

HDfLeX: Software for Flexible High Structural Resolution of Hydrogen/Deuterium-Exchange Mass Spectrometry Data

Neeleema Seetaloo, Monika Kish, and Jonathan J. Phillips*

Cite This: *Anal. Chem.* 2022, 94, 4557–4564

Read Online

ACCESS |



Metrics & More

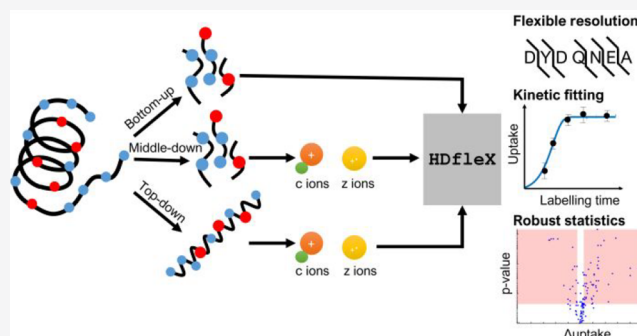


Article Recommendations



Supporting Information

ABSTRACT: Hydrogen/deuterium-exchange mass spectrometry (HDX-MS) experiments on protein structures can be performed at three levels: (1) by enzymatically digesting labeled proteins and analyzing the peptides (bottom-up), (2) by further fragmenting peptides following digestion (middle-down), and (3) by fragmenting the intact labeled protein (top-down) using soft gas-phase fragmentation methods, such as electron transfer dissociation (ETD). However, to the best of our knowledge, the software packages currently available for the analysis of HDX-MS data do not enable the peptide- and ETD-levels to be combined; they can only be analyzed separately. Thus, we developed HDfLeX, a standalone application for the analysis of flexible high structural resolution of HDX-MS data, which allows data at any level of structural resolution (intact protein, peptide, fragment) to be merged. HDfLeX features rapid experimental data fitting, robust statistical significance analyses, and optional methods for theoretical intrinsic calculations and a novel empirical correction for comparison between solution conditions.



Hydrogen/deuterium-exchange mass spectrometry (HDX-MS) is a widely used analytical technique that enables the probing of protein conformational dynamics and interactions on a wide timescale, from milliseconds to hours, by monitoring the isotopic exchange at the backbone amide hydrogens with the surrounding solvent.^{1,2} HDX-MS is growing in popularity as an orthogonal structural biology tool since it can analyze proteins that are not amenable to other techniques, such as antibodies^{3,4} and intrinsically disordered proteins.^{5,6} In order to localize the exchange kinetics to specific regions of a protein, it is crucial to obtain submolecular hydrogen/deuterium-exchange data. This is currently a major limitation in the mass spectrometry detection of hydrogen/deuterium-exchange and, until recently, the structural resolution of any bottom-up HDX-MS experiment relied on the number and position of proteolytic peptides identified.^{5,7} By spectral assignment of the digested peptides onto the protein sequence, the changes in deuterium uptake of each peptide can be mapped back to the primary sequence, thus giving insights into the behavior of localized regions of the protein.

Much recent interest has been directed to achieve higher structural resolution by fragmenting the digested peptides further using gas-phase soft fragmentation techniques, termed middle-down HDX-MS.^{8,9} However, it is important that the fragmentation does not induce any hydrogen/deuterium scrambling (H/D scrambling).^{10,11} Methods shown to be amenable to this include electron-based dissociation methods (ExD) with HDX-MS, such as electron-transfer dissociation

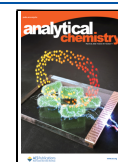
(ETD),^{3,12,13} electron-capture dissociation (ECD),^{11,14} and ultraviolet photodissociation (UVPD).^{15,16} These forms of fragmentation can be tuned to have analyte energy profiles that do not promote proton mobility and so result in low H/D scrambling under the right conditions.¹⁷ This has made HDX-MS possible in top-down (fragmentation of entire proteins into smaller fragments)¹³ and middle-down (fragmentation of digested peptides into even smaller fragments)⁸ fashions, which each provide varying levels of structural resolution.

Nevertheless, presently, each of these experiments and the resulting global and local HDX-MS data can only be analyzed and interpreted separately since there is currently no widely available process that allows the combination of data at all levels.^{18–21} This is especially confounding, as soft-fragmentation reactions are typically inefficient and yield sparse data which might usefully be complementary to conventional bottom-up data. In an effort to remedy this, we hereby present HDfLeX, software for flexible structural resolution of HDX-MS data, which allows combination of data from bottom-up, middle-down, and top-down experiments. HDfLeX can (1) perform

Received: December 9, 2021

Accepted: March 8, 2022

Published: March 9, 2022



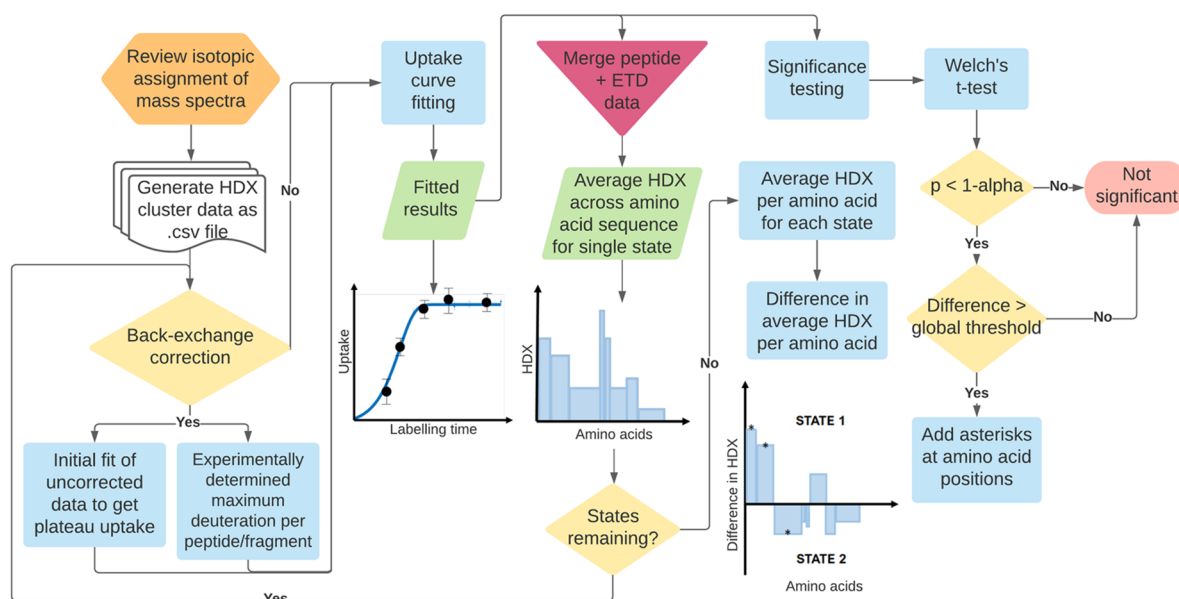


Figure 1. Flowchart summarizing the processes performed by HDfleX.

