

Software

Open Access

The Deuterator: software for the determination of backbone amide deuterium levels from H/D exchange MS data

BD Pascal¹, MJ Chalmers², SA Busby¹, CC Mader¹, MR Southern¹, NF Tsinoremas^{* 1,3} and PR Griffin^{* 1,2}

Address: ¹Translational Research Institute, The Scripps Research Institute – Scripps Florida, 5353 Parkside Drive, Jupiter, FL 33458, USA, ²Department of Biochemistry, The Scripps Research Institute – Scripps Florida, 5353 Parkside Drive, Jupiter, FL 33458, USA and ³Department of Informatics, The Scripps Research Institute – Scripps Florida, 5353 Parkside Drive, Jupiter, FL 33458, USA

Email: BD Pascal - bpascal@scripps.edu; MJ Chalmers - chalmers@scripps.edu; SA Busby - sbusby@scripps.edu; CC Mader - mader@scripps.edu; MR Southern - southern@scripps.edu; NF Tsinoremas* - tsinorem@scripps.edu; PR Griffin* - pgriffin@scripps.edu

* Corresponding authors

Published: 16 May 2007

Received: 11 January 2007

BMC Bioinformatics 2007, 8:156 doi:10.1186/1471-2105-8-156

Accepted: 16 May 2007

This article is available from: <http://www.biomedcentral.com/1471-2105/8/156>

© 2007 Pascal et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: The combination of mass spectrometry and solution phase amide hydrogen/deuterium exchange (H/D exchange) experiments is an effective method for characterizing protein dynamics, and protein-protein or protein-ligand interactions. Despite methodological advancements and improvements in instrumentation and automation, data analysis and display remains a tedious process. The factors that contribute to this bottleneck are the large number of data points produced in a typical experiment, each requiring manual curation and validation, and then calculation of the level of backbone amide exchange. Tools have become available that address some of these issues, but lack sufficient integration, functionality, and accessibility required to address the needs of the H/D exchange community. To date there is no software for the analysis of H/D exchange data that comprehensively addresses these issues.

Results: We have developed an integrated software system for the automated analysis and representation of H/D exchange data that has been titled "The Deuterator". Novel approaches have been implemented that enable high throughput analysis, automated determination of deuterium incorporation, and deconvolution of overlapping peptides. This has been achieved by using methods involving iterative theoretical envelope fitting, and consideration of peak data within expected *m/z* ranges. Existing common file formats have been leveraged to allow compatibility with the output from the myriad of MS instrument platforms and peptide sequence database search engines.

A web-based interface is used to integrate the components of The Deuterator that are able to analyze and present mass spectral data from instruments with varying resolving powers. The results, if necessary, can then be confirmed, adjusted, re-calculated and saved. Additional tools synchronize the curated calculation parameters with replicate time points, increasing throughput. Saved results can then be used to plot deuterium buildup curves and 3D structural overlays. The system has been used successfully in a production environment for over one year and is freely available as a web tool at the project home page <http://deuterator.florida.scripps.edu>.

Conclusion: The automated calculation and presentation of H/D exchange data in a user interface enables scientists to organize and analyze data efficiently. Integration of the different components of The Deuterator coupled with the flexibility of common data file formats allow this system to be accessible to the broadening H/D exchange community.

Background

Amide hydrogen/deuterium exchange (H/D exchange) coupled with mass spectrometry (MS) has proven to be a powerful technique for the study of protein structure and dynamics as well as protein-protein and protein-ligand interactions [1,2]. The rate at which an amide hydrogen exchanges with solvent deuterium reveals much about its local environment. Changes in protein structure or dynamics can result in changes in H/D exchange kinetics for a subset of amide hydrogens along the protein backbone. Measurement of these alterations in kinetics provides chemical sensors for changes in protein structure and dynamics.

To perform these experiments the target protein or protein complex of interest is subjected to a time course of incubation with deuterium containing buffer to initiate on-exchange of deuterium into the protein. At prescribed time intervals, aliquots of the on-exchange mixture are rapidly quenched by the addition of a low pH and low temperature buffer containing denaturants to prevent further on-exchange, minimize off-exchange of incorporated deuterium and help denature the protein. The quenched and denatured protein is then subjected to enzymatic digestion by acid stable proteases. The resultant peptides are collected on a reverse-phase HPLC column and then gradient eluted directly into a mass spectrometer equipped with an electrospray ionization source where the mass of each peptide is measured. Alternatively, peptides may be spotted onto a target and analyzed with matrix assisted laser desorption ionization (MALDI) mass spectrometry [3,4]. By repeating this process for all on-exchange time points the mass increase of each proteolytic peptide over the time course of deuterium exposure can be determined. From these data the rates of deuterium exchange can be determined for the corresponding areas of the intact protein.

Since the inception of H/D exchange mass spectrometry, substantial advancements in methodology, automation and instrumentation have been made, [5-7] however, processing, curation, and validation of the large number of data points produced in a typical H/D exchange experiment [2,8] presents a significant data processing and analysis challenge. These issues have in part been addressed by several existing software tools and algorithms. Interactive software to simplify the calculation of average molecular weight has been developed [9]. Other algorithms are implemented as command line software which automate the extraction of deuterium content with Fourier transform and natural isotopic abundance based calculations [10,11]. Proprietary software has also been described that automates the extraction of deuterium content in a standalone software application [7,8]. Peak width analysis approaches have been created [12] as well

as Excel tools leveraging this approach [13]. Deconvolution techniques have been developed to determine deuterium content in less straightforward situations [14-16]. These above mentioned approaches have shown success in given cases, and their efforts serve to illustrate the interest and challenge of this area, however, the processing and management of H/D exchange data remains a significant bottleneck in the technique.

While progress has been made in H/D exchange data processing, these solutions have not always been easy to employ, and thus have not been widely adopted. Some solutions are command line based, and others are not freely available or their methods have not been described in detail. Often users are required to run separate programs, covering different functions such as peptide location and centroiding. In addition, output from the wide variety of MS instruments and database search engines presents another hurdle in the creation of broadly applicable solutions. There are no published H/D exchange software solutions based on platform independent common mass spectrometry data file formats, such as mzXML. The large size and number of files in conventional studies overburdens users with organizing, migrating and storing data, a critical factor that has yet to be addressed. Integrated user interfaces are also required that allow for the presentation, curation and validation of the results. These factors illustrate the need for software that is precise, efficient, convenient to use, and freely available to the scientific community.

