


# SCIENTIFIC REPORTS



OPEN

## Computational method allowing Hydrogen-Deuterium Exchange Mass Spectrometry at single amide Resolution

Chris Gessner<sup>1</sup>, Wieland Steinchen<sup>5</sup>, Sabrina Bédard<sup>3</sup>, John J. Skinner<sup>4</sup>, Virgil L. Woods<sup>1</sup>, Thomas J. Walsh<sup>2</sup>, Gert Bange<sup>5</sup> & Dionysios P. Pantazatos<sup>1</sup> 

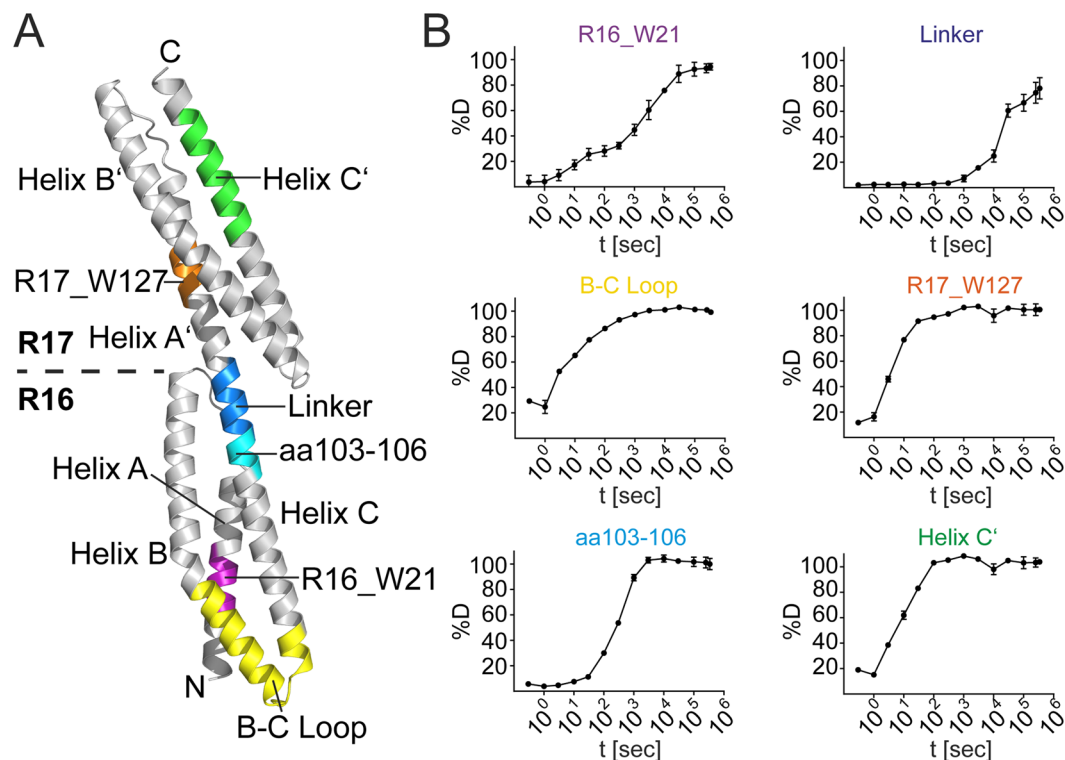
Hydrogen-deuterium exchange (HDX) coupled with mass spectrometry (HDXMS) is a rapid and effective method for localizing and determining protein stability and dynamics. Localization is routinely limited to a peptide resolution of 5 to 20 amino acid residues. HDXMS data can contain information beyond that needed for defining protein stability at single amide resolution. Here we present a method for extracting this information from an HDX dataset to generate a HDXMS protein stability fingerprint. High resolution (HR)-HDXMS was applied to the analysis of a model protein of a spectrin tandem repeat that exemplified an intuitive stability profile based on the linkage of two triple helical repeats connected by a helical linker. The fingerprint recapitulated expected stability maximums and minimums with interesting structural features that corroborate proposed mechanisms of spectrin flexibility and elasticity. HR-HDXMS provides the unprecedented ability to accurately assess protein stability at the resolution of a single amino acid. The determination of HDX stability fingerprints may be broadly applicable in many applications for understanding protein structure and function as well as protein ligand interactions.

Rapid characterization of protein conformation, dynamics, and mapping of protein-ligand interface sites is increasingly being performed by employing hydrogen-deuterium exchange mass spectrometry (HDXMS)<sup>1</sup>. This methodology is advantageous in terms of sensitivity, speed and capability of analyzing large proteins and protein-ligand complexes<sup>2</sup>. HDX is pH- and temperature-dependent<sup>3,4</sup> and involves the reversible exchange of backbone amide hydrogens in a protein with deuterium (D) from the solvent. HDX rates of proteins are determined by labeling of backbone amides with deuterium oxide (D<sub>2</sub>O) and are measured by liquid chromatography mass spectrometry (LCMS). In order for HDX to occur, the amide hydrogen must physically contact the D from the solvent through a local or global unfolding mechanism that may necessitate disruption of any hydrogen bonding.

The relationship of HDX and protein structure dynamics has been well-characterized<sup>5,6</sup>. It is an invaluable method for interrogating protein structure fluctuations and representing localized dynamics<sup>6</sup>. The HDX rate of protein amides is greatly impacted by their solvent accessibility in the context of the protein structure and may vary as many as 9 orders of magnitude on the time scale<sup>7-10</sup>. This wide range of HDX rates represents enormous differences in the protein stability spectrum and might also be expressed as the free energy of exchange ( $\Delta G_{ex}$ )<sup>8</sup>. Consequently, the stability fingerprint of a protein structure may then also be defined as the  $\Delta G_{ex}$  versus the number of amino acids.

The ultimate advantages of HDXMS in many areas of life sciences<sup>10-15</sup> has led to a rapid expansion of this technology<sup>1</sup> and fostered the development of software packages for high-throughput analysis<sup>16, 17</sup>. The HDXMS

<sup>1</sup>Indiana University, Department of Informatics and Computing, Bloomington, IN, USA. <sup>2</sup>Weill Cornell Medicine, Transplantation-Oncology Infectious Disease Program, Division of Infectious Diseases, 1300 York Ave, New York, NY, 10065, USA. <sup>3</sup>GlaxoSmithKline, Platform Technology & Science, Collegeville Road, Collegeville, Pennsylvania, 19426, United States. <sup>4</sup>iHuman Institute, ShanghaiTech University, 99 Haike Road, Pudong, Shanghai, China. <sup>5</sup>Philipps-University Marburg, Faculty of Chemistry & LOEWE Center for Synthetic Microbiology Hans-Meerwein-Strasse, 35043, Marburg, Germany. Virgil L. Woods is deceased. Correspondence and requests for materials should be addressed to D.P.P. (email: [dpantaza@gmail.com](mailto:dpantaza@gmail.com))



