ACS Partner Journal

Journal of the American Society for Mass Spectrometry

pubs.acs.org/jasms

Account & Perspective

Computational Modeling of Molecular Structures Guided by Hydrogen-Exchange Data

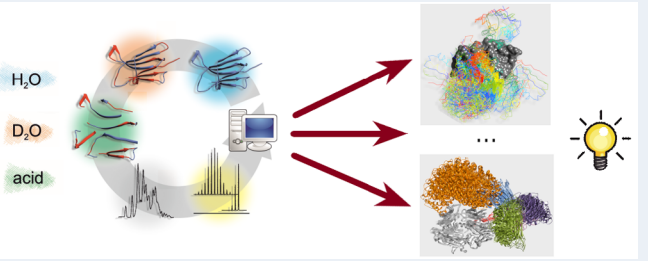
Didier Devaurs,* Dinler A. Antunes, and Antoni J. Borysik*

Cite This: *J. Am. Soc. Mass Spectrom.* 2022, 33, 215−237

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: Data produced by hydrogen-exchange monitoring experiments have been used in structural studies of molecules for several decades. Despite uncertainties about the structural determinants of hydrogen exchange itself, such data have successfully helped guide the structural modeling of challenging molecular systems, such as membrane proteins or large macromolecular complexes. As hydrogen-exchange monitoring provides information on the dynamics of molecules in solution, it can complement other experimental techniques in so-called integrative modeling approaches. However, hydrogen-exchange data have often only been used to qualitatively assess molecular structures produced by computational modeling tools. In this paper, we look beyond qualitative approaches and survey the various paradigms under which hydrogen-exchange data have been used to quantitatively guide the computational modeling of molecular structures. Although numerous prediction models have been proposed to link molecular structure and hydrogen exchange, none of them has been widely accepted by the structural biology community. Here, we present as many hydrogen-exchange prediction models as we could find in the literature, with the aim of providing the first exhaustive list of its kind. From purely structure-based models to so-called fractional-population models or knowledge-based models, the field is quite vast. We aspire for this paper to become a resource for practitioners to gain a broader perspective on the field and guide research toward the definition of better prediction models. This will eventually improve synergies between hydrogen-exchange monitoring and molecular modeling.



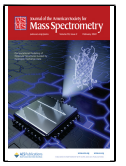
INTRODUCTION

Hydrogen−deuterium exchange (HDX) is a natural process by which hydrogen atoms in a molecule are exchanged with deuterium atoms of the solvent.[1,2] As this process is influenced by the three-dimensional shape of the molecule, it has been leveraged in experimental structural biology to obtain valuable information about molecular structures in solution.[3,4] However, contrary to other experimental techniques, such as cryo-electron microscopy or X-ray crystallography, monitoring HDX cannot fully reveal the three-dimensional structure of a molecule because of the lack of characterization of the structural determinants of HDX.[5] On the other hand, HDX experiments provide a valuable window into protein dynamics.[6,7] As a result, a strong interest has emerged in combining HDX monitoring with other experimental techniques that only produce information about molecular shapes, such as cryo-electron microscopy.[8] This idea fits within a larger trend, as it is now common to combine data from different experimental techniques within so-called *integrative modeling* approaches.[9−11]

The versatility of HDX monitoring has led to numerous applications for the analysis of protein structure and conformational changes, as well as protein folding and interactions.[12−14] The use of HDX monitoring has provided invaluable benefits to the study of challenging molecular systems, such as large protein complexes, glycoproteins, membrane proteins, or intrinsically disordered proteins.[15−25] Additionally, HDX experiments have had a tremendous impact on drug discovery and drug development, where they have helped characterize numerous biopharmaceuticals.[12,26,27] Testament to the success of HDX-related research is the significant growth of this field in the past decade, with a monthly rate of publication that has doubled between 2012 and 2020.[28]

Despite the obvious benefits of monitoring HDX for structural studies, HDX data are frequently difficult to interpret.[3,29] Raw HDX data are traditionally converted into *protection factors*, which are then often visualized on a protein *heat map* built using a structural model reported in the protein databank (PDB), when available.[5,30] Another classical way of making HDX data easier to interpret is to collect them for two

Received: November 5, 2021
Revised: December 29, 2021
Accepted: January 4, 2022
Published: January 25, 2022



different states of a protein (e.g., wild type vs mutant, or unbound vs bound to a ligand) and analyze relative differences in HDX, in so-called *differential HDX* experiments.[31−34]

One of the most promising ways to leverage experimental data would be to use them to guide computational modeling tools (e.g., for molecular docking or molecular dynamics) within integrative modeling approaches.[10,11] This would allow tackling the curse of dimensionality that plagues the computational modeling of large molecular systems. HDX data are a prime example of the kind of experimental data practitioners can benefit from in this context. However, HDX data are often only used in a qualitative fashion,[35] for example, when the aim of the HDX experiments is to corroborate three-dimensional structures produced by molecular docking[36,37] or molecular dynamics (MD) simulations.[34,38,39] Conversely, MD simulations are sometimes performed to elucidate structural properties inferred from HDX data[40−46] or to complement a HDX analysis by probing shorter time scales.[47−50] Additional examples of studies in which HDX data are used in combination with MD simulations, especially for membrane proteins, have been surveyed elsewhere.[51]

Looking beyond such qualitative approaches, this paper aims to survey various recent paradigms within which HDX monitoring experiments and computational modeling studies are performed in a synergistic fashion. In other words, we focus on research where experimental HDX data have been used to quantitatively guide computational molecular modeling. Our objective was to exhaustively survey the research carried out within this scope during the past 15 years, although we sometimes mention older work.

In general, guiding computational modeling with experimental HDX data requires defining a *HDX prediction model* to formalize a relationship between global and/or local (structural or other) properties of molecules and the level of HDX they undergo. Numerous such models have been proposed, but none of them has yet been widely accepted by the HDX community. Early work connecting HDX mechanisms with molecular structure, in the 1970s, was based on accessibility or penetration models. A common view was that solvent-accessible hydrogen atoms at the molecule's surface were exchanged more rapidly than buried hydrogen atoms. In other words, protection from HDX was believed to be positively correlated with burial, i.e., negatively correlated with solvent penetration. However, it is now widely accepted that atom burial is far from being the primary factor characterizing HDX.[15,52] Indeed, atoms involved in hydrogen bonds at the surface can be exchanged as slowly as deeply buried atoms. In the 1980s, HDX was mostly interpreted as being positively correlated with solvent accessible surface area. More recently, participation in hydrogen bonds has often been recognized as one of the strongest determinants of protection from HDX.[52,53] However, approaches that consider only hydrogen bonding to explain protection do not generalize well. In addition, cases have been documented where HDX protection has nothing to do with hydrogen bonding and is mostly due to hydrophobic clustering.[54] Overall, the quest for accurate HDX prediction is still ongoing, as various models have shown limitations in recent evaluation studies.[52,55]

In what follows, after summarizing the basic principles of hydrogen exchange monitoring experiments, we will present the various paradigms that have been proposed to use HDX data in computational studies of molecular structures. To start with the simplest paradigm, we will present approaches in which HDX data have been used to qualitatively guide computational modeling. Then we will move on to more sophisticated paradigms that rely on a HDX prediction model to quantitatively leverage HDX data. Each paradigm will be presented using the following template: (i) rationale, definition and history of the HDX prediction model(s) involved, (ii) overview of the computational method(s) in which HDX data have been leveraged, (iii) brief list of concrete applications. The first two paradigms involve a single structural feature of molecules: solvent accessibility or hydrogen bonding. The next paradigm is based on electrostatic properties of atoms. The two following paradigms involve various combinations of structural features, one being the widely used phenomenological approximation of HDX protection. Finally, we switch to completely different paradigms, from fractional-population models to knowledge-based approaches.

HYDROGEN EXCHANGE MONITORING

Hydrogen−deuterium exchange (HDX) is a chemical phenomenon in which hydrogen atoms in biomolecules are exchanged with deuterium atoms in D_2O solvent.[56,57] Intuitively, the extent to which different hydrogen atoms are subjected to exchange should be directly influenced by their accessibility to the solvent, such that assessing HDX could be seen as a way to characterize protein structure. In practice, however, using HDX monitoring for direct structural characterization has proved to be a considerable challenge due to the diverse nature of structural features that apparently act in concert to orchestrate HDX.[58,59]

The use of HDX to study biomolecules was first described in the 1950s.[1] However, biophysical approaches to characterize HDX only began to materialize in the 1970s with nuclear magnetic resonance (NMR) spectroscopy, thanks to its ability to exploit differences in the magnetic properties of ^1H and ^2H. Empirical characterization of near-neighbor effects in the 1990s further enhanced the appeal of HDX-NMR as it enabled the intrinsic exchange rates (k_{ch}) of backbone amide protons to be calculated.[60−62] Knowledge of k_{ch} allowed protein structural effects on HDX to be quantified from observed exchange rates (k_{obs}) and reported as protection factors (P) describing the extent to which the local structure of an amino acid corresponds to a random coil: $P = k_{ch}/k_{obs}$.

HDX-NMR found many applications throughout the 1990s, especially in areas of protein folding where it was used, for example, to characterize folding intermediates or misfolded protein conformations with potential links to disease.[63,64] The ability of HDX-NMR to characterize highly excited states with low populations along with advancements such as CLEANEX-PM (which enabled the characterization of fast exchange times) heightened the popularity of HDX-NMR throughout the 1990s and early 2000s.[65,66]

Mass spectrometry (MS) emerged in the 1980s as an alternative to NMR to characterize isotope exchange.[67,68] HDX-MS exploits the mass difference between ^1H and ^2H, allowing the extent of HDX to be assessed by accurate mass determination for any biomolecule. The foundations of modern HDX-MS were then established in the 1990s with the development of continuous labeling methods,[69−71] which remain the most widely used HDX-MS workflow today.

An interest in understanding the function of complex biomolecules such as multicomponent protein complexes and membrane proteins has seen HDX-MS become the dominant method in contemporary HDX monitoring.[19] This is due to

https://doi.org/10.1021/jasms.1c00328
J. Am. Soc. Mass Spectrom. 2022, 33, 215−237

the many advantages of MS over NMR, including lower restriction on protein mass, improved sensitivity, as well as enhanced throughput with regard to acquisition times, sample preparation, and analysis.[30] The straightforward workflow of continuous labeling HDX-MS also greatly facilitates automation, further enhancing throughput. Commercial availability of HDX-MS in the 2010s underpinned the rising importance of the technique for both academic and commercial practitioners. This has resulted in a surge of HDX-MS related research articles over the past decade.[28]

In a typical continuous labeling HDX-MS experiment, concentrated protein stocks in H_2O-based buffers are first diluted into D_2O to initiate the exchange of labile protons.[25,72,73] Aliquots of the protein solution are then taken at specified time points, and the exchange reaction is quenched by mixing with ice-cold acid. Quenched protein samples are then subjected to acid proteolysis, typically performed online using columns packed with pepsin. The resulting peptides undergo separation by reversed-phase chromatography, followed by mass determination. The extent of HDX in different protein regions can then be determined by characterizing the change in mass for each peptide over a range of isotope exposure times (Figure 1).

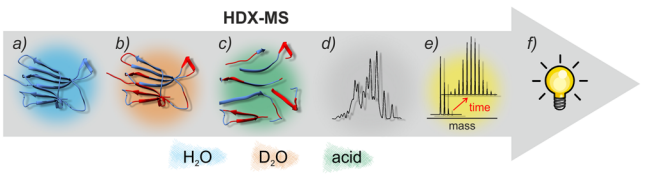


Figure 1. General workflow for continuous labeling HDX-MS. (a) Concentrated protein stock in H_2O-based buffer is prepared. (b) Protein sample is taken and diluted into deuterated buffer to initiate HDX. (c) Aliquots of protein in D_2O are taken at different isotope exposure times, and the reaction is quenched by diluting into ice-cold acid with the protein then digested online. (d) Peptides from acid proteolysis are desalted and then separated by reversed-phase chromatography. (e) Masses of the separated peptides are determined by MS to evaluate the mass relationship of different isotope exposure time. (f) Relative isotope incorporation of different peptide regions in the protein can be used to inform on protein structure/function.

Despite its clear predominance in recent years, HDX-MS is hampered by its low resolution and the absence of an established method to characterize exchange rates for individual amino acids (as with NMR). In bottom-up HDX-MS, resolution is limited to the size of proteolytically cleaved peptides, which vary in length around 5−20 amino acids depending on the extent of acid proteolysis. While fragmentation methods such as electron transfer dissociation (ETD) or electron capture dissociation (ECD) have been successfully incorporated into HDX-MS workflows, and can in some instances yield data resolved to the level of individual amino acids, limitations in ETD/ECD (such as the requirement for multiply charged ions) impose constraints on the applicability of the technique.[74,75] Although nonlinear programming (NLP) has also been applied to find a kinetic solution to HDX-MS data resolved at the residue level, this has proven challenging due to under-determination, as optimized exchange rates vary considerably depending on initial guess values for k_{obs}.[76,77] Nevertheless, recent developments in this area have demonstrated the possibility of validating NLP outcomes a priori, providing a more robust strategy to model exchange rates in HDX-MS data.[78,79] Note that other ideas have been suggested to obtain HDX-MS data at the amino acid level.[80−83]

An additional and often overlooked limitation of HDX-MS stems from the phenomenon of extraneous exchange, which encompasses both back exchange and forward exchange.[25] Extraneous exchange has the effect of compressing the dynamic window of HDX-MS data through the loss and gain of isotope during all chromatographic phases.[59,72] The consequence is that measurable isotope levels do not mirror those obtained in solution, and the effect can be significant, with more than 50% lost isotope not uncommon. Extraneous exchange also affects the character of the data because the magnitude of isotope that is lost or gained is unique to every peptide.[84]

Unlike HDX-NMR where quantitative metrics such as protection factors or free energy of HDX can be reported, the inability to quantify residue exchange rates by HDX-MS relegates it to reporting less meaningful qualitative outputs (Figure 2). A range of qualitative HDX-MS metrics are

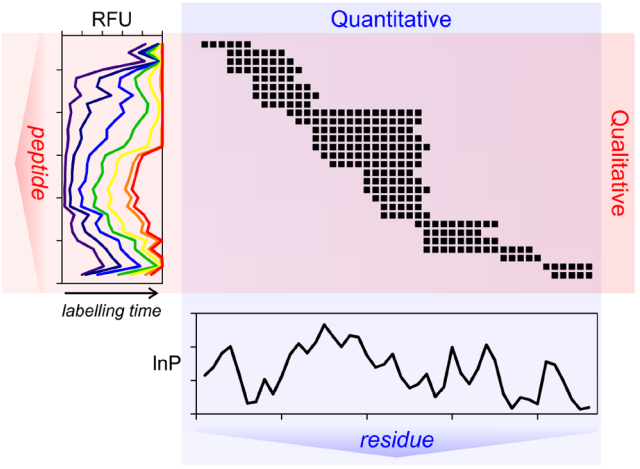


Figure 2. Relationship between resolution and HDX metrics: The central panel depicts part of a typical HDX-MS peptide map with each square representing an amino acid. HDX-MS is normally limited to qualitative metrics such as the relative fractional uptake (RFU) imposed by peptide resolution. Quantitative HDX metrics such as the natural logarithm of protection factors (ln P) require residue-resolved information normally only accessible by NMR.

typically reported, the most common being comparative and based on the change in mass (Δmass) between two experimental samples, normally calculated from the sum of all exposure times. The change in relative fractional uptake (RFU) is another common HDX-MS metric which normalizes the observed mass change to the maximum allowable change for each peptide. RFU calculations are possible due to preferential extraneous loss of isotope from amino acid side chains, such that the maximum number of deuterons for any peptide scales predictably with the number of amino acids. Noncomparative HDX-MS metrics such as the absolute mass or RFU require additional control experiments to quantify extraneous exchange for each peptide and are difficult to interpret meaningfully in isolation, such that they are used less frequently than difference outputs. The advantage of reporting comparative data is that it overcomes the requirement for control experiments, since it can be assumed that the degree of

https://doi.org/10.1021/jasms.1c00328
J. Am. Soc. Mass Spectrom. 2022, 33, 215−237

extraneous exchange will be identical for both the experimental and control data sets. However, it is instructive to note that uncorrected difference data have the potential to mislead, as the magnitude of any difference between control and experimental peptides vanishes to zero when back exchange is significant.

To facilitate the interpretation of HDX-MS data and better understand the conformational origins of changes in HDX, practitioners frequently overlay these data onto available 3-dimensional structures. This requires a nominal depiction of HDX rates at the residue level, which is typically performed through simple data averaging. The different reporting of HDX-NMR and HDX-MS data, as well as the range of available HDX metrics, impose additional considerations in the use of HDX data when modeling molecular structures. The predominance of comparative metrics in HDX-MS makes this technique particularly amenable to model protein assemblies, especially those involving binary interactions, which is the focus of the next section.

QUALITATIVE USE OF HDX DATA AS RESTRAINTS

Even without a HDX prediction model, it is possible to leverage HDX data to guide the computational modeling of molecular structures. However, as this can only be done qualitatively, the challenge is how to parametrize HDX data so that they can be utilized in a computational platform.

Computational Methods. In one of the first examples of qualitative use of HDX for protein modeling, HDX-MS data were converted into distance restraints to filter the results of docking simulations.[85] This was done by assuming that protected regions on two monomers should be located within a distance of less than 10 Å in a binary complex. Using this method, Anand et al. were able to reduce an initial 100000 docking solutions to just 15, allowing them to propose a structure for the complex. This work highlighted a natural synergy between HDX and molecular docking (Figure 3), which subsequently drove the development of other such computational platforms. Moreover, this synergy is even greater with HDX-MS, as docking algorithms only require information on differences between bound and unbound systems, which is readily available through the comparative HDX-MS metrics mentioned in the previous section.

As a more recent example, Pandit et al. use information derived from HDX-MS to guide protein−protein docking for epitope mapping.[87] HDX-MS experiments performed on an antibody−antigen complex allow for the identification of residues that are unlikely to belong to the epitope part of the antigen. These residues are then made ineligible for binding through a penalty term in the scoring function during computational docking performed with ZDOCK.[88] Using these HDX-derived constraints is shown to improve docking results by producing more tightly clustered binding modes.