A web-based software system called "The Deuterator" is presented here. It is freely available as an online service, and addresses many of the needs described above. The software can be used to precisely determine the deuterium content for protein segments in an integrated, high-throughput manner, taking steps toward instrument platform independence through the use of common data file formats. Novel tools have been implemented and are accessible through a user interface providing functionality such as the determination of percentage deuterium incorporation for a given peptide. The software is compatible with data from most common database search engines and is compatible with high and low-resolution mass spectral data. Although the software was developed for analyzing LC ESI-MS data, processing data from H/D experiments that do not incorporate a chromatographic step or data that were generated with MALDI MS instruments are supported equally. The output for each peptide included in the experiment is an accurate measurement of deuterium incorporation over time, which can then be plotted as deuterium vs. Log time graphs. Peptide sequences can be positioned next to the protein sequence to display the sequence coverage. Graphical representations of H/D exchange data can then be added by using

two-dimensional color maps to represent the data values, which may also be overlaid onto a high-resolution 3D protein structure. To develop and validate this web-based tool, we tested the system in a production environment for over one year with both low (~2000) and high (>60,000) resolution ESI-MS data.

Implementation

The Deuterator is available as a web-based application. The back-end architecture was built with Java and the online service was implemented with JSP. The MS data viewers were implemented using open source software [17,18]. Mass spectral and search result data are required in native format, and are converted into common file formats. Protein sequence data is required in FASTA format.

The broad range of mass spectrometers and sequence search tools has resulted in a large number of differing file formats, resulting in compatibility problems for many new and existing software packages. This issue has been addressed by recent efforts to create the common file formats mzXML [19] and mzData [20,21] for mass spectrometric data. One such effort [22] has developed the common formats mzXML and pepXML for mass spectral data and peptide identities respectively, and provides freely available mzXML conversion programs for various instrument formats. The Deuterator leverages these file formats and therefore takes steps toward greater compatibility with a broad range of mass spectrometers and protein search engines.

The components of The Deuterator are best viewed in a browser with tabbed browsing functionality, allowing for many time points to be viewed simultaneously. Results from the calculations are saved in a flat file format (CSV), which are easily viewed with applications such as Microsoft Excel.

Results and discussion

Workflow

Our H/D exchange workflow is shown in Figure 1 and starts with a typical proteomics experiment. The protein of interest is digested with an enzyme and the resulting peptides are analyzed by liquid chromatography tandem mass spectrometry (Figure 1A). Peptide identity is then established with the database search tool of choice (*e.g.* X!Tandem, Mascot or Sequest) (Figure 1B) and then visually confirmed with tools such as the Thermo-Finnigan Qual browser. Alternatively, peptides can be identified by using a peptide mass fingerprinting approach. It should be noted that the method for peptide identification, and the stringency by which the results are validated, are left to the end user. Subsequent solution phase amide H/D exchange experiments are performed for a number of pre-defined time periods (typically 8–10 time points span-

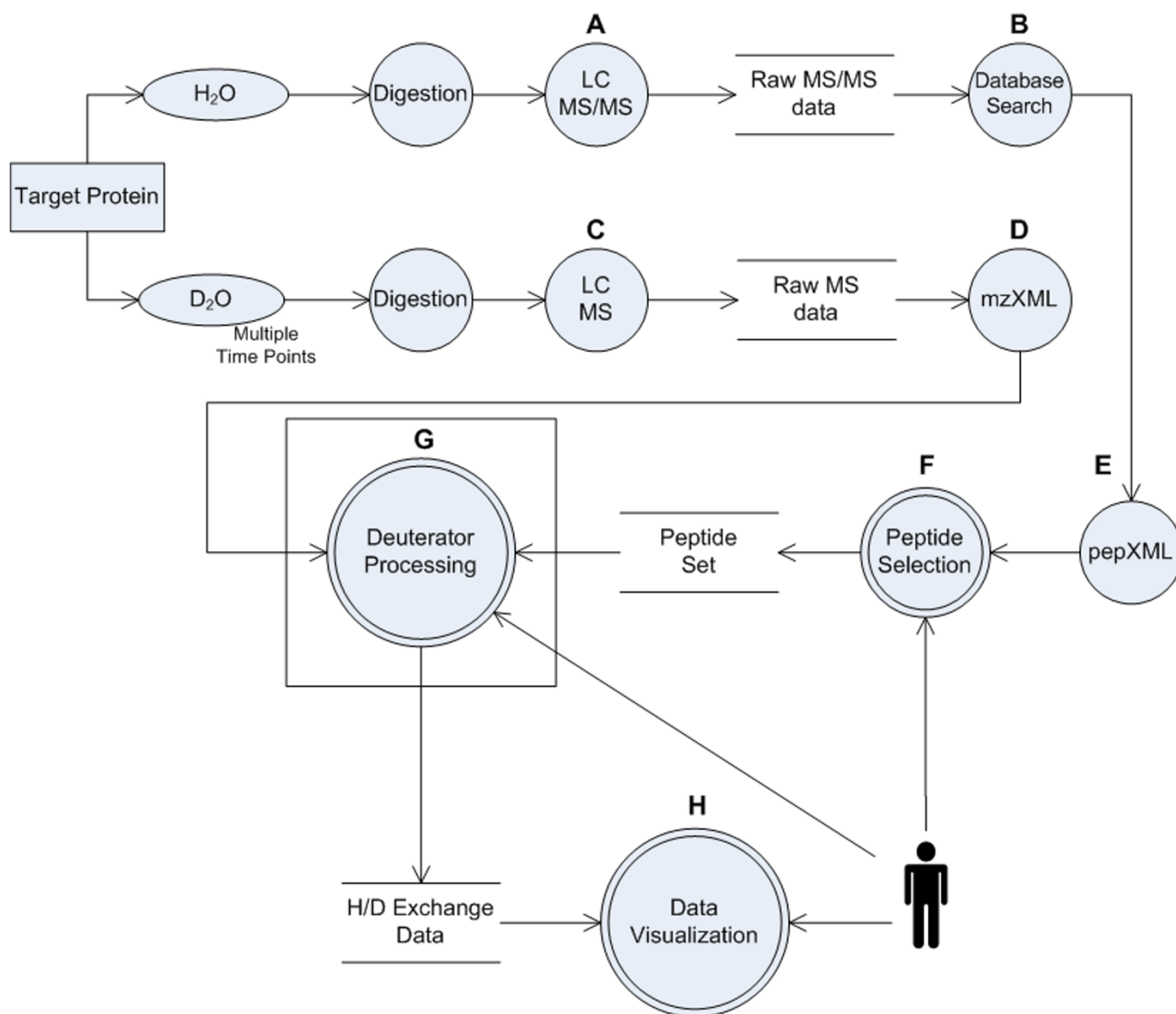
ning 1–10,000 s). Following incubation with heavy water (D_2O), the protein is digested and the resulting peptides are mass measured with LC-ESI-MS (Figure 1C). The raw MS data and search result files are then converted into the common file formats mzXML and PepXML, respectively (Figure 1D, 1E). The peptides of interest are selected from the search results to create the peptide set (Figure 1F). Some experiments may focus on a narrow region of the protein, in which case sequence coverage may not need to be comprehensive. Other experiments may require full protein sequence coverage, resulting in a larger peptide set. It is for these reasons that the validation and curation of the peptide set is currently the user's responsibility. The Deuterator then retrieves each search result, and uses the protein sequence data and mzXML files to proceed with positional determination, peak identification and centroid calculations (Figure 1G). The details of this method will be further discussed. Resulting data are then made available, and the curated results can then be sent to statistical interrogation and visualization tools (Figure 1H). Third party applications are then used to visualize the results such as graphs in Microsoft Excel and 3D structure overlays in PyMol [23]. Our software facilitates determination of H/D exchange rates and also facilitates differential H/D exchange analysis between multiple treatment conditions (*e.g.*, apo and ligand bound protein).