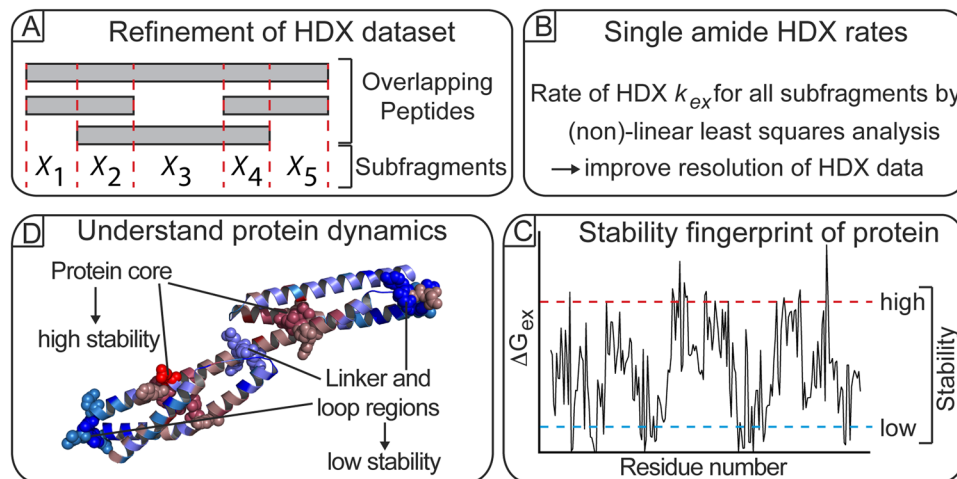
**Figure 1.** Conformational dynamics of R1617. **(A)** The crystal structure of repeats 16 & 17 of  $\alpha$ -spectrin (PDB: 1CUN<sup>23</sup>) is shown in cartoon representation. Regions shown in **(B)** are colored and indicated. **(B)** Deuterium uptake of regions of R1617: R16\_W21 (aa 15–21), B-C Loop (aa 61–84), aa 103–106, Linker (aa 108–114), R17\_W127 (aa 125–128) and Helix C' (aa 200–213).

workflow has been thoroughly described in numerous applications<sup>17–21</sup>. Briefly, proteins or their complexes are incubated in deuterated buffer and after defined time points the reaction is quenched by acidic quench buffer acids (pH 2.7). Subsequently, samples are digested proteolytically to generate peptide fragments<sup>17</sup>. Localization and quantification of D-incorporation into each peptide is performed by LCMS with resulting data analyzed using specialized software<sup>18</sup>. Ultimately, the HDXMS data are then either projected as the percentage of D-incorporation (%D) onto the amino acid sequence (ribbon heat maps) or onto a protein structure (3D heat maps)<sup>19,22</sup>. At present, HDXMS analysis is mainly restricted to peptide resolution governed by the length of the obtained peptides. Analysis of HDXMS data at single amino acid resolution would lead to a much greater insight and understanding of protein structure, dynamics and function. However, software tools enabling this task have yet to be (further) developed.

In this study, we present a software tool enabling HDXMS analysis at highest resolution (HR-HDXMS) by determining single amide HDX rates from acquired HDXMS centroid data. HR-HDXMS utilizes D-incorporation from multiple overlapping peptide fragments and multiple D-labeling time points to determine single amide HDX rates in proteins. These can be collectively presented as the HDX stability fingerprint of a given protein. To demonstrate the feasibility of our approach, we used the structurally and dynamically well-characterized  $\alpha$ -spectrin variant as a model protein<sup>23</sup>. Our software improved the spatial resolution of protein dynamics by providing a superior visualization of the spatial distribution of the stability profile of the structure. Our application demonstrates the utility of HR-HDXMS and HDX stability profiles for evaluating and understanding protein structure and function.

## Results

**Conventional HDXMS analysis of spectrin.** For the development of HR-HDXMS we chose a structurally well-characterized variant of  $\alpha$ -spectrin from chicken brain containing repeats 16 and 17 (further called R1617)<sup>23</sup>. R1617 is composed of two triple-helical bundles joined by a 5 amino acid linker with  $\alpha$ -helical features<sup>23</sup> (Fig. 1A). In order to generate a sufficient amount of overlapping peptides covering the complete amino acid sequence of spectrin, we employed a combination of the proteases pepsin and protease type XIII from *Aspergillus saitoi* (FPXIII) during LCMS analysis. We obtained approximately 200 peptides ranging from 4 to 41 amino acids and covering the entire amino acid sequence of spectrin (Fig. S1 and Table S1). Each peptide was manually reviewed for the correct assignment of charge state and the quality of the spectrum. The combined use of pepsin and FPXIII rather than each alone produced many small peptides with high overlap (Fig. S1). Moreover, this approach significantly increased the redundancy in coverage per amino acid in most regions greater than 2-fold and in some regions even greater than 5-fold. (Fig. S1). Furthermore, we were able to obtain many peptides with charge states



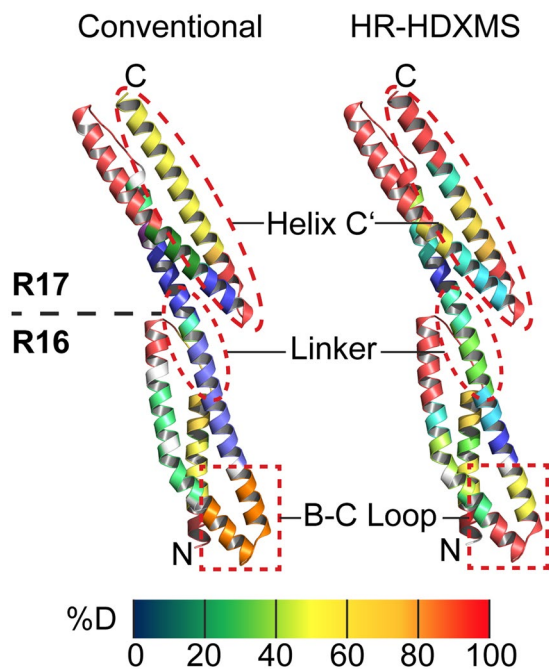
**Figure 2.** Workflow of HR-HDXMS. (A) Amino acid subfragments are determined from overlapping peptides generated by proteolytic digest of the protein. (B) The rate of hydrogen-deuterium exchange ( $k_{ex}$ ) of each subfragment is calculated by linear and non-linear least squares analysis. (C) The free energy value  $\Delta G_{ex}$  derived from  $k_{ex}$  is used to generate a stability fingerprint of the protein. (D) The protein fingerprint helps in understanding protein dynamics and allows predictions about residues that might be essential for protein stability and/or function.

ranging from +1 to +3. Multiple charge states are beneficial to estimate the degree of D-incorporation more accurately because the mass of the peptide should be the same at each charge state.

