake (also described in [2]).

## Isotopic Peak Identification

The first step in the analysis by the software is identification of the isotopic peaks in the spectrum. To aid in peak detection the spectra must be smoothed. (For peptide spectra acquired on a Waters QTOF2 with 7000 to 10,000 resolution, a $2 \times 4$ point Savitsky-Golay smoothing function is sufficient). The user specifies the charge state ($z$), a noise threshold, and an $m/z$ tolerance for peak detection. The algorithm [2] identifies the most intense isotopic peak and then searches from the maximum in both directions to identify peaks that are above the noise threshold. A point in the spectrum is identified as a peak if (1) it is the local maximum of its four nearest neighbors, (2) is $1/z$ ($\pm$ $m/z$ tolerance) from the previous peak, and (3) is above the noise threshold. If the isotopic peak detection feature is disabled, *HX-Express* skips this process.

## Peak Width Determination

For the spectra of multiply charged species, the width ($\Delta m/z$) of the isotopic distribution is measured at a specified percentage of base peak intensity (typically 20 or 50%). The peak width in mass units is then simply $\Delta m/z \times z$. For spectra with unresolved isotopic peaks such as proteins at $z > +4$ (see Figure 3a), peak width determination is simply a mater of measuring across the charge state envelope at the desired percentage of peak maximum. Linear interpolation between the pair of adjacent data points with intensities that span the % BPI value are used to find the $m/z$ value at the desired percentage of peak maximum. For spectra in which the individual isotopic peaks are resolved, an empirical peak-to-peak envelope is laid over the isotopic peaks (Figure 3b and c). When there are isotopic peaks with intensities less than the width threshold (Figure 3b), linear interpolation between the pair of peaks that span the threshold is used to find the $m/z$ value at the desired percentage of peak maximum (filled circles in Figure 3b). Some peptide spectra do not have isotopic peaks
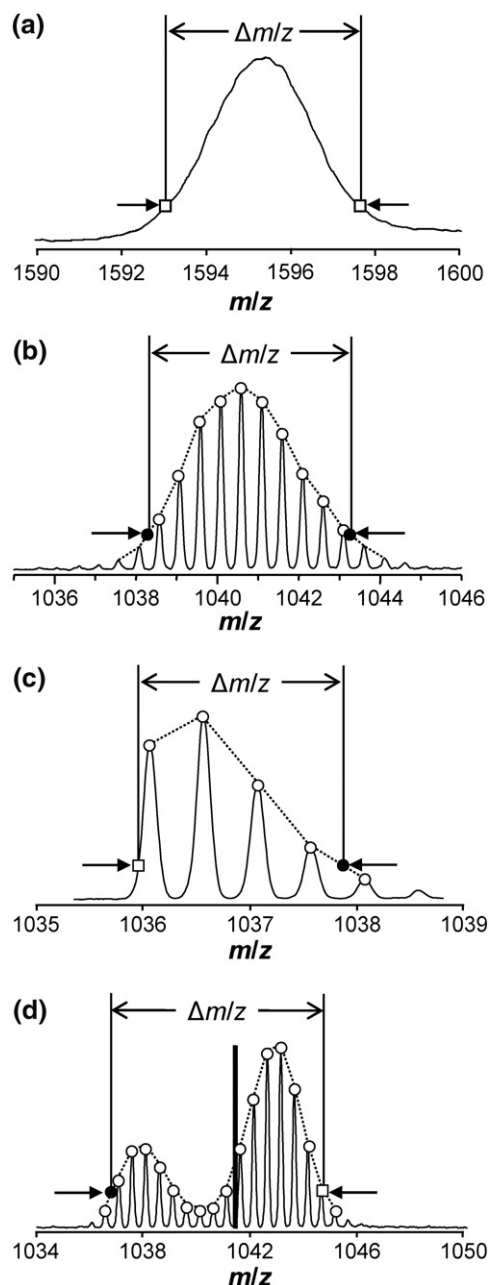