Data processing

There are three user inputs into The Deuterator: 1) The mass spectrometry data files (mzXML format); 2) The protein sequence file in FASTA format; and 3) A list of the peptides to be included in the H/D exchange experiment that we refer to as the "peptide set".

It is important to note that the comprehensive list of peptide identifications from a database search may not reflect the concise list of peptides of interest for H/D exchange experiments. For instance, some peptide identifications with poor search scores are often not correct, while other peptides identified may lie outside the experimental region of interest. For these reasons, we convert the PepXML search results into the common file format (CSV), and enable manual curation of the "peptide set" prior to import into The Deuterator (Figure 2).

Once input, the processing begins, and the "peptide set" is initially grouped according to sequence and charge and any redundancy is removed (multiple occurrences of the same peptide ions often occur within a single MS/MS data set). Peptide sequence, charge and retention time are the only data needed for processing, so in cases where these fields are identical, only one row is needed for representation. In cases where there are differing spectrum query times, the average of the group is preserved. Optionally, the score from the peptide search result, if available, can

**Figure 1**

Data Workflow. **A.** The protein of interest is digested with an enzyme and the resulting peptides are analyzed by LC MS. **B.** Peptide identity is then established with database search tools such as Sequest. **C.** The protein is incubated with D₂O at multiple time points, digested with an enzyme and the resulting peptides are analyzed by LC MS. **D.** Resulting raw files are converted to mzXML. **E.** Search results are converted into PepXML. **F.** User selects the peptides of interest from the search results to create the peptide set. **G.** Core Deuterator processing and interactive tools for H/D/exchange data curation. **H.** Data visualization components such as graphs in Microsoft Excel and 3D structure overlays in PyMol.

be preserved and represented in the Deuterator interface. The positional information for each peptide is obtained from the protein sequence located within the FASTA file and entered into *start* and *end* data fields. For each peptide the monoisotopic neutral mass is calculated and entered into the Mass data field. This is achieved by the summation of each residue mass, adding the mass of H₂O to the

result (mass = 18.0106, OH on the C-terminus and the H on the N-terminus). A chromatographic retention time (RT) window for each peptide is defined by *RTStart* and *RTEnd*. These values represent the query-spectrum retention time of the peptide in the pepXML file ± 0.15 minutes and function to allow for some temporal variability in the chromatographic step. All data fields are displayed to the

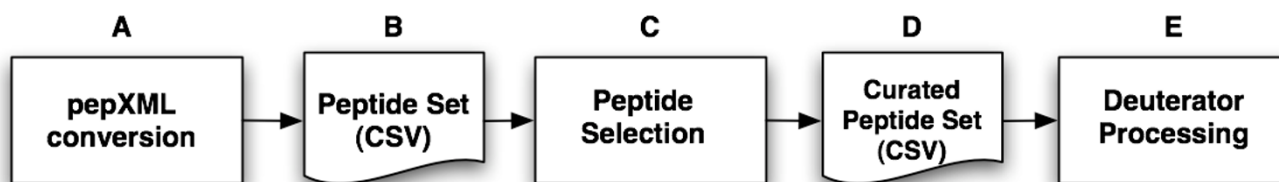


Figure 2
Peptide Set Process. **A.** The search results pepXML file is converted to CSV format to facilitate editing. **B.** The raw peptide set in CSV format. **C.** The user selects the peptides of interest. **D.** The final curated peptide set in CSV format. **E.** Deuterator processing.

user in the "Grid View" (Figure 3), which also contains additional data fields for each peptide; MZ_0 , MZ_{100} , MZ_{Start} , MZ_{End} , %D and Centroid.

Accounting for the charge state of the peptide, the 0% deuterium incorporation minimum, mz_0 , is defined as being the mass to charge ratio (m/z) value for the peptide's monoisotopic m/z . The 100% deuterium incorporation m/z limit, mz_{100} , for the given protein segment is then calculated according to equation where n = number of amino acids in the peptide, p = the number of non-N-terminal prolines, and z = charge.

$$mz_{100} = mz_0 + ((n - p - 2)/z) \tag{1}$$

The amino acid proline has no amide hydrogen and so they are subtracted from the equation. The subtraction of 2 in the equation is explained elsewhere [24]. The values

of mz_0 and mz_{100} are used in subsequent calculations and are presented in the interface to serve as positional markers. These values are different from MZ_{Start} and MZ_{End} , which are the user-defined limits of the m/z range, used to define the centroid calculation.

The software determines the average intensity of observed ions with given m/z values within the m/z and RT ranges defined by RT_{Start} and RT_{End} , MZ_{Start} and MZ_{End} . The average intensities of ions with given m/z values are derived by dividing the summed total of their observed intensities by the number of contributing spectra in native or profile mode. The end result is a series of mass-intensity pairs, called the "co-added" data that represents the isotopic distribution of the peptide ion of interest. The co-added data is displayed in the appropriate spectral viewer. Figure 4 shows examples of low (top), and high (bottom)