nonlinear regression fitting of deuterium uptake data against labeling time for each peptide or fragment, (2) correct for back-exchange using internal or external measures of maximal deuteration, (3) normalize between different pH and salt conditions, (4) calculate intrinsic rate and protection factor (P_f), (5) combine peptide and ExD fragments and visualize the absolute deuterium uptake as histograms for individual protein states, (6) perform hybrid significance testing using various methods of global significance testing and doing a Welch's t test, and (7) visualize significant differences on difference and volcano plots.

METHODS

HDfleX was written in MATLAB release 2021b (MathWorks, USA) and made into a graphical user interface with MATLAB App Designer. The software performs a series of steps summarized in Figure 1 and detailed below. An in-depth flowchart of the process can also be found in Figures S1–S3.

Data Import and Preparation. Prior to HDfleX analysis, an input file containing all the information about the HDX-MS experiment needs to be created. Following the review of isotopic assignments, a comma-separated variable (.csv) file is generated, which is the first user-input to HDfleX. The file must be comma-separated and have the nine columns in the order: protein name, sequence start number, sequence end number, sequence, modification, fragment, maximum possible uptake, mass of monoisotopic species, state name, exposure time, file name, charge, retention time, intensity, centroid. After HDfleX parses the input file, the user selects the protein and states to plot, type of back-exchange correction to perform (if at all), number of replicates collected, number of phases into which to fit the uptake curves, pH of reference state, pH of each experimental state, time format between milliseconds (ms), seconds (s), and minutes (min), and the time window over which the fitted uptake curve will be evaluated. In cases where the replicates are not named trivially (e.g., file_01, file_02, etc.), the user can manually assign files to the appropriate states and replicates. Charge deconvolution and deuterium uptake per peptide and ETD fragment are determined as described in Wales et al.²²

Uptake Curve Fitting and Area under the Uptake Curve.

Fitting the HDX-MS data from equilibrium experiments can provide crucial information about the exchange kinetics. We fit the data points for each peptide and ETD fragment either by interpolation or fitting to a stretched exponential. The deuterium uptake at labeling time t , is fitted to a stretched exponential given by eq 1,²³ where n_{exp} is a user-defined number of exponential phases, N is the maximum number of labile hydrogens, k_{obs} is the observed exchange rate constant, and β is a stretching factor.

$$D_t = N \sum_{i=1}^{n_{\text{exp}}} 1 - e^{-(k_{\text{obs},i}t)^{\beta_i}} \quad (1)$$

The nonlinear fitting is performed using the fit function from the Curve Fitting Toolbox in MATLAB. HDfleX fits all measurable n_{exp} and determines the best n_{exp} , using the adjusted R^2 value, which is used for the final fit. If $n_{\text{exp}} = 1$, N is fixed at the maximum number of labile hydrogens for that particular peptide/fragment when back-exchange correction has been applied. Otherwise, only an upper bound of N is applied. k_{obs} is constrained at >0 and β is limited between 0 and 1. β accounts for the differences in intrinsic exchange rates of the individual amino acids. The resulting fitted curve equation is subsequently evaluated and integrated over the user-defined time window to give the area under the uptake curve, HDX_{Area} . In the case of data that reach a plateau (i.e., maximal deuteration achieved at later time points), a further variable called X_{plat} (Figure S4) is calculated, which is used as the upper limit of the time window of integration. This prevents significant differences from being minimized when overly wide time windows have been used for any individual peptide.

There is also an option to fit the data points using an interpolation method. This method may be used for cases in which not enough time points are available to generate a robust fit, or where the exchange reaction is occurring at non-equilibrium conditions. Here, we used the Piecewise cubic Hermite interpolation (PCHIP from MATLAB's Curve Fitting Toolbox). However, this method does not provide a k_{obs} and β and thus cannot be used for empirical

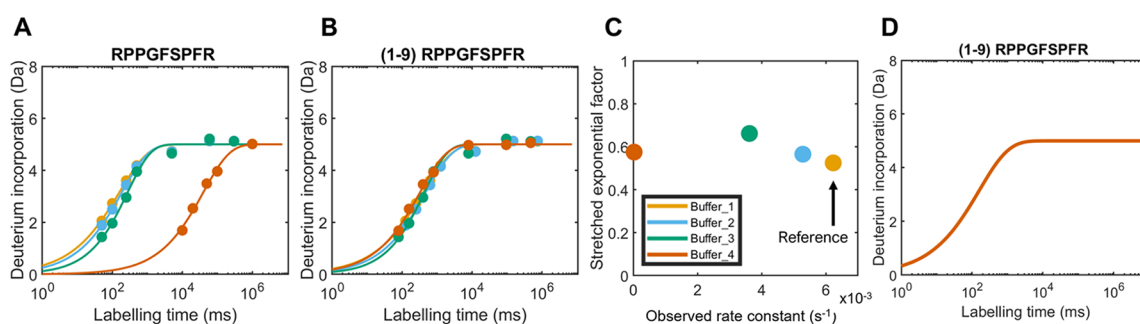


Figure 2. Calibrating the chemical exchange rate using bradykinin. (A) Bradykinin uptake curves in four states, back-exchange corrected and fitted in one exponential phase according to eq 1; (B) pH adjustment applied to the time points of bradykinin in each state prior to fitting; (C) stretching factor β against the observed rate constant k_{obs} for each buffer condition. (D) Empirical adjustment factor applied to bradykinin conditions, showing all four curves now perfectly overlaid ($\Delta\text{uptake} = 0$ Da).

adjustments, observed exchange rate constant, and protection factor analyses. The entire HDfleX workflow following uptake curve fitting, such as data flattening and statistical significance analysis, can be performed with uptake at a time t , uptake area up to a time t , the observed exchange rate constant or protection factors.