The percentage of D-incorporation over time for each peptide depends on the localization of the peptide within the protein structure. The use of 13 time points for the measurement of %D provided complete and accurate information about which regions of R1617 were less or more accessible to HDX and how fast the exchange occurred (Fig. 1A and B). We classified different regions of R1617 based on their velocity of HDX by determining the time required for 50% D-incorporation to occur on a peptide (Table S1). With this approach, peptides with 50% D-incorporation <1000 seconds were classified as fast exchanging (135 peptides), those with 50% D-incorporation between 1000–3000 seconds were intermediate exchanging (36 peptides), and peptides that reached 50% D-incorporation above 3000 seconds as slow exchanging (31 peptides). Examples for the fast exchange were the B-C Loop (<3 seconds), Helix C' (<10 seconds) and a peptide in repeat 17 spanning W127 (R17\_W127) residing in helix A' (<10 seconds, Fig. 1B). Also, amino acids 103–106 preceding the linker region of spectrin incorporate over 50%D after just 300 seconds, while the slow exchanging linker (*i.e.* amino acids 108–114) requires 100-times longer to reach similar deuterium incorporation (Fig. 1B). A peptide (R16\_W21) in repeat 16 that includes W21 residing in Helix A of R1617 displays an intermediate D-incorporation. However, two distinct kinetics of deuterium uptake were apparent from the time course with a rather fast and slow exchanging species being overlaid (Fig. 1B). Our data confirm that R1617 is a highly dynamic molecule. Nevertheless, with the conventional way of analyzing HDXMS data subtle differences within the dynamic behavior of spectrin cannot be identified or be attributed to single amino acid residues.

**Theoretical concept of HR-HDXMS.** To improve the resolution in visualizing HDXMS data, we aimed to develop a computational algorithm to determine the hydrogen-deuterium exchange rates and the thermodynamic stability of single amides from an HDXMS dataset of a protein (Fig. 2). Central to this approach is the generation of a peptide library containing multiple overlapping peptides that are produced by combined proteolysis of the protein using pepsin and FPXIII followed by the determination of subfragments between all peptides. In this a subfragment  $X$  is defined as the smallest common overlapping sequence between two given peptides (Fig. 2A). Subsequently, the amount of deuterium incorporated on each subfragment  $X$  is computationally determined by a linear least squares minimization of a global error (GE) function (Fig. 2B and Fig. S2). Deuterium shifts for each subfragment are then determined from the difference of D-incorporation of the subfragment between each successive labeling time point. Next, the HDX rates ( $k_{ex}$ ) for each subfragment are estimated by fitting the exponent in equation 1 using the D-shifts of the calculated subfragments (Fig. 2B and Fig. S3). Consequently,  $k_{ex}$  for each  $X$  represents an average of the exchange rates from all amides within  $X$ . The exchange rates from the subfragments are used to seed initial starting rates for the non-linear least squares optimization function that solves for the  $k_{ex}$  of all amides (Fig. 2B and Fig. S3). The so-obtained exchange rates  $k_{ex}$  of each amide can then be converted into the free energy of hydrogen-deuterium exchange ( $\Delta G_{ex}$ ) according to the Gibbs free energy equation allowing for a precise understanding of protein dynamics on the single amino acid level (Fig. 2C and D). Taken together, this procedure should enable the determination of  $k_{ex}$  values of single amides from HDXMS datasets containing a sufficiently large number of overlapping peptides with redundant sequence coverage.

**Subfragmentation improves the spatial resolution of HDXMS.** To learn more about subtle changes within the dynamics of R1617 that might not be visible by conventional HDXMS evaluation, we applied our HR-HDXMS software to the HDXMS dataset. Subfragments were determined from overlapping peptides



**Figure 3.** Improved resolution of R1617 conformational dynamics by HR-HDXMS. Deuterium uptake of R1617 mapped onto its crystal structure (PDB: 1CUN<sup>23</sup>). Data analysis by HR-HDXMS (*right*) provides a superior spatial resolution of deuterium uptake compared to ‘conventional’ HDX data analysis (*left*).

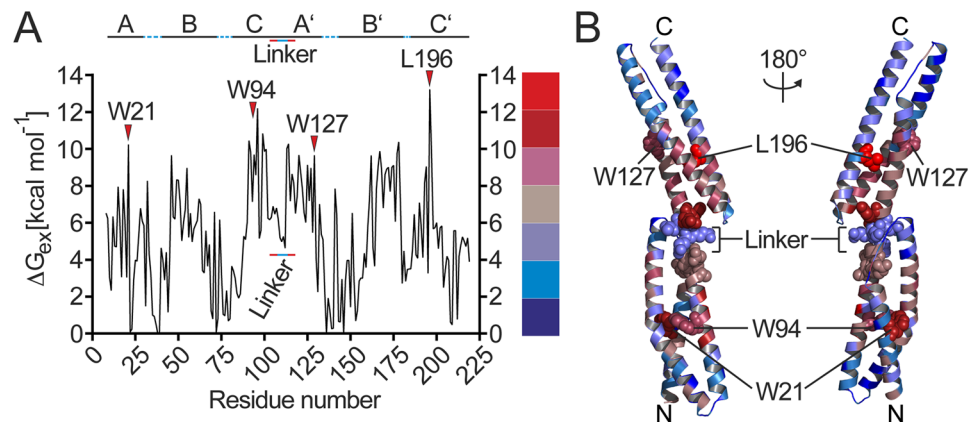
followed by determination of the deuterium incorporation and the rate of hydrogen-deuterium exchange  $k_{ex}$  of each subfragment by linear and non-linear least squares analysis (see above). These steps drastically improved the spatial resolution of the dynamics of R1617 (Fig. S4). In those areas of R1617 where many overlapping peptide fragments were available, deuterium uptake could even be determined for single amino acids (compare to Figs S1 and S4).

To visually inspect the improvements in resolution, we projected the results of ‘conventional’ HDXMS (see above) and HR-HDXMS data analysis on the crystal structure of R1617 (Fig. 3). Helix C’ could now be subdivided into four segments with different dynamic properties rather than two seen with the conventional approach (from the N- to C-terminus of helix C’: >90, 50, 20 and >90% D) (Fig. 3). Another example was found for the dynamics of the B-C Loop. While the conventional approach displayed intermediate-high dynamics (70% D), the HR-HDXMS heat map was further subdivided into three categories of dynamics including regions of very high (>90% D) over intermediate (50% D) to low (20% D) dynamics from the N- to the C-terminus of the B-C Loop (Fig. 3). Moreover, within the linker region the number of segments exhibiting different dynamics could be increased approximately two-fold allowing for a more detailed understanding (see below). Taken together, our approach of generating a subfragment map clearly refines analysis of HDXMS as exemplified by R1617 (Fig. 3).

**HR-HDXMS allows stability fingerprints at single amino acid residue resolution.** Following our theoretical considerations (see above), we wanted to take our analysis further, eventually to determine stability fingerprints at the single amide level. Therefore, the exchange rates from each subfragment were used to seed initial starting rates for the nonlinear least squares function solving for the exchange rate of each amide ( $k_{ex}$ ) (equation 7). The  $k_{ex}$  was converted into the free energy of exchange ( $\Delta G_{ex}$ ) according to the Gibbs Free Energy equation ( $\Delta G_{exchange,i} = -RT \ln(k_{ex,i}/k_{int,i})$ ) where  $k_{ex,i}$  and  $k_{int,i}$  are the experimental and intrinsic (random coil) exchange rates, respectively, at an amide  $i$  as determined from the intrinsic rates of random coil model peptides<sup>24,25</sup>. These  $\Delta G_{ex}$  for each amide were plotted versus the position of the amino acid residue (Fig. 4A). The  $\Delta G_{ex}$  values range from 0 kcal/mol (*i.e.* prolines lacking the amide hydrogen) to 14 kcal/mol. Generally, lower and higher  $\Delta G_{ex}$  values correlate to faster and slower HDX rates of the amide, respectively<sup>6,7,26</sup>. Therefore, lower  $\Delta G_{ex}$  values represent amides of amino acid residues that are more dynamic and, *vice versa*, higher  $\Delta G_{ex}$  represent amides of residues that are less dynamic.

