Probing universal protein dynamics using HDX-MS-derived residue-level Gibbs free energy

Jochem H. Smit, Srinath Krishnamurthy, Bindu Y. Srinivasu, Rinky Parakra, Spyridoula Karamanou and Anastassios Economou*

KU Leuven, Department of Microbiology, Immunology and Transplantation, Rega Institute of Medical Research, Laboratory of Molecular Bacteriology, 3000 Leuven, Belgium

* For correspondence: e-mail: tassos.economou@kuleuven.be

Abstract

Hydrogen Deuterium Exchange Mass Spectrometry (HDX-MS) is a powerful technique to monitor protein intrinsic dynamics. The technique provides high-resolution information on how protein intrinsic dynamics are altered in response to biological signals, such as ligand binding, oligomerization or allosteric networks. However, identification, interpretation and visualization of such events from HDX-MS datasets is challenging as these datasets consist of many individual datapoints collected across peptides, timepoints and experimental conditions. Here we present PyHDX, an open-source python package and web server, that allows the user to batch-extract the universal quantity Gibbs free energy at residue level over multiple protein conditions and homologues. The output is directly visualized on a linear map or 3D structures or is exported as .csv files or pymol scripts.

Keywords: HDX-MS, Mass spectrometry, near-residue resolution, Gibbs free energy, Normal mode analysis, proteins, protein intrinsic dynamics

Introduction

Intrinsic dynamics underlie protein function[1]. These dynamics are intrinsic to the protein polymer chain and arise as a consequence of its tertiary structure and are induced by thermal fluctuations. Hydrogen/deuterium exchange mass spectrometry (HDX-MS) is a powerful monitor of these dynamics[2], and despite the fact that challenges related to experimental reproducibility and interpretation remain[3,4], the method can resolve protein dynamics spanning several orders of magnitude[5] and approaches near-residue resolution[6,7].

In typical 'bottom-up' HDX-MS (Figure 1a)[3], proteins are D-labelled in deuterated buffers. Exchange is quenched to limit back-exchange (through lowered pH/temperature[8]), proteins are proteolyzed into multiple overlapping peptides that are analysed by liquid chromatography-MS. D-uptake on each peptide is calculated from average mass changes between the deuterated and undeuterated form[4].

D-uptake values are generally presented as heatmaps (Figure 1b) or curves (Figure S1). These 2D slices of the full 3D dataset (time, deuteration, peptide; Figure S1), fail to capture the full breadth of the experimental information. It is desirable to display HDX-MS datasets as a single value per residue, but direct single residue resolution HDX-MS is only possible using MS/MS-ETD ECD capable mass spectrometers[9]. Instead, the overlap between peptides can be exploited[7,10–12] since HDX-MS yields deuteration values of each peptide *in toto* and not those of its individual amino acids. Several software tools are currently available which exploit such overlap[12–18].

HDX-MS analysis faces four remaining hurdles. Specifically, a) the large parameter space (one fit parameter per residue, i.e. typical several hundreds) can render data analysis several hours-long b) experimental variations (temperature, pH, peptides) hamper comparison between multiple datasets. c) Analysis tools often have complicated installation or commercial licensing. d) for HDX-MS-derived protein dynamics to have a wider impact and interface with orthogonal methods, output should yield universal quantities (e.g. Gibbs free energy).

Here we present PyHDX that addresses many of these issues: it is fast, open source, with detailed documentation and derives Gibbs free energy (ΔG) at residue level. Data is input as 'HDX data' tables in CSV format (peptide list, D-exposure time and D-uptake), obtained by treatment of HDX-MS spectra with other software (eg DynamX, Waters) submitted either as a single or multiple experiments. The full analysis and Gibbs free energy level classification of residues and visualization is accomplished in a web interface (Figure S2), within minutes and exported as text or a script to colour 3D structures in PyMOL (Schrödinger, LLC).

Experimental section

Protein purification

*ec*SecB$_4$ was purified as described[19]. *ec*SecB$_2$ was generated by mutating 3 residues (Y109A, T115A, S119A) that form the dimer-dimer interface and purified by nickel affinity purification as described[19]. Mutations were introduced by the Quick-Change Mutagenesis System (Pfu turbo, Agilent) using plasmid pIMBB490 (pET16b *secB*) as a template and listed primers (Table S1).

Hydrogen/deuterium exchange mass spectrometry

D-exchange was initiated by diluting 100 pmol of SecB by 10-fold into D_2O buffer (50 mM Tris-HCl pH_{read} 8.0, 50 mM KCl, 1 mM $MgCl_2$, 4 µM $ZnSO_4$, 2 mM TCEP) reconstituted in 99.9% D_2O (Euroiso-top) to obtain a final D_2O concentration of 90%. Continuous deuterium labelling was carried out for 6 timepoints (10 s, 30 s, 1 min, 5 min, 10 min, 100 min) at 30 °C. The reaction was quenched by the addition of pre-chilled quench buffer (1.5% formic acid, 4 mM TCEP, 0.1% DDM) in a 1:1 ratio. The reaction was injected into a nanoAcquity UPLC system with HDX technology (Waters, UK) coupled to a Synapt G2 ESI-QTOF mass spectrometer. Protein digestion was carried out by an online home-packed immobilized pepsin (Sigma) column (2mm x 2 cm, Idex) at 18 °C. LC was done using described instrument parameters [20]. The 100% full deuteration (FD) control was obtained by incubating SecB in D_2O buffer containing 6 M $UreaD_4$ (98% D, Sigma) overnight at room temperature. Peptide identification was carried out in ProteinLynx Global Server (PLGS, Waters UK) and deuterium exchange data was analysed in DynamX 3.0 (Waters, UK). 'State data' was exported from DynamX in the form of csv files. *mt*SecB HDX-MS data were corrected for back-exchange by assuming a constant 28% back-exchange for all peptides. All other proteins were corrected for back-exchange using a FD control sample. HDX-MS data of other proteins were from previously published datasets (*h*PREP[20], *ec*SecA[21]) or will be described in detail elsewhere (*ec*TF, *h*Bcl-2, *ec*SctV *ec*PpiB, *ec*PpiA and *ec*MBP).

Theory

We used the commonly employed Linderstrøm-Lang model (Figure 1c) for H/D exchange[22,23]. In this model, backbone amides can be either in a D-exchange-incompetent 'closed state', with amide hydrogens hydrogen-bonded in secondary structures, or in a D-exchange-competent 'open state', with compromised hydrogen bonds:

$$\mathrm{NH_{closed}} \underset{k_{close}}{\overset{k_{open}}{\rightleftharpoons}} \mathrm{NH_{open}} \xrightarrow{k_{int}} \mathrm{ND}$$

Here, NH represents an amide hydrogen and ND an amide deuterium. The intrinsic exchange rate k_{int} (frequently also referred to as chemical exchange rate, k_{ch}) is dependent on the pH and temperature at which the deuterium labelling reaction takes place, as well as the primary structure of the peptide, and can be calculated accurately[24–27]. This intrinsic rate is a major influence on the observed kinetics of D-exchange as it

can vary up to three orders of magnitude (e.g. pH 6, 0°C, vs pH 8, 30°C). Using the rate of D-exchanged measured by HDX-MS (k_{obs}) the Gibbs free energy differences ΔG between the 'open' and 'closed' states can be calculated (Figure 1c).