Back-Exchange Correction. Despite efforts to minimize back-exchange in HDX-MS experiments,^{24–26} there is always some percentage of the on-exchanged deuterium that exchanges back to hydrogen. For some types of analysis, such as the calculation of protection factors,²⁷ this needs to be corrected for. Prior to curve fitting, the user can choose whether to perform back-exchange correction²⁸ or not. HDfleX allows for two methods of back-exchange correction from (1) an experimentally determined maximum deuteration (maxD) value, where the experimental maxD is included in the cluster data input file, and (2) a maximum deuteration value derived from the plateau value of the fitted uptake curve. The plateau maxD option can only be used on proteins and peptides that can undergo full exchange within the experimental time window.

pH Adjustment Factor. The rate of HDX is affected by several factors, including pH and ionic strength.^{27,29,30} When normalizing between different pH, HDfleX alters the experimental labeling time points according a pH adjustment factor, $\text{pH}_{\text{adjust_factor}}$, according to eq 2,³¹ where pH_{ref} is the reference state pH and pH_{exp} is the experimental pH of the state being compared. $\text{pH}_{\text{adjust_factor}}$ is subsequently multiplied to all the experimental labeling time points for that particular state. As this correction factor was derived from the base-catalyzed amide hydrogen exchange reaction, the pH range for which this pH adjustment factor is applicable is 5–10.

$$\text{pH}_{\text{adjust_factor}} = 10^{\text{pH}_{\text{exp}} - \text{pH}_{\text{ref}}} \quad (2)$$

To test the pH adjustment, we used bradykinin in four different buffers (Table S1, Buffers 1–4). We chose bradykinin peptide (Arg1-Pro2-Pro3-Gly4-Phe5-Ser6-Pro7-Phe8-Arg9) as it has been found to adopt random-coil conformational states in aqueous media.^{32–34} Due to its unstructured nature in aqueous medium, any changes in HDX in bradykinin between the conditions being compared can be attributed to the solution effects. HDX time courses on bradykinin with three replicates and seven time points were collected for each buffer condition being compared (see SI Methods in the Supporting Information for experimental details). The data was fitted and corrected for back-exchange as discussed above (Figure 2A). We first applied the pH adjustment to the four different conditions to normalize

the solution effects among them (Figure 2B). If the solution effects differed only by pH, we would expect to see the uptake curves to perfectly overlay each other. However, our data showed that despite applying the pH correction, there were still differences up to 0.43 Da between the uptake curves, which is above the calculated global threshold of 0.09 Da for this peptide. This confirms that the salts used also influenced the intrinsic exchange kinetics.³⁵

Empirical Adjustment Factor. As we saw in bradykinin, both pH and ionic composition can affect the intrinsic exchange rate.³⁶ In this case, it becomes difficult to discern the structural changes from the intrinsic exchange rate changes. To allow accurate comparison between conditions of different pH and ionic composition, we created a new empirical method which uses the unstructured peptide bradykinin to calibrate the intrinsic exchange rate effect.

Following back-exchange correction and curve-fitting, each state has two fitted parameters, k_{obs} and β , associated with it. In Figure 2C, a plot of the stretching factor β against the observed rate constant k_{obs} shows the k_{obs} and β for each buffer condition. The first step is to choose a reference state (Buffer 1). A 2D scaling factor is determined for each coordinate in Figure 2C to transform to the reference coordinate. This scaling factor, consisting of a value for k_{obs} and one for β is multiplied to the fitted parameters for all the buffer conditions. For bradykinin, this results in the uptake curves to perfectly overlay each other (Figure 2D), with differences of 0 Da between the curves, i.e., $\Delta\text{uptake} = 0$ Da. The empirical factor correction for each condition can then be applied to the uptake curves of the protein of interest, thereby normalizing the intrinsic exchange rate effects, allowing the structural effects to be clearly distinguished.

Statistical Significance Testing. It is important to carry out statistical significance analyses in differential HDX experiments to determine significant differences across the protein between different experimental conditions. HDfleX takes the hybrid significance testing approach brought forward by Hageman and Weis for statistical significance testing.³⁷

Obtaining Data Distribution by Separating the Replicates. In the first stages of HDfleX, all the replicates were pooled together to perform the calculations. To calculate the global significance threshold for the HDX parameters (uptake area, observed rate constant, and protection factors) except for the uptake, we need to know their distribution. Thus, it is necessary to perform curve fitting of different combinations of replicates. Here, we describe three methods by which HDfleX creates the distributions from the raw data for each peptide/

fragment. Figure S5 shows a graphical explanation of each method. (1) Simple per replicate number combinations: At each time point, one replicate is chosen in numerical order until all the replicates have been processed. In cases of missing replicates at any particular time points, the latter will not include that replicate number in the curve fitting. The mean and SD of each of these distributions is calculated. (2) Random replicate combinations: At each time point, a random replicate is chosen from the set of replicates available. This is repeated n times, where n is the original number of replicates in the experiment. The mean and SD of each of these distributions is calculated. (3) Bootstrapping method of combinations: All the possible combinations of time points and replicates are initially calculated. A number s of random combinations is sampled uniformly from all the possible combinations, with replacement, using the MATLAB function `datasample`. Thereafter, 10 000 bootstrap data samples are drawn from s using the `bootstrap` function, and the mean and SD of each of these distributions is calculated.

Global Significance Threshold. A global threshold is necessary to eliminate negligible differences that appear significant from the Welch's t test described in the next section, which would otherwise lead to false positives. Three methods of calculating the global significance threshold are available in HDflex. The first one is based on the calculation of the confidence interval from experimental standard deviations (SD) as described previously by Hageman and Weis.³⁷ We have extended their method to include ETD fragments, to remove outliers and to be also used on the area under the curve, in addition to the uptake. The second method of calculating the global threshold, applying only to (1) and (2) data distribution methods above, initially involves calculating the pairwise differences between replicates within each state for the experimental data (Figure S6A). The pairwise difference is calculated rather than the difference between each replicate and the mean because we did not want any outliers in the data set to skew the mean, which should be zero as the differences being calculated are within states. The outliers in the pairwise differences are optionally removed using the MATLAB function `isoutlier`,³⁸ with a method chosen by the user (mean, median, quartiles, grubbs, gesd), prior to fitting the pairwise replicate differences to a normal distribution. The upper confidence interval limit of this normal distribution is the global significance threshold. For the uptake area, the replicate differences and standard deviations are calculated from the integral of the fitted curves, whereas for the uptake, they are calculated from the experimental data points. When the bootstrapping method of combinations is selected (Figure S6B), the global threshold is calculated as simply being the upper limit of confidence interval of the all the means across the peptides and ETD fragments.