Peptide	Mono	Charge	Score	Start	End	Length	RT Start	RT End	m/z Start	m/z End	%d	Centroid	
<input type="radio"/> RALAKHLYDSY	1335.683	2	25.24	30	40	11	6.6	6.9	668.52	674	0.09	669.5629	0
<input type="radio"/> IRIFQGCQF	1110.554	2	25.27	97	105	9	11.47	11.7699995	555.95	558.6	0.0	556.7782	1
<input checked="" type="radio"/> YAKSIPGFVNL	1207.6497	2	25.27	117	127	11	11.200001	11.5	604.5	607.1	0.0	605.3153	2
<input type="radio"/> LKYGVHE	844.4338	2	25.27	136	142	7	9.35	9.65	423.49	427.59	0.61	424.88696	3
<input type="radio"/> ISEGQGFMTRE	1253.5605	2	25.27	159	169	11	8.05	8.349999	627.46	630.11	0.0	628.31805	4
<input type="radio"/> FLKSLRKPFQDF	1453.7977	3	25.27	170	181	12	9.700001	10.0	485.35	487.39	0.03	486.1492	5
<input type="radio"/> MEPKFEF	926.4103	2	25.27	182	188	7	10.85	11.15	463.88	466.48	0.01	464.6906	6
<input type="radio"/> AVKFNAL	761.433	2	25.27	189	195	7	9.150001	9.45	381.39	387	0.01	382.1263	7
<input type="radio"/> IILSGDRPGLL	1152.6763	2	25.27	209	219	11	10.150001	10.45	577.02	579.57	0.0	577.83075	8
<input type="radio"/> NVKPIED	813.4127	1	25.27	220	226	7	5.0299997	5.33	813.9	817.7	0.0	815.04	9
<input type="radio"/> QLKLNHPESQL	1392.7256	2	25.27	238	249	12	7.1	7.4	697.04	699.69	0.0	697.9071	10
<input type="radio"/> FAKLLQKMTDL	1306.7214	2	25.27	250	260	11	10.55	10.849999	654.04	657.14	0.02	655.13776	11
<input type="radio"/> RQIVTEHVQL	1221.6725	2	25.27	261	270	10	7.9500003	8.25	611.51	614.11	0.0	612.358	12
<input type="radio"/> LQVIKKTETDMSLHPLL	1965.0865	2	25.27	271	287	17	9.8	10.099999	983.22	990	0.0	985.63934	13
<input type="radio"/> QEIYKDLY	1070.518	2	25.27	288	295	8	4.06	4.36	535.94	541	0.0	537.688	14

Figure 3
Data Grid. Relevant H/D exchange data for each peptide are presented in the grid view. Selection of the peptide loads the associated data into the main spectral and extracted ion viewers.

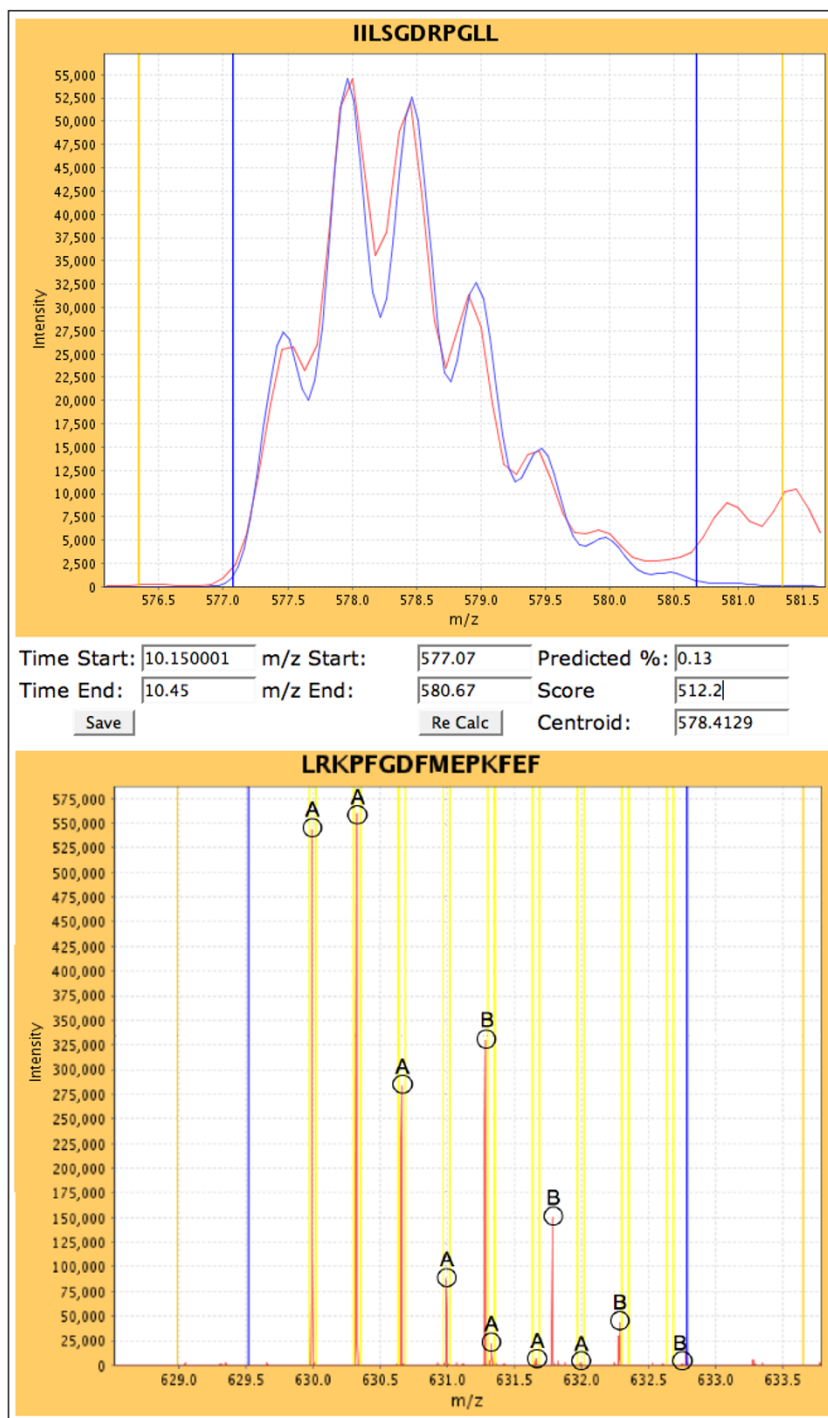


Figure 4
Main Spectral Viewer. Top. Co-added low-resolution spectral data within the selected retention time range is presented in the viewer. The theoretical isotopic envelope (blue) is overlaid on the observed data (red), showing the best-fit envelope used for calculation of deuterium content. **Bottom.** Co-added high-resolution spectral data within the selected retention time range is presented in the viewer. Correct peak data is extracted from calculated sub-ranges only, displayed as yellow bars. Correct peaks are extracted from the spectral data, enabling more accurate centroid measurements in areas with poor signal to noise ratios or overlapping peptides. 'A' Peaks belong to the peptide of interest and all reside within the sub ranges. 'B' peaks belong to another peptide (+I), which will be disregarded from centroid calculations.