To correct for back-exchange[3,28], a FD control sample is used. The experimentally determined FD provides the maximal degree of D-exchange possible for any given peptide. The corrected D-uptake for each peptide is then calculated as[28]:

$$D_{corr}(t) = \frac{D(t)}{D_{FD}} n_{labile} \quad (2)$$

Where *D(t)* is the experimentally measured D-uptake, D_{FD} the D-uptake of the fully deuterated control. *D(t)* and D_{FD} input values are optionally corrected by subtraction of with a non deuterated control sample. f_{FD} and $f_{FD(t)}$ represent the fractional D-content of the FD control buffer and the D-labelling buffer, respectively. n_{labile} is the number of exchange-competent amide groups in the peptide, equal to the number of residues in the peptide, minus the number of proline residues and minus the first (or first two, user configurable) residues as these residues have high intrinsic rates of exchange and can be assumed to fully exchange back.

Using the steady-state approximation, assuming that the fractional population of NH_{open} is small, the observed rate of formation of the deuterated residue ND is given by[29]:

$$k_{obs} = \frac{k_{open} k_{int}}{k_{open} + k_{close} + k_{int}} \quad (3)$$

Assuming that the protein dynamics are faster than the exchange reaction ($k_{open} + k_{close} >> k_{int}$) and introducing the substitution PF = k_{close} / k_{open}, the expression reduces to:

$$k_{obs} = \frac{k_{int}}{1 + PF} \quad (4)$$

Where PF is the protection factor[23,30] for this particular protein residue. This ratio of rates is equivalent to the system's Boltzmann factor and thus relates to the Gibbs free energy difference between the closed and open states:

$$PF = \frac{k_{close}}{k_{open}} = e^{\frac{\Delta G}{RT}} \quad (5)$$

Where the sign of ΔG is chosen such that ΔG is positive when the energy of the open state is higher compared to that of the closed state, as is generally true for rigidly structured proteins. A low value of ΔG

indicates highly dynamic or disordered proteins. The approximations made above are generally true in the so called EX2 regime of H/D-exchange (i.e. $k_{close} >> k_{int}$; Figure 1c)[31]. Typical proteins measured under native conditions show mostly EX2 behaviour where only disordered parts of the protein exchange according to EX1 kinetics[32]. The presence of cooperative EX1 kinetics can be identified in isotopic envelopes as their distribution becomes bimodal[33,34]. It is recommended to exclude such peptides from PyHDX analysis. Even without cooperative exchange, when $k_{close} << k_{int}$, the approximations made to derive equations 3 and 4 break down, thereby introducing errors. For highly disordered parts of the protein, PyHDX therefore only informs on differential dynamics qualitatively.

To find ΔG values which best describe the data, we formulate a Lagrangian (cost function), composed of the Mean Square Error (MSE) and two regularizers (Lagrange Multipliers)[17,18] (Figure 1d), that constrain the possible solutions of ΔG.

As amino acids of a peptide each have varying intrinsic H/D exchange rates that HDX-MS cannot determine directly, similar values of the Lagrangian can be satisfied with multiple combinations of ΔG assignments per amino acid, leading to non-identifiability[4]. To alleviate this, a regularizer λ_1 acts along the primary structure minimizing differences in ΔG between consecutive residues, unless overruled by experimental D-uptake values that support such differences.

Given a protein with N_r residues r and an HDX-MS experiment which yielded N_p peptides p at N_t timepoints t, each with an associated measured deuterium uptake $D_{\pi\tau}$ (Table S2), the Lagrangian is:

$$\mathcal{L}(\Delta G, \lambda_1) = \frac{1}{N_p N_t} \sum_{\pi\tau} \left[D_{\pi\tau} - \sum_{\rho} X_{\pi\rho} \left(1 - \exp\left\{ \frac{-k_{int,\rho} t_\tau}{1 + e^{\frac{\Delta G_\rho}{RT}}} \right\} \right) \right]^2 + \lambda_1 h(\Delta G) \tag{6}$$

With:

$$h(\Delta G) = \frac{\lambda}{N_r} \sum_{\rho=1}^{N_r - 1} |\Delta G_\rho - \Delta G_{\rho+1}| \tag{7}$$

Where $D_{\pi\tau}$ is the corrected D-uptake and X is a 'coupling matrix' describing to which residues each peptide corresponds:

$$X_{\pi\rho} = \begin{cases} 1 & , r_\rho \in p_\pi \\ 0 & , \text{otherwise} \end{cases} \quad (8)$$

Such that its elements are 1 when the corresponding residues are found in a given peptide.

HDX-MS can inform on a protein's dynamic response to external triggers, e.g. changes in oligomerization, mutations or ligands[35]. Such differential HDX-MS experiments[3,36], compare D-uptake of reference and test states. As PyHDX derives single ΔG/residue, differential dynamics can be obtained by simply subtracting two datasets without the need for comparison of matching peptides.

However, experimental variables (e.g. proteolysis, exchange timepoints and/or k_{int}) can lead to artefactual "differences" between datasets. To alleviate this, a second regularizer (λ_2) operates along the sample axis, minimizing differences between identical residues across datasets, unless experimental data support such a difference (Figure 1d).

When expanding to global fitting of N_s HDX-MS samples (e.g. liganded, oligomers, homologues, pH/temperature variation) the Lagrangian becomes:

$$\mathcal{L}(\Delta G, \lambda_1, \lambda_2) = \frac{1}{N_s N_p N_t} \sum_{\sigma\pi\tau} \left[D_{\sigma\pi\tau} - \sum_\rho X_{\sigma\pi\rho} \left(1 - \exp\left\{ \frac{-k_{int,\sigma\rho} t_{\sigma\tau}}{1 + e^{\frac{\Delta G_{\sigma\rho}}{RT}}} \right\} \right) \right]^2 + \lambda_1 h(\Delta G) + \lambda_2 g(\Delta G) \quad (9)$$

With

$$h(\Delta G) = \frac{1}{N_s N_r} \sum_\sigma \sum_{\rho=1}^{N_r - 1} |\Delta G_{\sigma\rho} - \Delta G_{\sigma(\rho+1)}| \quad (10)$$

And the second regularizer is given by:

$$g(\Delta G) = \frac{1}{N_s N_r} \sum_{\sigma\rho} |\overline{\Delta G}_\rho - \Delta G_{\sigma\rho}| \quad (11)$$

Covariances are obtained from the diagonal elements of the inverse of the Hessian, calculated from χ^2:

$$[\chi^2(\Delta G_i), \chi^2(\Delta G_i)] = \sqrt{\left|(-\mathcal{H})^{-1}_{ii}\right|} \quad \text{with} \quad \mathcal{H}_{ij} = \frac{\partial^2 \chi^2}{\partial(\Delta G_i)\partial(\Delta G_j)} \quad (12)$$

Here the Lagrangian is used without regualizer terms. Covariances reflect the shape of the Lagrangian landscape; higher covariances indicate a flat landscape and therefore represent more difficulty to locate minima.