The HDX stability fingerprint of R1617 as determined by our HR-HDXMS approach displays a jagged profile with an observable pattern of stability with a large window of variation in amide dynamics (Fig. 4A). Generally, higher  $\Delta G_{ex}$  values are observed around the centers of the  $\alpha$ -helices and lower  $\Delta G_{ex}$  values are observed in the loop regions connecting the  $\alpha$ -helices. Noteworthy, the amides of four amino acids (*i.e.* amino acids 109–112) constituting the linker that connects the repeats 16 and 17 exhibit a local minimum in  $\Delta G_{ex}$  while residues preceding and following the linker showed much higher  $\Delta G_{ex}$  values (Fig. 4A). The functional consequences of this keen discrimination between the linker region and the C- and N-terminal  $\alpha$ -helices of the spectrin repeats 16 and 17 are discussed below.





**Figure 4.** Stability fingerprint of R1617 determined by HR-HDXMS. **(A)** Plot of the free energy value  $\Delta G_{\text{ex}}$  for each amino acid residue of R1617. The linker region between repeats 16 and 17 and amino acids adjacent to the linker are indicated by solid blue and red lines, respectively. Amino acid residues exhibiting high stability are indicated. **(B)** The crystal structure of R1617 (PDB: 1CUN<sup>23</sup>) colored according to the  $\Delta G_{\text{ex}}$  values shown in **(A)**.

To challenge the validity of our approach, we examined the accuracy upon which the HDX stability fingerprint could correctly identify amino acids or locations on the structure of R1617 that have previously been shown to confer stability or impose flexibility<sup>23,27,28</sup>. For this, the  $\Delta G_{\text{ex}}$  values were projected onto the crystal structure of R1617 resulting in a superior spatial resolution of dynamics when compared to our previous subfragment analysis (Fig. 4B, and compare with Fig. 3). Our analysis revealed the amides of three tryptophan residues (*i.e.* W21, W94 within repeat 16 and W127 in repeat 17) and leucine at position 196 to exhibit very high  $\Delta G_{\text{ex}}$  values. These residues localize in the center of the triple helical bundles forming repeat 16 and 17 (ref. 23). Notably, W21 and W94 have already been shown to significantly confer stability of the triple helical core<sup>27,28</sup>, consolidating the validity of our HR-HDXMS approach to determine a stability fingerprint.

Taken together, these results clearly demonstrate that our HR-HDXMS approach allows unraveling the contribution of single amino acids to the overall dynamics of protein structures exemplified by spectrin.

**HR-HDXMS agrees with single amide HDX rates obtained by NMR.** Traditionally, single amide HDX rates have been determined by nuclear magnetic resonance (NMR)<sup>29,30</sup>. Therefore, we extended the validation of our HR-HDXMS method against data obtained by NMR. To do so, we performed HR-HDXMS on a dataset obtained for *Staphylococcus* nuclease (SNase) comprising 341 peptides and 10 sampled time points<sup>31</sup>. In the initial determination of deuterium incorporation by linear least squares analysis (equation 1), a large number of residues exhibited a high residual error (Fig. S5A). To circumvent this problem that might be caused by poor data quality or incorrect peptide assignment, we eliminated peptides with an error  $\geq 2$  deuterium. After exclusion of those peptides, the residual error for each peptide from the linear optimization (equation 2) could be reduced below 1 deuterium for the remaining 206 filtered fragments (Fig. S5B). Although approximately 40% of peptides are lost during this procedure reducing the obtainable resolution, the quality of deuterium incorporation assignment for the remaining peptides is significantly raised.

For comparison of the stability fingerprints obtained by HR-HDXMS and NMR, the analogous NMR exchange rate data of SNase served as a gold standard<sup>9,26</sup>. At first, the  $k_{\text{ex}}$  rates obtained by NMR were used to determine the stability fingerprint of SNase by the HR-HDXMS algorithm. The  $\Delta G_{\text{ex}}$  values obtained by global error minimization applying HR-HDXMS are almost identical to the  $\Delta G_{\text{ex}}$  values obtained from the experimental data (Fig. S6A). Also, the stability fingerprint of SNase obtained by determination of  $k_{\text{ex}}$  values from HDXMS dataset resulted in a good overlay with the NMR-derived stability fingerprint (Fig. S6B). These results further demonstrate the feasibility of HR-HDXMS even when suboptimal HDX data are available (see above) and show that the accuracy of the fingerprint relies on the quality of the seeding rates in the nonlinear solver.

## Discussion

**HR-HDXMS as tool to refine HDX data to single amide resolution.** In this work, we present HR-HDXMS as suitable tool for refining HDXMS data to single amide resolution. Therefore, HR-HDXMS opens a new route to delineate the individual contribution of single amino acids to the overall dynamics of proteins and their macromolecular assemblies. To this end, suitability of HR-HDXMS was tested with repeats 16 & 17 of chicken brain  $\alpha$ -spectrin (R1617), an enormously well characterized protein in terms of structure<sup>23,32</sup>, biochemical characteristics<sup>27,33</sup>, and postulated mechanisms of function. As exemplified, HR-HDXMS was able to recapitulate all known features of spectrin structure and dynamics, but much more provided further insights into the dynamic features of spectrin (see below). However, whether these are of functional relevance remains to be further investigated.

The HR-HDXMS approach involves a simple two-phase numerical technique. A linear least squares fit is performed for the initial estimation of amide exchange rates from overlapping peptide fragments followed by a non-linear least squares fit solving for individual amide exchange rates within each peptide. We generated an

HDXMS dataset of repeats 16 & 17 of chicken brain  $\alpha$ - spectrin, which provided 2600 overlapping constraints that were a product of 200 peptides and 13 time points of deuterium labeling for performing the linear and non-linear fits. Our study showed that successful application of HR-HDXMS critically relies on: the generation of multiple overlapping peptide fragments, general data quality, correction for D-loss into the solvent by incorporating residue-specific back-exchange rates in the non-linear fit, and imposing appropriate boundaries for HDX rates during non-linear optimization to serve as constraints for the fit. We believe that the generation of multiple overlapping peptides (allowing the subfragment assignment) is most critical for the success of HR-HDXMS. This issue can be addressed by the use of the two proteases pepsin and FPXIII. Overall, HR-HDXMS seems to offer a promising way to analyze HDXMS datasets at single amide resolution. However, further studies are required to challenge the suitability of our approach especially with respect to larger proteins and macromolecular assemblies where the amount of overlapping peptides and redundant sequence coverage becomes a greater challenge.

**Advantages of HR-HDXMS over other methods and alternatives.** Traditionally, single amide HDX rates have been determined by nuclear magnetic resonance (NMR)<sup>29,30</sup>. While NMR is powerful to accurately measure single amide HDX rates, it is usually limited in protein size (<30 kDa molecular weight). A great advantage of HDXMS is its ability to analyze proteins and macromolecular assemblies extending molecular weights over 100 kDa. Moreover, our study shows that HR-HDXMS is able to recapitulate single amide HDX rates obtained by NMR as exemplified by SNase.