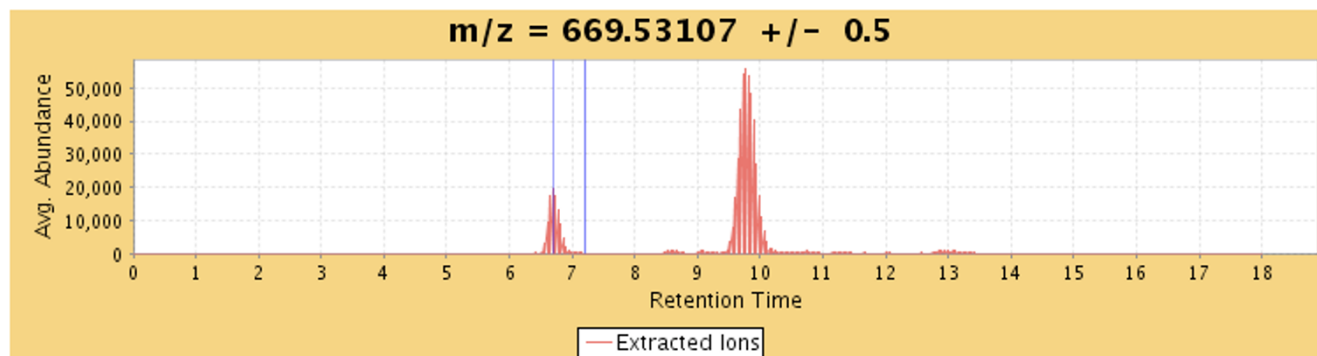


Figure 5
Extracted ion chromatogram. The averaged ion intensities for a centroid mass ± 0.5 (Da) are displayed across all scans allowing the user to graphically determine the retention time range that will produce the best data results. In the case above, the co-added data may yield an isotopic profile with a higher signal-to-noise ratio if the retention time start and end positions were reduced slightly.

resolution data. To assess the accuracy of the RT window we present an extracted ion chromatogram above the spectral viewer (Figure 5).

Deuterium exchange calculations

The calculation of the amount of backbone deuterium exchange can be accomplished using one of two methods, the "centroid" method or the "theoretical fit method".

For the theoretical fit method we followed the approach previously described [11], where the theoretical isotopic envelope is fit to the observed data at varying levels of deuterium incorporation. A number of approaches to calculate the natural isotopic envelope have been described [25-28]. Our method determines the natural isotopic envelope of the undeuterated peptide ion, by first determining its chemical formula, and submitting it to the available software package qmass [26]. The isotopic abundance calculations from this software are based on the natural abundance ratios of each element, which are stored in a file (ISOTOPE.DAT). To produce a distribution representing a fixed percentage of deuterium incorporation, an entry in the isotope distributions file is created to represent the amide hydrogens, which we label as 'D'. The ratios of 'D' are altered to reflect the desired percentage of deuterium incorporation. For example, to find the isotopic distribution for the peptide sequence "FRVSE" at 10% deuterium incorporation, one would first determine the chemical formula for the sequence. The resulting formula (C₂₈H₄₄N₈O₉) is then modified to account for the number of exchangeable hydrogens (D₄), to C₂₈H₄₀D₄N₈O₉. The ¹H to ²H ratio is then changed to 0.9/0.1, reflecting the 10% deuterium incorporation. The resulting qmass theoretical distribution is then scaled to the observed high peak intensity within the predicted theoretical m/z range, smoothed using the normal distribu-

tion formula, and aligned according to the resolution of the instrument. Least squares is then used to fit the envelope to the observed spectral data by measuring the difference between the Y coordinates (intensity) of the theoretical and observed plots, squaring them, and summing the result of all Y pairs. This process is repeated, iterating through increasing levels of deuterium incorporation (Figure 6), and the lowest score is assumed to be the best fit. The percent deuterium incorporation that produced the best fit is populated into the %D field within the data grid and can be used for downstream calculations. This approach has been incorporated for low-resolution data analysis and is currently being implemented for high-resolution data.

A different approach is taken for the "centroid" method. Variables m/z minimum and m/z maximum are initially assigned to $mz_0 - 1$ Da and $mz_{100} + 1$ Da, and are adjustable. XY data between m/z minimum and m/z maximum are used to calculate the intensity weighted mean m/z centroid value. The final centroid result is then used to calculate the percentage of deuterium incorporation with methods outlined previously [29].

For high resolution MS data, the isotopic peaks can be expected to reside within much narrower m/z ranges allowing for precise determination of the weighted average m/z value. Peptide ion isotopes are spaced at equal m/z steps according to the charge of the ion. For a doubly charged ion, its isotopes are spaced 0.5 m/z apart and for a triply charged ion, its isotopes are spaced 0.33 m/z apart. Combining this information with the known monoisotopic mass and charge state of the peptide ion allows the algorithm to predict the central m/z of the first isotope in the distribution, as well as the m/z spacing of the isotopic cluster. An approach using m/z ranges, which we will