As another example, Roberts et al. compare two ways of integrating experimental HDX-MS data into rigid-body protein−protein docking by trying to reproduce a given molecular complex.[89] In the first approach, HDX-MS data are used to filter an ensemble of binding poses obtained through unbiased docking by transforming these data into a set of distance constraints. More specifically, the presence of a minimum number of amide nitrogens in the protein−protein interface is enforced for peptides showing different HDX behaviors between the bound and unbound states. In the second approach, HDX-derived distance constraints are

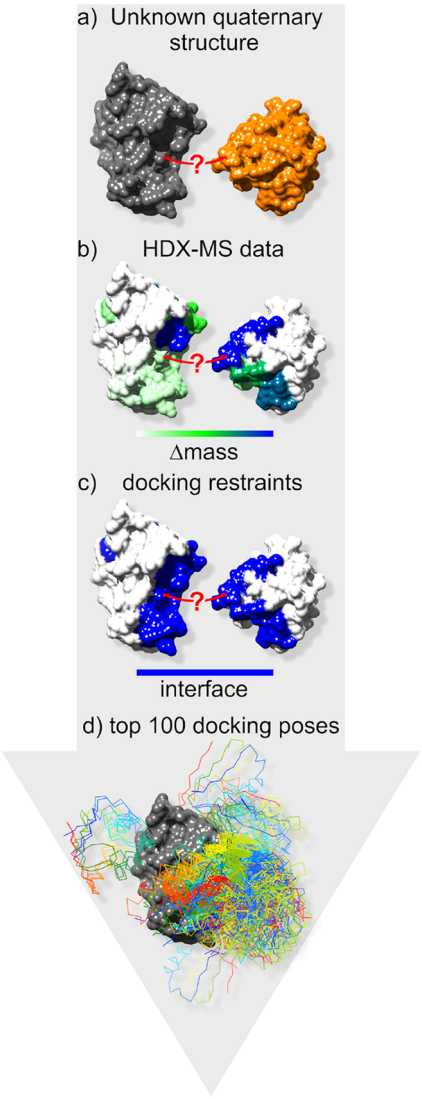


Figure 3. Macromolecular structure determination by HDX-MS-guided docking:[86] (a) Monomeric proteins forming a binary complex where only the structures of the respective monomers are known. (b) Information on the binding interface (projected onto each monomer) obtained by differential HDX-MS where the Δmass reports on differences between the bound and unbound proteins. (c) HDX-MS Δmass reduced to binary information on the binding interface to guide molecular docking. (d) Example poses from a docking simulation guided by HDX-MS restraints demonstrating potential for ambiguity in these approaches.

incorporated as a favorable potential in the scoring function of the docking tool to bias its search. Results of this comparison show that the ensemble-filtering method is superior to the search-biasing method.

Following a different approach, Rey et al. have developed Mass Spec Studio,[9] a platform that can extract modeling restraints from HDX-MS data directly for use with the docking tool HADDOCK.[90] The specificity of HADDOCK is to encode a broad range of biophysical data into ambiguous interaction restraints (AIRs) for restraint-based flexible protein−protein docking. Using biophysical data, HADDOCK can determine which residues are likely to be at the protein interface. The method involves sophisticated optimization procedures that can deal with ambiguity, which is especially

https://doi.org/10.1021/jasms.1c00328
J. Am. Soc. Mass Spectrom. 2022, 33, 215−237

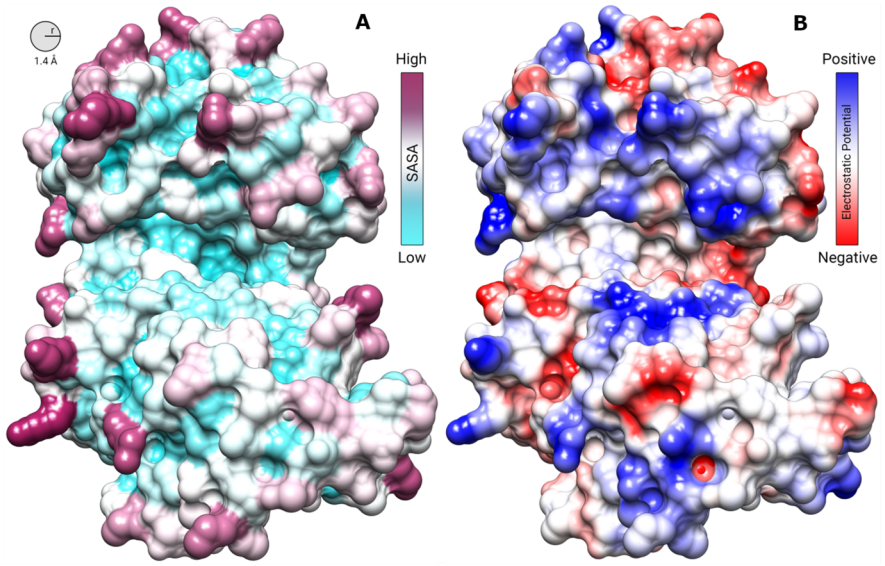


Figure 4. Protein surface analysis. Surface representations of a mitogen-activated protein kinase (with PDB code 3HEC). (A) Colors indicate a per-residue estimation of the solvent accessible surface area (SASA), ranging from cyan (low SASA) to maroon (high SASA). SASA predictions were obtained using UCSF Chimera, with a probe radius of 1.42 Å. (B) Colors indicate a per-residue estimation of the electrostatic potential over the protein surface, ranging from red (negative) to blue (positive), as computed by the Coulombic surface coloring of UCSF Chimera.[102]

relevant for HDX as it cannot distinguish between residues that are affected by binding or allosteric effects. Ambiguity is managed by designating some residues as active (i.e., known to be involved in protein interactions) and others as passive (i.e., possibly involved in interactions).[91] Mass Spec Studio allows one to define qualitative criteria for the selection of active and passive residues. Then HADDOCK uses solvent accessibility to automatically identify active and passive residues in docking outcomes.

More recently, Merkle et al. have developed another customized version of HADDOCK, following the idea that protein−protein docking can be improved by exploiting experimental HDX-MS data.[92] In their approach, AIRs are also defined using differential HDX between bound and unbound states. Interestingly, not all restraints have to be satisfied, but violations lead to penalties in the docking score.

Given the high complementarity between HDX-MS and molecular docking, it is likely that further developments will continue to be made in this area. Nevertheless, limitations with regard to the assignment of active amino acids from peptide data in HDX-MS need to be addressed. Docking simulations guided by HDX-MS can also frequently result in the generation of many structurally distinct models (Figure 3). Ranking these models may require more dynamic methods utilizing a greater proportion of HDX data by involving more sophisticated HDX prediction models.

Applications. Using HADDOCK, Snijder et al. generate structural models of complexes involving Kai proteins.[93] From the HDX-MS data collected on bound and unbound Kai proteins, they derive AIRs (associated with lists of active residues) that can drive HADDOCK into producing models where KaiB binds either to the CI or CII domain of KaiC. In this work (based on an integrative strategy involving native-MS, ion-mobility MS, HDX-MS and molecular docking), they conclude that the CII domain is the most likely binding site.

Eron et al. use an in-house HDX-guided docking approach to model degrader-induced ternary complexes.[94] Such complexes are formed when a small molecule (the degrader) promotes the binding of a target protein and the E3 ligase, which leads to ubiquitination and then degradation of the target by the proteasome. Eron et al. generate thousands of conformations for a ternary complex of interest and then filter these conformations using a HDX score based on differential HDX-MS data between bound and unbound proteins. They show that using the HDX score allows for the selection of conformations that are consistent with the experimental data, which is not necessarily the case for conformations ranked first by the regular docking score.

Finally, it is important to note that molecular docking is not the only modeling approach that can benefit from the integration of qualitative HDX-based restraints; molecular dynamics is another example. In recent work, Martens et al. have used HDX-MS and MD simulations together to study lipid-modulated conformational changes in membrane proteins.[95] HDX-MS allows them to identify which lipid mix influences the conformational dynamics of a given membrane protein. MD simulations are then performed to uncover the mechanistic aspects of the protein−membrane interactions. Finally, HDX-MS is used to qualitatively validate these interactions by monitoring protein mutants in the relevant lipid mix.

SOLVENT ACCESSIBILITY

We will now move on to quantitative paradigms that rely on the definition of HDX prediction models. The first such paradigm is based solely on the concept of solvent accessibility.

HDX Models. Solvent-accessible residues at the protein's surface are known to be, in general, less protected from HDX than residues in the protein's core. Indeed, the exchange of core hydrogens is limited by the slow rates of global unfolding and solvent penetration.[52] Therefore, rapid HDX rates are often observed for surface residues, with even sometimes a continuum of exchange from a protein's core to its surface.[96−98]

The solvent-accessible surface area (SASA) of an amide group is most often evaluated using probe-based techniques

https://doi.org/10.1021/jasms.1c00328
J. Am. Soc. Mass Spectrom. 2022, 33, 215−237

(Figure 4A). In this context, it is sometimes found to correlate well with HDX rates,[99,100] but not always.[52,53] For example, a lack of correlation is observed for the globular proteins thrombin and CheB, with SASA predicted from crystal structures.[99] On the contrary, a good correlation is observed between SASA and the number of amides exchanged after 2 min of HDX-MS for the nonglobular protein IκBα, especially for peptides in surface loops.[101] However, this correlation is not observed for the backbone SASA but only for the total residue SASA, which includes the residue side chain, and thus indirectly accounts for the packing density around its amide group (in addition to its solvent exposure).

Instead of being evaluated using a single-crystal structure, the SASA of amide groups has often been derived from all-atom MD simulations, with better results. For example, following this approach, Sheinerman and Brooks find a good agreement between predicted and experimental protection factors obtained for segment B1 of streptococcal protein G.[103] More recently, Shan et al. have used a 27 μs-long all-atom MD simulation of epidermal growth factor receptor kinase to evaluate the correlation between the HDX experienced by each amide group and its SASA, calculated as an average over a sequence of consecutive snapshots spanning 1 μs of the simulation.[104]

Following a different approach, Ma and Nussinov evaluate solvent accessibility without using a probe-based method.[105] They derive so-called NH solvation factors (or solvent exposure ratios) from an MD simulation, by calculating the average number of water molecules that are in the proximity of each backbone amide hydrogen atom. Then, they assess the correlation between these solvation factors and experimental HDX-NMR protection factors.

Petruk et al. adopt a different strategy, while combining all the ideas mentioned above.[106] For each amide group, they calculate (i) the number of water molecules in its first solvation shell and (ii) its SASA averaged over the frames of a 100 ns MD simulation of the mitogen-activated kinase ERK2 protein. Petruk et al. observe good correlations between the values of these two parameters (when they are aggregated at the peptide level) and the number of exchanged hydrogens for peptides monitored in previous HDX-MS experiments.[106] However, a limitation of this study is that the HDX data are analyzed at only one time point. Another potential issue is that values of both parameters increase with peptide size, which might facilitate the observed correlation.

Eventually, a drawback of all SASA-based HDX prediction models is the difficulty to define the granularity at which SASA should be evaluated. Indeed, averaging SASA over all atoms of a residue only provides a very coarse estimate of the HDX protection experienced by an amide hydrogen. On the other hand, considering the SASA of this amide hydrogen alone is generally not sensitive enough to make accurate predictions.[107]

Computational Methods. SASA-based evaluations of HDX rates have mostly been derived from crystal structures, NMR ensembles, or MD simulations. In a very different approach, Marsh et al. use the HDX protection factor of a single residue in a qualitative manner to define a quantitative restraint associated with SASA.[108] Then they use their ENSEMBLE software to assign weights to a population of pregenerated protein conformers. These weights are iteratively adjusted to minimize a sum of pseudoenergy terms evaluating the goodness-of-fit between predicted (as ensemble averages over the conformer population) and experimental values for several structural features serving as restraints. In a later study, Marsh and Forman-Kay expand their approach by using the HDX protection factors of all amide groups to define a larger set of SASA-based restraints.[109]

More recently, Zhang et al. have used HDX-MS data to guide an iterative homology-modeling procedure.[110] After each modeling round, they evaluate the best obtained model by comparing its secondary and tertiary structures to the experimental HDX-MS data by classifying peptides into four categories based on their HDX behavior. Correlations are assessed by plotting the time-averaged deuterium uptake of each peptide against its SASA, calculated as an average of the backbone SASA of all its residues. In regions of disagreement, the structural template is manually adjusted before performing another round of homology modeling and so on, until the best match to the HDX-MS data is achieved.

Applications. Using their ENSEMBLE software, Marsh et al. study the unfolded state of the *Drosophila* drk N-terminal SH3 domain, which naturally exists in equilibrium with its folded state.[108] They obtain conformer ensembles featuring a mix of native and non-native local structures.[109]

MD-based approaches have also proven useful. For example, Ma and Nussinov investigate structural models for intermediate aggregates involving amyloid β peptides, which have been linked to neurodegenerative diseases, such as Alzheimer's disease.[105] In another application, Petruk et al. analyze differences in the structures and dynamics of the phosphorylated active and unphosphorylated inactive states of the mitogen-activated kinase ERK2 protein, both in the presence or absence of bound ATP.[106]

More recently, Zhang et al. have used their HDX-guided homology modeling methodology to obtain structures for the reduced and oxidized states of a diheme cytochrome *c* protein from *Heliobacterium modesticaldum*.[110] Their analysis shows that the oxidized form of this protein is more flexible. They hypothesize that the more compact structure of this protein's reduced state improves its function, which relates to electron transfer.

HYDROGEN BONDING

HDX Models. Going beyond solvent-accessibility models, HDX protection has been classically interpreted as being due to the presence of hydrogen bonds (H-bonds) involving amide hydrogens.[111] Note that these can be H-bonds with a protein backbone or side chains (whether native or non-native ones) as well as H-bonds with water molecules (Figure 5). Indeed, Skinner et al. argue that H-bonds formed with water molecules held in place by protein interactions, such as those observed in crystal structures, can block HDX.[58] More generally, HDX is often interpreted as directly monitoring the H-bond content of specific protein states.[112]

Following this paradigm, Ma and Nussinov attribute protection from HDX solely to H-bonding and assess it by calculating two terms: (i) the number of H-bonds between an amide group and water oxygens, and (ii) the number of H-bonds between an amide group and backbone oxygens.[113] Using this prediction model, they compare experimental HDX protection factors reported in the literature with protection factors derived from MD simulations as ensemble averages.

On the other hand, Anderson et al. have argued that trying to explain the protection experienced by some solvent-exposed amide hydrogens in terms of steric blocking due to crystallographic water molecules is not realistic.[114]

https://doi.org/10.1021/jasms.1c00328
J. Am. Soc. Mass Spectrom. 2022, 33, 215−237

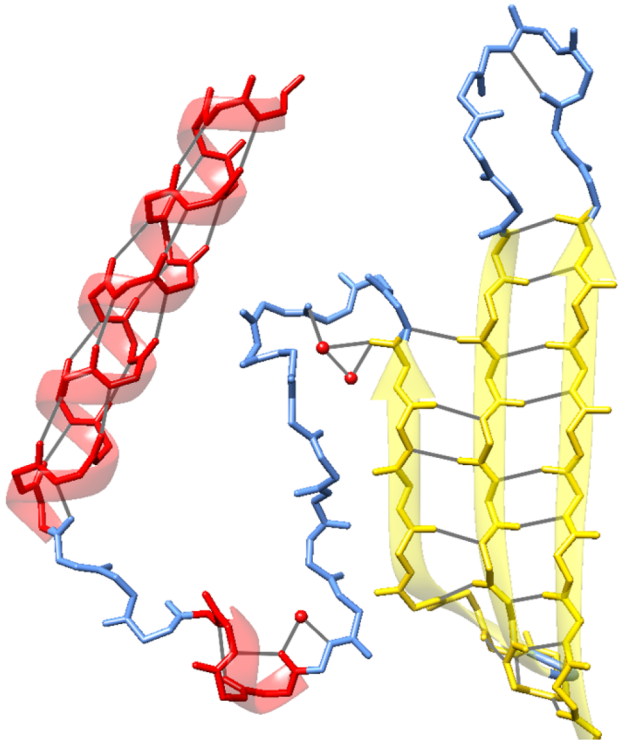


Figure 5. Illustration of the role of hydrogen bonds (shown as gray lines), which protect participating hydrogens from HDX. This stick representation of the main-chain atoms of an arbitrary protein segment includes different secondary structure elements (highlighted in transparent cartoon), such as α-helices (in red), β-sheets (in yellow), and unstructured coil regions (in blue). Red spheres represent oxygen atoms belonging to water molecules that might be "trapped" in specific regions of the protein and might even contribute to HDX protection. Image rendering and hydrogen bonds prediction were produced with UCSF Chimera.[102]

In a very detailed study, McAllister and Konermann perform a 1 μs all-atom MD simulation of ubiquitin to assess correlations between H-bonding (as well as other structural factors) and available HDX-NMR protection factors (i.e., ln P).[115] They conclude that the HDX behavior of 57 amide groups (out of 72 existent in ubiquitin) can be explained by the presence of main-chain H-bonds (for 42 protected residues), the presence of side-chain H-bonds (for five protected residues), a low solvent accessibility (for three protected residues), or none of the above (for seven unprotected residues). To test the robustness of their results, they replicate them with different force fields and water models. While trying to explain the HDX behavior of the 15 remaining residues, they refute the idea that HDX protection might be due to the presence of crystallographically identified water molecules or partially immobilized water networks (so-called H-bonded water cages). Eventually, they do not provide any explanation for the HDX behavior of these 15 residues.

To directly quote McAllister and Konermann, the results of their analysis can be summarized as follows:[115]

1. H-bonding always leads to HDX protection; this includes H-bonds to backbone carbonyls, side chains, as well as bifurcated H-bonds.
2. For NH sites that are not H-bonded, low SASA values are often (but not always) associated with HDX protection.
3. A lack of H-bonding at solvent-accessible amides does not imply that the corresponding sites are unprotected; instead, many of these amides are characterized by ln $P \gg 0$.

In subsequent sections, we will see that most HDX prediction models that consider H-bonding as a factor of protection against HDX do not consider it as being the only factor, but rather combine it with other important factors.

Computational Method. Most research based on considering H-bonding as the main source of HDX protection have involved MD simulations.[112,113,116] Notable exceptions are studies trying to connect HDX protection with crystallographic water molecules.[58]

Applications. Using their H-bonding-based HDX model, Ma and Nussinov analyze several structural arrangements of Alzheimer amyloid β42 peptides in polymorphic aggregates.[113] They show that arrangements involving a triple β-sheet motif for the peptide provide the best fit to experimental HDX protection factors.