Mass spectrometry is also capable of determining the D-incorporation at a single amide through stepwise peptide fragmentation employing electron capture dissociation (ECD) and electron transfer dissociation (ETD)<sup>34–37</sup>. Collision-induced dissociation (CID) fragmentation approaches have commonly been used to improve coverage and sublocalization of D-labeling. However, this approach typically suffers from deuterium scrambling<sup>38</sup> among adjacent amides making reliable sublocalization challenging. HR-HDXMS alleviates the drawbacks of CID, because it relies on the total D-incorporation on the peptide irrespective of any scrambling occurring during CID fragmentation.

Fragmentation methods such as ECD<sup>39</sup> and ETD<sup>36,40</sup> also allow for determination of single amide D-incorporation by inducing stepwise fragmentation. However, the conventional use of these methods to determine D-incorporation at the single amide level is limited. ETD requires peptides with higher charge states to work effectively<sup>41</sup>. Therefore, not all peptides will be amenable to ETD fragmentation.

Other mathematical and analytical approaches have recently been developed for determining single amide rates and/or improving the resolution in HDXMS datasets<sup>31,42–46</sup>. These approaches include a mathematical model for predicting the assignment of HDX rate classes to amino acid segments from overlapping peptides<sup>46</sup>, a statistical approach to quantify the difference of HDX between two states<sup>42</sup>, residue averaging of the mid-point values of HDX<sup>43</sup>, and utilization of the shape of the isotopic envelope to determine the distribution of deuterium on the amides of the peptide by the software HDSite<sup>31</sup>. HDSite demonstrates the feasibility of determining exchange rates at single amide resolution, but it functions optimally only with high-resolution data providing enough information to produce a well-defined isotopic envelope. Implementation of HDSite involves a separate iteration for each peptide to determine the amide D-occupancy in the peptide and a separate fit to an exponential to determine the HDX rate for each residue based on the previously determined amount of D on the amide at each time point. In contrast to HDSite, HR-HDXMS performs a single global linear optimization iteration to simultaneously determine the D level on all subfragments from overlapping peptides and then a single global non-linear iteration using a non-linear optimizer to simultaneously determine all HDX rates for all amides. One drawback of the HR-HDXMS software is that it assumes all HDX to be governed by EX2 kinetics, and in its present version does not take EX1 kinetics into account.

Taken together, both HDSite and HR-HDXMS depend on a sufficiently high number of overlapping fragments, number of time points, length of the sequence, and number of replicates. These issues represent the most challenging limitation of both approaches. Nevertheless, our validation of HR-HDXMS demonstrated its robustness to correctly calculate the stability profile of a protein.

**Lessons learned on the dynamics of spectrin by HR-HDXMS.** Our HR-HDXMS approach improved the spatial resolution of the dynamics of the  $\alpha$ -spectrin tandem repeat (R1617) to the single amide level. Based on the crystal structure of R1617, two models of elasticity and flexibility were proposed<sup>23</sup>. The first model described the flexibility imposed by the linker<sup>23,47</sup> through helix bending at the linker region. The second one suggested high flexibilities along the  $\alpha$ -helices and their connecting loops due to their conformational rearrangement<sup>18</sup>. Interestingly, our HR-HDXMS analysis supports both models and extends them further. We found substantial stability gradients at the center of the  $\alpha$ -helices that gradually decrease towards the helical ends (Fig. 4B). Probably, gradients of stability near the helical ends may facilitate the structural transition from helical propensity into loop geometry. With respect to the linker, conventional analysis of our HDXMS supports the highly dynamic behavior of the linker as already inferred from the crystal structure (Fig. 3). However, the refinement in spatial resolution provided by our HR-HDXMS approach revealed that residues 104–106 preceding the linker exhibited higher  $\Delta G_{ex}$  than the linker residues 108–112 that exhibited a local minimum in  $\Delta G_{ex}$ . Thereby, we could identify specific amino acids that possibly confer less or more stability within the linker region. Therefore, our HR-HDXMS analysis of the spectrin tandem repeat R1617 supports a model combining the ones previously postulated and suggests a much more complex scenario in which all the structural elements of spectrin contribute to the overall dynamics required for its functionality. Moreover, our analysis clearly shows that the triple helical bundles of repeats 16 and 17 require structural stabilization that is provided through intramolecular contacts by a number of central tryptophane residues<sup>27,48,49</sup>.

Taken together, HR-HDXMS might represent a powerful tool to analyze protein dynamics at single amide resolution.

## Materials and Methods

**Protein preparation.** The cDNA of repeats 16 and 17 from chicken brain  $\alpha$ -spectrin was cloned and expressed in *Escherichia coli* BL21 (DE3) (Novagen) and purified as previously described<sup>23</sup>.

**Sample preparation and fragmentation optimization.** Conditions for optimal protein digestion were determined by adding 30  $\mu$ l of quench solution with increasing concentrations of Guanidine Hydrochloride (GuHCl) (0/0.8/1.6/3.2/6.4 M GuHCl in 0.8% (v/v) formic acid) to 20  $\mu$ l of sample (final concentration 0.5% (v/v) formic acid, 0/0.5/1/2/4 M GuHCl) containing 10–15  $\mu$ g of protein in Tris-buffered saline (TBS: 50 mM Tris-HCl, pH 7.0 and 150 mM NaCl). Samples were then frozen on dry ice within one minute after addition of quench solution and stored at  $-80^{\circ}\text{C}$  until LCMS analysis. Sequest (Thermo Finnigan Inc) was used for peptide identification and peptides were verified using the “DXMS Explorer” software (Sierra Analytics, LLC, Modesto, CA)<sup>19</sup>.

**Deuterium labeling of proteins.** Exchange-deuterated samples of spectrin were prepared by the addition of 15  $\mu$ l of  $\text{D}_2\text{O}$ , containing 50 mM Tris, pD 2.7 and 150 mM NaCl to 5  $\mu$ l of protein stock (3 mg/ml in TBS pH 7.0) and incubated at room temperature  $25^{\circ}\text{C}$  for 3, 10, 30, 100, 300,  $10^3$ ,  $3 \times 10^3$ ,  $10^4$ ,  $2.5 \times 10^5$  and  $3.4 \times 10^5$  seconds, with the reaction quenched by addition of 30  $\mu$ l quench solution (0.8% (v/v) formic acid, 0.8 M GuHCl) at  $0^{\circ}\text{C}$ , and samples immediately placed on dry ice and frozen at  $-80^{\circ}\text{C}$  until further processed (as described above). Time points of 3 and 10 seconds were incubated on ice to simulate 0.3 and 1 seconds deuteration time as there would be a 10-fold decrease in the exchange rate for every  $25^{\circ}\text{C}$  decrease in temperature<sup>51</sup>. Data on the deuterated sample sets were acquired in a single automated 8 hour run, and subsequent data reduction was performed on the data reduction software (see above). Corrections for loss of D-labelling by individual fragments during HDXMS analysis (after “quench”) were made by determination of loss of deuterium from reference samples that had been fully deuterated by overnight incubation under denaturing conditions, as previously described<sup>52–55</sup>.