Welch's t Test. A Welch's t test is performed between the peptides and fragments of the two states being compared to give a p -value. The MATLAB function `ttest2` is used with the Satterthwaite's approximation for the effective degrees of freedom.³⁹ To be significant, any differences between the peptides and fragments of the two states being compared must be above the global significance threshold and have a p -value lower than the alpha confidence level (e.g., $p < 0.05$).³⁷

Calculation of Intrinsic Rates and Protection Factors. The intrinsic rates and protection factors were fitted and calculated as described previously.⁴⁰ Briefly, intrinsic exchange constants are calculated and generated for each amide in a

peptide as described by Bai et al. and adapted from the Excel sheet provided by Englander lab (available online here <http://hx2.med.upenn.edu/>). This allows simulation of the degree of deuterium incorporation as a function of labeling time at the reference pH and temperature. An intrinsic rate constant, k_{int} is then obtained by fitting the simulated deuterium uptake according to eq 1. The protection factor is obtained using eq 3 where P_f is the protection factor, k_{int} is the intrinsic exchange rate derived from the intrinsic fitting, and k_{obs} is the experimental observed rate constant derived from the experimental fitting. The P_f can be visualized per protein state, per peptide or flattened per amino acid as explained in the next section.

$$P_f = \frac{k_{\text{int}}}{k_{\text{obs}}} \quad (3)$$

Data Flattening over Amino Acid Sequence. Instead of analyzing the HDX-MS on a per peptide/fragment basis, HDflex flattens the data across all the available peptides or fragments to give high structural resolution HDX-MS data. The uptake area or uptake at time t at any peptide/fragment is divided by the number of exchangeable hydrogens (q). This number is subsequently applied across the amino acid residues covering that peptide as shown in Figure S7, except for the N-terminal residue and any prolines, which are fixed at 0. Note, for the ETD fragments, we attempted an intuitive subtraction of the consecutive fragments in the c/z series to obtain the deuterium uptake/area of highly resolved overlapping fragments; however, this introduced large errors and was ultimately rejected (Figure S8).

Visualizations of Differences Generated by HDflex. HDflex provides several ways to visualize the difference between states: difference and volcano plots at peptide resolution (unflattened data) and amino acid residue resolution (flattened data). The differences between the means of the uptake area or uptake at any time t are calculated by subtraction of one state from the other at each amino acid/peptide. Those differences are then plotted against the peptide/amino acid number to give a difference plot. Volcano plots for both flattened and unflattened data are also generated for the p -value vs difference in uptake/uptake area.³⁷

3D Structural Visualization of HDX Differences. The differences calculated by HDflex for uptake, uptake area, protection factor, and observed rate constant are commonly visualized on the protein structure. HDflex generates .pml script files at user-selected time points that can be run in PYMOL.⁴¹

RESULTS

To test HDflex, we used data sets containing HDX-MS peptide and ETD information on the purified wild-type intrinsically disordered protein alpha-synuclein (aSyn) in various solution conditions (Table S1, Buffers 1–4). Briefly, the millisecond time scale HDX was performed using a fully automated, millisecond HDX labeling and online quench-flow instrument, ms2 min (Applied Photophysics, U.K.), connected to an HDX manager (Waters, USA). Mass spectrometry (MS) data were acquired on a Waters Synapt G2-Si Q-IM-TOF instrument with ETD capabilities. Both bottom-up and middle-down data were collected.

Improved Structural Resolution when ETD Fragments Included with Peptides. The results of HDflex processing of HDX-MS data for aSyn shows an increase in structural resolution of MS data when ETD data are included with peptide

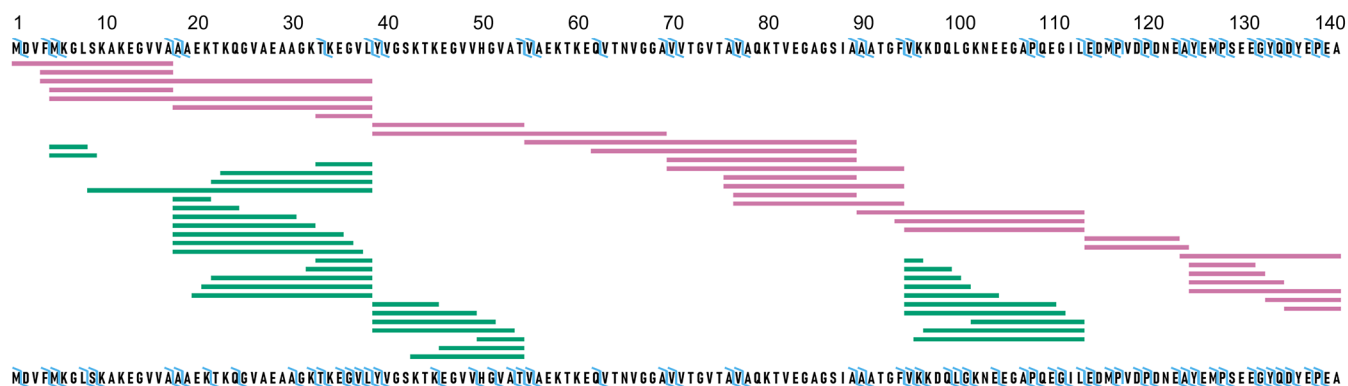


Figure 3. Inclusion of ETD fragments to peptide data improves structural resolution for sample data of aSyn HDX. Top sequence corresponds to peptide-only data (pink) and bottom sequence corresponds to peptide + ETD fragments data (green). The blue ticks represent the merged resolution obtained by combining peptides and *c/z* fragments.

data (Figure 3). Bottom-up data were combined with middle-down ETD data, resulting in a 50% increase (from 20 to 30) of single amino acids with fully resolved HDX data and a 150% and 230% increase in HDX data resolved at the dipeptides and tripeptides level, respectively (Figure 4). Overall, the merging of

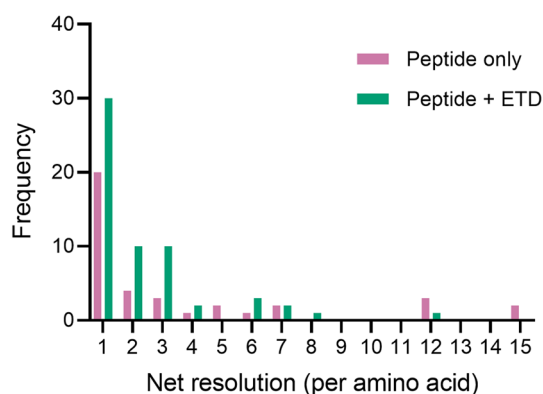


Figure 4. Comparison of the merged resolution between peptide only (pink) and peptide + ETD (green) data.

the ETD fragments and peptides have greatly improved the structural resolution. It was not possible to test any top-down or ECD data for this study, but these data should in principle be compatible with the HDfLeX analysis to provide even higher structural resolution.