Skinner et al. evaluate the denatured state produced by an MD simulation of a variant of protein G.[112] They show that this denatured state contains too many H-bonds when assessed against free energy values derived from experimental HDX-NMR data. Indeed, these data suggest that this state contains no stable H-bond and is highly solvated. Therefore, this study highlights the need to improve existing force fields with respect to H-bonding and backbone hydration.

In an attempt to correct this overestimation of the number of intramolecular H-bonds, Sung evaluates the effect of including the dielectric screening effect of the electronic polarization.[117] This study involves scaling down the peptide atomic charges and running several MD simulations of an α-helical peptide and a β-hairpin peptide. The results demonstrate that the number of intramolecular H-bonds is then reduced.

Finally, in a study of cytochrome c, Scrosati et al. use HDX-MS data and MD simulations to explain differences in deuteration patterns between the Fe(II) and Fe(III) species.[116] Their analysis reveals that the increased deuteration experienced by the central portion of Fe(III) cytochrome c cannot simply be explained by the reversible rupture of the distal M80−Fe(III) bond because this rupture does not result in a significant decrease in H-bonding. Instead, this increased deuteration could be due to changes in population distribution: a so-called 4-coordinate state showing significant decrease in H-bonding is likely to be more populated for the Fe(III) species than for the Fe(II) species.

■ ELECTROSTATIC CALCULATIONS

To explain the fact that even solvent-exposed hydrogens can sometimes exchange very slowly, other protein properties have been investigated, such as electrostatic effects (Figure 4B).

HDX Model. Anderson et al. have extensively studied the solvent-accessible amide hydrogens of *Pyrococcus furiosus* rubredoxin that are not involved in H-bonds with a backbone carbonyl oxygen, according to a crystal structure of this protein.[118] They observe that these amide hydrogens experience vastly different exchange rates, with one of them exchanging nearly a billion-fold slower than the others. The range of these exchange rates clearly shows the limitations of the solvent accessibility model of HDX protection. Anderson et al. hypothesize that these exchange rates could correlate with

https://doi.org/10.1021/jasms.1c00328
J. Am. Soc. Mass Spectrom. 2022, 33, 215−237

the electrostatic solvation free energies of the corresponding amide groups.[118] Their argument is that, beyond nearest neighbor effects observed between amino acids, tertiary interactions can produce an electrostatic environment influencing HDX of amide groups.

Indeed, numerous studies have shown that electrostatic interactions can modulate HDX.[119−128] For example, Avbelj and Baldwin have reported a strong correlation between available HDX rates and electrostatic free energies calculated using partial charges and a continuum solvent.[129] Hernández et al. have reported the connection between electrostatic stabilization and an example of general base catalysis within a protein.[130] Shaw et al. have shown that increasing the net negative charge at the surface of a protein can increase the overall number of unexchanged hydrogens (as observed at several time points by HDX-MS) without affecting its secondary or tertiary structure.[131]

The HDX model proposed by Anderson et al. is underpinned by the fact that most solvent-exposed amide groups at the surface of a globular protein have lower acidities than model peptides because of the presence of the low-dielectric protein interior.[118] To evaluate this model, they perform electrostatic calculations using nonlinear Poisson−Boltzmann (i.e., continuum dielectric) methods. They simulate the peptide anions that form during the hydroxide-catalyzed HDX reaction by removing the amide hydrogen from each relevant residue. Anderson et al. show that the correlation between energies and HDX rate constants obtained via HDX-NMR is good when using the CHARMM22 electrostatic parameters for atomic charge and radius, with an internal dielectric value of 3. Using the parameters of PARSE or AMBER parm99 instead, or changing the dielectric value to 2 or 4, leads to a much poorer correlation.

Then, Hernández et al. study the solvent-exposed amide hydrogens of three additional proteins: human FK506 binding protein, human ubiquitin, and chymotrypsin inhibitor 2 (CI2) from barley.[132] To assess electrostatic solvation free energies, they carry out nonlinear Poisson−Boltzmann calculations on high-resolution crystal structures of these proteins, using the parameters of CHARMM22, OPLS-AA, PARSE, AMBER parm99, and ff03. Using OPLS or CHARMM parameters (which are very similar), good correlations are observed between the HDX rate constants of amide hydrogens and their electrostatic solvation free energies, while using PARSE or AMBER yields poorer correlations. However, as their study involves only crystal structures, Hernández et al. recognize that conformational flexibility at a protein's surface is likely to deteriorate these correlations.

To broaden their analysis, LeMaster et al. determine the HDX rate constants of all backbone amide groups in ubiquitin under physiological conditions through HDX-NMR.[133] For electrostatic calculations, they use a linear approximation of the Poisson−Boltzmann continuum dielectric model, with the CHARMM22 nonpolarizable parameters for atomic charges and radii. By performing these calculations on ubiquitin conformers extracted from two available ensembles produced by NMR-restrained MD, LeMaster et al. obtain (as ensemble averages) predicted HDX rate constants that they can compare to experimentally determined rate constants. Note that both ensembles are expected to constitute Boltzmann-weighted representations of ubiquitin's native state and, thus, to characterize its inherent flexibility. Predictions obtained for amide hydrogens that are highly exposed to solvent are more accurate when using these ensembles than when using a high-resolution crystal structure. Predictions obtained for less exposed amides are almost as accurate, but only when using the NMR relaxation-restrained ensemble and not the residual dipolar coupling-restrained ensemble.

HDX protection factors for a protein's residues can be derived from experimentally observed HDX rates and theoretical HDX rates for unstructured model peptides.[60−62] However, Hernández et al. argue that calculations performed for these model peptides are not accurate enough because they do not account for the conformational flexibility of these peptides.[134] Indeed, electrostatic calculations have indicated that the HDX rate of an amide group is strongly dependent on the relative orientation of the adjacent amino acid residues.[125,128] Using a protein coil library to model the Boltzmann-weighted distribution of conformations in unstructured peptides, Anderson et al. have predicted HDX rates for a few implicitly solvated model peptides.[135] They show that, despite the extreme variations in predictions between various backbone conformations of a given peptide, differential HDX rates can be accurately predicted for model peptides as averages over conformer ensembles. Then, Anderson et al. predict similar differential HDX rates using side-chain conformers of model peptides.[136] As the distribution of conformations determines HDX rates, these studies point to the limitations of the classical intrinsic HDX rates of amino acids.[60−62]

Anderson et al. also show that, contrary to what is commonly believed, it is incorrect to calculate the protection factor of a given peptide by adding the protection factors of its residues,[136] i.e., using the classical additive approach.[60−62] This lack of validity of the classical approach was already discussed in previous work.[129,137] Another limitation of theoretical HDX rates is that they were assessed for each residue using only its direct neighbors, therefore neglecting local cooperativity and domain dynamics.[138] Finally, related work reveals that errors might appear when neglecting dielectric shielding produced by electronic polarizability in standard nonpolarizable force fields.[139]

In a different kind of study, Abdolvahabi et al. have also illustrated the importance of electrostatic effects on HDX protection.[140] First, they explain how the acetylation of lysine residues at the surface of a protein can increase HDX protection, despite decreasing secondary structure. Then, they show how certain free salt ions can reverse this protective effect, again without acting on the protein's structure. They also show that other salt ions do not have such an impact. Therefore, these salt-mediated effects would be challenging to account for in a HDX prediction model based on electrostatic mechanisms.

Finally, Dass et al. show that an electrostatic potential calculated with a hybrid mean-field approach can accurately estimate HDX protection.[141] They evaluate this approach on the human protein α-synuclein, in which no secondary structure element can explain differences in HDX protection along the backbone, as this protein is known to be intrinsically disordered. Results show a good agreement between predicted protection factors and protection factors derived from HDX-NMR data. Specifically, the highly acidic C-terminal tail of α-synuclein is the only region showing HDX protection.

Computational Methods. The HDX model proposed by Anderson et al. has been used to estimate HDX rate constants of static solvent-exposed amide groups from a single-crystal

https://doi.org/10.1021/jasms.1c00328
J. Am. Soc. Mass Spectrom. 2022, 33, 215−237

structure[132] or from a conformer ensemble produced by an MD simulation.[133] More recently, Anderson et al. have considered the fact that, even when using a single protein conformation, sources other than X-ray crystallography are available.[114] For that, they analyze four proteins: *P. furiosus* rubredoxin, barley CI2, human FK506 binding protein, and human ubiquitin. They show that crystal structures, and especially high-resolution ones, provide better predictions for HDX rate constants than NMR or homology models. They also show that the resolution of the crystal structure has a limited yet beneficial impact on the quality of these predictions.

Finally, Anderson et al. insist on the fact that a single conformation can provide statistically significant predictions, almost as accurately as a conformer ensemble.[114] Note that it should be expected that this conformation lies near or within the most highly populated region of the Boltzmann distribution.[142] This is important because in many applications it is useful to consider a single representative conformation of a protein in solution.

The model proposed by Anderson et al. has shown successes, but also limitations. For example, a recent evaluation study has found no correlation between the HDX rate constants of surface amide groups and their estimated electrostatic solvation free energies.[52] Electrostatic effects can obviously not explain HDX protection on their own, but they could contribute to more general approaches. Some obstacles to their integration in computational methods, however, are the computational cost of their assessment and their strong dependence on force-field parameters.[132] The main challenges that would have to be addressed are the imperfections of existing force fields in describing dielectric shielding, electronic polarizability, and salt-mediated screening.

Applications. The body of work reviewed by Hernández et al. shows that Poisson−Boltzmann electrostatic calculations of peptide acidity can predict HDX rates in some cases.[134] However, these calculations are too computationally expensive to be integrated as restraints in molecular simulations. On the other hand, molecular simulations are routinely employed to try and generate the Boltzmann-weighted conformational distribution of a protein's native state. In this context, the methodology of Hernández et al. can be useful to assess whether a given conformer ensemble is consistent with the properly weighted Boltzmann distribution.[134]

For example, by predicting HDX rate constants as ensemble averages for three conformer ensembles of ubiquitin, Hernández et al. assess their ability to constitute a proper Boltzmann-weighted ensemble of ubiquitin's native state.[142] Only two ensembles show good agreement between the computationally derived data and the experimental data. From the third ensemble, Hernández et al. extract a sub-ensemble having a significantly better consistency with the experimental data.

Using this methodology, Hernández et al. analyze model ensembles produced by unconstrained as well as NMR-constrained MD simulations of ubiquitin.[143] They provide structural evidence for the divergence in predictions of the conformational dynamics of a functionally important region. More specifically, they show that the tight turn preceding the major site of proteasome-directed polyubiquitylation (Lys 48) undergoes conformational changes more rarely than previously predicted.

Hernández et al. also reevaluate the stability of MD-predicted conformational basins of bovine pancreatic trypsin inhibitor (BPTI), and show it had been overestimated for the two most populated basins.[143]

PHENOMENOLOGICAL APPROXIMATION OF HDX PROTECTION

HDX Model. In this popular approach, protection from local or subglobal HDX (i.e., the natural logarithm of the protection factor of residue i is attributed to a combination of amide burial and H-bonding within secondary structure elements. The first version of this model was based on a phenomenological equation involving the number of contacts between residue i and surrounding residues (N_i^c) as a proxy for burial or packing density, as well as the number of H-bonds formed by the amide group of residue i (N_i^h).[144] These two terms were then weighted by coefficients (β^c and β^h, respectively) estimated by fitting experimental HDX protection factors collected for several proteins whose native structures were known:

$$\ln P_i = \beta^h N_i^h + \beta^c N_i^c \quad (1)$$

In the second version of this model, the terms of the phenomenological equation approximating the natural logarithm of the protection factor of a residue are evaluated slightly differently.[145] For each residue, amide burial is estimated by counting heavy atoms in the vicinity of its amide nitrogen; H-bonding is estimated by counting native H-bonds involving its amide nitrogen. Including non-native H-bonds did not appear to make any difference, as the authors focused on protein native states. An additional term estimating SASA was evaluated, but it was not retained, as it did not improve results. The weights of the two terms in the phenomenological equation are estimated from seven proteins for which experimental ln P and crystal structures are available. The procedure involves minimizing the root-mean-square deviation between the experimental protection factors and protection factors predicted as ensemble averages from unbiased 1 ns native state MD simulations of these proteins.

In work by another group, the phenomenological equation is evaluated on a small bacterial cytochrome *c* protein.[146] The logarithm of each amide protection factor is calculated from the ensemble produced by a 3 ns unbiased MD simulation under native conditions, as well as from an available NMR ensemble and a crystal structure. The fit to experimental HDX protection factors is slightly better when using the MD ensemble than when using the NMR ensemble or the crystal structure. In addition, it appears that replacing one of the two terms in the phenomenological equation by similar terms derived from the MD simulation has little impact on the goodness-of-fit. More specifically, Kieseritzky et al. replace the simple H-bond count by H-bond occupancy (i.e., the fraction of MD simulation time during which a bond is formed), the average H-bond survival time, or the inverse of fluctuations in H-bond length.[146] With slightly less success, they also replace the number of contacts by the inverse of backbone atom fluctuations, as it is also related to packing density.

More recently, Xu et al. have used one of these variants in which the basic H-bond count is replaced by the inverse of the root-mean-square fluctuation (RMSF) of H-bond lengths in an MD simulation.[147] They also propose a modification of the phenomenological equation itself, which consists of defining

https://doi.org/10.1021/jasms.1c00328
J. Am. Soc. Mass Spectrom. 2022, 33, 215−237

two sets of weights in the equation: one for the slow exchange regime and one for the rapid exchange regime.

Going back to the second version of the phenomenological equation, Devaurs et al. have refined the way its terms are assessed.[148] Indeed, some studies have shown that this model can sometimes yield poor correlations between experimental and predicted protection factors.[52] For best results, only H-bonds maintaining secondary structure elements are considered, whether they are native or not. More precisely, one should count main-chain acceptor oxygens within a cutoff distance of 2.4 Å from the amide hydrogen of residue i, ignoring oxygens in residues $i - 2$, ..., $i + 2$. Packing density is estimated as the number of contacts experienced by residue i, i.e., the number of heavy atoms within a cutoff distance of 6.5 Å from its amide hydrogen, ignoring atoms in residues $i - 2$, ..., $i + 2$ (Figure 6). The weights of the H-bonding and packing density terms are 2 and 0.35, respectively.

Finally, Wan et al. have recently added a cooperativity term to the phenomenological equation to compensate for correlations between the usual terms.[150] They determine optimal parameters for this model using Bayesian inference to fit ultralong MD simulations of BPTI and ubiquitin in their native states.

Computational Methods. In the first version of the phenomenological approximation approach, a protein's conformational space was explored using Monte Carlo sampling biased by a pseudoenergy function.[144] As this function evaluates differences between the predicted and experimental HDX data, these data work as experimental restraints on the simulation. Note that, in practice, HDX protection factors are predicted as averages over several replicas of the Monte Carlo simulation.

Biasing a computational simulation has been the most common way of leveraging HDX protection factors predicted by the phenomenological approximation approach. This has been done with different computational simulation techniques and biasing strategies. In the second version of this approach, HDX data are combined within an empirical force field, to act as low-resolution restraints in MD simulations.[145] Again, several replicas of the protein are used to predict protection factors. Their root-mean-square deviation to experimental protection factors is used as pseudoenergy term in a detailed force field.

In more recent work, Radou et al. have modified this approach to predict HDX-MS profiles of peptides as Boltzmann averages using frames extracted from an unrestrained equilibrium MD simulation.[151] They claim that predicting HDX-MS data from an MD ensemble, instead of a crystal structure, produces a better fit to experimental data. Adhikary et al. similarly show that correlation between experimental protection factors of peptides and predicted ones is better when deriving them from an MD ensemble than from a crystal structure.[152] Note that this requires to let the estimated protection factors converge, which translates into running MD simulations for around 200 ns.

This common dichotomy between ensemble and single conformation has recently been questioned by the work of Devaurs et al.[153] Indeed, they show that using a single conformation obtained through sampling can produce a better fit to experimental HDX-MS data than using an MD ensemble. This work involves a coarse-grained conformational sampling approach that integrates sampling-based motion-planning algorithms, initially proposed in the field of robotics.[148] Also,

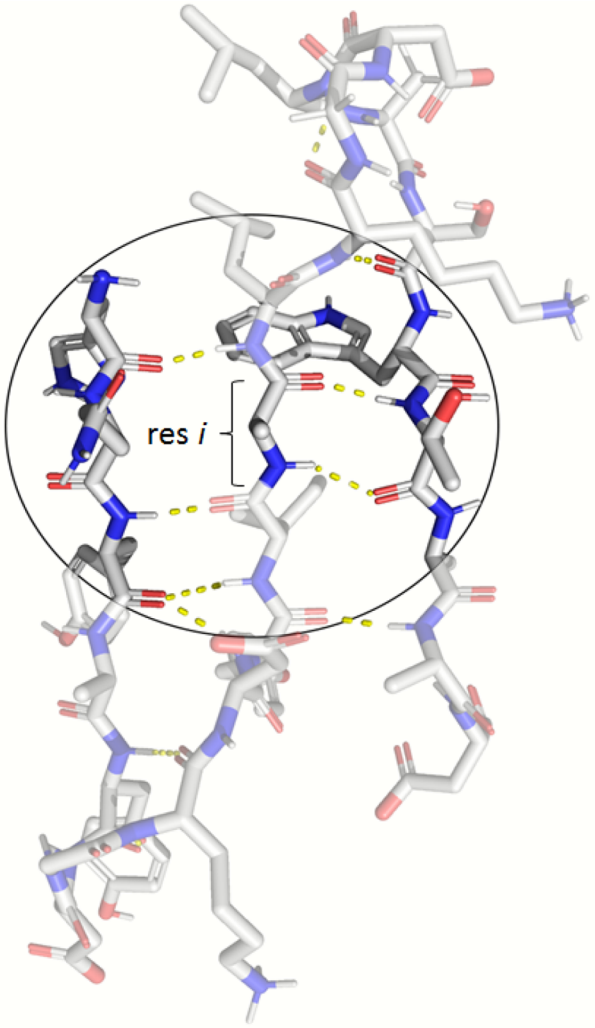


Figure 6. Illustration of the phenomenological approximation of HDX protection. In this stick representation of an arbitrary protein segment, atom colors indicate different atom types: oxygen (red), nitrogen (blue), hydrogen (white), and carbon (gray). Hydrogen bonds involving amide hydrogens are indicated by dashed yellow lines. The phenomenological equation is based on assessing hydrogen bonding and packing density for each residue i within a defined distance (represented with a black circle), excluding its immediate neighbors (e.g., residues $i - 2$, $i - 1$, $i + 1$, and $i + 2$). Image rendering and hydrogen bond predictions were produced with PyMOL.[149]

contrary to previous work, this approach does not rely on a constant biasing of conformational sampling, but on an adaptive biasing scheme based on an incremental protocol.[154] At the first step of this incremental process, conformational sampling is performed unbiased and the resulting conformations are filtered based on their fit to the experimental HDX data to select a starting point for the next step. Then, at every subsequent step, protein regions where the fit is the worst are sampled more heavily than others.