Implementation

In PyHDX, all quantities are PyTorch[37] Tensors or Numpy[38] arrays, with shapes as indicated in Table S3, such that the Lagrangian can be computed through matrix multiplications, where in the 3D case multiplication is done in batch along the first axis, according to Python's PEP465 convention.

For minimization of the Lagrangian, we use the PyTorch machine learning framework, such that its *autograd* automatic differentiation engine can be used to accelerate the process. The Stochastic-Gradient Descent (SGD) method is used by default.

To further ensure convergence to the correct solution, the ΔG is initialized with guess values obtained from weighted averaging (by inverse peptide length) of all peptides for a given timepoint. This procedure yields a kinetic uptake curve per residue from which apparent exchange rates are determined.

PyHDX was built on top of the scientific python ecosystem. Computation is done using the packages numpy[38], pandas[39], scipy[40], scikit-image[41] and symfit[42]. Fitting of ΔG is implemented on the machine learning platform PyTorch[37]. Computationally intensive tasks are scheduled to be processed in parallel through Dask[43]. Intrinsic exchange rates are calculated as previously described[24–27] and implemented by HDXRate[44]. Graphical output is generated with either matplotlib[45], ProPlot[46] or bokeh[47]. PyHDX features an API for data analysis in Jupyter notebooks[48] and a web application implemented in panel[49] using NGL[50,51] to visualize proteins.

Results and Discussion

We assessed PyHDX on the tetrameric *E. coli* chaperone SecB (*ec*SecB$_4$)[52]. D-uptake was measured across six timepoints (10 sec-100 min; 30°C, pH$_{read}$ 8; peptide heatmap (t=30s; Figure 1b; all timepoints in Figure S1). PyHDX-calculated Gibbs free energies for all residues (Figure 1e; regularizer $\lambda_1 = 2$). Fit curves for each peptide can be autogenerated in a pdf report file and are shown in Figure S3. Setting λ_1 sufficiently low allows the algorithm to extract features at high-resolution (Figure S4). Covariances are shown as error bars,

where a high covariance indicates that the protein's flexibility in these regions lies outside of the range of ΔG values resolved by the experiment (as determined by temperature, pH and particularly timepoints). In these regions, no sufficient change in D-uptake values is measured over the duration of the experiment to be able to accurately determine its exchange kinetics, either because the peptide is (almost) fully deuterated at the first timepoint or the peptide is (almost) fully undeuterated at the last timepoint. The obtained energies were classified into three regimes of relative flexibility: 'rigid' (40kJ mol^{-1}, blue), 'flexible', (25 kJ mol^{-1}, green) and 'hyper-flexible' (10 kJ mol^{-1}, red), and assigned colours by linear interpolation between them. The resulting dynamics landscape was visualized onto the structure of *ec*$SecB_4$ (Figure 1g).

To highlight the importance of overlapping residues to resolve ΔG per residue, we randomly sub-sampled the peptides in the *ec*SecB dataset thrice with different fractions of the original dataset (0.75, 0.50 and 0.25, Figure S5). The figure shows that the overall ΔG profile remains well resolved even at 50% of the original peptides. However, as the peptide overlap decreases some high-resolution features are lost. For example, in Figure S5 IIIc, residues 60-75, only a single peptide is present while from the temporal information residues with different degrees of flexibility are resolved. As there is only one peptide, it is impossible to determine to which residues the different ΔG values should be assigned. When more peptides are added, the region is resolved into a 'rigid' and 'flexible' part.

Therefore, when interpreting ΔG results obtained by PyHDX, it is important to refer to the original peptide coverage map to verify if sufficient peptide overlap is present to support biological conclusions relying on (near) residue-level resolution.

ΔG values were also obtained for the dimeric mutant *ec*$SecB_2$ (Figure 2f; $\lambda_1 = 2$, $\lambda_2 = 2$) and values from the two states (ΔΔG) were subtracted and coloured differently (10kJ mol^{-1}, dark purple-increased rigidity; 0 kJ mol^{-1}, white-no change; 10 kJ mol^{-1}; dark green-increased dynamics). In *ec*$SecB_4$, the internalized multimerization helix and the preceding β4 strand are most rigidified. ΔΔG values of near zero indicate that data is insufficient to substantiate dynamics differences between two states. Specifically, in regions of extreme dynamics (e.g. highly disordered/rigid), experimental D-exchange timepoints must be chosen to adequately resolve differences if present. In the absence of such measurements, the regularizer

λ_2 ensures that ΔΔG values in these regions show no change (e.g. *ec*SecB$_4$ C-tail, residues 140-160, Figure S6).

Next, we tested the dynamics of the 901-residue SecA (Figure 2a), in 6 different biochemical contexts (monomer, dimer, ΔC-tail; +/- ADP)[21]. Despite the larger computational challenge (8883 peptide-timepoints; 5364 fit parameters), PyHDX converged to a solution within 6.5 minutes. PyHDX fits several states in parallel, at manageable computation times (e.g. 25 minutes on an i7-9750H CPU to test 26 conditions, 23244 fit parameters; not shown). ΔG and ΔΔG values were mapped onto linear maps of SecA$_1$ or SecA$_2$ or SecA-ΔC$_2$ (Figure 2a) and on the SecA$_1$ structure (Figures 2a and 2b). Nucleotide binding decreased dynamics/increased ΔΔG mainly in helicase motifs[21,53] (Figure 2c).

We next compared two structural homologues, analysed at different experimental conditions: *ec*SecB and *mt*SecB, from *M. tuberculosis* showing modest sequence conservation (13% identity/27% similarity), measured at 20 °C, pD 6, 4 timepoints 10s – 30 min[54]. When aligned by secondary structure[54], both flexibility profiles show a large degree of similarity (Mean absolute ΔG difference 8.2 kJ/mol, Figure S7a). Initial apparent differences decrease after applying a small value of λ_2 = 1 between aligned residues (Figure 2d, Figure S7b, Figure S8; 4.7 kJ/mol). This effect is most apparent at the ΔG resolution limit of the experiment (Figure S7, yellow shaded regions) where experimental differences (pH, temperature) would result in artefactual ΔG differences between the proteins, despite being unsubstantiated by the experimental data. The regularizer λ_2 can thus be used to select for the most significant differences between datasets (see also Figure S6).