Validation of Global Significance Threshold Methods.

The importance of a global threshold for significance has recently been stressed by various groups for the correct statistical analysis of HDX-MS data.^{30,37,42} HDfLeX calculates the global threshold for the uptake (in Da) in the same way devised by Hageman and Weis.³⁷ For the uptake area global threshold, three methods are available: calculation of a confidence interval from (1) the standard deviations,^{37,43} (2) the pairwise differences between replicates within states, and (3) bootstrapped replicates.

To validate the three methods, we carried out a null experiment on the aSyn protein in Buffer 2 for four replicates and created six groups of two replicates from the same data set (Figure S9). The difference in uptake area between the groups were calculated for all peptides, shown as scatter density plots in Figure 5. By definition, for a null experiment, the mean of the differences should be zero and this is what we observe. The

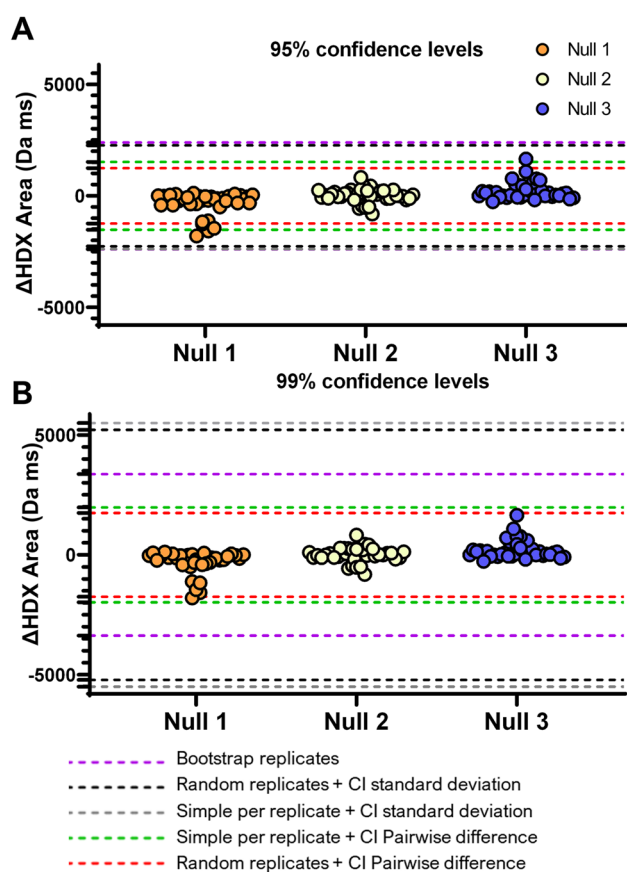


Figure 5. Comparison of the three methods of global threshold calculation on three null data sets derived from the protein aSyn in one condition only. The circles represent the difference in uptake between the two groups in the null data set. Each null comparison was unique. The dashed lines represent the global thresholds calculated from each of the methods available in HDfLeX. CI: confidence interval.

spread of the differences is affected by outliers, as can be seen from the Null 3 data set, for example. The presence of outliers alters the calculated global threshold for significant differences. HDfLeX allows for the optional removal of outliers from the calculation of the global threshold using the MATLAB function `isoutlier`.³⁸ Without removing any outliers, we can see that for the 95% confidence intervals (Figure 5A), the bootstrap replicates and standard deviation-based confidence interval