In another example of experimental guide without constant bias of a computational method, Borysik proposes the following for molecular docking.[86] After running a docking simulation, HDX-MS profiles are generated for all docking poses using the phenomenological expression and experimental peptide maps. Simulated HDX-MS difference data are then prepared for each monomer in a binary complex, and all

https://doi.org/10.1021/jasms.1c00328
J. Am. Soc. Mass Spectrom. 2022, 33, 215−237

docking poses are ranked on the basis of agreement between simulated and experimental HDX-MS difference data. As with other HDX methods for molecular docking previously discussed, this approach benefits from applicability with uncorrected HDX-MS data, with the added advantage of leveraging whole HDX-MS profiles, unlike AIRs. Allosteric effects induced by binding can also be considered by factoring in changes in HDX-MS profiles between unbound and bound conformations by normal-mode analysis. The method benefits from a high throughput and an ability to successfully identify native assemblies from just a single protein in the complex.

Leveraging the phenomenological expression, Harris et al. quantify the ability of experimental HDX-MS data to select the native fold of different proteins from a pool of decoys.[155] This study aims to assess the ability of HDX for "sample and select" in protein modeling where computationally generated structures could be ranked according to experimental data. Since HDX-MS profiles are simulated from single conformations, this approach provides much higher throughput than alternative approaches relying on lengthy sampling, such as MD or fractional-population methods. Harris et al. report high accuracy for the identification of native folds for monomeric proteins. However, they report low diagnostic power for homomeric protein multimers, which can be explained by the incapacity to differentiate between protein interfaces and the protein core of each chain in these assemblies. This problem can be overcome in heteromeric binary complexes, since HDX data can be recorded for each protein in the bound and unbound state with the resulting difference data highlighting interfaces.

Starting from a conformational ensemble generated by a molecular simulation method (MD, Monte Carlo, etc.), Bradshaw et al. propose an approach to reweigh conformations in this ensemble so that it fits experimental HDX-MS data.[156] Their approach, called HDX ensemble reweighing (HDXer), is based on applying a maximum-entropy bias to assign statistical weights to conformations in the initial ensemble, thereby adjusting populations in this ensemble. This allows ensemble-averaged protection factors or deuterium uptake values to fit experimental data with a certain level of uncertainty. In other words, HDXer can identify structural ensembles reflected in HDX-MS data, therefore offering mechanistic interpretations of these data. To derive protection factors from a single conformation, Bradshaw et al. use the phenomenological expression.[145] They evaluate HDXer on artificial data generated for TeaA, a binding protein characterized by two major conformational states. Then, they use HDXer on real experimental data to reweigh a conformational ensemble associated with the membrane protein LeuT. They also assess the robustness of HDXer by analyzing the impact of various sources of error and noise in the data.

Finally, also inspired by the maximum entropy principle, Wan et al. define a new restraint potential involving protection factors, which can be used to bias MD simulations and produce structural ensembles consistent with experimental data.[150] After running several MD simulations, Wan et al. construct multiensemble Markov state models and then use their Bayesian inference of conformational populations (BICePs) algorithm to find the set of conformational populations that best agree with the experimental data.

Applications. The phenomenological approximation of HDX protection has been used in numerous instances to help interpret HDX data with respect to a specific protein state or to bias conformational sampling. Using native HDX data to bias sampling allows one to determine structures characterizing the variability of a protein's native state, which includes even rare fluctuations. This was first done for human α-lactalbumin by Vendruscolo et al.[144] Then, Best and Vendruscolo determine an ensemble of structures representing the native state of CI2.[145] Khorvash et al. perform an HDX-restrained MD to analyze the conformational fluctuations of the human prion protein.[157] Through coarse-grained sampling of the complement protein C3d, Devaurs et al. obtain a conformation whose strong similarity to C3d's crystal structure confirms the lack of flexibility of its native state.[153]

HDX data predicted by unbiased computational simulations can be compared to experimental HDX data to analyze different protein states. For example, Radou et al. apply this approach to a hexameric viral helicase P4.[151] A comparison between the free and capsid-bound hexamers reveals that the free hexamer rapidly switches between closed and open conformations, whereas the capsid-bound hexamer only adopts a closed conformation. Adhikary et al. compare the inward-open and outward-open conformations of the neurotransmitter:sodium symporter LeuT within a phospholipid bilayer nanodisc.[152] They show that the protection factors of peptides predicted from MD simulations match well experimental HDX-MS data, except for pathway-lining regions in contact with lipids, which might reflect a limitation of the phenomenological expression.

Biasing a computational simulation with non-native HDX data can yield useful information on a non-native protein state. For example, Gsponer et al. produce an ensemble of structures to characterize a folding intermediate of the bacterial immunity protein Im7 that undergo significant reorganization.[158] Bemporad et al. study the initial steps of protein aggregation for the acylphosphatase from the archaebacterium *Sulfolobus solfataricus* and observe an increased flexibility of specific segments without local unfolding.[159] Xu et al. study the impact of phosphorylation on the activity of an enzyme, the α-D-phosphohexomutase.[147]

In general, (partially) biasing a computational simulation with HDX data allows generating atomic-resolution structural models for protein states previously undescribed or described only by low-resolution structural models. For example, Devaurs et al. produce a structural model for the unbound state of the ligand-binding domain of the vitamin D receptor, for which only structures of the bound state had been reported in the PDB, and show that it is more compact than the bound state.[154] Devaurs et al. also create an atomic-resolution structural model for the native state of the complement protein iC3b, thus resolving a previous disagreement between two contradicting medium-resolution models.[154]

Finally, Kihn et al. combine HDXer and enhanced-sampling MD to study the conformational heterogeneity of PhuS, the cytoplasmic heme binding protein from *Pseudomonas aeruginosa*.[160] They compare the native states of apo-PhuS and holo-PhuS (i.e., heme-bound PhuS) and explain differences observed in their respective HDX-MS data. These differences contradicted the fact that crystal structures of apo- and holo-PhuS were nearly identical; they were also too large to be attributed to heme binding or conformational fluctuations alone.[161] Using HDXer, Kihn et al. reweigh a conformational ensemble produced by aggregating several enhanced-sampling MD simulations, so that it fits the HDX-MS data collected for apo-PhuS. This analysis reveals that the native state of apo-

https://doi.org/10.1021/jasms.1c00328
J. Am. Soc. Mass Spectrom. 2022, 33, 215−237

PhuS encompasses several substates, some showing significant levels of unfolding.[160]

OTHER COMBINATIONS OF MOLECULAR FEATURES

As no single molecular feature has ever appeared to correlate well enough with HDX protection, the literature is rich with HDX prediction models combining several molecular features. The phenomenological expression proposed by Vendruscolo et al. was among the first ones,[144] but many have followed. It is interesting to note that some of these models do not involve "classical" structural features; others include features that are not even structural in nature.

HDX Models. Brand et al. evaluate the HDX rate constants of human ubiquitin's residues using NMR diffusion experiments.[162] They also assess correlations between the exchange rates of the six most-rapidly exchanging amide hydrogens and several molecular features: the RMSF and the H-bonding of amide hydrogens obtained from a 10 ns MD simulation, the SASA of amide groups derived from a crystal structure, the charges of amide nitrogens, and the orders of the amide N−H bonds estimated via quantum chemical studies. Eventually, they find that no single parameter provides a good correlation but that a bilinear model involving the amide RMSF and the N−H bond orders yields a decent correlation.

In a study combining rigidity analysis and solvent accessibility analysis, Sljoka and Wilson try to predict HDX-NMR protection factors in a protein.[163] Using conformations from an NMR ensemble, they assess the average local rigidity of the protein by determining the strength of H-bonds. Then, they calculate the SASA of each amide group as an ensemble average, and combine it with the rigidity information. Unfortunately, this analysis is performed in a qualitative manner and excludes α-helices.

Pester et al. derive HDX rate constants from an MD trajectory using intrinsic exchange rates, the probability for each amide H-bond to be open, and the concentration of the exchange catalyst.[164] The amide H-bond is considered open if the amide hydrogen is at a distance greater than 3 Å from the closest intrahelical carbonyl oxygen in an MD frame. The local hydroxide concentration is defined as the product between the bulk hydroxide concentration and the ratio of local to bulk water concentration. The local water concentration is calculated by counting water molecules within 7 Å of the amide hydrogen in MD frames.

Khakinejad et al. perform HDX-MS experiments on various ions of a model peptide.[165] Then, they run MD simulations of these ions and extract as representative conformations low-energy structures whose collision cross sections match experimental ones. They find a good correlation between the residue-level experimental deuterium incorporation and data derived from these representative ion conformations using their so-called "two-distances" model. This model involves the distance between a charge site and a less basic carbonyl group, as well as the distance between this carbonyl group and the exchange site, with a threshold of 7 Å. Their rationale for this choice is that HDX happens through a relay mechanism involving both a charge site and a less basic site.

In a first modification of their model, Khakinejad et al. propose to determine the contribution of each residue to the deuterium uptake by calculating its hydrogen accessibility score.[166] This score is defined based on sums of inverse distances between charge sites and carbonyl oxygens (i.e., initial incorporation sites) as well as inverse distances between carbonyl oxygens and exchange sites. Then, in a second modification of their model, Khakinejad et al. propose to account for the protection from the protein fold.[167] For that, they scale the score of each exchange site by its solvent excluded surface area.

Following a different approach, Mohammadiarani et al. propose an empirical model combining, via a power function, the SASA of amide hydrogens and the distance of each amide hydrogen to the closest polar atom that is not in a neighboring residue.[55]

Recently, Marzolf et al. have defined a model combining solvent exposure and local flexibility to determine the level of HDX protection of residues.[168] Solvent exposure of a given residue is assessed through a combination of relative SASA and amide neighbor count (i.e., number of oxygen atoms neighboring the amide proton), while local flexibility is determined by combining H-bonding energy (of the backbone amide group) and order score (which measures residue-resolved disorder), all being provided by the Rosetta modeling tool.[169] Marzolf et al. use these four terms to assess how native-like a residue conformation is by scoring their deviation from the distributions expected for HDX-protected residues, as determined by an analysis of crystal structures of proteins listed in the Start2Fold database.[170] Local sequence context is also accounted for by assessing these HDX-based scores using intrinsic exchange rates.[62] The overall score for a protein conformation combines the four HDX-based scores within the traditional Rosetta score.

Finally, Peng et al. predict protection from HDX using a model based on hydrogen bonding and backbone burial.[171] As their predictions are derived from conformational ensembles generated by their in-house coarse-grained modeling approach, called Upside, the HDX model is also defined in a coarse-grained manner. Specifically, the burial level of an amide nitrogen is assessed through contributions from backbone heavy atoms and side-chain beads. The criteria for hydrogen bonding are borrowed from Persson and Halle, who considered nearby water molecules.[172]

Computational Method. Wu et al. have developed a tool, called DEXANAL, to derive the HDX rate constants of a protein's residues from a structural model.[173] The HDX protection factor of a given residue is calculated as the product of two quantities. First, the so-called backbone accessibility index of this residue involves the ratio between the SASA of its amide hydrogen atom in the protein's structural model and that in a random coil conformation. To account for local interactions, this ratio is multiplied by a negative exponential factor making it inversely proportional to the length of potential H-bonds. Second, the so-called residue accessibility index involves the ratio between the SASA of this residue in the protein's model and that in a random coil configuration.

Later, Wu et al. proposed an improved version of DEXANAL.[174] In this version, the ratio involved in the residue accessibility index is multiplied by a term assessing this residue's hydrophobicity to indirectly account for protein dynamics effects. Finally, Gogonea et al. have used snapshots extracted from an MD simulation, instead of a single static structure, to calculate HDX rate constants as ensemble averages.[175]

Marzolf et al. have developed a methodology for ab initio protein structure prediction using the Rosetta modeling platform.[168] They integrate four HDX-based scoring terms

https://doi.org/10.1021/jasms.1c00328
J. Am. Soc. Mass Spectrom. 2022, 33, 215−237

into the traditional Rosetta score to assess how native-like generated decoys are, based on their agreement with available HDX-NMR data. They show that using these terms improve native structure prediction for 38 small proteins from the Start2Fold database.[170] However, this approach is limited by the qualitative nature of the HDX data it involves. To address this limitation, Nguyen et al. have recently enhanced this approach to allow for the use of quantitative HDX data, namely residue protection factors obtained via HDX-NMR.[176] Linear regression is used to predict the four HDX-based scoring terms from protection factors. These predicted terms are then compared to those derived from generated decoys to score these decoys.

Applications. With the help of DEXANAL, Wu et al. have produced an all-atom model (the so-called solar flares model) of nascent high-density lipoprotein using experimental HDX-MS data to guide this model's refinement.[173] Using additional experimental data, Wu et al. have later proposed a better model, which they name the double superhelix model.[174] Then, Gogonea et al. have compared structural data derived from an MD simulation of this complex with the corresponding experimental data to validate their structural model.[175] Finally, Högel et al. have recently applied the model developed by Pester et al.[164] to the study of trans-membrane helices.[177]

FRACTIONAL-POPULATION MODELS

HDX Models. Adopting a totally different paradigm than the approaches presented so far, some models aim to predict HDX protection by estimating the free energy of exchange. For that, they assess the difference in free energy between the closed (i.e., folded, protected, exchange-competent) state and the open (i.e., unfolded, unprotected, exchange-incompetent) state of each amide group (Figure 7). Some studies suggest that these so-called fractional-population models provide more robust predictions than other models.[55]

Among the oldest such approaches, COREX is based on a hybrid structural−thermodynamic model of local unfolding in which each residue is either in a closed or open state.[178] In this approach, the protection factor of a residue is thus derived from the Boltzmann probability of finding it in an open conformation at thermodynamic equilibrium. In addition, for each residue, COREX allows calculating a stability value and an energetic cost of exposing it to solvent. Unfortunately, this approach has been reported to have a limited predictive power with respect to HDX protection factors.[52,133,139] Another limitation is that COREX cannot measure changes due to ligand binding or post-transductional modifications, such as phosphorylation.[106]

A more recent approach, called DXCOREX, uses the COREX algorithm to specifically predict the HDX protection factors of a protein's residues.[179] The protection factor of a residue is calculated as the ratio of the sum of probabilities of microstates in which it is folded over the sum of probabilities of microstates in which it is unfolded. In addition, COREX is modified so that the exchange rate of a residue is calculated based on the solvent exposure of its amide hydrogen instead of its amide nitrogen. Liu et al. evaluate DXCOREX by predicting HDX-MS data for 13 proteins with available high-resolution structural models.[179]

In a very different approach, Craig et al. derive HDX protection factors from coarse-grained structure-based model simulations (with implicit solvent) involving perfectly funneled

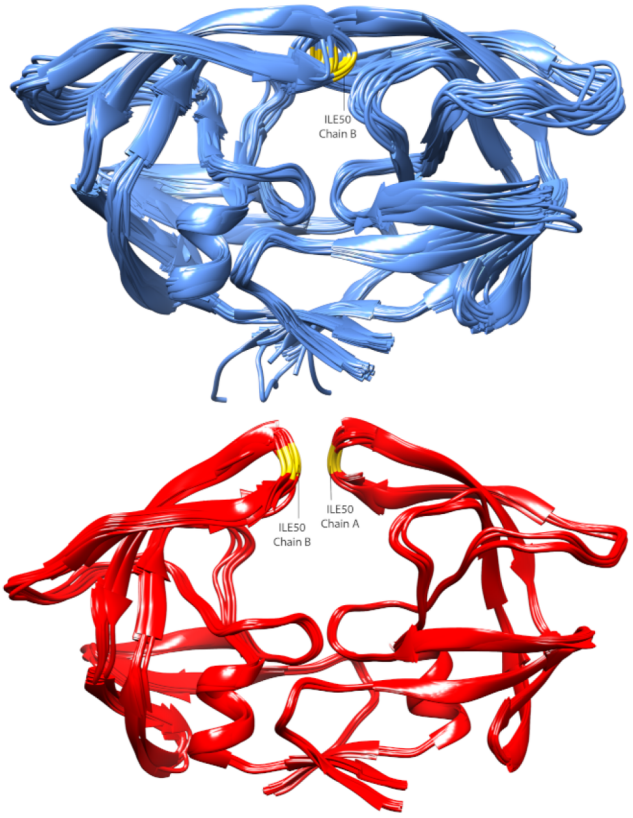


Figure 7. Illustrative representation of fractional-population models. Cartoon representation of ensembles of experimentally determined structures of HIV-1 protease. In solution, this protein can alternate between closed (blue) and open (red) conformations. The HDX protection for residues in the moving "flaps" (such as Ile50, depicted in yellow) will differ significantly between both states. Images produced with PyMOL.[149]

energy functions.[180] For that, they evaluate the probabilities for a residue to be in an open or a closed state, in each conformation of the generated ensemble. These probabilities are calculated by integration over a global reaction coordinate, using a weighted histogram analysis. They are based on the number of native contacts in which a residue is involved and on its H-bonding state, considering that residues are not represented with all atoms but in a coarse-grained fashion. The H-bonding state is assessed by comparing the distance between residues that form a H-bond in the native state and their distance in conformations of the ensemble. As observed in previous studies,[53,181] an increase of about 2−3 Å in distance between Cβ atoms is sufficient to allow for HDX. Using three proteins studied under native conditions (human ubiquitin, CI2 from barley and *Staphylococcal* nuclease), Craig et al. show that their predicted protection factors correlate well with experimental ones.[180] They also observe that the local environment of a residue has to be significantly distorted for it to enter an exchange-competent state.