**LCMS analysis.** Frozen quenched samples were placed in a Spectrapysics AS3000 autosampler kept at  $-45^{\circ}\text{C}$  with dry ice powder, rapidly thawed within 30 seconds followed by immediate injection into LC sample loop pre-cooled at  $0^{\circ}\text{C}$ . LCMS analysis was performed at  $0^{\circ}\text{C}$  and pH 2.3 as previously described<sup>56,57</sup>. Online proteolysis involved the sample passing through an initial immobilized pepsin (30 mg/ml) column (66  $\mu$ l column bed volume) that was coupled to a 20AL resin from PerSeptive Biosystems in 0.05% (w/v) trifluoroacetic acid (TFA) followed by passage through a column (66  $\mu$ l column bed volume) containing immobilized protease type XIII from *Aspergillus saitoi* (20 mg/ml)<sup>58</sup> at 250  $\mu$ l/min flow rate. Reversed-phase separation of digested peptides utilized a C18 column, eluted by a linear acetonitrile gradient (5 to 45% B in 30 minutes; 50  $\mu$ l/min; solvent A, 0.05% TFA; solvent B, 80% acetonitrile, 20% water, 0.01% TFA), with data acquisition using a Thermo Finnigan Classic LCQ electrospray ion trap type mass spectrometer or an electrospray Micromass Q-TOF mass spectrometer operated with a capillary temperature at  $200^{\circ}\text{C}$ .

**HDXMS analysis.** Mass spectrometry data was analyzed using DXMS data reduction software (Sierra Analytics, Version 1.01 (beta))<sup>19</sup>. Deuterated peptides were identified in DXMS and pasted into an excel macro that determines the D-incorporation of each peptide fragment at each timepoint. Protein fragmentation map, ribbon heat maps, and D-incorporation labeling curves were generated as previously described<sup>56</sup>. All structure figures were generated with Pymol (Delano Scientific).

**Linear least squares analysis.** Given  $m$  overlapping fragments spanning residues  $a \rightarrow b$ , a subfragment  $x$  can be defined as the overlapping and non-overlapping region(s) between fragments (Fig. S2). Given a number of subfragments  $N$  within a given fragment  $F$ , the total amount of experimentally measured deuterium  $D_{F_n,t}^{\text{exp}}$  incorporated on fragment  $n$  at each timepoint  $t$  can be represented as the sum of the deuterium incorporated on each subfragment plus a residual error  $E_{F_n,t}^{\text{lin}}$  as depicted in equation 1, where  $j$  is the defined subfragment,  $N$  is the total number of subfragments spanning fragment  $n$ , and  $p$  is the number of prolines that are subtracted from the number of exchangeable amides as they do not contain amide hydrogens.

$$D_{F_n,t}^{\text{exp}} = \sum_{j=1}^{N_n} x_{j,t} - p + E_{F_n,t}^{\text{lin}} \quad (1)$$

The application of the linear program based on approximation of initial exchange rates of subfragments is illustrated in Fig. S2. The three peptide fragments  $F_1, F_2, F_3$  spanning the first 29 amino acids of spectrin define subfragments  $(x_1, \dots, x_3)$ . The amount of D on each subfragment at each timepoint is determined by linear least square optimization to minimize the sum of the residual error over all fragments ( $M$ ), and all time points ( $T$ ) as defined by the global error function in equation 2.

$$GE^{\text{lin}} = \sum_{t=1}^T \sum_{n=1}^M (E_{F_n,t}^{\text{lin}})^2 \quad (2)$$

**Non-Linear Least Squares Analysis.** The collective exchange rate  $k_{\text{ex}}$  of the amides within each subfragment is determined by simultaneously fitting the D-incorporation data to the first-order reaction kinetics for D-incorporation of an amide in equation 3, where  $D_{i,t}$  is the amount of D incorporated on residue  $i$  at time  $t$ .

$$D_{i,t} = (1 - e^{-k_{\text{ex},i}t}) \quad (3)$$

The rate constant  $k_{ex,i}$  is dependent on pH, temperature, protein sequence, and protein conformation. In a completely unstructured polypeptide chain, all peptide amide hydrogens are freely accessible to water and exchange at their maximal possible rate, with a half-life of exchange of approximately one second at 0 °C and pH 7.0 (ref. 25). Exact off-exchange rates for amide hydrogens in a fully unstructured protein sequence can be reliably calculated from knowledge of the temperature, pH and primary amino acid sequence involved<sup>24,25</sup>. The precise rate of HDX of a particular amide in random coil can vary more than 30-fold from the average rates for all amides in a peptide under such conditions, with the precise rate depending upon the identity of the two amino acids flanking the particular amide bond. Because the deuterated samples are subjected to quench conditions during the LC process, a correction for back-exchange must be made to account for the loss of label from the time the sample is injected to the time it is eluted, also referred to as the lag time  $t_{lag}$ . We can model the back-exchange  $D_i^{bk}$  on subfragment  $x$  as the fraction of remaining deuterium after a lag time of an amide at quench conditions through:

$$D_i^{bk} = e^{-k_{bk,i} t_{lag}} \quad (4)$$

where  $k_{bk,i}$  is the calculated back-exchange rate<sup>24,25</sup> of amino acid  $i$  under quench conditions at lag time  $t_{lag}$ .

**Optimizing on-exchange rates by incorporating back-exchange rates.** The amount of D on a single amide  $i$ , in a peptide fragment corrected for back-exchange can be described as the product of the amount of D-incorporated on the amide and the back-exchange according to:

$$D_{i,t}^{calc} = (1 - e^{-k_{ex,i} t})(e^{-k_{bk,i} t_{lag}}) \quad (5)$$

It follows that the experimental D-accumulation of a given peptide fragment  $F$ , including back-exchange corrections is the sum of the individual exchange rates of each amide plus a residual error  $E_{F,n,t}^{nonlin}$  defined by equation 6.

$$D_{F,n,t}^{exp} = \sum_{i=m}^l (1 - e^{-k_{ex,i} t})(e^{-k_{bk,i} t_{lag}}) + E_{F,n,t}^{nonlin} \quad (6)$$

Here,  $D_{i,t}^{calc}$  is summed over all amide position  $I$ ,  $m < i < l$ , where  $m$  and  $l$  are the first and last residues of fragment  $n$ , respectively. The non-linear least squares algorithm solves for the rate constants  $k_{ex}$  that minimize the squared difference between the fitted and experimental deuteration level of all fragments. This concept is illustrated in Fig. S3. The objective function aims at minimizing the global error for all fragments ( $n = 1, \dots, M$ ) at all time points ( $t = 1, \dots, T$ ):

$$GE^{nonlin} = \sum_{t=1}^T \sum_{n=1}^M (E_{F,n,t}^{nonlin})^2 \quad (7)$$

The success of HR-HDXMS relies on the extent of overlapping fragments and the number of sampled on-exchange time points. Upper and lower boundaries are set as constraints for the fit using the intrinsic back-exchange rates determined by Molday *et al.*<sup>24</sup> from spreadsheet calculations obtained at <http://hx2.med.upenn.edu/download.html>. The exchange rate of all prolines was set to 0 during the non-linear fit. Validation of HR-HDXMS was performed by determination of the HDX stability profile of SNase from HDXMS data and comparison with the NMR-derived stability profile. Acquisition of NMR data was performed as previously described<sup>9</sup>.