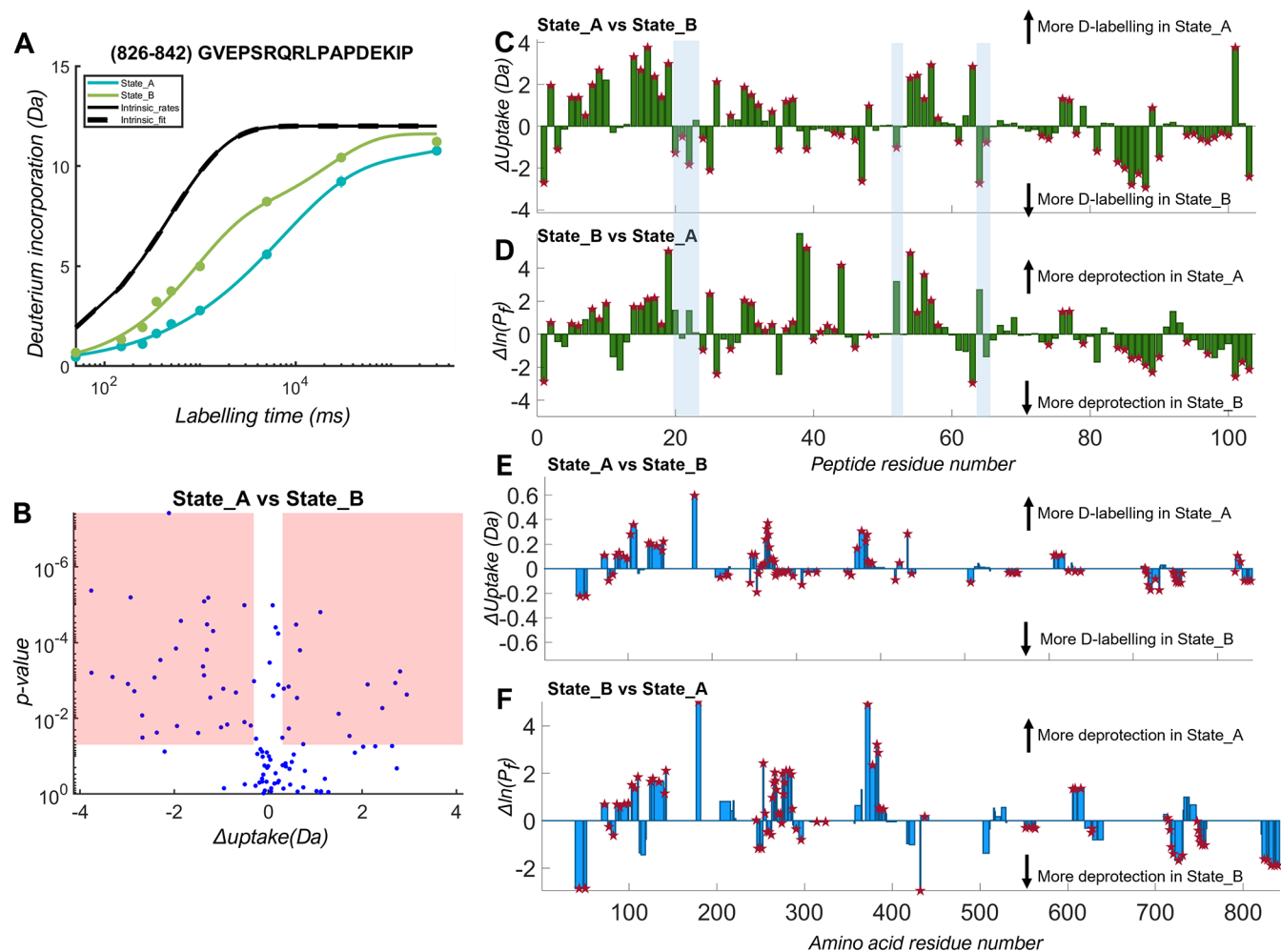


Figure 6. Glycogen phosphorylase HDX-MS data analyzed with HDfLeX. (A) Example deuterium uptake plot showing the experimental and theoretical fitting of the glycogen phosphorylase State_A and State_B; (B) volcano plot representation of the hybrid significance testing comparison of State_A vs State_B; (C) uptake difference plot of State_A vs State_B at peptide resolution; (D) $\ln(P_i)$ difference plot of State_A vs State_B at peptide resolution; (E) flattened uptake difference plot of State_A vs State_B at amino acid residue resolution; (F) flattened $\ln(P_i)$ difference plot of State_A vs State_B at amino acid residue resolution. Significant differences are denoted by a red asterisk.

calculations result in no false positives. The other methods show significant differences due to false positives. Increasing the confidence levels to 99% (Figure 5B) removes false positives from the pairwise difference-based calculation using the simple per replicate data distribution method. However, the increase in confidence interval disproportionately alters the standard deviation and bootstrapping-based calculations and could result in false negatives, causing significant differences to be missed. Therefore, we consider that the bootstrap replicates and standard deviation-based calculations are highly robust at removing false positives while minimizing the risk of false negatives.

Empirical Adjustment as a Novel Normalization Method for HDX-MS Experiments. In order to normalize data between different buffer conditions (i.e., salts) that would alter intrinsic rates in a manner with no established solution, we used the unstructured peptide bradykinin^{32,33} as an internal standard. In this way, we aim to remove significant differences that are a result of buffer-derived changes to intrinsic rates and better differentiate those derived from changes to protein structural dynamics. The protein aSyn was on-labeled in the same buffers as bradykinin in separate experiments and the data

kinetically fitted in HDfLeX to obtain a rate constant (k) and stretch term (β), shown in Table S2. The empirical adjustment factors for k and β to convert to a reference state (here Buffer 1) are globally set values derived from bradykinin that are multiplied to the fitted parameters of each peptide and fragment of aSyn (Figure S10). It is important to note that this method is entirely empirical and has only been tested on the unstructured peptide bradykinin and intrinsically disordered protein aSyn in single-exponential phases. Unlike structured proteins, intrinsically disordered proteins experience very little protection from changes in local structure (i.e., solvent accessible surface area and hydrogen-bonding). Therefore, we do not expect this to be a large effect on highly structured regions of folded proteins.

Intrinsic Rate Calculation and Protection Factors. HDfLeX also calculates the theoretical intrinsic rates for peptides⁴⁰ at the experimental pH and temperature. aSyn is incompatible with the Bai and Englander model for the calculation of intrinsic rates^{27,44} as all the peptides exchanged faster than the predicted rates. To demonstrate the ability of HDfLeX to calculate intrinsic rates and protection factors, we analyzed previously described data sets on glycogen phosphorylase a and b (State_A and State_B, respectively).⁴⁰ The uptake

plot in Figure 6A shows the expected theoretical (black) and experimentally measured (blue and green) deuterium uptake for a selected peptide for both protein states, from which the P_f is calculated. There is an upper limit of detection for P_f as determined by the measured labeling times (for this data set $\ln(P_f) < 10$), which we have described elsewhere.⁴⁰ Therefore, a P_f threshold can be set to discard unquantifiable, highly protected, slow-exchanging peptides. Following this filtering, 103 peptides were selected for the differential HDX analysis. To determine the significant differences between the two states, we performed hybrid significance testing as described earlier, the results of which are represented as a volcano plot (Figure 6B). Significant peptide uptake differences were found in 53% of the peptides and fall in the red shaded region. Furthermore, differences between State_A and State_B can be visualized as difference plots per peptide (Figure 6C,D) and per amino acid residue (Figure 6E,F), with significant differences denoted with a red asterisk. After flattening, it is important to note that the values derived from the stretched exponential fitting are averaged over the amino acids on the peptide. We compared the uptake area difference plot per peptide generated by HDfLeX to the butterfly plot generated by DynamX (Figure S11). Multiple features were found to be (in)significant by HDfLeX, which would have the opposite evaluation without fitting and statistical analysis. Here, we show an important distinction when using protection factors is that “hidden” features with low uptake but with high protection can be detected (shown as blue shaded regions in Figure 6C,D).

CONCLUSION

In this paper, we presented HDfLeX, a new standalone application for the analysis of HDX-MS data postprocessing, allowing the combination of data at different levels: peptides and ETD fragments. Other currently available programs do not allow the combination of data at different fragmentation levels, limiting the analysis of HDX-MS data and the resulting structural resolution. With HDfLeX, we saw an improvement in structural resolution of the HDX-MS data for areas where ETD data was collected, up to a 50% increase in single residues.