**Figure 3.** Illustrations of *HX-Express* processing for the determination of peak width and centroid for (**a**) intact protein without the use of isotopic peak detection, (**b**) a deuterated peptide, (**c**) an undeuterated peptide, and (**d**) a deuterated peptide showing a bimodal isotope pattern. The open circles are isotopic peaks identified by *HX-Express* with a noise threshold of 5% of the peak maximum. The dashed line shows a peak-to-peak empirical envelope laid onto the data. The width of each spectrum at 20% of the base peak intensity ($\Delta m/z$) is shown. The $m/z$ limits were determined by linear interpolation of the empirical envelope (filled circles) or by linear interpolation of the spectral data points (open squares) as described in the text. The vertical bar in (**d**) shows the centroid of the distribution.

with intensities below the threshold or isotopic peaks that straddle threshold exist only on one side of the maximum (typically, the high $m/z$ side) (Figure 3c). Since the monoisotopic peak is the maximum peak,

there are no peaks on the low $m/z$ side of the peak maximum. In these cases, the interpolation scheme automatically switches over from linear interpolation of the empirical envelope to direct linear interpolation of the spectral data (open square in Figure 3c). To avoid becoming entrapped in local minima such as in the biomodal distribution shown in Figure 3d, the algorithm defines the width of the isotopic distribution by searching from the low and high $m/z$ extremes towards the maximum. Thus, the width of the full distribution is measured and the centroid is calculated for the entire distribution (vertical bar in Figure 3d).

### *Centroid Determination*

The *HX-Express* determines the centroid

$$\frac{\sum_i (m/z)_i \times I_i}{\sum_i I_i} \qquad (4)$$

(where $I_i$ is the spectral intensity of the $i$th channel in the spectrum) from the portion of the spectrum that has an intensity greater than or equal to a user-specified fraction of the peak maximum intensity. This $m/z$ range is determined using the same algorithm used to determine the peak width.

## Discussion and Conclusions

Manual analysis, as described in the Introduction section, and *HX-Express* analysis of peptide datasets give identical results for the centroid values and peak-width analysis (results not shown). After the peptide spectra have been identified, averaged, and smoothed, an adept user can enter 10 spectra into *HX-Express* in only a few minutes. Importing data directly from MassLynx takes only a few seconds. *HX-Express* processing of 20 spectra requires ~5 s. Manual processing of this data, including production of the deuterium uptake curves and peak width plots [2] requires ~20 to 30 min. The agreement between the two methods is excellent.

*HX-Express* does not fully automate the analysis of HXMS data. We intentionally do not want total, black-box, automated analysis. Because selection of the peaks to process is the most difficult step in any computer-driven data analysis method, we choose to perform this step manually. Human peak identification is vastly superior to automated methods and is far less prone to errors. Furthermore, one can have high confidence in the quality of the data, the peak selection process, and the final data analysis if the peaks are first visualized before processing begins.

The benefits of *HX-Express* are the substantial savings of time, its reproducibility, and the preservation of a record of the data processing so that the data can be reinterpreted later if necessary. *HX-Express* provides a visual and numeric record of the processing (Figures 1 and 2) making it completely transparent to the user: each spectrum can be quickly inspected to insure accurate data processing simply by tabbing through the worksheets, and the entire dataset can be easily reprocessed if the results of the first analysis are not satisfactory.

## References

1. Wales, T. E.; Engen, J. R. Hydrogen Exchange Mass Spectrometry for the Analysis of Protein Dynamics. *Mass Spectrom. Rev.* **2006,** *25,* 158–170.
2. Weis, D. D., Hotchko, M., Wales, T. E., Ten Eyck, L. F., Engen, J. R. Identification and Characterization of EX1 Kinetics in H/D Exchange Mass Spectrometry by Peak Width Analysis. *J. Am. Soc. Mass Spectrom.* **2006,** in press.
3. Garcia, R. A.; Pantazatos, D.; Villarreal, F. J. Hydrogen/Deuterium Exchange Mass Spectrometry for Investigating Protein–Ligand Interactions. *Assay Drug Dev. Technol.* **2004,** *2,* 81–91.
4. Hotchko, M.; Anand, G. S.; Komives, E. A.; Ten Eyck, L. F. Automated Extraction of Backbone Deuteration Levels from Amide H/(2)H Mass Spectrometry Experiments. *Protein Sci.* **2006,** *15,* 583–601.
5. Chik, J. K.; Vande 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.