HR-HDXMS was written in MatLab utilizing the optimization toolbox implementing the linear and non-linear functions. The MatLab algorithm is available under <http://www.hr-hdxms.us>.

## References

- Pirrone, G. F., Iacob, R. E. & Engen, J. R. Applications of hydrogen/deuterium exchange MS from 2012 to 2014. *Anal Chem* **87**, 99–118 (2015).
- Wales, T. E. & Engen, J. R. Hydrogen exchange mass spectrometry for the analysis of protein dynamics. *Mass Spectrometry Reviews* **25**, 158–170 (2006).
- Linderstrom-Lang, K. The pH-dependence of the deuterium exchange of insulin. *Biochim Biophys Acta* **18**, 308 (1955).
- Zhang, Z. & Smith, D. L. Determination of amide hydrogen exchange by mass spectrometry: a new tool for protein structure elucidation. *Protein Sci* **2**, 522–31 (1993).
- Berger, A. & Linderstrom-Lang, K. Deuterium exchange of poly-DL-alanine in aqueous solution. *Arch Biochem Biophys* **69**, 106–118 (1957).
- Englander, S. W., Mayne, L., Bai, Y. & Sosnick, T. R. Hydrogen exchange: the modern legacy of Linderstrom-Lang. *Protein Science* **6**, 1101–9 (1997).
- Maity, H., Lim, W. K., Rumbley, J. N. & Englander, S. W. Protein hydrogen exchange mechanism: local fluctuations. *Protein Sci* **12**, 153–60 (2003).
- Milne, J. S., Mayne, L., Roder, H., Wand, A. J. & Englander, S. W. Determinants of protein hydrogen exchange studied in equine cytochrome *c*. *Protein Sci* **7**, 739–45 (1998).
- Skinner, J. J., Lim, W. K., Bedard, S., Black, B. E. & Englander, S. W. Protein hydrogen exchange: testing current models. *Protein Sci* **21**, 987–95 (2012).
- Houde, D., Berkowitz, S. A. & Engen, J. R. The utility of hydrogen/deuterium exchange mass spectrometry in biopharmaceutical comparability studies. *J Pharm Sci* **100**, 2071–86 (2011).
- Wei, H. *et al.* Hydrogen/deuterium exchange mass spectrometry for probing higher order structure of protein therapeutics: methodology and applications. *Drug Discov Today* **19**, 95–102 (2014).
- Barclay, L. A. *et al.* Inhibition of Pro-apoptotic BAX by a noncanonical interaction mechanism. *Mol Cell* **57**, 873–86 (2015).
- Melnyk, R. A. *et al.* Structural determinants for the binding of anthrax lethal factor to oligomeric protective antigen. *J Biol Chem* **281**, 1630–5 (2006).



14. Hamuro, Y. *et al.* Hydrogen/deuterium-exchange (H/D-Ex) of PPARgamma LBD in the presence of various modulators. *Protein Sci* **15**, 1883–92 (2006).
15. Ramachandran, G. & Udgaonkar, J. B. Difference in fibril core stability between two tau four-repeat domain proteins: a hydrogen-deuterium exchange coupled to mass spectrometry study. *Biochemistry* **52**, 8787–9 (2013).
16. Hamuro, Y., Weber, P. C. & Griffin, P. R. High-throughput analysis of protein structure by hydrogen/deuterium exchange mass spectrometry. *Methods Biochem Anal* **45**, 131–57 (2005).
17. Gallagher, E. S. & Hudgens, J. W. Mapping Protein-Ligand Interactions with Proteolytic Fragmentation, Hydrogen/Deuterium Exchange-Mass Spectrometry. *Methods Enzymol* **566**, 357–404 (2016).
18. Mayne, L. *et al.* Many overlapping peptides for protein hydrogen exchange experiments by the fragment separation-mass spectrometry method. *J Am Soc Mass Spectrom* **22**, 1898–905 (2011).
19. Garcia, R. A., Pantazatos, D. & Villarreal, F. J. Hydrogen/deuterium exchange mass spectrometry for investigating protein-ligand interactions. *Assay Drug Dev Technol* **2**, 81–91 (2004).
20. Steinchen, W. *et al.* Catalytic mechanism and allosteric regulation of an oligomeric (p)ppGpp synthetase by an alarmone. *Proc Natl Acad Sci USA* **112**, 13348–53 (2015).
21. Schuhmacher, J. S. *et al.* MinD-like ATPase FlhG effects location and number of bacterial flagella during C-ring assembly. *Proc Natl Acad Sci USA* **112**, 3092–7 (2015).
22. Hamuro, Y. *et al.* Rapid analysis of protein structure and dynamics by hydrogen/deuterium exchange mass spectrometry. *J Biomol Tech* **14**, 171–82 (2003).
23. Grum, V. L., Li, D., MacDonald, R. I. & Mondragon, A. Structures of two repeats of spectrin suggest models of flexibility. *Cell* **98**, 523–35 (1999).
24. Molday, R. S. & Englander, S. W. Primary structure effects on peptide group hydrogen exchange. *Biochemistry* **11**, 150 (1972).
25. Bai, Y., Milne, J. S., Mayne, L. & Englander, S. W. Primary structure effects on peptide group hydrogen exchange. *Proteins* **17**, 75–86 (1993).
26. Skinner, J. J., Lim, W. K., Bedard, S., Black, B. E. & Englander, S. W. Protein dynamics viewed by hydrogen exchange. *Protein Sci* **21**, 996–1005 (2012).
27. Pantazatos, D. P. & MacDonald, R. I. Site-directed mutagenesis of either the highly conserved Trp-22 or the moderately conserved Trp-95 to a large, hydrophobic residue reduces the thermodynamic stability of a spectrin repeating unit. *J Biol Chem* **272**, 21052–9 (1997).
28. MacDonald, R. I., Musacchio, A., Holmgren, R. A. & Saraste, M. Invariant tryptophan at a shielded site promotes folding of the conformational unit of spectrin. *Proc Natl Acad Sci USA* **91**, 1299–303 (1994).
29. Cortajarena, A. L., Mochrie, S. G. & Regan, L. Mapping the energy landscape of repeat proteins using NMR-detected hydrogen exchange. *J Mol Biol* **379**, 617–26 (2008).
30. Henry, G. D., Weiner, J. H. & Sykes, B. D. Backbone dynamics of a model membrane protein: measurement of individual amide hydrogen-exchange rates in detergent-solubilized M13 coat protein using <sup>13</sup>C NMR hydrogen/deuterium isotope shifts. *Biochemistry* **26**, 3626–34 (1987).
31. Kan, Z. Y., Walters, B. T., Mayne, L. & Englander, S. W. Protein hydrogen exchange at residue resolution by proteolytic fragmentation mass spectrometry analysis. *Proc Natl Acad Sci USA* **110**, 16438–43 (2013).
32. Kusunoki, H., Minasov, G., Macdonald, R. I. & Mondragon, A. Independent movement, dimerization and stability of tandem repeats of chicken brain alpha-spectrin. *J Mol Biol* **344**, 495–511 (2004).
33. Li, D. *et al.* Structural and Thermodynamic Studies of Cloned Fragments of Spectrin. *Cell Mol Biol Lett* **6**, 218 (2001).
34. Rand, K. D., Pringle, S. D., Morris, M., Engen, J. R. & Brown, J. M. ETD in a traveling wave ion guide at tuned Z-spray ion source conditions allows for site-specific hydrogen/deuterium exchange measurements. *J Am Soc Mass Spectrom* **22**, 1784–93 (2011).
35. Rand, K. D., Zehl, M., Jensen, O. N. & Jorgensen, T. J. Protein hydrogen exchange measured at single-residue resolution by electron transfer dissociation mass spectrometry. *Anal Chem* **81**, 5577–84 (2009).
36. Zehl, M., Rand, K. D., Jensen, O. N. & Jorgensen, T. J. Electron transfer dissociation facilitates the measurement of deuterium incorporation into selectively labeled peptides with single residue resolution. *J Am Chem Soc* **130**, 17453–9 (2008).
37. Konermann, L., Pan, J. & Liu, Y. H. Hydrogen exchange mass spectrometry for studying protein structure and dynamics. *Chem Soc Rev* **40**, 1224–34 (2011).
38. Ferguson, P. L. & Konermann, L. Nonuniform isotope patterns produced by collision-induced dissociation of homogeneously labeled ubiquitin: implications for spatially resolved hydrogen/deuterium exchange ESI-MS studies. *Anal Chem* **80**, 4078–86 (2008).
39. Abzalimov, R. R., Kaplan, D. A., Easterling, M. L. & Kaltashov, I. A. Protein conformations can be probed in top-down HDX MS experiments utilizing electron transfer dissociation of protein ions without hydrogen scrambling. *J Am Soc Mass Spectrom* **20**, 1514–7 (2009).
40. Rand, K. D. *et al.* Gas-phase hydrogen/deuterium exchange in a traveling wave ion guide for the examination of protein conformations. *Anal Chem* **81**, 10019–28 (2009).
41. Abzalimov, R. R. & Kaltashov, I. A. Controlling hydrogen scrambling in multiply charged protein ions during collisional activation: implications for top-down hydrogen/deuterium exchange MS utilizing collisional activation in the gas phase. *Anal Chem* **82**, 942–50 (2010).
42. Saltzberg, D. J. *et al.* A Residue Resolved Bayesian Approach to Quantitative Interpretation of Hydrogen Deuterium Exchange from Mass Spectrometry: Application to Characterizing Protein-Ligand Interactions. *J Phys Chem B* (2016).
43. Keppel, T. R. & Weis, D. D. Mapping residual structure in intrinsically disordered proteins at residue resolution using millisecond hydrogen/deuterium exchange and residue averaging. *J Am Soc Mass Spectrom* **26**, 547–54 (2015).
44. Hu, W. *et al.* Stepwise protein folding at near amino acid resolution by hydrogen exchange and mass spectrometry. *Proc Natl Acad Sci USA* **110**, 7684–9 (2013).
45. Fajer, P. G., Bou-Assaf, G. M. & Marshall, A. G. Improved sequence resolution by global analysis of overlapped peptides in hydrogen/deuterium exchange mass spectrometry. *J Am Soc Mass Spectrom* **23**, 1202–8 (2012).
46. Althaus, E. *et al.* Computing H/D-exchange rates of single residues from data of proteolytic fragments. *BMC Bioinformatics* **11**, 424 (2010).
47. Kusunoki, H., MacDonald, R. I. & Mondragon, A. Structural insights into the stability and flexibility of unusual erythroid spectrin repeats. *Structure* **12**, 645–56 (2004).
48. MacDonald, R. I. & Cummings, J. A. Stabilities of folding of clustered, two-repeat fragments of spectrin reveal a potential hinge in the human erythroid spectrin tetramer. *Proc Natl Acad Sci USA* **101**, 1502–7 (2004).
49. Yan, Y. *et al.* Crystal structure of the repetitive segments of spectrin. *Science* **262**, 2027–30 (1993).
50. Krezel, A. & Bal, W. A formula for correlating pKa values determined in D<sub>2</sub>O and H<sub>2</sub>O. *J Inorg Biochem* **98**, 161–6 (2004).
51. Englander, J. J., Calhoun, D. B. & Englander, S. W. Measurement and calibration of peptide group hydrogen-deuterium exchange by ultraviolet spectrophotometry. *Anal Biochem* **92**, 517–24 (1979).
52. Woods-Jr., V. L. & Hamuro, Y. H. 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* **37**, 89–98 (2001).
53. Woods-Jr., V. L. Methods for identifying hot- spot residues of binding proteins and small compounds that bind to the same. in *US Patent # 6,599,707*. (USA, 2003).