Considering the role of water, Park et al. have assessed HDX protection by analyzing H-bonds in snapshots of an MD simulation.[107] They check whether an amide group forms H-bonds with other residues or with water molecules, as proxies to assess HDX incompetence or competence, respectively. The so-called "closed propensity" of this amide group is then the fraction of MD snapshots in which it forms H-bonds with

https://doi.org/10.1021/jasms.1c00328
J. Am. Soc. Mass Spectrom. 2022, 33, 215−237

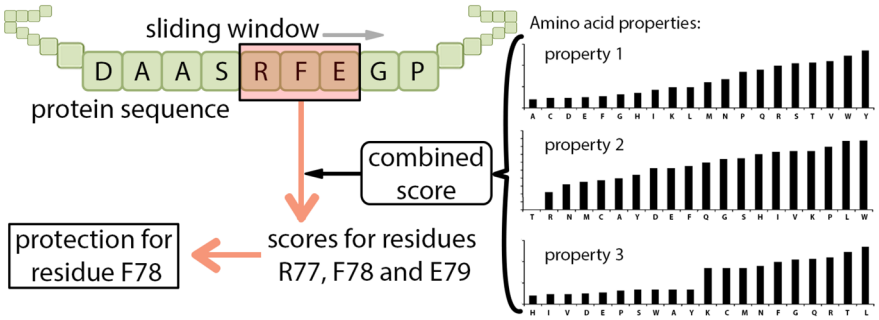


Figure 8. Illustrative representation of knowledge-based prediction models. Using various biochemical properties of amino acids, scores are calculated for all residues in a sliding window and then combined into a HDX protection factor for the residue at the center of the window. Sliding the window over the whole protein sequence produces HDX protection factors for all protein residues.

other residues; its so-called "open propensity" is the fraction of MD snapshots in which it forms H-bonds with water molecules. Protection factors are then derived from these statistics. Park et al. evaluate their model on several proteins,[107] and compare it to other existing models.[144,179,180] This comparison reveals that none of these approaches is clearly superior to the others.

Following yet another approach, Persson and Halle define the open state of an amide as a locally distorted conformation with two water molecules coordinated to this amide.[172] Their specific criterion is that the amide hydrogen should have at least two water oxygens within 2.6 Å. This criterion is made more stringent by additionally requiring that no other polar protein atom should be within 2.6 Å of the amide hydrogen, except within the same or adjacent residues. This definition is validated by analyzing a 262 μs-long MD simulation of fully solvated, native BPTI. Protection factors are estimated as the ratio of frames exhibiting the closed vs open state, and compared to experimental ones. The authors observe a good correlation, better than when using SASA and intramolecular H-bonding. However, the need for long MD simulations with explicit water molecules limit the practicality of their approach.

Finally, Mohammadiarani et al. propose a fractional-population model involving the SASA of an amide hydrogen and its distance to the closest polar atom.[55] As in other fractional-population models, protection is defined as the ratio of conformations in the closed state to conformations in the open state.

Computational methods. In COREX, a conformational ensemble is generated for a protein by a combinatorial algorithm exhaustively unfolding regions within a sliding window containing a few amino acid residues.[178] To each protein is therefore associated a statistical thermodynamic ensemble comprising various conformational microstates. Conformational energies are calculated using the mean field approximation of a Gibbs free energy functional, which allows evaluating the probability of each microstate. This functional involves a conformational entropy term and a solvation entropy term evaluated from changes in SASA. This solvation term was parametrized using calorimetric measures (of polar and apolar heat capacity contributions) characterizing the thermodynamics of unfolding for a limited set of globular proteins.

Despite being quite coarse, the COREX energy function has been shown to provide reasonable energy estimates.[182] Because of its combinatorial algorithm, the COREX method is very computationally intensive.[151] Therefore, in the more recent DXCOREX, a Monte Carlo sampling method is used to produce conformational ensembles in a more efficient manner.[179]

In general, coarse-grained simulations allow for a more extensive sampling of a protein's conformational space than all-atom simulations. Therefore, instead of studying only small fluctuations around the native state, Craig et al. can obtain information about large-scale unfolding transitions associated with HDX.[180] This is important because transitions between the closed and open states are exceedingly rare in all-atom simulations.[172]

Applications. COREX has been used to study allosteric binding effects[183−185] as well as the impacts on protein stability of pH[186,187] and temperature.[188,189] It has allowed energetic profiling of protein folds,[190] characterizing the determinants of fold specificity,[191] as well as more generally, describing models of folding[192] and pathological misfolding.[193] Finally, COREX has been applied to the description of protein cooperativity, intrinsic disorder, and evolutionary conservation of fluctuations.[194,195]

In a different approach, Hromić-Jahjefendić et al. have used the HDX model proposed by Park et al.[107] to correlate experimental HDX-MS data and MD simulation data.[196] They investigate the structure and dynamics of an enzyme, called dipeptidyl peptidase, which is important for the growth of *Porphyromonas gingivalis*, a pathogen responsible for periodontitis. When running MD simulations of this enzyme, they observe that the correlation between HDX-MS data and MD data improves over simulation time. From this, they conclude that this enzyme adopts a closed or semiclosed conformation in solution.

KNOWLEDGE-BASED PREDICTIONS

HDX Models. Strongly departing from the aforementioned approaches, there have been attempts to derive HDX data directly from the amino acid sequence of proteins, without using any structural information (Figure 8).

Tartaglia et al. propose a method that predicts the coefficients of the Fourier transform of the protection factor profile associated with a given protein.[197] This method involves an artificial neural network trained on 2000 protein structures for which protection factors were estimated using the phenomenological expression.[145] From this training, Tartaglia et al. note that the contribution of nonpolar residues to protection factor profiles is slightly larger than that of polar residues. Eventually, an evaluation involving 12 proteins whose protection factors were previously determined shows that the

https://doi.org/10.1021/jasms.1c00328
J. Am. Soc. Mass Spectrom. 2022, 33, 215−237

accuracy of the prediction is only around 58%. However, this sequence-based method generally produces better predictions than an alternative approach based on generating a structural model from a protein's sequence and deriving protection factors from this model using the phenomenological expression. As a side note, Tartaglia et al. show that the correlation between experimental protection factors and B factors is usually relatively weak.

Following a different approach, Dovidchenko and Galzitskaya investigate the possibility of predicting protection from HDX on seven proteins, but only in a qualitative manner: they perform a binary separation of residues between so-called protected and unprotected ones.[198] They compare their approach to that of Tartaglia et al. and show that it provides similar results. Then, using reported structures of 14 proteins, Dovidchenko et al. assess the predictive power of several structural parameters, which they calculate for each residue: the number of contacts with other residues, the H-bond energy, the secondary structure type, and the B factor.[199] They also assess the predictive power of equivalent parameters derived directly from protein sequences: the predicted number of contacts with other residues, the predicted probability of H-bond formation, the predicted secondary structure type, and the predicted probability of lack of secondary structure. The secondary structure data is produced for each residue of the tested proteins by the PsiPred program.[200] Typical contact and H-bonding data is obtained for the 20 natural residue types using a database containing 3769 structures collected from the PDB, therefore disregarding information on a specific residue's environment. The parameter providing the best predictions is the structurally derived H-bond energy. However, all parameters show a rather poor predictive power, especially the sequence-derived ones, but the structurally-derived B factor is the worst one. Note that Dovidchenko et al. do not try to combine several parameters in their predictions.[199]

More recently, Lobanov et al. have released an improved version of this method as a web server.[201] Again, the method involves calculating, for each amino acid type, the expected number of contacts and the H-bonding probability of its amide group. The novelty is that, to account for a residue's neighboring amino acids, an average of these statistics is calculated within a sliding window and associated with the central residue. Residues with a number of contacts or H-bonding probability above specific thresholds are predicted as being protected. After an evaluation involving 45 proteins with experimentally determined HDX behavior, it appears that the predictions of this web server are quite poor.

Claesen and Politis propose another method, called POPPeT, to predict protection factors based on protein motions that make amide hydrogens exchange-competent.[202] These protein motions are divided into four categories: local, unfolding (detected by adding denaturant), EX1 (i.e., unfolding detected by increasing pH level), and unfolding with EX1. Unfortunately, this type of information might be difficult to obtain. Therefore, Claesen and Politis also consider information derived from crystal structures: secondary structure (i.e., helix, β-sheet, or none), H-bonding (with a side-chain or main-chain oxygen, a water molecule, or none), and burial. They define a log−linear model similar to the phenomenological expression.[144] They show that POPPeT yields better predictions than COREX or the phenomenological expression, at least for *Staphylococcal* nuclease and equine cytochrome *c*.

With a very different objective in mind, Start2Fold has been curated by Pancsa et al. as an openly accessible and searchable database of HDX data reported in the literature.[170] It contains information about the investigated proteins and relevant experimental procedures, such as pulsed labeling, quench flow, or HDX-NMR. Protein residues are classified based on their level of protection (early, intermediate, or late) and/or stability (strong, medium or weak).

Computational Methods. Raimondi et al. have developed EFoldMine, a computational tool that can predict from a protein sequence which amino acids are likely to be involved in early folding events.[203] It is based on a support-vector machine (SVM) algorithm using secondary structure propensity and backbone/side-chain dynamics as features. It is trained on HDX data about early folding, collected through NMR pulsed labeling experiments, and available in the Start2Fold database.

Using a random forest classifier, Wang et al. propose a machine learning method that can predict whether a given residue will have undergone HDX, based on experimental conditions and protein structure.[204] From the Start2Fold database, they obtain information on the HDX status of residues, as well as experimental temperature and pH. From the protein structure, they extract features related to secondary structure, H-bonding, packing density, solvent accessibility, and Coulombic interactions. Their results show that the estimated strength of H-bonds is critical for prediction success.

PERSPECTIVE

As this paper illustrates, numerous approaches have been proposed to leverage experimental HDX data for computational protein modeling. The variety of approaches developed over the last 20 years is a testament to the complexity of predicting protein HDX behavior and the need for more fundamental research defining the structural determinants of HDX. The capacity to accurately generate biophysical signals from protein structures is the foundation of molecular modeling with many techniques such as Small Angle X-ray Scattering and NMR. As such links remain elusive for HDX, this represents a considerable barrier toward the development of effective HDX-based modeling strategies. Establishing the fundamental basis for protection from HDX has particular challenges because exchange rates are a characteristic of the protein ensemble, and therefore depend on protein motions both within and between local and global conformational minima. No single state should therefore be expected to fully capture the unique HDX signal of a protein. Indeed, the ability of HDX to provide insight into protein motions is one of its defining characteristics and is why integrative structural approaches involving the technique are so appealing. Despite the absence of well-defined structural origins for HDX, new and innovative applications of the method for protein modeling continue to be developed. This is an indication of the broad utility of the technique with diverse protein systems and of the enduring interest of practitioners to model molecular structures using HDX.

One of the simplest paradigms facilitating the direct use of HDX data for structural modeling is molecular docking. The determination of structures for interacting proteins is highly amenable to HDX, especially for binary complexes where experimental data can be acquired for each chain in the bound and unbound states. The ability to characterize protein monomers both in complex form and in isolation is significant because it permits the acquisition of difference data for each

https://doi.org/10.1021/jasms.1c00328
J. Am. Soc. Mass Spectrom. 2022, 33, 215−237

Table 1. Overview of All Paradigms under Which Experimental Hydrogen-Exchange Data Have Been Used to Guide the Computational Modeling of Molecular Structures

paradigm (and main references)	main characteristics
solvent accessibility[101,104−106,109]	intuitive interpretation, but weak correlation with HDX data, and challenges in defining SASA granularity
hydrogen bonding[58,112,113,115,117]	widely recognized as the main factor driving HDX protection, but not enough on its own to fully characterize protection
electrostatic calculations[118,132,133,140]	can explain the protection of solvent-accessible residues not involved in H-bonds, but force-field dependent and computationally intensive
phenomenological expression[144−148,150]	combines hydrogen bonding and atom burial; one of the most popular HDX prediction models, but can show poor correlation with HDX data
combination of molecular features[55,162−164,167,174]	can involve nonstructural features (e.g., bond order, catalyst concentration, or hydrophobicity), but often just increases model complexity
fractional-population models[55,107,172,179,180,182]	estimate the free energy of exchange by calculating the probabilities of protected vs unprotected states, but very computationally intensive
knowledge-based predictions[197,201,202,204]	derive HDX protection directly from protein sequence, but usually show very weak correlation with HDX data

chain, which is highly enriched with information on quaternary structure, and thus constitutes a footprint guiding the formation of the native assembly. Although docking techniques typically utilize HDX data qualitatively, and do not therefore rely explicitly on a HDX prediction model, they cannot be excluded from discussions on HDX-guided modeling, given their prevalence. From the simple definition of a penalty term or favorable potential in a scoring function[87,89] to the more sophisticated generation of ambiguous interaction restraints,[9,92,93] molecular docking is a prime example of how a computational tool can benefit greatly from being guided by HDX data. Nevertheless, this area may profit from developments in HDX prediction models to permit more extensive use of experimental data and ranking of the various docking poses.[86] The synergy between HDX and molecular docking could see methods emerge that are able to provide accurate models of protein assemblies where classical techniques have failed. However, these methods will need to address areas such as allostery, where interactions between proteins alter HDX kinetics in regions remote from the protein−protein interface, compounding interpretation or even producing counterintuitive consequences, such as a widespread increase in HDX instead of a more localized decrease.[205]

Focusing on quantitative paradigms, from early atom burial models to recent knowledge-based approaches deriving HDX data directly from protein sequence, the spectrum of HDX prediction models is vast and diverse (Table 1). As some of the fastest exchanging sites occupy the protein surface, early HDX prediction models sought to link protection from HDX with atom/residue burial in the native protein structure and aligned with early solvent penetration theories on HDX. However, the literature holds many examples where atom/residue burial cannot be correlated with experimental HDX data, and these theories are now generally considered too simplistic to fully account for protection from HDX.[15,52,118] They have not been abandoned completely, however,[206] and persist in combination with other factors in the form of related concepts, such as packing density. Another model in keeping with early solvent penetration theories is solvent accessibility, usually assessed as SASA via probe-based techniques, but also sometimes as solvent exposure ratios[105] or a combination of solvation-related terms.[106] While good correlations are sometimes observed between SASA and HDX protection factors, this is generally not the case, especially when SASA is assessed using crystal structures.[52,118] Although SASA-based prediction methods tend to perform better when derived from MD ensembles, this does not really redeem SASA-based methods but rather reaffirms that better HDX predictions are usually produced by MD ensembles than by crystal structures.[153]

Hydrogen bonding is often considered the main structural feature driving protection from HDX.[53,111,112] Indeed, the exchange competent state is typically depicted as emerging via H-bond breakage, such that the importance of H-bonds in orchestrating HDX kinetics is rooted in our understanding of the technique. HDX prediction models based solely on H-bonding have been proposed, but H-bonds are more commonly found in more elaborate models, combined with other structural features. One of the most well-known such models is the phenomenological approximation of HDX protection, in which H-bonding is combined with packing density.[144] This model uses weights that were determined empirically by fitting a limited library of experimental protection factors.[145] Originally conceived to identify rare fluctuations in proteins by MD, this approximation of HDX protection was never fully intended for modeling protein structures directly. Nevertheless, it remains one of the most popular HDX prediction models because of its accessibility and throughput, allowing HDX-guided generation of thousands of structures in relatively short time without advanced computing. The phenomenological model has benefited diverse applications, including analyzing inherent protein variability,[153,157] deciphering changes in protein conformation,[147,158,159] comparing different protein states,[151,152] and generating atomic-resolution structural models for undescribed protein states.[154] The ability of the phenomenological HDX approximation to successfully identify native folds based on HDX simulations generated from single structures is reportedly high, although it remains to be established if this ability is unique to this model.[155] While able to perform relatively well at deriving HDX data from MD ensembles[151,152] or coarse-grained simulation ensembles,[148,153] poor accuracy of the model has been reported when data are derived from crystal structures.[52] Several variants of this model have been proposed, but none of them yield any particular improvements in predictive power.[146,147,150] While H-bond parametrization has yet to yield an accurate HDX model, the involvement of H-bonds in dictating exchange rates is unquestioned, and the success of future HDX predictors will likely depend on an accurate representation of these structural features.

Some of the limitations of structure-based HDX prediction models lie with inconsistencies in the molecular features that govern protection from HDX. These inconsistencies originate from the various types of conformational change that promote exchange, including global unfolding, subglobal motions

https://doi.org/10.1021/jasms.1c00328
J. Am. Soc. Mass Spectrom. 2022, 33, 215−237

between local minima, and breathing motions within minima.[58] The paradigm referred to as the fractional population model was proposed to circumvent difficulties associated with defining structure-based HDX prediction models. It is based on the idea of estimating the difference in free energy between the HDX-competent and the HDX-incompetent states of each amide group. The most well-known approaches following this paradigm are COREX[182] and its newer variant DXCOREX,[179] which are based on a hybrid structural-thermodynamic model of local unfolding. Other approaches in which HDX predictions are based on evaluating protein folding landscapes also exist but utilize coarse-grained molecular simulations[180] or conventional MD[55,107,172] to calculate free energy differences. The various fractional population models are distinct in their definition of protected and unprotected groups, but as with other models, these definitions are linked to specific structural features in proteins. Nevertheless, fractional population models are expected to have the edge with regard to HDX calculations, since the molecular features on which the predictions are based are aggregated over large conformational ensembles intended to mimic the solution state of a protein. This argument is supported in a comprehensive comparison of different HDX predictors, in which fractional population methods were found to be performing best.[55] However, these results were established on the basis of how well each model could reconstruct experimental HDX-MS data for which no account of extraneous exchange was made, which may undermine them. Indeed, a different study of HDX predictors found the accuracy of fractional population methods to be largely consistent with other methods.[52] A limitation of computationally expensive methods such as fractional population models is poor throughput, which renders them unable to guide computational modeling tasks. HDX data are typically leveraged to either impose a bias during a conformational search[89,144,145,150,157] or filter conformations following unbiased sampling.[86,155] While there are opposing merits for either approach, or indeed adaptive biasing that can leverage the advantages of both,[154] all are unfeasible when computationally expensive ensemble generation is required to simulate data. Fractional population methods appear best suited to analyze specific molecular structures; it is unrealistic to expect these models to guide the computational exploration of a protein conformational space. Nevertheless, fractional population methods remain the only HDX prediction models to base their calculations on estimates of protein free energy landscapes.[207] In this regard, they share a unique and innate link with molecular events that drive the exchange reaction.