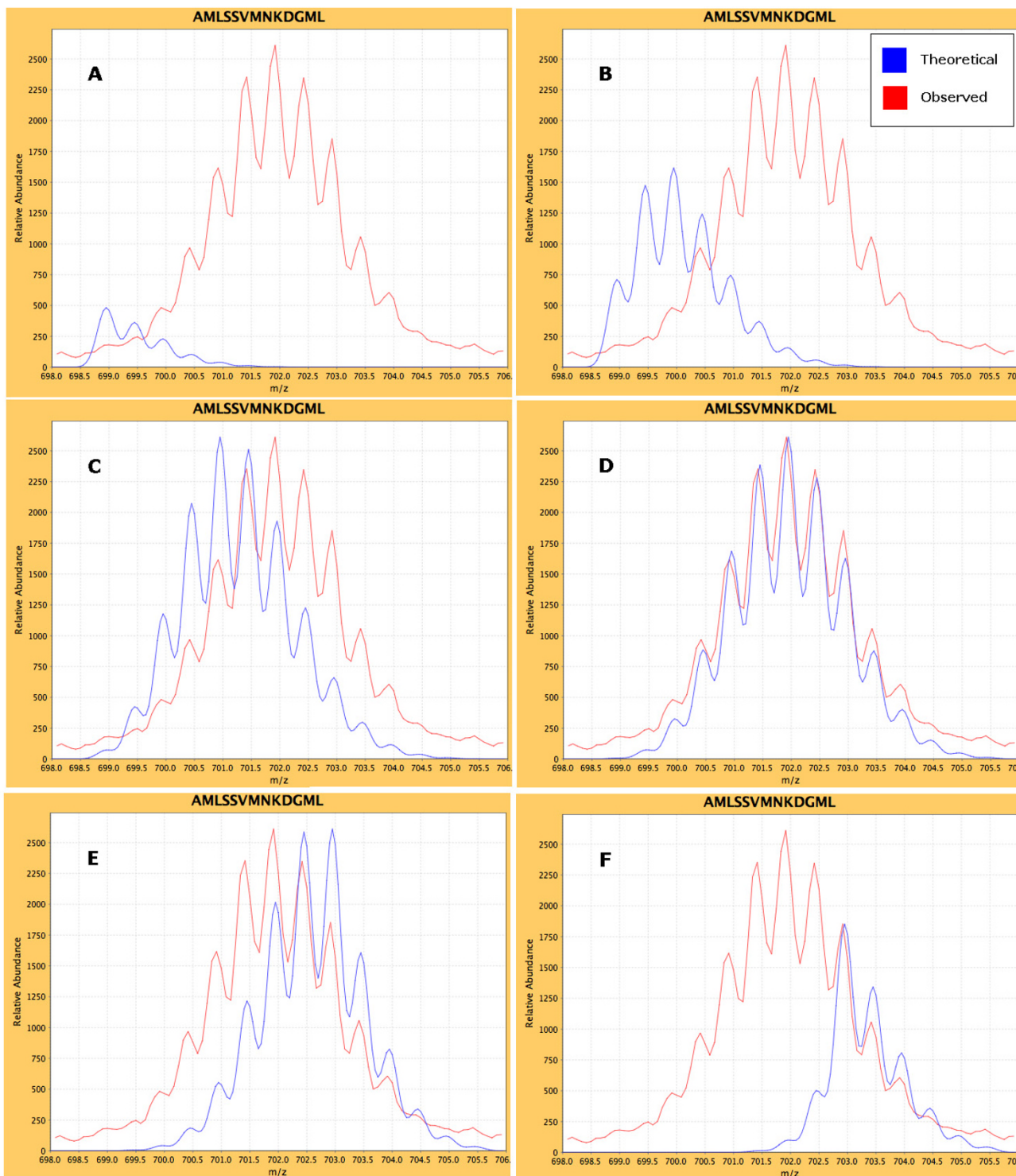


Figure 6
Iterative Theoretical Isotope Approach. 1. The natural isotopic envelope is calculated for the peptide at 0% deuterium content and scaled to the highest observed intensity within its m/z range. The chi-square fit score is noted. 2. The process is repeated in 1% increments to 100%. 3. The best (lowest) score is used in the results. The panels above show some of the iterations at increasing levels of deuterium incorporation. A = 0%, B = 10%, C = 25%, D = 43%, E = 65%, F = 90%. The best chi-square score was for panel D at 43% deuterium incorporation.

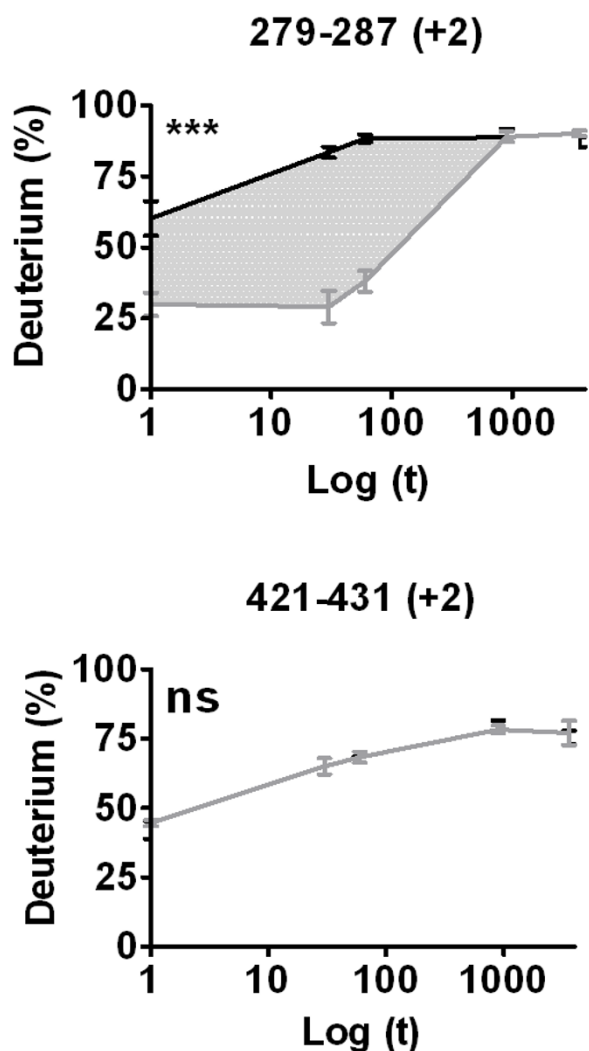


Figure 7
Deuterium vs. Log Time (s). Differential H/D exchange data can be exported and plotted as %D vs. Log time (s) plots. Here we show data from a differential H/D exchange experiment to investigate perturbation in H/D exchange rates upon binding of a ligand (rosiglitazone) to PPAR γ . Black line = PPAR γ , Grey line = PPAR γ + ligand. **Top.** The region between residues 279 and 287 exhibits a significant slow down in exchange rate following ligand binding. **Bottom.** Exchange rates for the region spanning residues 421–431 show no change. Statistical summary from a 2-way ANOVA between apo and ligand bound data; *** = $P < 0.001$, ns = not significant.

henceforth call " m/z sub-ranges", is presented here, where the algorithm only considers data residing within a predefined m/z distance from the expected peptide ion center. The sub-range centers are defined by the following equation:

$$f_i = f_0 + \frac{i}{z} \quad (2)$$

$i=0, \dots, i_{max}$

where $f_0 = mz_0$, z = charge state, i is an integer, and i_{max} is defined as:

$$i_{max} = z (\alpha - f_0) \quad (3)$$

where α = end m/z . The resulting windows are set to a width based on the desired distance from the peptide ion center to accommodate for the resolution of the instrument. A $0.025 m/z$ distance has been shown to be effective for a resolution of $\sim 60,000$, that generates a baseline peak width of $\sim 0.015 m/z$. The Deuterator uses this number as default. Ions with m/z values outside of these sub-ranges cannot arise from the peptide ion of interest and thus they are removed from further consideration. The weighted average of the data within these expected m/z sub-ranges is then calculated. The interface also allows the limits of the m/z sub-ranges to be adjusted, which enables the correct peaks to be extracted from the spectral data. This approach has been shown to correctly distinguish peptides in complex situations such as overlapping peptides or peptide ions with poor signal-to-noise (S/N) ratios. The limits of m/z sub-ranges are displayed in the user interface as yellow bars, allowing the visual confirmation of the approach (Figure 4 bottom).

Main analysis page

It is often beneficial to visually validate and manually curate the data. The main analysis page provides the necessary tools to facilitate the validation, curation and recalculation of the %D and Centroid values. This page has three sections: the data grid, the spectral viewer, and the extracted ion chromatogram.

