Site-resolved energetic information from HX/MS experiments

Chenlin Lu[1]*, Malcolm L. Wells[1]*, Andrew Reckers[1], and Anum Glasgow[1]
[1] Department of Biochemistry and Molecular Biophysics, Columbia University Irving Medical Center, New York, NY, United States
* These authors made equal contributions.

Abstract

While bioinformatics reveals patterns in protein sequences and structural biology methods elucidate atomic details of protein structures, it is difficult to attain equally high-resolution energetic information about protein conformational ensembles. We present PIGEON-FEATHER, a method for calculating free energies of opening (ΔG_{op}) at single- or near-single-amino acid resolution for protein ensembles of all sizes from hydrogen exchange/mass spectrometry (HX/MS) data. PIGEON-FEATHER disambiguates and reconstructs all experimentally measured isotopic mass envelopes using a Bayesian Monte Carlo sampling approach. We applied PIGEON-FEATHER to reveal how *E. coli* and human dihydrofolate reductase orthologs (ecDHFR, hDHFR) have evolved distinct ensembles tuned to their catalytic cycles, and how two competitive inhibitors of ecDHFR arrest its ensemble in different ways. Extending the method to a large protein-DNA complex, we mapped ligand-induced ensemble reweighting in the *E. coli lac* repressor to understand the functional switching mechanism crucial for transcriptional regulation.

Proteins can perform biological functions because they adopt conformational ensembles: sets of rapidly interconverting conformational states with different populations and energies.[1–3] For a given process, the thermodynamic ensemble of a protein or residue has an ensemble energy determined by the distribution of conformational states that it occupies. To respond to molecular perturbations like ligand binding, proteins reweight their ensembles. Disease-causing mutations disrupt ensembles. Because of this key relationship between a protein's ensemble and function, measuring ensemble energies at amino acid resolution is crucial to deeply understanding protein behavior.

However, biophysical techniques are limited in their capacity to measure ensemble energies. X-ray crystallography and cryo-electron microscopy (cryo-EM) capture only the lowest-energy conformational states of proteins and do not report ensemble energies. Nuclear magnetic resonance (NMR) spectroscopy monitors shifts among conformational states but is limited by experimental throughput and protein size. Molecular dynamics (MD) simulations can reveal the details of the earliest stages in a conformational transition, but are constrained by simulation timescales, inaccuracies in force fields, and the availability of structural data.

We set out to measure amino acid-level ensemble energies using hydrogen exchange with mass spectrometry (HX/MS). With the advent of integrated robotics[4,5] and accurate protein structure prediction methods[6,7], HX/MS presents a powerful strategy for probing protein ensembles in solution. HX/MS measures the rate of exchange of labile hydrogen atoms with deuterium in a protein's backbone amide groups to report changes in its ensemble at the peptide level (Fig. 1a).[8] For hydrogen exchange to occur in a folded protein at equilibrium (EX2 conditions, see section 1 of the SI), the protein must transiently expose its backbone amide hydrogens to solvent, which incurs an energetic cost for each exchangeable site: the free energy of opening, or ΔG_{op} (Fig. 1b).[9] The ΔG_{op} is an ensemble energy derived from the hydrogen exchange rate.

Because of a lack of methods to determine exchange rates for each amino acid, HX/MS data are typically interpreted as relative differences in total hydrogen exchange between states (*e.g.*, apo vs. holo proteins) for selected peptides. This approach is flawed for three reasons. First, total hydrogen exchange is biased by the timecourse, because exchange half-lives can span years but the experiment is performed within hours.[10] Second, the effects of protein sequence on exchange rates are ignored. The ΔG_{op} depends on the intrinsic rate of exchange for each amino acid, or how it exchanges in an unstructured polypeptide (Fig. 1b), which depends on its own and its neighbors' identities[11]; but the peptide resolution of HX/MS data precludes assignments of amino acid-specific exchange rates. Third, the manner of curating, combining, and visualizing the peptide-level total hydrogen exchange data influences how we understand protein ensembles.[12,13]

We therefore developed PIGEON-FEATHER *(Peptide ID Generation for Exchange Of Nuclei - Free Energy Assignment Through Hydrogen Exchange Rates)*, an HX/MS analysis method to assign ΔG_{op} for almost all the amino acids in a protein (Fig. 1c). PIGEON-FEATHER disambiguates peptides based on tandem MS fragment coverage and fits hydrogen exchange rates via Bayesian sampling to reconstruct full isotopic mass envelopes, capitalizing on information-rich raw HX/MS data for ΔG_{op} determination. Benchmarking shows that PIGEON-FEATHER calculates exchange rates with near-perfect accuracy, consistently outperforming existing methods. We applied the method to bacterial and human enzyme orthologs to quantify how evolutionary divergence in ΔG_{op} permits catalysis in different organisms, and we measured how two inhibitors differentially impact the bacterial enzyme's ensemble to discern why only one of them is effective in the human ortholog. Finally, we used PIGEON-FEATHER to connect ligand-induced allosteric changes in the regulatory domain of a large transcription factor with local unfolding in its DNA-binding domain.