We have implemented a robust significance testing method, adapting the hybrid significance testing approach by Hageman and Weis.³⁷ The adjustments we included, with respect to outlier removal make the calculation of the global significance testing more robust to the presence of outliers. Throughout, we use the “uptake area”, as opposed to the more commonly used “uptake”, as our main parameter. Other groups have proposed the use of uptake area in differential HDX studies^{37,45} but with no global threshold calculation method available due to the added complexity of the different combinations of replicate data points; a problem now solved by HDfLeX. As such, not only did we expand the hybrid significance testing method to work on uptake areas, but we created two new methods of calculating the global threshold for the uptake area (based on replicate variability and bootstrapping). This enables users the option to use uptake area easily and effortlessly in their analyses while maintaining robust elimination of false positives and false negatives.

HDfLeX generates the theoretical intrinsic uptake plots and protection factor, based on the intrinsic rates devised by Bai et al.²⁷ We also introduced a novel empirical approach to normalize between various conditions (e.g., salts) and for proteins (e.g., intrinsically disordered proteins) in a way not accounted for by established correction.²⁷ We tested this

approach on the unstructured peptide standard bradykinin and the intrinsically disordered aSyn. We hope that this new method, when used with scrutiny, will expand the types of differential HDX-MS experiments that can be done, as at present, it is rare to see comparisons between factors that affect the intrinsic rate (such as pH, temperature, ionic strength).^{31,46}

HDfLeX is available to download from <http://hdl.handle.net/10871/127982> and does not require a MATLAB license to use.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.analchem.1c05339>.

Additional method details (bradykinin HDX-MS experimental details); flowcharts describing import and fitting stage, data flattening and statistical significance analysis; plateau back-exchange determination; methods of data distribution by separating the replicates; error distribution of the entire data set; schematic of data flattening process; testing weights for data flattening; deuterium uptake plots for null experiment on aSyn; application of empirical adjustment to an aSyn peptide; comparison of difference plots generated by DynamX vs HDfLeX; bradykinin buffer chemical compositions; and determination of empirical adjustment factor from bradykinin fitted parameters (PDF)

AUTHOR INFORMATION

Corresponding Author

Jonathan J. Phillips – *Living Systems Institute, Department of Biosciences, University of Exeter, Exeter EX4 4QD, U.K.; Alan Turing Institute, British Library, London NW1 2DB, U.K.;* orcid.org/0000-0002-5361-9582; Email: jj.phillips@exeter.ac.uk

Authors

Neeleema Seetaloo – *Living Systems Institute, Department of Biosciences, University of Exeter, Exeter EX4 4QD, U.K.*
Monika Kish – *Living Systems Institute, Department of Biosciences, University of Exeter, Exeter EX4 4QD, U.K.*

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acs.analchem.1c05339>

Notes

The authors declare no competing financial interest.

HDfLeX is available to download from <http://hdl.handle.net/10871/127982> and does not require a MATLAB license to use.

ACKNOWLEDGMENTS

N.S. is funded by the University Council Diamond Jubilee Scholarship. J.J.P. and M.K. are supported by a UKRI Future Leaders Fellowship [Grant Number MR/T02223X/1]. We are grateful for fruitful discussions with Dr. S. Subramanian.

REFERENCES

- (1) Katta, V.; Chait, B. T.; Carr, S. *Rapid Commun. Mass Spectrom.* **1991**, *5* (4), 214–217.
- (2) Englander, J. J.; Del Mar, C.; Li, W.; Englander, S. W.; Kim, J. S.; Stranz, D. D.; Hamuro, Y.; Woods, V. L., Jr *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (12), 7057–7062.
- (3) Phillips, J. J.; Buchanan, A.; Andrews, J.; Chodorge, M.; Sridharan, S.; Mitchell, L.; Burmeister, N.; Kippen, A. D.; Vaughan, T. J.; Higazi, D. R.; Lowe, D. *Anal. Chem.* **2017**, *89* (4), 2361–2368.