A Clustal sequence alignment yielded a similar result (Figure S7c). We ensured that λ_2 did not artefactually remove differences, by confirming that flexibility differences remain in non-aligned sequences (Figure S7d). These observations imply that protein flexibility might be evolutionarily selected for prior to individual amino acids. This conservation of flexibility profiles is a common feature in protein superfamilies[55–57].

To test the applicability of PyHDX to a wider protein space, we derived ΔG/residue values for a total of 9 proteins and visualized their collective distribution independently of sequence position (Figure 2f). ΔG values per residue indicate a wide range of flexibility (0-40 kJ mol^{-1}) being distributed within each protein in

two, rarely three, distinct populations, predominantly a rigid and a hyper-flexible one (Figure 2e), that quantitatively shift between them when additional interactions take place (e.g. dimerization; Figure 2f).

Conclusions

In summary, PyHDX rapidly processes single and multiple HDX-MS datasets and visualizes residue-level Gibbs free energies on linear and 3D structures. Residue level energies open up previously unavailable possibilities for evolutionary, structural and functional studies and a universal description of protein flexibility. PyHDX is fully open source and its and documentation are available online. A companion interactive web application returns results within minutes, allowing users to get feedback on obtained fit results and directly interact with and explore their data. The PyHDX infrastructure established as part of this work will allow us to implement additional models which go beyond steady-state EX2 approximation and the Linderstrøm-Lang model. Future development of PyHDX could include extending models taking into account EX1 kinetics[29,58] and temperature dependence[59]. We anticipate that these updates will allow us to extract more information on the cooperative behaviour of global- and local unfolding events, which lie at the basis of H/D exchange kinetics.

Figures

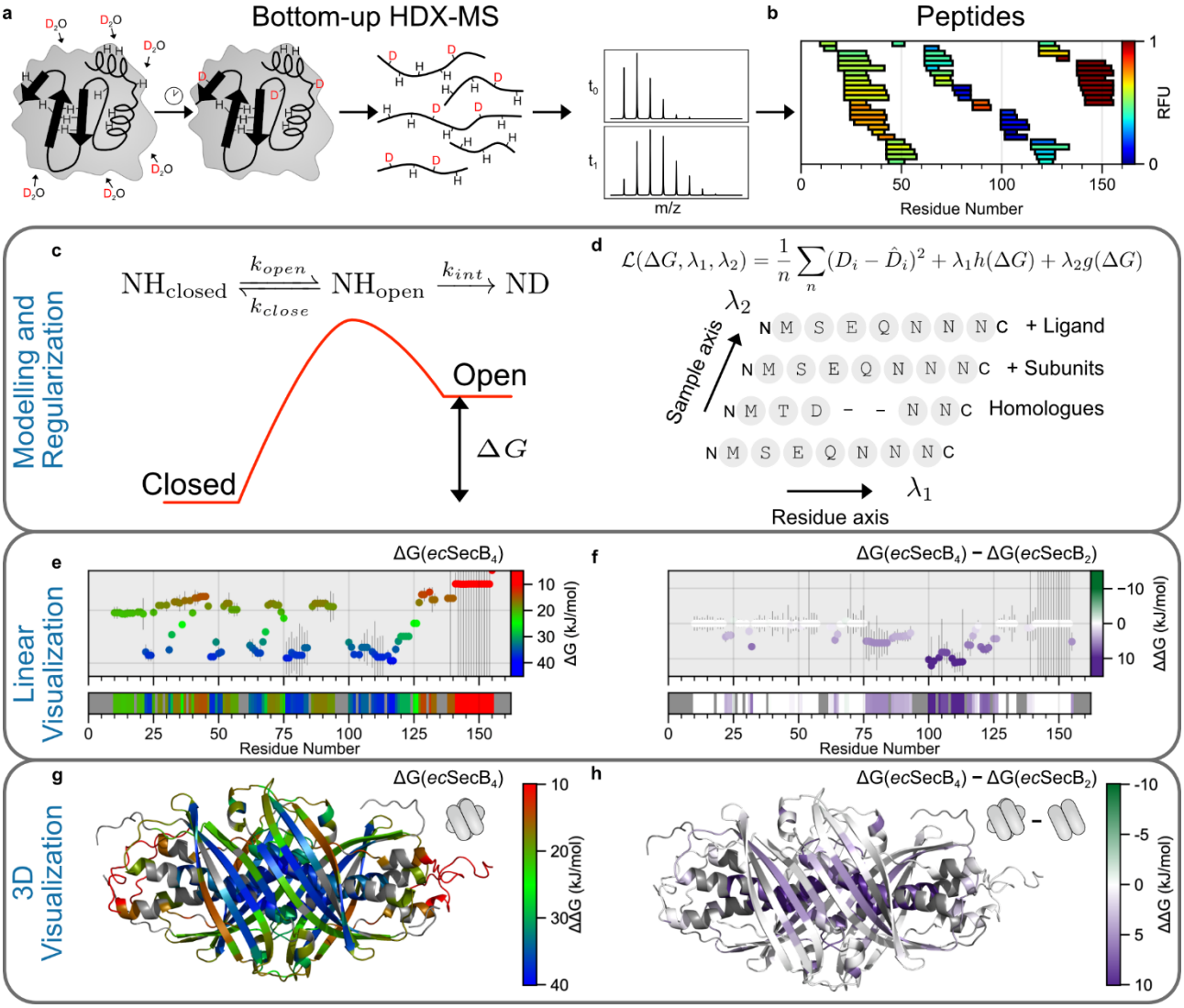

Figure 1 | Pipeline of a bottom up HDX-MS experiment (top) analysed with PyHDX (bottom). a, HDX-MS experimental workflow consisting of deuterium exposure, quenching and pepsin digestion, and identification of peptides by LC/MS[60]. **b,** Overlapping peptides obtained as in **a** for *ec*SecB (UniProt P0AG86) showing a D-uptake heatmap relative to a fully deuterated control sample for all derivative peptides at t=30 s. **c,** Linderstrøm-Lang model[22,23] of H/D exchange used in PyHDX to describe exchange in terms of Gibbs free energy (ΔG) between the closed non-exchanging and the open exchange-competent states. **d,** Two regularizers (λ_1, λ_2) are applied across two axes: the residue axis (λ_1), minimizing variation in ΔG between consecutive residues, and the sample axis (λ_2), minimizing variation in ΔG between residues along a set of HDX conditions (e.g. ligands, oligomeric state) or homologues. **e**, Output ΔG values per residue plotted against the linear sequence of *ec*SecB (top) coloured according to a gradient colour map (right). Residue colours are additionally shown as a linear bars (bottom), regions without peptide coverage are coloured grey. Error bars are covariances (see Methods). **f**, Differential dynamics between *ec*SecB$_4$ and *ec*SecB$_2$ shown as differences in ΔG of both states (ΔΔG). Regions in purple are rigidifying in *ec*SecB$_4$ compared to the dimeric state **g**, 3D structure of *ec*SecB$_4$ coloured according to ΔG/residue from **e**. (PDB 5JTR[61], ligand removed) **h**, 3D structure of *ec*SecB$_4$ coloured according to the per residue ΔΔG from **f**.