54. Hamuro, Y. *et al.* Dynamics of cAPK Type IIb Activation Revealed by Enhanced Amide H/2H Exchange Mass Spectrometry (DXMS). *J. Mol. Biol.* **327**, 1065–1076 (2003).
55. Englander, J. J. *et al.* Protein structure change studied by hydrogen-deuterium exchange, functional labeling, and mass spectrometry. *Proc. Nat. Acad. Sci.* **100**, 7057–7062 (2003).
56. Pantazatos, D., Gessner, C. R., Woods, V. L. Jr. & Gilbert, G. E. Changes in the Factor VIII C2 domain upon membrane binding determined by hydrogen-deuterium exchange MS. *Biochem J* **461**, 443–51 (2014).
57. Garcia, R. A. *et al.* Molecular interactions between matrilysin and the matrix metalloproteinase inhibitor doxycycline investigated by deuterium exchange mass spectrometry. *Mol Pharmacol* **67**, 1128–36 (2005).
58. Cravello, L., Lascoux, D. & Forest, E. Use of different proteases working in acidic conditions to improve sequence coverage and resolution in hydrogen/deuterium exchange of large proteins. *Rapid Commun Mass Spectrom* **17**, 2387–93 (2003).

## Acknowledgements

This work is dedicated to the memory of Dr. Virgil L. Woods. We thank David Wemmer, Pat Jennings, David Baker, Russell Doolittle, Philip Bourne, Halbert White, and Karin Maxson for support and guidance. We thank Pavel Pevzner, Neil Jones, and Bryant Forsgren for their guidance in the design and development of the precursor of the HR-HDXMS stability profile algorithm, Eleftheria Mavridou for assistance in preparation of the figures and manuscript, Spiro Pantazatos, Ruby I. MacDonald and Michel Alain Cuendet for their valuable comments and review of the manuscript. This work was supported by NIH grant CA099835 (VLW) and by grants from the University of California BioStar and LSI programs, grants S97-90, S99-44, L98-30 (VLW). The LOEWE excellence program of the state of Hessen (Germany) supported this work (GB). We gratefully acknowledge funding from the Hellenic American Foundation (DP).

## Author Contributions

S.B. performed NMR and HDXMS experiments on SNase. J.S. performed the fitting for the NMR data and its analysis. D.P. designed, performed and analyzed HDX experiments and wrote the MatLab code. C.G. assisted in the development of the MatLab code and performed computational validation studies with the NMR and HDXMS data. V.L.W. conceived, co-designed and supported experiments and conducted data interpretation. T.J.W. provided insight in manuscript preparation and supported the study. G.B., W.S. and D.P. prepared the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at doi:[10.1038/s41598-017-03922-3](https://doi.org/10.1038/s41598-017-03922-3)

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017