Attempts to derive HDX data directly from protein sequences have largely proved unsuccessful,[197,199,201] but these methods have been included in this article for completeness. On the other hand, the HDX database Start2Fold[170] has been used to develop machine learning algorithms that can predict the likelihood of amino acids being involved in early folding events.[203,204] This could form the foundation of methods capable of predicting HDX data from the primary structure, although the challenge facing knowledge-based predictions are still considerable and possibly exceed those faced by protein folding algorithms. The Start2Fold database raises an additional and important point, however, regarding an overall deficiency of adequately curated information on experimental HDX data for proteins with known structures. The success of future methods able to predict HDX data accurately will hinge upon the availability of these data. Unfortunately, many current models were developed using sparse data sets obtained more than 20 years ago, as well as more recent HDX-MS data largely lacking correction for extraneous exchange and thus unsuitable for model development. The appeal of HDX-guided protein modeling lies with its potential to uncover structures of proteins that escape conventional structural biology. Accordingly, a reliance on less-established methods to extract exchange kinetics may be required, with NMR limited to a relatively narrow window of protein systems. The ability of emerging HDX-MS optimization methods[78,82,208] or fast fragmentation techniques, such as electron transfer dissociation,[74,209] to provide this information may become especially useful as they permit an understanding of exchange rates for proteins of high biological importance which thwart classical structural methods. As the availability of experimental data usable for model development increases, so will the likelihood of an accurate HDX predictor and the effectiveness of HDX-based protein modeling. A recent survey of HDX practitioners revealed 96% to regard accurate protein modeling by HDX as a realistic aim, with 72% believing it to be feasible within the next ten years. (Data resulting from a poll carried out in 2021 by the International Society for HDX-MS (www.hdxms.net) as part of the Online Seminars in HDX-MS.) This will require continued cooperation between experimental and computational researchers working together on both fundamental and theoretical aspects of HDX, as well as sustained ambition of the HDX community.

AUTHOR INFORMATION

Corresponding Authors

Didier Devaurs − *MRC Institute of Genetics and Cancer, University of Edinburgh, Edinburgh EH4 2XU, U.K.*; orcid.org/0000-0002-3415-9816; Email: ddevaurs@ed.ac.uk

Antoni J. Borysik − *Department of Chemistry, King's College London, London SE1 1DB, U.K.*; orcid.org/0000-0002-4509-9580; Email: antoni.borysik@kcl.ac.uk

Author

Dinler A. Antunes − *Department of Biology and Biochemistry, University of Houston, Houston, Texas 77005, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/jasms.1c00328

Funding

D.D. is a cross-disciplinary postdoctoral fellow supported by funding from the University of Edinburgh and Medical Research Council (MC_UU_00009/2). A.B. is supported by funding from the Biotechnology and Biological Sciences Research Council (BB/R006792/1).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors thank Dr. Emanuele Paci for his feedback on an early version of this manuscript.

REFERENCES

(1) Englander, S. W.; Mayne, L.; Bai, Y.; Sosnick, T. R. Hydrogen exchange: The modern legacy of Linderstrøm-Lang. *Protein Sci.* **1997**, *6*, 1101−1109.

(2) Engen, J. R.; Wales, T. E.; Shi, X. In *Encyclopedia of Analytical Chemistry*; Meyers, R., Ed.; John Wiley & Sons, 2011.
(3) Weis, D. D., Ed. *Hydrogen Exchange Mass Spectrometry of Proteins: Fundamentals, Methods, and Applications*; John Wiley & Sons, 2016.
(4) Claesen, J.; Burzykowski, T. Computational methods and challenges in hydrogen/deuterium exchange mass spectrometry. *Mass Spectrom Rev.* **2017**, *36*, 649−667.
(5) Huang, R. Y.; Chen, G. Higher order structure characterization of protein therapeutics by hydrogen/deuterium exchange mass spectrometry. *Anal Bioanal Chem.* **2014**, *406*, 6541−6558.
(6) Konermann, L.; Tong, X.; Pan, Y. Protein structure and dynamics studied by mass spectrometry: H/D exchange, hydroxyl radical labeling, and related approaches. *J. Mass Spectrom* **2008**, *43*, 1021−1036.
(7) Witten, J.; Ruschak, A.; Poterba, T.; Jaramillo, A.; Miranker, A. D.; Jaswal, S. S. Mapping protein conformational landscapes under strongly native conditions with hydrogen exchange mass spectrometry. *J. Phys. Chem. B* **2015**, *119*, 10016−10024.
(8) Engen, J. R.; Komives, E. A. Complementarity of hydrogen/deuterium exchange mass spectrometry and cryo-electron microscopy. *Trends Biochem. Sci.* **2020**, *45*, 906−918.
(9) Rey, M.; Sarpe, V.; Burns, K. M.; Buse, J.; Baker, C. A.; van Dijk, M.; Wordeman, L.; Bonvin, A. M.; Schriemer, D. C. Mass Spec Studio for integrative structural biology. *Structure* **2014**, *22*, 1538−1548.
(10) Vallat, B.; Webb, B.; Westbrook, J.; Sali, A.; Berman, H. M. Archiving and disseminating integrative structure models. *J. Biomol NMR* **2019**, *73*, 385−398.
(11) Dokholyan, N. V. Experimentally-driven protein structure modeling. *J. Proteomics* **2020**, *220*, 103777.
(12) Pirrone, G. F.; Iacob, R. E.; Engen, J. R. Applications of hydrogen/deuterium exchange MS from 2012 to 2014. *Anal. Chem.* **2015**, *87*, 99−118.
(13) Englander, S. W.; Mayne, L.; Kan, Z.-Y.; Hu, W. Protein folding−how and why: By hydrogen exchange, fragment separation, and mass spectrometry. *Annu. Rev. Biophys* **2016**, *45*, 135−152.
(14) Ramirez-Sarmiento, C. A.; Komives, E. A. Hydrogen-deuterium exchange mass spectrometry reveals folding and allostery in protein-protein interactions. *Methods* **2018**, *144*, 43−52.
(15) Konermann, L.; Pan, J.; Liu, Y.-H. Hydrogen exchange mass spectrometry for studying protein structure and dynamics. *Chem. Soc. Rev.* **2011**, *40*, 1224−1234.
(16) Mehmood, S.; Domene, C.; Forest, E.; Jault, J.-M. Dynamics of a bacterial multidrug ABC transporter in the inward- and outward-facing conformations. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 10832−10836.
(17) Balasubramaniam, D.; Komives, E. A. Hydrogen-exchange mass spectrometry for the study of intrinsic disorder in proteins. *Biochim. Biophys. Acta* **2013**, *1834*, 1202−1209.
(18) Harrison, R. A.; Engen, J. R. Conformational insight into multi-protein signaling assemblies by hydrogen−deuterium exchange mass spectrometry. *Curr. Opin Struct Biol.* **2016**, *41*, 187−193.
(19) Trabjerg, E.; Nazari, Z. E.; Rand, K. D. Conformational analysis of complex protein states by hydrogen/deuterium exchange mass spectrometry (HDX-MS): Challenges and emerging solutions. *Trends Anal Chem.* **2018**, *106*, 125−138.
(20) Redhair, M.; Clouser, A. F.; Atkins, W. M. Hydrogen-deuterium exchange mass spectrometry of membrane proteins in lipid nanodiscs. *Chem. Phys. Lipids* **2019**, *220*, 14−22.
(21) Giladi, M.; Khananshvili, D. Hydrogen-deuterium exchange mass-spectrometry of secondary active transporters: From structural dynamics to molecular mechanisms. *Front Pharmacol* **2020**, *11*, 70.
(22) Martens, C.; Politis, A. A glimpse into the molecular mechanism of integral membrane proteins through hydrogen-deuterium exchange mass spectrometry. *Protein Sci.* **2020**, *29*, 1285−1301.
(23) Hodge, E. A.; Benhaim, M. A.; Lee, K. K. Bridging protein structure, dynamics, and function using hydrogen/deuterium-exchange mass spectrometry. *Protein Sci.* **2020**, *29*, 843−855.
(24) Narang, D.; Lento, C.; Wilson, D. J. HDX-MS: An analytical tool to capture protein motion in action. *Biomedicines* **2020**, *8*, 224.
(25) James, E. I.; Murphree, T. A.; Vorauer, C.; Engen, J. R.; Guttman, M. Advances in hydrogen/deuterium exchange mass spectrometry and the pursuit of challenging biological systems. *Chem. Rev.* **2021**, DOI: 10.1021/acs.chemrev.1c00279.
(26) Deng, B.; Lento, C.; Wilson, D. J. Hydrogen deuterium exchange mass spectrometry in biopharmaceutical discovery and development − A review. *Anal. Chim. Acta* **2016**, *940*, 8−20.
(27) Masson, G. R.; Jenkins, M. L.; Burke, J. E. An overview of hydrogen deuterium exchange mass spectrometry (HDX-MS) in drug discovery. *Exp Op Drug Disc* **2017**, *12*, 981−994.
(28) Engen, J. R.; Botzanowski, T.; Peterle, D.; Georgescauld, F.; Wales, T. E. Developments in hydrogen/deuterium exchange mass spectrometry. *Anal. Chem.* **2021**, *93*, 567−582.
(29) Gallagher, E. S.; Hudgens, J. W. Isotope Labeling of Biomolecules - Applications. In *Methods in Enzymology*; Kelman, Z., Ed.; Elsevier, 2016; Vol. *566*; pp 357−404.
(30) Jaswal, S. S. Biological insights from hydrogen exchange mass spectrometry. *Biochim. Biophys. Acta* **2013**, *1834*, 1188−1201.
(31) Calabrese, A. N.; Schiffrin, B.; Watson, M.; Karamanos, T. K.; Walko, M.; Humes, J. R.; Horne, J. E.; White, P.; Wilson, A. J.; Kalli, A. C.; Tuma, R.; Ashcroft, A. E.; Brockwell, D. J.; Radford, S. E. Inter-domain dynamics in the chaperone SurA and multi-site binding to its outer membrane protein clients. *Nat. Commun.* **2020**, *11*, 2155.
(32) Lau, A. M.; Claesen, J.; Hansen, K.; Politis, A. Deuteros 2.0: Peptide-level significance testing of data from hydrogen deuterium exchange mass spectrometry. *Bioinformatics* **2021**, *37*, 270−272.
(33) Chalmers, M. J.; Busby, S. A.; Pascal, B. D.; West, G. M.; Griffin, P. R. Differential hydrogen/deuterium exchange mass spectrometry analysis of protein−ligand interactions. *Expert Rev. Proteomics* **2011**, *8*, 43−59.
(34) Mao, Y.; Zlatic, C. O.; Griffin, M. D.; Howlett, G. J.; Todorova, N.; Yarovsky, I.; Gooley, P. R. Hydrogen/deuterium exchange and molecular dynamics analysis of amyloid fibrils formed by a D69K charge-pair mutant of human apolipoprotein C-II. *Biochemistry* **2015**, *54*, 4805−4814.
(35) Hamuro, Y. Quantitative hydrogen/deuterium exchange mass spectrometry. *J. Am. Soc. Mass Spectrom.* **2021**, *32*, 2711−2727.
(36) Zhang, M. M.; Beno, B. R.; Huang, R. Y.; Adhikari, J.; Deyanova, E. G.; Li, J.; Chen, G.; Gross, M. L. An integrated approach for determining a protein-protein binding interface in solution and an evaluation of hydrogen-deuterium exchange kinetics for adjudicating candidate docking models. *Anal. Chem.* **2019**, *91*, 15709−15717.
(37) Bastos, V. A.; Gomes-Neto, F.; Rocha, S. L.; Teixeira-Ferreira, A.; Perales, J.; Neves-Ferreira, A. G.; Valente, R. H. The interaction between the natural metalloendopeptidase inhibitor BJ46a and its target toxin jararhagin analyzed by structural mass spectrometry and molecular modeling. *J. Proteomics* **2020**, *221*, 103761.
(38) Petersen, M.; Madsen, J. B.; Jørgensen, T. J.; Trelle, M. B. Conformational preludes to the latency transition in PAI-1 as determined by atomistic computer simulations and hydrogen/deuterium-exchange mass spectrometry. *Sci. Rep* **2017**, *7*, 6636.
(39) Murcia Rios, A.; Vahidi, S.; Dunn, S. D.; Konermann, L. Evidence for a partially stalled γ rotor in F1-ATPase from hydrogen−deuterium exchange experiments and molecular dynamics simulations. *J. Am. Chem. Soc.* **2018**, *140*, 14860−14869.
(40) Huang, L.; So, P.-K.; Yao, Z.-P. Protein dynamics revealed by hydrogen/deuterium exchange mass spectrometry: Correlation between experiments and simulation. *Rapid Commun. Mass Spectrom.* **2019**, *33*, 83−89.
(41) Makepeace, K. A.; Brodie, N. I.; Popov, K. I.; Gudavicius, G.; Nelson, C. J.; Petrotchenko, E. V.; Dokholyan, N. V.; Borchers, C. H. Ligand-induced disorder-to-order transitions characterized by structural proteomics and molecular dynamics simulations. *J. Proteomics* **2020**, *211*, 103544.
(42) Frame, N. M.; Kumanan, M.; Wales, T. E.; Bandara, A.; Fändrich, M.; Straub, J. E.; Engen, J. R.; Gursky, O. Structural basis

for lipid binding and function by an evolutionarily conserved protein, serum amyloid A. *J. Mol. Biol.* **2020**, *432*, 1978−1995.
(43) Paço, L.; Zarate-Perez, F.; Clouser, A. F.; Atkins, W. M.; Hackett, J. C. Dynamics and mechanism of binding of androstenedione to membrane-associated aromatase. *Biochemistry* **2020**, *59*, 2999−3009.
(44) Pacheco, S.; Widjaja, M. A.; Gomez, J. S.; Crowhurst, K. A.; Abrol, R. The complex role of the N-terminus and acidic residues of HdeA as pH-dependent switches in its chaperone function. *Biophys Chem.* **2020**, *264*, 106406.
(45) Devaurs, D.; Antunes, D. A.; Kavraki, L. E. Computational analysis of complement inhibitor compstatin using molecular dynamics. *J. Mol. Model* **2020**, *26*, 231.
(46) Jia, R.; Martens, C.; Shekhar, M.; Pant, S.; Pellowe, G. A.; Lau, A. M.; Findlay, H. E.; Harris, N. J.; Tajkhorshid, E.; Booth, P. J.; Politis, A. Hydrogen-deuterium exchange mass spectrometry captures distinct dynamics upon substrate and inhibitor binding to a transporter. *Nat. Commun.* **2020**, *11*, 6162.
(47) Redhair, M.; Hackett, J. C.; Pelletier, R. D.; Atkins, W. M. Dynamics and location of the allosteric midazolam site in cytochrome P4503A4 in lipid nanodiscs. *Biochemistry* **2020**, *59*, 766−779.
(48) Huang, L.; So, P.-K.; Chen, Y. W.; Leung, Y.-C.; Yao, Z.-P. Conformational dynamics of the helix 10 region as an allosteric site in class A β-lactamase inhibitory binding. *J. Am. Chem. Soc.* **2020**, *142*, 13756−13767.
(49) Lee, S.-M.; Jeong, Y.; Simms, J.; Warner, M. L.; Poyner, D. R.; Chung, K. Y.; Pioszak, A. A. Calcitonin receptor N-glycosylation enhances peptide hormone affinity by controlling receptor dynamics. *J. Mol. Biol.* **2020**, *432*, 1996−2014.
(50) Medina, E.; Villalobos, P.; Hamilton, G. L.; Komives, E. A.; Sanabria, H.; Ramírez-Sarmiento, C. A.; Babul, J. Intrinsically disordered regions of the DNA-binding domain of human FoxP1 facilitate domain swapping. *J. Mol. Biol.* **2020**, *432*, 5411−5429.
(51) Lau, A. M.; Jia, R.; Bradshaw, R. T.; Politis, A. Structural predictions of the functions of membrane proteins from HDX-MS. *Biochem. Soc. Trans.* **2020**, *48*, 971−979.
(52) Skinner, J. J.; Lim, W. K.; Bédard, S.; Black, B. E.; Englander, S. W. Protein hydrogen exchange: Testing current models. *Protein Sci.* **2012**, *21*, 987−995.
(53) 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.* **1998**, *7*, 739−745.
(54) Sowole, M. A.; Alexopoulos, J. A.; Cheng, Y.-Q.; Ortega, J.; Konermann, L. Activation of ClpP protease by ADEP antibiotics: Insights from hydrogen exchange mass spectrometry. *J. Mol. Biol.* **2013**, *425*, 4508−4519.
(55) Mohammadiarani, H.; Shaw, V. S.; Neubig, R. R.; Vashisth, H. Interpreting hydrogen-deuterium exchange events in proteins using atomistic simulations: Case studies on regulators of G-protein signaling proteins. *J. Phys. Chem. B* **2018**, *122*, 9314−9323.
(56) Brier, S.; Engen, J. R. In *Mass Spectrometry Analysis for Protein-Protein Interactions and Dynamics*; Chance, M., Ed.; John Wiley & Sons, Inc., 2008; pp 11−43.
(57) Chance, M., Ed. *Mass Spectrometry Analysis for Protein-Protein Interactions and Dynamics*; John Wiley & Sons, 2008.
(58) Skinner, J. J.; Lim, W. K.; Bédard, S.; Black, B. E.; Englander, S. W. Protein dynamics viewed by hydrogen exchange. *Protein Sci.* **2012**, *21*, 996−1005.
(59) Wei, H.; Tymiak, A. A.; Chen, G. In *Characterization of Protein Therapeutics Using Mass Spectrometry*; Chen, G., Ed.; Springer, 2013; pp 305−341.
(60) Bai, Y.; Milne, J. S.; Mayne, L.; Englander, S. W. Primary structure effects on peptide group hydrogen exchange. *Proteins* **1993**, *17*, 75−86.
(61) Connelly, G. P.; Bai, Y.; Jeng, M.-F.; Englander, S. W. Isotope effects in peptide group hydrogen exchange. *Proteins* **1993**, *17*, 87−92.
(62) Nguyen, D.; Mayne, L.; Phillips, M. C.; Englander, S. W. Reference parameters for protein hydrogen exchange rates. *J. Am. Soc. Mass Spectrom.* **2018**, *29*, 1936−1939.
(63) Udgaonkar, J. B.; Baldwin, R. L. Early folding intermediate of ribonuclease A. *Proc. Natl. Acad. Sci. U. S. A.* **1990**, *87*, 8197−8201.
(64) Morozova, L. A.; Haynie, D. T.; Arico-Muendel, C.; Van Dael, H.; Dobson, C. M. Structural basis of the stability of a lysozyme molten globule. *Nat. Struct. Biol.* **1995**, *2*, 871−875.
(65) Hwang, T.-L.; Mori, S.; Shaka, A. J.; van Zijl, P. C. Application of phase-modulated CLEAN chemical exchange spectroscopy (CLEANEX-PM) to detect water−protein proton exchange and intermolecular NOEs. *J. Am. Chem. Soc.* **1997**, *119*, 6203−6204.
(66) Baldwin, A. J.; Kay, L. E. NMR spectroscopy brings invisible protein states into focus. *Nat. Chem. Biol.* **2009**, *5*, 808−814.
(67) Sethi, S. K.; Smith, D. L.; McCloskey, J. A. Determination of active hydrogen content by fast atom bombardment mass spectrometry following hydrogen-deuterium exchange. *Biochem. Biophys. Res. Commun.* **1983**, *112*, 126−131.
(68) Tüchsen, E.; Hayes, J. M.; Ramaprasad, S.; Copie, V.; Woodward, C. Solvent exchange of buried water and hydrogen exchange of peptide NH groups hydrogen bonded to buried waters in bovine pancreatic trypsin inhibitor. *Biochemistry* **1987**, *26*, 5163−5172.
(69) Katta, V.; Chait, B. T.; Carr, S. Conformational changes in proteins probed by hydrogen-exchange electrospray-ionization mass spectrometry. *Rapid Commun. Mass Spectrom.* **1991**, *5*, 214−217.
(70) Zhang, Z.; Smith, D. L. Determination of amide hydrogen exchange by mass spectrometry: A new tool for protein structure elucidation. *Protein Sci.* **1993**, *2*, 522−531.
(71) Johnson, R. S.; Walsh, K. A. Mass spectrometric measurement of protein amide hydrogen exchange rates of apo- and holomyoglobin. *Protein Sci.* **1994**, *3*, 2411−2418.
(72) Mayne, L. Isotope Labeling of Biomolecules - Applications. In *Methods in Enzymology*; Kelman, Z., Ed.; Academic Press, 2016; Vol. *566*; pp 335−356.
(73) Oganesyan, I.; Lento, C.; Wilson, D. J. Contemporary hydrogen deuterium exchange mass spectrometry. *Methods* **2018**, *144*, 27−42.
(74) Rand, K. D.; Zehl, M.; Jensen, O. N.; Jørgensen, T. J. Protein hydrogen exchange measured at single-residue resolution by electron transfer dissociation mass spectrometry. *Anal. Chem.* **2009**, *81*, 5577−5584.
(75) Rand, K. D.; Pringle, S. D.; Morris, M.; Brown, J. M. Site-specific analysis of gas-phase hydrogen/deuterium exchange of peptides and proteins by electron transfer dissociation. *Anal. Chem.* **2012**, *84*, 1931−1940.
(76) Althaus, E.; Canzar, S.; Ehrler, C.; Emmett, M. R.; Karrenbauer, A.; Marshall, A. G.; Meyer-Bäse, A.; Tipton, J. D.; Zhang, H.-M. Computing H/D-exchange rates of single residues from data of proteolytic fragments. *BMC Bioinform* **2010**, *11*, 424.
(77) Gessner, C.; Steinchen, W.; Bédard, S.; Skinner, J. J.; Woods, V. L.; Walsh, T. J.; Bange, G.; Pantazatos, D. P. Computational method allowing hydrogen-deuterium exchange mass spectrometry at single amide resolution. *Sci. Rep* **2017**, *7*, 3789.
(78) Salmas, R. E.; Borysik, A. J. HDXmodeller: An online webserver for high-resolution HDX-MS with auto-validation. *Commun. Biol.* **2021**, *4*, 199.
(79) Salmas, R. E.; Borysik, A. J. Characterization and management of noise in HDX-MS data modeling. *Anal. Chem.* **2021**, *93*, 7323−7331.
(80) 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.* **2012**, *23*, 1202−1208.
(81) 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. U. S. A.* **2013**, *110*, 16438−16443.
(82) Saltzberg, D. J.; Broughton, H. B.; Pellarin, R.; Chalmers, M. J.; Espada, A.; Dodge, J. A.; Pascal, B. D.; Griffin, P. R.; Humblet, C.; Sali, A. A residue-resolved Bayesian approach to quantitative interpretation of hydrogen−deuterium exchange from mass spec-