- (4) Kim, M.; Sun, Z.-Y. J.; Rand, K. D.; Shi, X.; Song, L.; Cheng, Y.; Fahmy, A. F.; Majumdar, S.; Ofek, G.; Yang, Y.; Kwong, P. D.; Wang, J.-H.; Engen, J. R.; Wagner, G.; Reinherz, E. L. *Nat. Struct. Mol. Biol.* **2011**, *18* (11), 1235–1243.
- (5) Keppel, T. R.; Weis, D. D. *J. Am. Soc. Mass Spectrom.* **2015**, *26* (4), 547–554.
- (6) Stephens, A. D.; Zacharopoulou, M.; Moons, R.; Fusco, G.; Seetaloo, N.; Chiki, A.; Woodhams, P. J.; Mela, I.; Lashuel, H. A.; Phillips, J. J.; De Simone, A.; Sobott, F.; Schierle, G. S. K. *Nat. Commun.* **2020**, *11*, 2820.
- (7) Masson, G. R.; Maslen, S. L.; Williams, R. L. *Biochem. J.* **2017**, *474* (11), 1867–1877.
- (8) Karch, K. R.; Coradin, M.; Zandarashvili, L.; Kan, Z.-Y.; Gerace, M.; Englander, S. W.; Black, B. E.; Garcia, B. A. *Structure* **2018**, *26* (12), 1651–1663.e3.
- (9) Pan, J.; Zhang, S.; Chou, A.; Borchers, C. H. *Chem. Sci.* **2016**, *7* (2), 1480–1486.
- (10) Rand, K. D.; Jørgensen, T. J. D. *Anal. Chem.* **2007**, *79* (22), 8686–8693.
- (11) Rand, K. D.; Adams, C. M.; Zubarev, R. A.; Jørgensen, T. J. D. *J. Am. Chem. Soc.* **2008**, *130* (4), 1341–1349.
- (12) Zehl, M.; Rand, K. D.; Jensen, O. N.; Jørgensen, T. J. D. *J. Am. Chem. Soc.* **2008**, *130* (51), 17453–17459.
- (13) Abzalimov, R. R.; Kaplan, D. A.; Easterling, M. L.; Kaltashov, I. A. *J. Am. Soc. Mass Spectrom.* **2009**, *20* (8), 1514–1517.
- (14) Zubarev, R. A.; Horn, D. M.; Fridriksson, E. K.; Kelleher, N. L.; Kruger, N. A.; Lewis, M. A.; Carpenter, B. K.; McLafferty, F. W. *Anal. Chem.* **2000**, *72* (3), 563–573.
- (15) Mistarz, U. H.; Bellina, B.; Jensen, P. F.; Brown, J. M.; Barran, P. E.; Rand, K. D. *Anal. Chem.* **2018**, *90* (2), 1077–1080.
- (16) Brodie, N. I.; Huguet, R.; Zhang, T.; Viner, R.; Zabrouskov, V.; Pan, J.; Petrotchenko, E. V.; Borchers, C. H. *Anal. Chem.* **2018**, *90* (5), 3079–3082.
- (17) Rand, K. D.; Pringle, S. D.; Morris, M.; Engen, J. R.; Brown, J. M. *J. Am. Soc. Mass Spectrom.* **2011**, *22* (10), 1784.
- (18) Kan, Z.; Ye, X.; Skinner, J. J.; Mayne, L.; Englander, S. W. *Anal. Chem.* **2019**, *91* (11), 7474–7481.
- (19) Lau, A. M. C.; Ahdash, Z.; Martens, C.; Politis, A. *Bioinformatics* **2019**, *35* (17), 3171–3173.
- (20) Hamuro, Y.; Coales, S. J.; Southern, M. R.; Nemeth-Cawley, J. F.; Stranz, D. D.; Griffin, P. R. *J. Biomol. Technol.* **2003**, *14* (3), 171–182.
- (21) Pascal, B. D.; Willis, S.; Lauer, J. L.; Landgraf, R. R.; West, G. M.; Marciano, D.; Novick, S.; Goswami, D.; Chalmers, M. J.; Griffin, P. R. *J. Am. Soc. Mass Spectrom.* **2012**, *23* (9), 1512–1521.
- (22) Wales, T. E.; Eggertson, M. J.; Engen, J. R. *Methods Mol. Biol.* **2013**, *1007*, 263–288.
- (23) Chetty, P. S.; Mayne, L.; Lund-Katz, S.; Stranz, D.; Englander, S. W.; Phillips, M. C. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (45), 19005–19010.
- (24) Hamuro, Y.; Coales, S. J. *J. Am. Soc. Mass Spectrom.* **2018**, *29* (3), 623–629.
- (25) Walters, B. T.; Ricciuti, A.; Mayne, L.; Englander, S. W. *J. Am. Soc. Mass Spectrom.* **2012**, *23* (12), 2132–2139.
- (26) Venable, J. D.; Okach, L.; Agarwalla, S.; Brock, A. *Anal. Chem.* **2012**, *84* (21), 9601–9608.
- (27) Bai, Y.; Milne, J. S.; Mayne, L.; Englander, S. W. *Proteins* **1993**, *17* (1), 75–86.
- (28) Zhang, Z.; Smith, D. L. *Protein Sci.* **1993**, *2* (4), 522–531.
- (29) Marcisin, S. R.; Engen, J. R. *Anal. Bioanal. Chem.* **2010**, *397* (3), 967–972.
- (30) Masson, G. R.; Burke, J. E.; Ahn, N. G.; Anand, G. S.; Borchers, C.; Brier, S.; Bou-Assaf, G. M.; Engen, J. R.; Englander, S. W.; Faber, J.; Garlish, R.; Griffin, P. R.; Gross, M. L.; Guttman, M.; Hamuro, Y.; Heck, A. J. R.; Houde, D.; Iacob, R. E.; Jørgensen, T. J. D.; Kaltashov, I. A.; Klinman, J. P.; Konermann, L.; Man, P.; Mayne, L.; Pascal, B. D.; Reichmann, D.; Skehel, M.; Snijder, J.; Strutzenberg, T. S.; Underbakke, E. S.; Wagner, C.; Wales, T. E.; Walters, B. T.; Weis, D. D.; Wilson, D. J.; Wintrod, P. L.; Zhang, Z.; Zheng, J.; Schriemer, D. C.; Rand, K. D. *Nat. Methods* **2019**, *16* (7), 595–602.
- (31) Li, J.; Rodnina, M. V.; Ladokhin, A. S.; Gross, M. L. *Biochemistry* **2014**, *53* (43), 6849–6856.
- (32) Bonechi, C.; Ristori, S.; Martini, G.; Martini, S.; Rossi, C. *Biochim. Biophys. Acta - Biomembr.* **2009**, *1788* (3), 708–716.
- (33) Kyle, D. J.; Martin, J. A.; Farmer, S. G.; Burch, R. M. *J. Med. Chem.* **1991**, *34* (3), 1230–1233.
- (34) Cai, X.; Dass, C. *Curr. Org. Chem.* **2003**, *7* (18), 1841–1854.
- (35) Christoffersen, M.; Bolvig, S.; Tüchsen, E. *Biochemistry* **1996**, *35* (7), 2309–2315.
- (36) Murphree, T. A.; Vorauer, C.; Brzoska, M.; Guttman, M. *Anal. Chem.* **2020**, *92* (14), 9830–9837.
- (37) Hageman, T. S.; Weis, D. D. *Anal. Chem.* **2019**, *91* (13), 8008–8016.
- (38) MathWorks. *isoutlier*, <https://uk.mathworks.com/help/matlab/ref/isoutlier.html> (accessed 2021-05-11).
- (39) MathWorks. *ttest2*, https://uk.mathworks.com/help/stats/ttest2.html?s_tid=doc_ta (accessed 2021-05-11).
- (40) Kish, M.; Smith, V.; Subramanian, S.; Vollmer, F.; Lethbridge, N.; Cole, L.; Bond, N. J.; Phillips, J. J. Allosteric Regulation of Glycogen Phosphorylase Solution Phase Structural Dynamics at High Spatial Resolution. *bioRxiv* July 16, **2019**, DOI: 10.1101/654665.
- (41) L, S. *The PyMOL Molecular Graphics System*, 2015.
- (42) Houde, D.; Berkowitz, S. A.; Engen, J. R. *J. Pharm. Sci.* **2011**, *100* (6), 2071–2086.
- (43) Hageman, T. S.; Weis, D. D. *Anal. Chem.* **2019**, *91* (13), 8017–8024.
- (44) Del Mar, C.; Greenbaum, E. A.; Mayne, L.; Englander, S. W.; Woods, V. L. *J. Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (43), 15477–15482.
- (45) Mazur, S. J.; Weber, D. P. *J. Am. Soc. Mass Spectrom.* **2017**, *28* (5), 978–981.
- (46) Tajoddin, N. N.; Konermann, L. *Anal. Chem.* **2020**, *92* (14), 10058–10067.