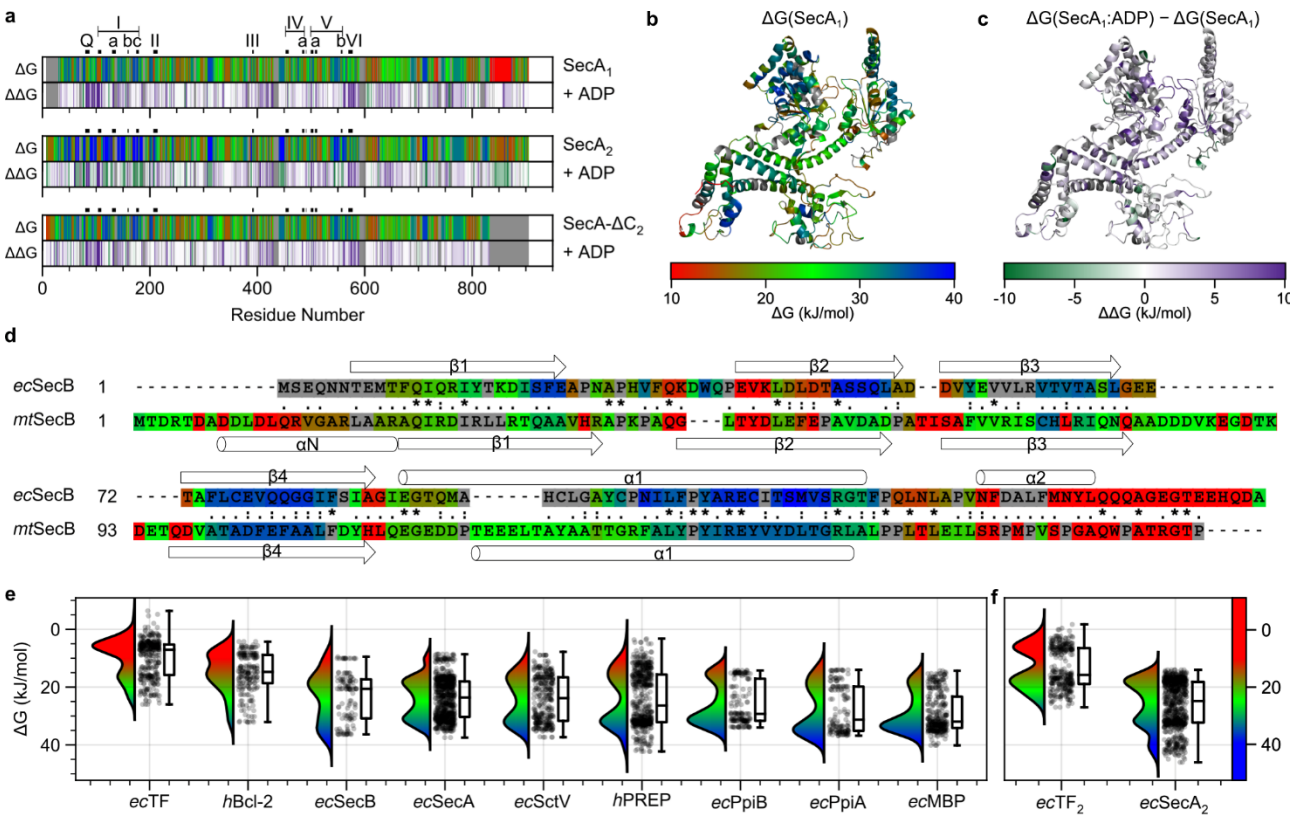

Figure 2 | PyHDX analysis applied to a wider protein space. a, Protein flexibility of monomeric, dimeric and C-tail deleted SecA as coloured linear bars (ΔG, top bars; blue: rigid/red:flexible) and their differential dynamics upon ADP binding (ΔΔG, bottom bars; ADP-rigidified regions in purple). Helicase motifs (Q, I, Ia, Ib, Ic, II, III, IV, IVa, V, Va, Vb and VI; critical regions in ATP hydrolysis and function)[21,53] are indicated. **b**, Gibbs free energies (ΔG) of protein flexibility from HDX-MS for $SecA_1$ apoprotein mapped onto its 3D structure (PDB ID 2VDA, ligand removed[62]). **c**, ADP-driven differential dynamics in monomeric SecA shown as differences in ΔG (ΔΔG). Regions in purple rigidify upon ADP binding. **d**, Alignment of *ec*SecB (top) and *mt*SecB (bottom) based on both sequence alignment and secondary structure as performed in[54]. Residue similarity is indicated in the middle: * = identical residues, : = strongly similar, . = weakly similar (according to the Gonnet PAM 250 matrix). Colours indicate ΔG/residue. **e** and **f,** Raincloud plots[63] of ΔG/residue for the indicated proteins (**e**) and two dimeric derivatives (**f**).

Code Availability

Code for PyHDX analysis software is open source and released under the MIT license and available at: www.github.com/jhsmit/pyhdx.

Code for generating figure panels in the paper will be made available on publication.

Data Availability

*ec*SecB HDX-MS state data is available on the PyHDX GitHub repository. SecA HDX-MS state data was published previously[21].

Supporting Information

Supplementary figures S1-S8: Example of an HDX-MS dataset, screenshot of PyHDX web application, peptides uptake curves and fit curves, influence of regularizers λ_1 and λ_2, influence of peptide redundancy on ΔG, effect of different alignments between *ec*SecB and *mt*SecB. Supplementary Tables 1-3: Primers for generating *ec*$SecB_2$, Lagrangian symbols, indices, variables and shapes.

Acknowledgements

We are grateful to: J. Marcoux, P. Geneveux and L. Mourey for generously sharing *mt*SecB HDX-MS data; J. Claesen for discussions; A.Portaliou for constructs; the open source software community for their support and advice. Research in our lab was funded by grants (to AE): ProFlow (FWO/F.R.S.-FNRS "Excellence of Science - EOS" programme grant #30550343), CARBS (#G0C6814N; FWO) and PROFOUND (WOG; W002421N; FWO) and (to AE and SK): FOscil (ZKD4582 - C16/18/008; KU Leuven). SKr was a FWO [PEGASUS]^2 MSC postdoctoral fellow. JHS is a PDM, KU Leuven postdoctoral fellow. This project has received funding from the Research Foundation – Flanders (FWO) and the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 665501.

Competing Interests

The authors declare they have no competing financial interests or other conflicts of interest.

Author Contributions

JHS conceived all mathematical analysis and developed and implemented software and web interface. SKr, BYS, RP and SK provided HDX-MS data and analysis. SKr guided optimization of software parameters and validated output. JHS wrote the first draft with contributions from SKr and AE. All authors reviewed and approved the final manuscript. AE conceived and managed the project.