https://doi.org/10.1021/jasms.1c00328
J. Am. Soc. Mass Spectrom. 2022, 33, 215−237

trometry: Application to characterizing protein−ligand interactions. *J. Phys. Chem. B* **2017**, *121*, 3493−3501.
(83) Skinner, S. P.; Radou, G.; Tuma, R.; Houwing-Duistermaat, J. J.; Paci, E. Estimating constraints for protection factors from HDX-MS data. *Biophys. J.* **2019**, *116*, 1194−1203.
(84) Sheff, J. G.; Rey, M.; Schriemer, D. C. Peptide-column interactions and their influence on back exchange rates in hydrogen/deuterium exchange-MS. *J. Am. Soc. Mass Spectrom.* **2013**, *24*, 1006−1015.
(85) Anand, G. S.; Law, D.; Mandell, J. G.; Snead, A. N.; Tsigelny, I.; Taylor, S. S.; Ten Eyck, L. F.; Komives, E. A. Identification of the protein kinase A regulatory RIα-catalytic subunit interface by amide H/^{2}H exchange and protein docking. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 13264−13269.
(86) Borysik, A. J. Simulated isotope exchange patterns enable protein structure determination. *Angew. Chem. Int. Ed* **2017**, *56*, 9396−9399.
(87) Pandit, D.; Tuske, S. J.; Coales, S. J.; Sook, Y. E.; Liu, A.; Lee, J. E.; Morrow, J. A.; Nemeth, J. F.; Hamuro, Y. Mapping of discontinuous conformational epitopes by amide hydrogen/deuterium exchange mass spectrometry and computational docking. *J. Mol. Recognit* **2012**, *25*, 114−124.
(88) Chen, R.; Li, L.; Weng, Z. ZDOCK: An initial-stage protein-docking algorithm. *Proteins* **2003**, *52*, 80−87.
(89) Roberts, V. A.; Pique, M. E.; Hsu, S.; Li, S. Combining H/D exchange mass spectrometry and computational docking to derive the structure of protein-protein complexes. *Biochemistry* **2017**, *56*, 6329−6342.
(90) Dominguez, C.; Boelens, R.; Bonvin, A. M. HADDOCK: A protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* **2003**, *125*, 1731−1737.
(91) de Vries, S. J.; van Dijk, M.; Bonvin, A. M. The HADDOCK web server for data-driven biomolecular docking. *Nat. Protoc* **2010**, *5*, 883−897.
(92) Merkle, P. S.; Irving, M.; Hongjian, S.; Ferber, M.; Jørgensen, T. J.; Scholten, K.; Luescher, I.; Coukos, G.; Zoete, V.; Cuendet, M. A.; Michielin, O.; Rand, K. D. The T-cell receptor can bind to the peptide-bound major histocompatibility complex and uncomplexed β2-microglobulin through distinct binding sites. *Biochemistry* **2017**, *56*, 3945−3961.
(93) Snijder, J.; Burnley, R. J.; Wiegard, A.; Melquiond, A. S.; Bonvin, A. M.; Axmann, I. M.; Heck, A. J. Insight into cyanobacterial circadian timing from structural details of the KaiB−KaiC interaction. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 1379−1384.
(94) Eron, S. J.; Huang, H.; Agafonov, R. V.; Fitzgerald, M. E.; Patel, J.; Michael, R. E.; Lee, T. D.; Hart, A. A.; Shaulsky, J.; Nasveschuk, C. G.; Phillips, A. J.; Fisher, S. L.; Good, A. Structural characterization of degrader-induced ternary complexes using hydrogen-deuterium exchange mass spectrometry and computational modeling: Implications for structure-based design. *ACS Chem. Biol.* **2021**, *16*, 2228−2243.
(95) Martens, C.; Shekhar, M.; Lau, A. M.; Tajkhorshid, E.; Politis, A. Integrating hydrogen−deuterium exchange mass spectrometry with molecular dynamics simulations to probe lipid-modulated conformational changes in membrane proteins. *Nat. Protoc* **2019**, *14*, 3183−3204.
(96) Gemmecker, G.; Jahnke, W.; Kessler, H. Measurement of fast proton exchange rates in isotopically labeled compounds. *J. Am. Chem. Soc.* **1993**, *115*, 11620−11621.
(97) Jeng, M.-F.; Dyson, H. J. Comparison of the hydrogen-exchange behavior of reduced and oxidized Escherichia coli thioredoxin. *Biochemistry* **1995**, *34*, 611−619.
(98) Dharmasiri, K.; Smith, D. L. Mass spectrometric determination of isotopic exchange rates of amide hydrogens located on the surfaces of proteins. *Anal. Chem.* **1996**, *68*, 2340−2344.
(99) Hughes, C. A.; Mandell, J. G.; Anand, G. S.; Stock, A. M.; Komives, E. A. Phosphorylation causes subtle changes in solvent accessibility at the interdomain interface of methylesterase CheB. *J. Mol. Biol.* **2001**, *307*, 967−976.
(100) Yan, X.; Watson, J.; Ho, P. S.; Deinzer, M. L. Electrospray ionization charge states and hydrogen-deuterium exchange for determining protein structures and their conformational changes. *Mol. Cell Proteom* **2004**, *3*, 10−23.
(101) Truhlar, S. M.; Croy, C. H.; Torpey, J. W.; Koeppe, J. R.; Komives, E. A. Solvent accessibility of protein surfaces by amide H/^{2}H exchange MALDI-TOF mass spectrometry. *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 1490−1497.
(102) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera−A visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605−1612.
(103) Sheinerman, F. B.; Brooks, C. L. Molecular picture of folding of a small α/β protein. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 1562−1567.
(104) Shan, Y.; Arkhipov, A.; Kim, E. T.; Pan, A. C.; Shaw, D. E. Transitions to catalytically inactive conformations in EGFR kinase. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 7270−7275.
(105) Ma, B.; Nussinov, R. Polymorphic C-terminal β-sheet interactions determine the formation of fibril or amyloid β-derived diffusible ligand-like globulomer for the Alzheimer Aβ42 dodecamer. *J. Biol. Chem.* **2010**, *285*, 37102−37110.
(106) Petruk, A. A.; Defelipe, L. A.; Rodríguez Limardo, R. G.; Bucci, H.; Marti, M. A.; Turjanski, A. G. Molecular dynamics simulations provide atomistic insight into hydrogen exchange mass spectrometry experiments. *J. Chem. Theory Comput* **2013**, *9*, 658−669.
(107) Park, I.-H.; Venable, J. D.; Steckler, C.; Cellitti, S. E.; Lesley, S. A.; Spraggon, G.; Brock, A. Estimation of hydrogen-exchange protection factors from MD simulation based on amide hydrogen bonding analysis. *J. Chem. Inf Model* **2015**, *55*, 1914−1925.
(108) Marsh, J. A.; Neale, C.; Jack, F. E.; Choy, W.-Y.; Lee, A. Y.; Crowhurst, K. A.; Forman-Kay, J. D. Improved structural characterizations of the drkN SH3 domain unfolded state suggest a compact ensemble with native-like and non-native structure. *J. Mol. Biol.* **2007**, *367*, 1494−1510.
(109) Marsh, J. A.; Forman-Kay, J. D. Structure and disorder in an unfolded state under nondenaturing conditions from ensemble models consistent with a large number of experimental restraints. *J. Mol. Biol.* **2009**, *391*, 359−374.
(110) Zhang, Y.; Majumder, E. L.-W.; Yue, H.; Blankenship, R. E.; Gross, M. L. Structural analysis of diheme cytochrome c by hydrogen-deuterium exchange mass spectrometry and homology modeling. *Biochemistry* **2014**, *53*, 5619−5630.
(111) Fazelinia, H.; Xu, M.; Cheng, H.; Roder, H. Ultrafast hydrogen exchange reveals specific structural events during the initial stages of folding of cytochrome *c*. *J. Am. Chem. Soc.* **2014**, *136*, 733−740.
(112) Skinner, J. J.; Yu, W.; Gichana, E. K.; Baxa, M. C.; Hinshaw, J. R.; Freed, K. F.; Sosnick, T. R. Benchmarking all-atom simulations using hydrogen exchange. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 15975−15980.
(113) Ma, B.; Nussinov, R. Polymorphic triple β-sheet structures contribute to amide hydrogen/deuterium (H/D) exchange protection in the Alzheimer amyloid β42 peptide. *J. Biol. Chem.* **2011**, *286*, 34244−34253.
(114) Anderson, J. S.; Hernández, G.; LeMaster, D. M. Assessing the chemical accuracy of protein structures via peptide acidity. *Biophys Chem.* **2013**, *171*, 63−75.
(115) McAllister, R. G.; Konermann, L. Challenges in the interpretation of protein H/D exchange data: A molecular dynamics simulation perspective. *Biochemistry* **2015**, *54*, 2683−2692.
(116) Scrosati, P. M.; Yin, V.; Konermann, L. Hydrogen/deuterium exchange measurements may provide an incomplete view of protein dynamics: A case study on cytochrome c. *Anal. Chem.* **2021**, *93*, 14121−14129.
(117) Sung, S.-S. Dielectric screening effect of electronic polarization and intramolecular hydrogen bonding. *Protein Sci.* **2017**, *26*, 2003−2009.

https://doi.org/10.1021/jasms.1c00328
J. Am. Soc. Mass Spectrom. 2022, 33, 215−237

(118) Anderson, J. S.; Hernández, G.; LeMaster, D. M. A billion-fold range in acidity for the solvent-exposed amides of *Pyrococcus furiosus* rubredoxin. *Biochemistry* **2008**, *47*, 6178−6188.
(119) Englander, S. W.; Kallenbach, N. R. Hydrogen exchange and structural dynamics of proteins and nucleic acids. *Q. Rev. Biophys.* **1983**, *16*, 521−655.
(120) Tüchsen, E.; Woodward, C. Hydrogen kinetics of peptide amide protons at the bovine pancreatic trypsin inhibitor protein-solvent interface. *J. Mol. Biol.* **1985**, *185*, 405−419.
(121) Dempsey, C. E. Hydrogen bond stabilities in the isolated alamethicin helix: pH-dependent amide exchange measurements in methanol. *J. Am. Chem. Soc.* **1995**, *117*, 7526−7534.
(122) Forsyth, W. R.; Robertson, A. D. Intramolecular electrostatic interactions accelerate hydrogen exchange in diketopiperazine relative to 2-piperidone. *J. Am. Chem. Soc.* **1996**, *118*, 2694−2698.
(123) Kim, P. S.; Baldwin, R. L. Influence of charge on the rate of amide proton exchange. *Biochemistry* **1982**, *21*, 1−5.
(124) Delepierre, M.; Dobson, C. M.; Karplus, M.; Poulsen, F. M.; States, D. J.; Wedin, R. E. Electrostatic effects and hydrogen exchange behaviour in proteins: The pH dependence of exchange rates in lysozyme. *J. Mol. Biol.* **1987**, *197*, 111−122.
(125) Fogolari, F.; Esposito, G.; Viglino, P.; Briggs, J. M.; McCammon, J. A. pK_a shift effects on backbone amide base-catalyzed hydrogen exchange rates in peptides. *J. Am. Chem. Soc.* **1998**, *120*, 3735−3738.
(126) Perrin, C. L.; Chen, J.-H.; Ohta, B. K. Amide proton exchange in micelles. *J. Am. Chem. Soc.* **1999**, *121*, 2448−2455.
(127) Radkiewicz, J. L.; Zipse, H.; Clarke, S.; Houk, K. N. Neighboring side chain effects on asparaginyl and aspartyl degradation: An ab initio study of the relationship between peptide conformation and backbone NH acidity. *J. Am. Chem. Soc.* **2001**, *123*, 3499−3506.
(128) Avbelj, F.; Baldwin, R. L. Origin of the neighboring residue effect on peptide backbone conformation. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 10967−10972.
(129) Avbelj, F.; Baldwin, R. L. Origin of the change in solvation enthalpy of the peptide group when neighboring peptide groups are added. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 3137−3141.
(130) Hernández, G.; Anderson, J. S.; LeMaster, D. M. Electrostatic stabilization and general base catalysis in the active site of the human protein disulfide isomerase a domain monitored by hydrogen exchange. *ChemBioChem.* **2008**, *9*, 768−778.
(131) Shaw, B. F.; Arthanari, H.; Narovlyansky, M.; Durazo, A.; Frueh, D. P.; Pollastri, M. P.; Lee, A.; Bilgicer, B.; Gygi, S. P.; Wagner, G.; Whitesides, G. M. Neutralizing positive charges at the surface of a protein lowers its rate of amide hydrogen exchange without altering its structure or increasing its thermostability. *J. Am. Chem. Soc.* **2010**, *132*, 17411−17425.
(132) Hernández, G.; Anderson, J. S.; LeMaster, D. M. Polarization and polarizability assessed by protein amide acidity. *Biochemistry* **2009**, *48*, 6482−6494.
(133) LeMaster, D. M.; Anderson, J. S.; Hernández, G. Peptide conformer acidity analysis of protein flexibility monitored by hydrogen exchange. *Biochemistry* **2009**, *48*, 9256−9265.
(134) Hernández, G.; Anderson, J. S.; LeMaster, D. M. Protein NMR Techniques. In *Methods in Molecular Biology*; Shekhtman, A., Burz, D., Eds.; Humana Press, 2012; Vol. *831*; pp 369−405.
(135) Anderson, J. S.; Hernández, G.; LeMaster, D. M. Backbone conformational dependence of peptide acidity. *Biophys Chem.* **2009**, *141*, 124−130.
(136) Anderson, J. S.; Hernández, G.; LeMaster, D. M. Sidechain conformational dependence of hydrogen exchange in model peptides. *Biophys Chem.* **2010**, *151*, 61−70.
(137) Avbelj, F.; Baldwin, R. L. Limited validity of group additivity for the folding energetics of the peptide group. *Proteins* **2006**, *63*, 283−289.
(138) Barnes, C. A.; Shen, Y.; Ying, J.; Takagi, Y.; Torchia, D. A.; Sellers, J. R.; Bax, A. Remarkable rigidity of the single α-helical domain of myosin-VI as revealed by NMR spectroscopy. *J. Am. Chem. Soc.* **2019**, *141*, 9004−9017.
(139) Anderson, J. S.; Hernández, G.; LeMaster, D. M. Conformational electrostatics in the stabilization of the peptide anion. *Curr. Org. Chem.* **2010**, *14*, 162−180.
(140) Abdolvahabi, A.; Gober, J. L.; Mowery, R. A.; Shi, Y.; Shaw, B. F. Metal-ion-specific screening of charge effects in protein amide H/D exchange and the Hofmeister series. *Anal. Chem.* **2014**, *86*, 10303−10310.
(141) Dass, R.; Corlianò, E.; Mulder, F. A. The contribution of electrostatics to hydrogen exchange in the unfolded protein state. *Biophys. J.* **2021**, *120*, 4107−4114.
(142) Hernández, G.; Anderson, J. S.; LeMaster, D. M. Assessing the native state conformational distribution of ubiquitin by peptide acidity. *Biophys Chem.* **2010**, *153*, 70−82.
(143) Hernández, G.; Anderson, J. S.; LeMaster, D. M. Experimentally assessing molecular dynamics sampling of the protein native state conformational distribution. *Biophys Chem.* **2012**, *163*−*164*, 21−34.
(144) Vendruscolo, M.; Paci, E.; Dobson, C. M.; Karplus, M. Rare fluctuations of native proteins sampled by equilibrium hydrogen exchange. *J. Am. Chem. Soc.* **2003**, *125*, 15686−15687.
(145) Best, R. B.; Vendruscolo, M. Structural interpretation of hydrogen exchange protection factors in proteins: Characterization of the native state fluctuations of CI2. *Structure* **2006**, *14*, 97−106.
(146) Kieseritzky, G.; Morra, G.; Knapp, E.-W. Stability and fluctuations of amide hydrogen bonds in a bacterial cytochrome *c*: A molecular dynamics study. *J. Biol. Inorg. Chem.* **2006**, *11*, 26−40.
(147) Xu, J.; Lee, Y.; Beamer, L. J.; Van Doren, S. R. Phosphorylation in the catalytic cleft stabilizes and attracts domains of a phosphohexomutase. *Biophys. J.* **2015**, *108*, 325−337.
(148) Devaurs, D.; Antunes, D. A.; Papanastasiou, M.; Moll, M.; Ricklin, D.; Lambris, J. D.; Kavraki, L. E. Coarse-grained conformational sampling of protein structure improves the fit to experimental hydrogen-exchange data. *Front Mol. Biosci* **2017**, DOI: 10.3389/fmolb.2017.00013.
(149) Schrödinger, L. L. C. *PyMOL molecular graphics system*, version 1.8, 2015.
(150) Wan, H.; Ge, Y.; Razavi, A.; Voelz, V. A. Reconciling simulated ensembles of apomyoglobin with experimental hydrogen/deuterium exchange data using Bayesian inference and multiensemble Markov state models. *J. Chem. Theory Comput* **2020**, *16*, 1333−1348.
(151) Radou, G.; Dreyer, F. N.; Tuma, R.; Paci, E. Functional dynamics of hexameric helicase probed by hydrogen exchange and simulation. *Biophys. J.* **2014**, *107*, 983−990.
(152) Adhikary, S.; Deredge, D. J.; Nagarajan, A.; Forrest, L. R.; Wintrode, P. L.; Singh, S. K. Conformational dynamics of a neurotransmitter:sodium symporter in a lipid bilayer. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, E1786−E1795.
(153) Devaurs, D.; Papanastasiou, M.; Antunes, D. A.; Abella, J. R.; Moll, M.; Ricklin, D.; Lambris, J. D.; Kavraki, L. E. Native state of complement protein C3d analysed via hydrogen exchange and conformational sampling. *Int. J. Comput. Biol. Drug Des* **2018**, *11*, 90−113.
(154) Devaurs, D.; Antunes, D. A.; Kavraki, L. E. Revealing unknown protein structures using computational conformational sampling guided by experimental hydrogen-exchange data. *Int. J. Mol. Sci.* **2018**, *19*, 3406.
(155) Harris, M. J.; Raghavan, D.; Borysik, A. J. Quantitative evaluation of native protein folds and assemblies by hydrogen deuterium exchange mass spectrometry (HDX-MS). *J. Am. Soc. Mass Spectrom.* **2019**, *30*, 58−66.
(156) Bradshaw, R. T.; Marinelli, F.; Faraldo-Gómez, J. D.; Forrest, L. R. Interpretation of HDX data by maximum-entropy reweighting of simulated structural ensembles. *Biophys. J.* **2020**, *118*, 1649−1664.
(157) Khorvash, M.; Lamour, G.; Gsponer, J. Long-time scale fluctuations of human prion protein determined by restrained MD simulations. *Biochemistry* **2011**, *50*, 10192−10194.