When a peptide row is selected from the data grid, the co-added spectral data for the calculated m/z range of the peptide are presented in the main spectral viewer (Figure 4). The best-fit theoretical isotopic envelope is then overlaid and the sub-ranges are presented for high-resolution data. The indicator bars display the mz_0 , mz_{100} , m/z , and centroid range positions. Adjustment of the m/z range controls allow for zooming in and out of the spectral data. Navigation controls are available to further refine the sub-range widths (for high resolution data), retention time range, and centroid range of the selected peptide, with the %D and centroid values being automatically recalculated after each refinement. The extracted ion chromatogram (Figure 5) is also populated when a peptide is selected, and the averaged ion intensities for a given mass across all scans is presented. Blue bars display the retention time ranges used for the extracted ion plots, and can be adjusted for refinement of the scan range with the best

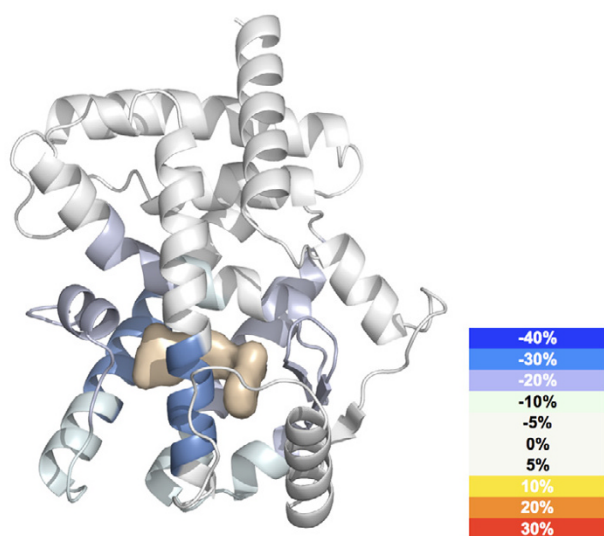


Figure 8
3D Overlay. H/D exchange data can be plotted over the three-dimensional protein structure. Here we show the % difference in H/D exchange rate for PPAR γ following binding of rosiglitazone plotted over 2PRG.pdb (ligand is represented as a surface rendering). The magnitude of any change in H/D exchange rate upon ligand binding is plotted according to the color key.

abundance for a given peptide's centroid mass. For example, when the retention time start and end values are equal, the view will represent spectral data from the nearest single scan. When the user has completed data curation, the results can be saved.

Synchronization function

Adjusting data in the main page can be a lengthy task, especially if there are a large number of curations needed for several experimental repetitions. Retention time and m/z ranges for like peptides across replicates are usually consistent. By synchronizing across the replicate data sets the number of manual curations needed is significantly reduced and automated reprocessing of centroid calculations is enabled. The synchronization page provides this functionality by allowing the selection of a "core" time point, a list of "sync" time points, and data column values. The process then adjusts the appropriate ranges and the centroids are recalculated for each peptide row. This has demonstrated to save substantial amounts of time by automating the curation process. A typical experiment produced approximately 100 peptides at eight timepoints. Each experiment was replicated four times for consistency. In this case 3200 peptides, each taking one minute to curate with the software, would cost the end user around 53 hours. In our tests, the synchronization function has

been shown to be effective on roughly 20% of peptides saving over 10 hours of manual curation time.

Output

Once the *Centroid* and %*D* values for all of the peptides in the peptide set have been determined and saved, the entire grid data can be exported to a common file format. From this point the deuterium content of each peptide is readily calculated using known algorithms [29]. Generation of deuterium buildup curves using tools such as Microsoft Excel (Figure 7) is facilitated. High-resolution 3D structural overlays (Figure 8) can be accomplished using tools such as PyMol [23].

The Deuterator in practice

The Deuterator has been used extensively on a number of projects over the last year, one of which has been published [30]. Another example of these projects was the measurement of perturbation of H/D exchange rates for the nuclear receptor PPAR γ upon binding of synthetic and endogenous ligands. The aim of this research was to establish the different mechanisms of activation of the receptor upon binding of full and partial agonists. The system has been tested with high and low-resolution raw mass spectral data from Thermo Electron LTQ Orbitrap (resolution $\sim 60,000$) as well as LTQ linear Ion-Trap (resolution $\sim 1,500$). The centroid calculation is a simple weighted average function, and so the validation of the centroid results has been straightforward. The software has been compared to existing tools such as Magtran [9] and the results have been shown to be identical. Native search engine data from the database search programs Mascot and Sequest, and X!Tandem have also been used with the system. The in-house installation is tightly integrated with commercial LIMS software (Proteus, Genologics, Canada), which captures the project files and streamlines automated processing.

Future work

Rendering components will be built that provide extended graphical capabilities, 2D and 3D color gradient maps, and data grids that convey changes in deuterium incorporation between different experiments. Sequence coverage can be presented in the user interface with heat maps, color-coding the rate of exchange. Improvements in the preprocessing phase will be implemented which will improve precision and reduce further the need for manual intervention. Ongoing tests to increase compatibility with other instrument data formats will be conducted. Any future developments to The Deuterator will be made accessible to the community.

Conclusion

The Deuterator is a web application presenting integrated and novel functionality in H/D exchange analysis and it is

compatible with multiple file formats. The interface presents the centroid results and associated data for each peptide and allows for "drilling-down" and visualization of the mass spectral data. Developed to aid scientists at The Scripps Research Institute in the analysis of H/D exchange data, it is freely available online and eventually will be offered as a standalone open source solution.

Availability and requirements

Project name: The Deuterator

Project home page: <http://deuterator.florida.scripps.edu>

Operating system: Platform Independent

Programming language: Java/JSP

Other requirements: N/A

License: available free of charge to academic, government and non-profit users for non-commercial use

Abbreviations

Da – Dalton

ESI – Electrospray ionization

H/D exchange – Hydrogen/deuterium exchange

JSP – Java Server Pages

LC – Liquid chromatography

MS – Mass spectrometry

m/z – Mass-to-charge ratio

NMR – Nuclear magnetic resonance

RT – Retention time

Competing interests

The author(s) declare that they have no competing interests.

Authors' contributions

BDP developed the software and wrote the manuscript. MJC provided core requirements and feedback. SAB tested the software and provided requirements. MJC and SAB acquired all the H/D exchange test data sets. CCM managed the overall software development process and design for the project, and assisted with development. MRS contributed development to the implementation of the external website as well as Excel files for rendering H/D exchange data. NFT provided essential direction and over-

saw all informatics aspects of the project. PRG conceived and initiated the H/D exchange program, and was the project manager driving the development of the application. All authors have read and approved of the manuscript.

Acknowledgements

We would like to thank Robert Driscoll and Mohammad Fallahi for assistance with formulas and overall review of the manuscript. The Scripps Research Institute, Scripps Florida is supported by the State of Florida.