References

(1) Karplus, M.; Kuriyan, J. Molecular Dynamics and Protein Function. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (19), 6679–6685. https://doi.org/10.1073/pnas.0408930102.

(2) Engen, J. R.; Botzanowski, T.; Peterle, D.; Georgescauld, F.; Wales, T. E. Developments in Hydrogen/Deuterium Exchange Mass Spectrometry. *Anal. Chem.* **2021**, *93* (1), 567–582. https://doi.org/10.1021/acs.analchem.0c04281.

(3) Masson, G. R.; Burke, J. E.; Ahn, N. G.; Anand, G. S.; Borchers, C.; Brier, S.; Bou-Assaf, G. M.; Engen, J. R.; Englander, S. W.; Faber, J.; Garlish, R.; Griffin, P. R.; Gross, M. L.; Guttman, M.; Hamuro, Y.; Heck, A. J. R.; Houde, D.; Iacob, R. E.; Jørgensen, T. J. D.; Kaltashov, I. A.; Klinman, J. P.; Konermann, L.; Man, P.; Mayne, L.; Pascal, B. D.; Reichmann, D.; Skehel, M.; Snijder, J.; Strutzenberg, T. S.; Underbakke, E. S.; Wagner, C.; Wales, T. E.; Walters, B. T.; Weis, D. D.; Wilson, D. J.; Wintrode, P. L.; Zhang, Z.; Zheng, J.; Schriemer, D. C.; Rand, K. D. Recommendations for Performing, Interpreting and Reporting Hydrogen Deuterium Exchange Mass Spectrometry (HDX-MS) Experiments. *Nat. Methods* **2019**, *16* (7), 595–602. https://doi.org/10.1038/s41592-019-0459-y.

(4) Claesen, J.; Burzykowski, T. Computational Methods and Challenges in Hydrogen/Deuterium Exchange Mass Spectrometry. *Mass Spectrom. Rev.* **2017**, *36* (5), 649–667. https://doi.org/10.1002/mas.21519.

(5) Oganesyan, I.; Lento, C.; Wilson, D. J. Contemporary Hydrogen Deuterium Exchange Mass Spectrometry. *Methods* **2018**, *144*, 27–42. https://doi.org/10.1016/j.ymeth.2018.04.023.

(6) Masson, G. R.; Maslen, S. L.; Williams, R. L. Analysis of Phosphoinositide 3-Kinase Inhibitors by Bottom-up Electron-Transfer Dissociation Hydrogen/Deuterium Exchange Mass Spectrometry. *Biochem. J.* **2017**, *474* (11), 1867–1877. https://doi.org/10.1042/BCJ20170127.

(7) 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* (7), 1202–1208. https://doi.org/10.1007/s13361-012-0373-3.

(8) Walters, B. T.; Ricciuti, A.; Mayne, L.; Englander, S. W. Minimizing Back Exchange in the Hydrogen Exchange-Mass Spectrometry Experiment. *J. Am. Soc. Mass Spectrom.* **2012**, *23* (12), 2132–2139. https://doi.org/10.1021/jasms.8b04189.

(9) Zehl, M.; Rand, K. D.; Jensen, O. N.; Jørgensen, T. J. D. Electron Transfer Dissociation Facilitates the Measurement of Deuterium Incorporation into Selectively Labeled Peptides with Single Residue Resolution. *J. Am. Chem. Soc.* **2008**, *130* (51), 17453–17459. https://doi.org/10.1021/ja805573h.

(10) 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 Bioinformatics* **2010**, *11* (1), 424. https://doi.org/10.1186/1471-2105-11-424.

(11) Hamuro, Y.; Zhang, T. High-Resolution HDX-MS of Cytochrome c Using Pepsin/Fungal Protease Type XIII Mixed Bed Column. *J. Am. Soc. Mass Spectrom.* **2019**, *30* (2), 227–234. https://doi.org/10.1007/s13361-018-2087-7.

(12) Kan, Z.; Ye, X.; Skinner, J. J.; Mayne, L.; Englander, S. W. ExMS2: An Integrated Solution for Hydrogen–Deuterium Exchange Mass Spectrometry Data Analysis. *Anal. Chem.* **2019**, *91* (11), 7474–7481. https://doi.org/10.1021/acs.analchem.9b01682.

(13) Gessner, C.; Steinchen, W.; Bédard, S.; J. Skinner, 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* (1), 3789. https://doi.org/10.1038/s41598-017-03922-3.

(14) 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* (7), 1194–1203. https://doi.org/10.1016/j.bpj.2019.02.024.

(15) 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 Spectrometry: Application to Characterizing Protein–Ligand Interactions. *J. Phys. Chem. B* **2017**, *121* (15), 3493–3501. https://doi.org/10.1021/acs.jpcb.6b09358.

(16) Salmas, R. E.; Borysik, A. J. HDXmodeller: An Online Webserver for High-Resolution HDX-MS with Auto-Validation. *Commun. Biol.* **2021**, *4* (1), 1–8. https://doi.org/10.1038/s42003-021-01709-x.

(17) Zhang, Z.; Zhang, A.; Xiao, G. Improved Protein Hydrogen/Deuterium Exchange Mass Spectrometry Platform with Fully Automated Data Processing. *Anal. Chem.* **2012**, *84* (11), 4942–4949. https://doi.org/10.1021/ac300535r.

(18) Zhang, Z. Complete Extraction of Protein Dynamics Information in Hydrogen/Deuterium Exchange Mass Spectrometry Data. *Anal. Chem.* **2020**, *92* (9), 6486–6494. https://doi.org/10.1021/acs.analchem.9b05724.

(19) De Geyter, J.; Portaliou, A. G.; Srinivasu, B.; Krishnamurthy, S.; Economou, A.; Karamanou, S. Trigger Factor Is a Bona Fide Secretory Pathway Chaperone That Interacts with SecB and the Translocase. *EMBO Rep.* **2020**, *21* (6), e49054. https://doi.org/10.15252/embr.201949054.

(20) Tsirigotaki, A.; Elzen, R. V.; Veken, P. V. D.; Lambeir, A.-M.; Economou, A. Dynamics and Ligand-Induced Conformational Changes in Human Prolyl Oligopeptidase Analyzed by Hydrogen/Deuterium Exchange Mass Spectrometry. *Sci. Rep.* **2017**, *7* (1), 2456. https://doi.org/10.1038/s41598-017-02550-1.

(21) Krishnamurthy, S.; Eleftheriadis, N.; Karathanou, K.; Smit, J. H.; Portaliou, A. G.; Chatzi, K. E.; Karamanou, S.; Bondar, A.-N.; Gouridis, G.; Economou, A. A Nexus of Intrinsic Dynamics Underlies Translocase Priming. *Structure* **2021**. https://doi.org/10.1016/j.str.2021.03.015.