(158) Gsponer, J.; Hopearuoho, H.; Whittaker, S. B.-M.; Spence, G. R.; Moore, G. R.; Paci, E.; Radford, S. E.; Vendruscolo, M. Determination of an ensemble of structures representing the intermediate state of the bacterial immunity protein Im7. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 99−104.
(159) Bemporad, F.; De Simone, A.; Chiti, F.; Dobson, C. M. Characterizing intermolecular interactions that initiate native-like protein aggregation. *Biophys. J.* **2012**, *102*, 2595−2604.
(160) Kihn, K. C.; Wilson, T.; Smith, A. K.; Bradshaw, R. T.; Wintrode, P. L.; Forrest, L. R.; Wilks, A.; Deredge, D. J. Modeling the native ensemble of PhuS using enhanced sampling MD and HDX-ensemble reweighting. *Biophys. J.* **2021**, *120*, 5141−5157.
(161) Deredge, D. J.; Huang, W.; Hui, C.; Matsumura, H.; Yue, Z.; Moënne-Loccoz, P.; Shen, J.; Wintrode, P. L.; Wilks, A. Ligand-induced allostery in the interaction of the *Pseudomonas aeruginosa* heme binding protein with heme oxygenase. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 3421−3426.
(162) Brand, T.; Cabrita, E. J.; Morris, G. A.; Günther, R.; Hofmann, H.-J.; Berger, S. Residue-specific NH exchange rates studied by NMR diffusion experiments. *J. Magn. Reson.* **2007**, *187*, 97−104.
(163) Sljoka, A.; Wilson, D. Probing protein ensemble rigidity and hydrogen-deuterium exchange. *Phys. Biol.* **2013**, *10*, 056013.
(164) Pester, O.; Barrett, P. J.; Hornburg, D.; Hornburg, P.; Pröbstle, R.; Widmaier, S.; Kutzner, C.; Dürrbaum, M.; Kapurniotu, A.; Sanders, C. R.; Scharnagl, C.; Langosch, D. The backbone dynamics of the amyloid precursor protein transmembrane helix provides a rationale for the sequential cleavage mechanism of γ-secretase. *J. Am. Chem. Soc.* **2013**, *135*, 1317−1329.
(165) Khakinejad, M.; Kondalaji, S. G.; Maleki, H.; Arndt, J. R.; Donohoe, G. C.; Valentine, S. J. Combining ion mobility spectrometry with hydrogen-deuterium exchange and top-down MS for peptide ion structure analysis. *J. Am. Soc. Mass Spectrom.* **2014**, *25*, 2103−2115.
(166) Khakinejad, M.; Kondalaji, S. G.; Tafreshian, A.; Valentine, S. J. Gas-phase hydrogen-deuterium exchange labeling of select peptide ion conformer types: A per-residue kinetics analysis. *J. Am. Soc. Mass Spectrom.* **2015**, *26*, 1115−1127.
(167) Khakinejad, M.; Kondalaji, S. G.; Donohoe, G. C.; Valentine, S. J. Ion mobility spectrometry-hydrogen deuterium exchange mass spectrometry of anions: Part 3. Estimating surface area exposure by deuterium uptake. *J. Am. Soc. Mass Spectrom.* **2016**, *27*, 462−473.
(168) Marzolf, D. R.; Seffernick, J. T.; Lindert, S. Protein structure prediction from NMR hydrogen−deuterium exchange data. *J. Chem. Theory Comput* **2021**, *17*, 2619−2629.
(169) Alford, R. F.; et al. The Rosetta all-atom energy function for macromolecular modeling and design. *J. Chem. Theory Comput* **2017**, *13*, 3031−3048.
(170) Pancsa, R.; Varadi, M.; Tompa, P.; Vranken, W. F. Start2Fold: A database of hydrogen/deuterium exchange data on protein folding and stability. *Nucleic Acids Res.* **2016**, *44*, D429−D434.
(171) Peng, X.; Baxa, M.; Faruk, N.; Sachleben, J. R.; Pintscher, S.; Gagnon, I. A.; Houliston, S.; Arrowsmith, C. H.; Freed, K. F.; Rocklin, G. J.; Sosnick, T. R. Prediction and validation of a protein's free energy surface using hydrogen exchange and (importantly) its denaturant dependence. *J. Chem. Theory Comput* **2022**, *18*, 550.
(172) Persson, F.; Halle, B. How amide hydrogens exchange in native proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 10383−10388.
(173) Wu, Z.; Wagner, M. A.; Zheng, L.; Parks, J. S.; Shy, J. M.; Smith, J. D.; Gogonea, V.; Hazen, S. L. The refined structure of nascent HDL reveals a key functional domain for particle maturation and dysfunction. *Nat. Struct Mol. Biol.* **2007**, *14*, 861−868.
(174) Wu, Z.; Gogonea, V.; Lee, X.; Wagner, M. A.; Li, X.-M.; Huang, Y.; Undurti, A.; May, R. P.; Haertlein, M.; Moulin, M.; Gutsche, I.; Zaccai, G.; DiDonato, J. A.; Hazen, S. L. Double superhelix model of high density lipoprotein. *J. Biol. Chem.* **2009**, *284*, 36605−36619.
(175) Gogonea, V.; Wu, Z.; Lee, X.; Pipich, V.; Li, X.-M.; Ioffe, A. I.; DiDonato, J. A.; Hazen, S. L. Congruency between biophysical data from multiple platforms and molecular dynamics simulation of the double-super helix model of nascent high-density lipoprotein. *Biochemistry* **2010**, *49*, 7323−7343.
(176) Nguyen, T. T.; Marzolf, D. R.; Seffernick, J. T.; Heinze, S.; Lindert, S. Protein structure prediction using residue-resolved protection factors from hydrogen-deuterium exchange NMR. *Structure* **2021**, DOI: 10.1016/j.str.2021.10.006.
(177) Högel, P.; Götz, A.; Kuhne, F.; Ebert, M.; Stelzer, W.; Rand, K. D.; Scharnagl, C.; Langosch, D. Glycine perturbs local and global conformational flexibility of a transmembrane helix. *Biochemistry* **2018**, *57*, 1326−1337.
(178) Hilser, V. J.; Freire, E. Structure-based calculation of the equilibrium folding pathway of proteins: Correlation with hydrogen exchange protection factors. *J. Mol. Biol.* **1996**, *262*, 756−772.
(179) Liu, T.; Pantazatos, D.; Li, S.; Hamuro, Y.; Hilser, V. J.; Woods, V. L. Quantitative assessment of protein structural models by comparison of H/D exchange MS data with exchange behavior accurately predicted by DXCOREX. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 43−56.
(180) Craig, P. O.; Lätzer, J.; Weinkam, P.; Hoffman, R. M.; Ferreiro, D. U.; Komives, E. A.; Wolynes, P. G. Prediction of native-state hydrogen exchange from perfectly funneled energy landscapes. *J. Am. Chem. Soc.* **2011**, *133*, 17463−17472.
(181) Sessions, R. B.; Gibbs, N.; Dempsey, C. E. Hydrogen bonding in helical polypeptides from molecular dynamics simulations and amide hydrogen exchange analysis: Alamethicin and melittin in methanol. *Biophys. J.* **1998**, *74*, 138−152.
(182) Hilser, V. J.; García-Moreno, B.; Oas, T. G.; Kapp, G.; Whitten, S. T. A statistical thermodynamic model of the protein ensemble. *Chem. Rev.* **2006**, *106*, 1545−1558.
(183) Freire, E. The propagation of binding interactions to remote sites in proteins: Analysis of the binding of the monoclonal antibody D1.3 to lysozyme. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 10118−10122.
(184) Liu, T.; Whitten, S. T.; Hilser, V. J. Functional residues serve a dominant role in mediating the cooperativity of the protein ensemble. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 4347−4352.
(185) Pan, H.; Lee, J. C.; Hilser, V. J. Binding sites in *Escherichia coli* dihydrofolate reductase communicate by modulating the conformational ensemble. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 12020−12025.
(186) Whitten, S. T.; Wooll, J. O.; Razeghifard, R.; García-Moreno, B. E.; Hilser, V. J. The origin of pH-dependent changes in *m*-values for the denaturant-induced unfolding of proteins. *J. Mol. Biol.* **2001**, *309*, 1165−1175.
(187) Whitten, S. T.; García-Moreno, B. E.; Hilser, V. J. In Biophysical Tools for Biologists, Vol. 1. In Vitro Techniques. In *Methods in Cell Biology*; Correia, J., Detrich, H., Eds.; Elsevier, 2008; Vol. *84*; pp 871−891.
(188) Babu, C. R.; Hilser, V. J.; Wand, A. J. Direct access to the cooperative substructure of proteins and the protein ensemble via cold denaturation. *Nat. Struct Mol. Biol.* **2004**, *11*, 352−357.
(189) Whitten, S. T.; Kurtz, A. J.; Pometun, M. S.; Wand, A. J.; Hilser, V. J. Revealing the nature of the native state ensemble through cold denaturation. *Biochemistry* **2006**, *45*, 10163−10174.
(190) Vertrees, J.; Wrabl, J. O.; Hilser, V. J. In Biothermodynamics, Part A. In *Methods in Enzymology*; Johnson, M., Holt, J., Ackers, G., Eds.; Academic Press, 2009; Vol. *455*; pp 299−327.
(191) Wrabl, J. O.; Larson, S. A.; Hilser, V. J. Thermodynamic environments in proteins: Fundamental determinants of fold specificity. *Protein Sci.* **2002**, *11*, 1945−1957.
(192) Wang, S.; Gu, J.; Larson, S. A.; Whitten, S. T.; Hilser, V. J. Denatured-state energy landscapes of a protein structural database reveal the energetic determinants of a framework model for folding. *J. Mol. Biol.* **2008**, *381*, 1184−1201.
(193) Cremades, N.; Sancho, J.; Freire, E. The native-state ensemble of proteins provides clues for folding, misfolding and function. *Trends Biochem. Sci.* **2006**, *31*, 494−496.

(194) Gu, J.; Hilser, V. J. Predicting the energetics of conformational fluctuations in proteins from sequence: A strategy for profiling the proteome. *Structure* **2008**, *16*, 1627−1637.
(195) Wrabl, J.; Gu, J.; Liu, T.; Schrank, T.; Whitten, S.; Hilser, V. The role of protein conformational fluctuations in allostery, function, and evolution. *Biophys Chem.* **2011**, *159*, 129−141.
(196) Hromić-Jahjefendić, A.; Jajčanin Jozić, N.; Kazazić, S.; Grabar Branilović, M.; Karačić, Z.; Schrittwieser, J. H.; Das, K. M.; Tomin, M.; Oberer, M.; Gruber, K.; Abramić, M.; Tomić, S. A novel Porphyromonas gingivalis enzyme: An atypical dipeptidyl peptidase III with an ARM repeat domain. *PLoS One* **2017**, *12*, No. e0188915.
(197) Tartaglia, G. G.; Cavalli, A.; Vendruscolo, M. Prediction of local structural stabilities of proteins from their amino acid sequences. *Structure* **2007**, *15*, 139−143.
(198) Dovidchenko, N. V.; Galzitskaya, O. V. Prediction of residue status to be protected or not protected from hydrogen exchange using amino acid sequence only. *Open Biochem J.* **2008**, *2*, 77−80.
(199) Dovidchenko, N. V.; Lobanov, M. Y.; Garbuzynskiy, S. O.; Galzitskaya, O. V. Prediction of amino acid residues protected from hydrogen-deuterium exchange in a protein chain. *Biochemistry (Moscow)* **2009**, *74*, 888−897.
(200) Jones, D. T. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **1999**, *292*, 195−202.
(201) Lobanov, M. Y.; Suvorina, M. Y.; Dovidchenko, N. V.; Sokolovskiy, I. V.; Surin, A. K.; Galzitskaya, O. V. A novel web server predicts amino acid residue protection against hydrogen-deuterium exchange. *Bioinformatics* **2013**, *29*, 1375−1381.
(202) Claesen, J.; Politis, A. POPPeT A new method to predict the protection factor of backbone amide hydrogens. *J. Am. Soc. Mass Spectrom.* **2019**, *30*, 67−76.
(203) Raimondi, D.; Orlando, G.; Pancsa, R.; Khan, T.; Vranken, W. F. Exploring the sequence-based prediction of folding initiation sites in proteins. *Sci. Rep.* **2017**, *7*, 8826.
(204) Wang, B.; Perez-Rathke, A.; Li, R.; Liang, J. A general method for predicting amino acid residues experiencing hydrogen exchange. *Proc. IEEE EMBS Int. Conf Biomed Health Info.* **2018**, 341−344.
(205) Sowole, M. A.; Konermann, L. Effects of protein−ligand interactions on hydrogen/deuterium exchange kinetics: Canonical and noncanonical scenarios. *Anal. Chem.* **2014**, *86*, 6715−6722.
(206) Malmstroem, L.; Hou, L.; Atkins, W. M.; Goodlett, D. R. On the use of hydrogen/deuterium exchange mass spectrometry data to improve *de novo* protein structure prediction. *Rapid Commun. Mass Spectrom.* **2009**, *23*, 459−461.
(207) Devaurs, D.; Vaisset, M.; Siméon, T.; Cortés, J. A multi-tree approach to compute transition paths on energy landscapes. *Proc. Workshop on Artificial Intelligence and Robotics Methods in Computational Biology (AAAI)*, 2013.
(208) Zhang, Z. Complete extraction of protein dynamics information in hydrogen/deuterium exchange mass spectrometry data. *Anal. Chem.* **2020**, *92*, 6486−6494.
(209) Wollenberg, D. T.; Pengelley, S.; Mouritsen, J. C.; Suckau, D.; Jørgensen, C. I.; Jørgensen, T. J. Avoiding H/D scrambling with minimal ion transmission loss for HDX-MS/MS-ETD analysis on a high-resolution Q-TOF mass spectrometer. *Anal. Chem.* **2020**, *92*, 7453−7461.

https://doi.org/10.1021/jasms.1c00328
J. Am. Soc. Mass Spectrom. 2022, 33, 215−237