References

- Engen JR, Smith DL: **Investigating protein structure and dynamics by hydrogen exchange MS.** *Analytical Chemistry* 2001, **73(9)**:256A-265A.
- Englander SW: **Hydrogen Exchange Mass Spectrometry: A Historical Perspective.** *Journal of the American Society for Mass Spectrometry* 2006, **17(11)**:1481-1489.
- Buijs J, Costa Vera C, Ayala E, Steensma E, Hakansson P, Oscarsson S: **Conformational stability of adsorbed insulin studied with mass spectrometry and hydrogen exchange.** *Analytical Chemistry* 1999, **71(15)**:3219-3225.
- Mandell JG, Falick AM, Komives EA: **Measurement of Amide Hydrogen Exchange by MALDI-TOF Mass Spectrometry.** *Analytical Chemistry* 1998, **70(19)**:3987-3995.
- Chalmers MJ, Busby SA, Pascal BD, He Y, Hendrickson CL, Marshall AG, Griffin PR: **Probing protein ligand interactions by automated hydrogen/deuterium exchange mass spectrometry.** *Analytical Chemistry* 2006, **78(4)**:1005-1014.
- Dai SYF M.C.: **Accuracy of SUPREX (Stability of Unpurified Proteins from Rates of H/D Exchange) and MALDI Mass Spectrometry-Derived Protein Unfolding Free Energies Determined Under Non-EX2 Exchange Conditions.** *J Am Soc Mass Spectrom* 2006, **17(11)**:1535-1542.
- Woods VLJ and Hamuro, Y.: **High resolution, high-throughput amide deuterium exchange-mass spectrometry (DXMS) determination of protein binding site structure and dynamics: utility in pharmaceutical design.** *Journal of Cellular Biochemistry* 2001, **Suppl 37**:89-98.
- Hamuro Y, Coales SJ, Southern MR, Nemeth-Cawley JF, Stranz DD, Griffin PR: **Rapid analysis of protein structure and dynamics by hydrogen/deuterium exchange mass spectrometry.** *Journal of Biomolecular Techniques: JBT* 2003, **14(3)**:171-182.
- Zhang ZM A: **A Universal Algorithm for Fast Automated Charge State Deconvolution of Electrospray Mass-to-Charge Ratio Spectra.** *Journal of the American Society for Mass Spectrometry* 1998, **9**:225-233.
- Hotchko M, Anand GS, Komives EA, Ten Eyck LF: **Automated extraction of backbone deuteration levels from amide H/2H mass spectrometry experiments.** *Protein Science* 2006, **15(3)**:583-601.
- Palmblad M Buijs, J, and Hakansson, P.: **Automatic Analysis of Hydrogen/Deuterium Exchange Mass Spectra of Peptides and Proteins Using Calculations of Isotopic Distributions.** *Journal of the American Society for Mass Spectrometry* 2001, **12**:1153-1162.
- Weis DD, Hotchko M, Wales TE, Ten Eyck LF, Engen JR: **Identification and characterization of EX1 kinetics in H/D exchange mass spectrometry by peak width analysis.** *Journal of the American Society for Mass Spectrometry* 2006, **17(11)**:1498-1509.
- Weis DD, Engen JR, Kass IJ: **Semi-Automated Data Processing of Hydrogen Exchange Mass Spectra Using HX-Express.** *Journal of the American Society for Mass Spectrometry* 2006, **17(12)**:1700-1703.
- Abzalimov RR, Kaltashov IA: **Extraction of Local Hydrogen Exchange Data from HDX CAD MS Measurements by Deconvolution of Isotopic Distributions of Fragment Ions.** *Journal of the American Society for Mass Spectrometry* 2006, **17(11)**:1543-1551.
- Chik JK, Vande Graaf JL, Schriemer DC: **Quantitating the statistical distribution of deuterium incorporation to extend the**

- utility of H/D exchange MS data. *Analytical Chemistry* 2006, **78(1)**:207-214.
16. Zhang Z, Li W, Logan M, Li M, Marshall AG: **Human recombinant [C22A] FK506-binding protein amide hydrogen exchange rates from mass spectrometry match and extend those from NMR.** *Protein Science* 1997, **6(10)**:2203-2217.
 17. **JFreeChart.** :[http://www.jfree.org/jfreechart].
 18. **CEWolf.** :[http://cewolf.sourceforge.net].
 19. Pedrioli PG Eng J.K., Hubley R, Vogelzang M, Deutch E.W., Raught B., Pratt B., Nilsson, E., Angeletti R.H., Apweiler R., Cheung K, Costello CE, Hermjakob H, Huang S, Julian RK, Kapp E, McComb ME, Oliver SG, Omenn G, Paton NW, Simpson R, Smith R, Taylor CF, Zhu W, Aebersold R: **A common open representation of mass spectrometry data and its application to proteomics research.** *Nature Biotechnology* 2004, **22**:1459-1466.
 20. Orchard S, Hermjakob H, Binz PA, Hoogland C, Taylor CF, Zhu W, Julian RK Jr., Apweiler R: **Further steps towards data standardisation: the Proteomic Standards Initiative HUPO 3(rd) annual congress, Beijing 25-27(th) October, 2004.** *Proteomics* 2005, **5(2)**:337-339.
 21. Orchard S, Hermjakob H, Julian RK Jr., Runte K, Sherman D, Wojcik J, Zhu W, Apweiler R: **Common interchange standards for proteomics data: Public availability of tools and schema.** *Proteomics* 2004, **4(2)**:490-491.
 22. **The Sashimi Project.** :[http://sashimi.sourceforge.net].
 23. **Pymol.** :[http://pymol.sourceforge.net].
 24. Bai Y, Milne JS, Englander SW: **Primary Structure Effects on Peptide Group Hydrogen Exchange.** *Proteins: Struct Funct, Genet* 1993, **17(1)**:75-86.
 25. Clauser KR, Baker P, Burlingame AL: **Role of accurate mass measurement (+/- 10 ppm) in protein identification strategies employing MS or MS/MS and database searching.** *Analytical Chemistry* 1999, **71(14)**:2871-2882.
 26. Rockwood AL, Haimi P: **Efficient calculation of accurate masses of isotopic peaks.** *Journal of the American Society for Mass Spectrometry* 2006, **17(3)**:415-419.
 27. Rockwood AL, Van Orman JR, Dearden DV: **Isotopic compositions and accurate masses of single isotopic peaks.** *Journal of the American Society for Mass Spectrometry* 2004, **15(1)**:12-21.
 28. Yergey JA: **A general-approach to calculating isotopic distributions for mass spectrometry.** *Int J Mass Spectrom Ion Proc* 1983, **52**:337-349.
 29. Zhang ZS D. L.: **Determination of amide hydrogen exchange by mass spectrometry: a new tool for protein structure elucidation.** *Protein Science* 1993, **2(4)**:522-531.
 30. Quint P, Ayala I, Busby SA, Chalmers MJ, Griffin PR, Rocca J, Nick HS, Silverman DN: **Structural mobility in Human manganese superoxide dismutase.** *Biochemistry* 2006, **45**:8209-8215.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