(22) Linderstrøm-Lang, K. Deuterium Exchange between Peptides and Water. *Chem Soc Spec Publ* **1955**, *2*, 1–20.

(23) Englander, S. W.; Sosnick, T. R.; Englander, J. J.; Mayne, L. Mechanisms and Uses of Hydrogen Exchange. *Curr. Opin. Struct. Biol.* **1996**, *6* (1), 18–23. https://doi.org/10.1016/S0959-440X(96)80090-X.

(24) Bai, Y.; Milne, J. S.; Mayne, L.; Englander, S. W. Primary Structure Effects on Peptide Group Hydrogen Exchange. *Proteins Struct. Funct. Bioinforma.* **1993**, *17* (1), 75–86. https://doi.org/10.1002/prot.340170110.

(25) Connelly, G. P.; Bai, Y.; Jeng, M.-F.; Englander, S. W. Isotope Effects in Peptide Group Hydrogen Exchange. *Proteins Struct. Funct. Genet.* **1993**, *17* (1), 87–92. https://doi.org/10.1002/prot.340170111.

(26) Mori, S.; Zijl, P. C. M. van; Shortle, D. Measurement of water–amide proton exchange rates in the denatured state of staphylococcal nuclease by a magnetization transfer technique. *Proteins Struct. Funct. Bioinforma.* **1997**, *28* (3), 325–332. https://doi.org/10.1002/(SICI)1097-0134(199707)28:3<325::AID-PROT3>3.0.CO;2-B.

(27) Nguyen, D.; Mayne, L.; Phillips, M. C.; Walter Englander, S. Reference Parameters for Protein Hydrogen Exchange Rates. *J. Am. Soc. Mass Spectrom.* **2018**, *29* (9), 1936–1939. https://doi.org/10.1021/jasms.8b05911.

(28) Zhang, Z.; Smith, D. L. Determination of Amide Hydrogen Exchange by Mass Spectrometry: A New Tool for Protein Structure Elucidation. *Protein Sci. Publ. Protein Soc.* **1993**, *2* (4), 522–531.

(29) Hvidt, A.; Nielsen, S. O. Hydrogen Exchange in Proteins. In *Advances in Protein Chemistry*; Anfinsen, C. B., Anson, M. L., Edsall, J. T., Richards, F. M., Eds.; Academic Press, 1966; Vol. 21, pp 287–386. https://doi.org/10.1016/S0065-3233(08)60129-1.

(30) Roder, H.; Elove, G. A.; Englander, S. W. Structural Characterization of Folding Intermediates in Cytochrome c by Ff-Exchange Labelling and Proton NMR. 5.

(31) Englander, S. W.; Kallenbach, N. R. Hydrogen Exchange and Structural Dynamics of Proteins and Nucleic Acids. *Q. Rev. Biophys.* **1983**, *16* (4), 521–655. https://doi.org/10.1017/S0033583500005217.

(32) Ferraro, D. M.; Lazo, N. D.; Robertson, A. D. EX1 Hydrogen Exchange and Protein Folding. *Biochemistry* **2004**, *43* (3), 587–594. https://doi.org/10.1021/bi035943y.

(33) Xiao, H.; Hoerner, J. K.; Eyles, S. J.; Dobo, A.; Voigtman, E.; Mel'čuk, A. I.; Kaltashov, I. A. Mapping Protein Energy Landscapes with Amide Hydrogen Exchange and Mass Spectrometry: I. A Generalized Model for a Two-State Protein and Comparison with Experiment. *Protein Sci.* **2005**, *14* (2), 543–557. https://doi.org/10.1110/ps.041001705.

(34) Weis, D. D.; Wales, T. E.; Engen, J. R.; Hotchko, M.; Ten Eyck, L. F. Identification and Characterization of EX1 Kinetics in H/D Exchange Mass Spectrometry by Peak Width Analysis. *J. Am. Soc. Mass Spectrom.* **2006**, *17* (11), 1498–1509. https://doi.org/10.1016/j.jasms.2006.05.014.

(35) Pirrone, G. F.; Iacob, R. E.; Engen, J. R. Applications of Hydrogen/Deuterium Exchange MS from 2012 to 2014. *Anal. Chem.* **2015**, *87* (1), 99–118. https://doi.org/10.1021/ac5040242.

(36) 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* (1), 43–59. https://doi.org/10.1586/epr.10.109.

(37) Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; Desmaison, A.; Kopf, A.; Yang, E.; DeVito, Z.; Raison, M.; Tejani, A.; Chilamkurthy, S.; Steiner, B.; Fang, L.; Bai, J.; Chintala, S. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems 32*; Wallach, H., Larochelle, H., Beygelzimer, A., Alché-Buc, F. d\textquotesingle, Fox, E., Garnett, R., Eds.; Curran Associates, Inc., 2019; pp 8024–8035.

(38) Harris, C. R.; Millman, K. J.; van der Walt, S. J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N. J.; Kern, R.; Picus, M.; Hoyer, S.; van Kerkwijk, M. H.; Brett, M.; Haldane, A.; del Río, J. F.; Wiebe, M.; Peterson, P.; Gérard-Marchant, P.; Sheppard, K.; Reddy, T.; Weckesser, W.; Abbasi, H.; Gohlke, C.; Oliphant, T. E. Array Programming with NumPy. *Nature* **2020**, *585* (7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2.

(39) team, T. pandas development. *Pandas-Dev/Pandas: Pandas*; Zenodo, 2020. https://doi.org/10.5281/zenodo.3509134.

(40) Virtanen, P.; Gommers, R.; Oliphant, T. E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; van der Walt, S. J.; Brett, M.; Wilson, J.; Jarrod Millman, K.; Mayorov, N.; Nelson, A. R. J.; Jones, E.; Kern, R.; Larson, E.; Carey, C.; Polat, İ.; Feng, Y.; Moore, E. W.; Vand erPlas, J.; Laxalde, D.; Perktold, J.; Cimrman, R.; Henriksen, I.; Quintero, E. A.; Harris, C. R.; Archibald, A. M.; Ribeiro, A. H.; Pedregosa, F.; van Mulbregt, P.; Contributors, S. 1. 0. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods* **2020**. https://doi.org/10.1038/s41592-019-0686-2.

(41) Walt, S. van der; Schönberger, J. L.; Nunez-Iglesias, J.; Boulogne, F.; Warner, J. D.; Yager, N.; Gouillart, E.; Yu, T. Scikit-Image: Image Processing in Python. *PeerJ* **2014**, *2*, e453. https://doi.org/10.7717/peerj.453.

(42) Martin Roelfs; Peter C Kroon. *TBuLi/Symfit: Symfit 0.5.3*; Zenodo, 2020. https://doi.org/10.5281/zenodo.4020630.

(43) Dask Development Team. *Dask: Library for Dynamic Task Scheduling*; 2016.

(44) Jochem Smit. *Jhsmit/HDXrate: Zenodo Release*; Zenodo, 2021. https://doi.org/10.5281/zenodo.4630881.

(45) Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **2007**, *9* (3), 90–95.

(46) Davis, L. L. B. *ProPlot*; Zenodo, 2021. https://doi.org/10.5281/zenodo.5234192.

(47) Bokeh Development Team. *Bokeh: Python Library for Interactive Visualization*; 2020.

(48) Kluyver, T.; Ragan-Kelley, B.; Pérez, F.; Granger, B.; Bussonnier, M.; Frederic, J.; Kelley, K.; Hamrick, J.; Grout, J.; Corlay, S.; Ivanov, P.; Avila, D.; Abdalla, S.; Willing, C.; Jupyter development team. Jupyter Notebooks – a Publishing Format for Reproducible Computational Workflows. In *Positioning and Power in Academic Publishing: Players, Agents and Agendas*; Loizides, F., Scmidt, B., Eds.; IOS Press, 2016; pp 87–90. https://doi.org/Kluyver, Thomas, Ragan-Kelley, Benjamin, Pérez, Fernando, Granger, Brian, Bussonnier, Matthias, Frederic, Jonathan, Kelley, Kyle, Hamrick, Jessica, Grout, Jason, Corlay, Sylvain, Ivanov, Paul, Avila, Damián, Abdalla, Safia, Willing, Carol and Jupyter development team, (2016) Jupyter Notebooks – a publishing format for reproducible computational workflows. Loizides, Fernando and Scmidt, Birgit (eds.) In Positioning and Power in Academic Publishing: Players, Agents and Agendas. IOS Press. pp. 87-90 . (doi:10.3233/978-1-61499-649-1-87 <http://dx.doi.org/10.3233/978-1-61499-649-1-87>).

(49) Philipp Rudiger; Xavier Artusi; James A. Bednar; Marc Skov Madsen; Chris B; Julia Signell; Hoxbro; Jean-Luc Stevens; Maxime Liquet; Jon Mease; Andrew; Arne; Mateusz Paprocki; kbowen; Ed Jung; Hugues-Yanis Amanieu; Jacob Barhak; Julius Winkelmann; Leopold Talirz; Pav A; Achim Randelhoff; Bane Sullivan; Dhruv Balwada; Nestor Ghenzi; Ryan Morshead; hoseppan; kleavor; miliante; Ravi Mulpuri. *Holoviz/Panel: Version 0.11.3*; Zenodo, 2021. https://doi.org/10.5281/zenodo.4692827.

(50) Rose, A. S.; Hildebrand, P. W. NGL Viewer: A Web Application for Molecular Visualization. *Nucleic Acids Res.* **2015**, *43* (W1), W576–W579. https://doi.org/10.1093/nar/gkv402.

(51) Rose, A. S.; Bradley, A. R.; Valasatava, Y.; Duarte, J. M.; Prlić, A.; Rose, P. W. NGL Viewer: Web-Based Molecular Graphics for Large Complexes. *Bioinformatics* **2018**, *34* (21), 3755–3758. https://doi.org/10.1093/bioinformatics/bty419.

(52) Sala, A.; Bordes, P.; Genevaux, P. Multitasking SecB Chaperones in Bacteria. *Front. Microbiol.* **2014**, *5*. https://doi.org/10.3389/fmicb.2014.00666.

(53) Papanikou, E.; Karamanou, S.; Economou, A. Bacterial Protein Secretion through the Translocase Nanomachine. *Nat. Rev. Microbiol.* **2007**, *5* (11), 839–851. https://doi.org/10.1038/nrmicro1771.

(54) Guillet, V.; Bordes, P.; Bon, C.; Marcoux, J.; Gervais, V.; Sala, A. J.; Dos Reis, S.; Slama, N.; Mares-Mejía, I.; Cirinesi, A.-M.; Maveyraud, L.; Genevaux, P.; Mourey, L. Structural Insights into Chaperone Addiction of Toxin-Antitoxin Systems. *Nat. Commun.* **2019**, *10* (1), 782. https://doi.org/10.1038/s41467-019-08747-4.

(55) Echave, J. Why Are the Low-Energy Protein Normal Modes Evolutionarily Conserved? *Pure Appl. Chem.* **2012**, *84* (9), 1931–1937. https://doi.org/10.1351/PAC-CON-12-02-15.

(56) Micheletti, C. Comparing Proteins by Their Internal Dynamics: Exploring Structure-Function Relationships beyond Static Structural Alignments. *Phys. Life Rev.* **2013**, *10* (1), 1–26. https://doi.org/10.1016/j.plrev.2012.10.009.

(57) Zhang, S.; Li, H.; Krieger, J. M.; Bahar, I. Shared Signature Dynamics Tempered by Local Fluctuations Enables Fold Adaptability and Specificity. *Mol. Biol. Evol.* **2019**, *36* (9), 2053–2068. https://doi.org/10.1093/molbev/msz102.

(58) Pedersen, T. G.; Thomsen, N. K.; Andersen, K. V.; Madsen, J. Chr.; Poulsen, F. M. Determination of the Rate Constants K1 and K2 of the Linderstrøm-Lang Model for Protein Amide Hydrogen Exchange: A Study of the Individual Amides in Hen Egg-White Lysozyme. *J. Mol. Biol.* **1993**, *230* (2), 651–660. https://doi.org/10.1006/jmbi.1993.1176.

(59) Tajoddin, N. N.; Konermann, L. Analysis of Temperature-Dependent H/D Exchange Mass Spectrometry Experiments. *Anal. Chem.* **2020**. https://doi.org/10.1021/acs.analchem.0c01828.

(60) Vadas, O.; Burke, J. E. Probing the Dynamic Regulation of Peripheral Membrane Proteins Using Hydrogen Deuterium Exchange–MS (HDX–MS). *Biochem. Soc. Trans.* **2015**, *43* (5), 773–786. https://doi.org/10.1042/BST20150065.

(61) Huang, C.; Rossi, P.; Saio, T.; Kalodimos, C. G. Structural Basis for the Antifolding Activity of a Molecular Chaperone. *Nature* **2016**, *537* (7619), 202–206. https://doi.org/10.1038/nature18965.

(62) Gelis, I.; Bonvin, A. M. J. J.; Keramisanou, D.; Koukaki, M.; Gouridis, G.; Karamanou, S.; Economou, A.; Kalodimos, C. G. Structural Basis for Signal-Sequence Recognition by the Translocase Motor SecA as Determined by NMR. *Cell* **2007**, *131* (4), 756–769. https://doi.org/10.1016/j.cell.2007.09.039.

(63) Allen, M.; Poggiali, D.; Whitaker, K.; Marshall, T. R.; van Langen, J.; Kievit, R. A. Raincloud Plots: A Multi-Platform Tool for Robust Data Visualization. *Wellcome Open Res.* **2021**, *4*, 63. https://doi.org/10.12688/wellcomeopenres.15191.2.