Results

Peptide disambiguation using PIGEON

In hydrogen-deuterium (H-D) exchange/MS experiments, following the exchange timecourse, proteins are proteolytically digested and the resulting peptides are measured for their level of D uptake. We first identify each MS peak by performing an MS/MS experiment on the undeuterated protein and matching the

monoisotopic masses and *y*- and *b*- ion fragments to theoretical masses based on the protein sequence. The observed mass, charge, and chromatographic retention time are used to identify the corresponding peaks in the deuterated sample spectra. However, the combination of monoisotopic mass and fragment coverage is frequently inadequate to uniquely specify one peptide (Fig. 2a-c).[14–16] This complicates the analysis because considering only peptides with strong fragment coverage can reduce sequence coverage, but including less well-supported peptides introduces the risk of spurious identifications (Fig. 2d).

PIGEON enables accurate peptide identification (Fig. 2e). First, PIGEON removes systematic errors from *e.g.* insufficient calibration (Figs. 2e, ii, S1a-c). Next, PIGEON removes all duplicate matches except for one consensus peak per charge state per peptide by identifying those with the highest fragment ion coverage and/or peak signal (Fig. S1d). Finally, PIGEON corrects double assignments in the peptide list by checking for their monoisotopic *m*/*z* and retention times, discarding those without fragment support (Fig. 2e, iii, 2g). For co-eluting peptides where both mass duplicates have strong fragment support, the user may decide to keep all, none, or only ambiguous peptides with substantial sequence overlap (Fig. 2e, iv, 2h). In this way, PIGEON allows flexible treatment for peptides for which the true assignment is ambiguous and the proportion of each peptide in the signal unknown, particularly those that are identical up to a small sequence difference. PIGEON uses sequence overlap as a discriminator because highly overlapping peptides typically report on the same set of amino acids, so uptake behavior is likely to be consistent. The remaining assignments form a peptide pool for HX analysis using FEATHER (Fig. 2e, v, 2h).

We evaluated how PIGEON impacts peptide assignments for HX/MS on *E. coli* dihydrofolate reductase (ecDHFR), an essential enzyme that reduces 7,8-dihydrofolate (DHF) to 5,6,7,8-tetrahydrofolate (THF) via NADPH-catalyzed hydride transfer. There are 2,107 possible peptides of length 3-15 amino acids in the ecDHFR sequence, many of which are degenerate in mass, especially when considering instrument sensitivity limitations (Fig. 2a, b). We assessed the impact of each filtering step and each level of disambiguation stringency for co-eluting peptides on the number of assigned peaks from MS/MS (Fig. 2h, Table S1). In one case, for example, we discarded 152 of 747 initial peptide assignments as duplicates.

ΔG_{op} assignment using FEATHER

Following peptide disambiguation from the MS/MS experiment, HX/MS data can be analyzed using FEATHER. FEATHER is a Bayesian method to assign protection factors (PFs), the ratio of intrinsic and observed exchange rates for each residue (Figs. 1b, 3). As a first step, peptides with N/C terminal overlaps are subtracted from one another at the mass envelope level to produce explicit mini-peptides at an enhanced resolution as a data augmentation for PF fitting (Fig. 3a).

While there are several methods to calculate centroid values from isotopic mass envelopes to fit PFs[17–19], the degenerate nature of MS centroids causes considerable information loss, as many envelopes can yield identical centroid values.[16] By contrast, direct envelope fitting introduces additional constraints, thereby minimizing calculation uncertainties and enhancing accuracy.[20] Building on a centroid-based Bayesian method[17], we implemented sampling improvements and isotopic mass envelope fitting.[20] First, FEATHER initiates full isotopic mass envelope reconstruction for each mini-peptide at each timepoint with preliminary PF guesses. The PFs are refined through iterative Markov Chain Monte Carlo (MCMC) sampling (Fig. 3b, c). PFs are explored on discrete grids on the logarithmic scale within a log(PF) range of 0 to 14. The full theoretical kinetic exchange curves for each residue are built from single exponential functions given the PF value. D-to-H back-exchange, an effect of the chromatography steps in HX/MS, is corrected at each timepoint, followed by a probability distribution calculation to determine the isotopic envelope of deuteration. The isotopic envelope of deuteration and experimental isotopic envelope for each amino acid are then convolved to obtain a theoretical envelope. Then, the error and score based on the comparison between the theoretical and observed MS are calculated. Finally, FEATHER performs an MCMC simulation to reduce the score by randomly sampling the log(PF)s on discrete grids. FEATHER continues sampling until the absolute error between the theoretical and

measured envelopes is minimized and the model converges on optimized PFs for each amino acid or mini-peptide (Fig. 3b). To enhance the sampling, we implemented an explicit PF swapping step (Fig. 3c), motivated by the switchable nature of exchange rates within a peptide that can yield the same mass spectrum. During each swapping step, a subset of sampled residues may swap their log(PF) values with adjacent amino acids. The final PF values are derived by combining results from multiple bootstrap replicates. The standard deviation of PFs within a cluster aggregated across multiple bootstrapping runs is used as a measurement of the error. We also implemented two optional Bayesian priors: a *structural prior*, which leverages a protein structure to bias sampling, and an *uptake prior*, which compares calculated exchange rates with centroid-level uptake (see Methods).

Simulated data benchmarks

To assess FEATHER accuracy, we prepared simulated datasets by assigning exchange rates to all non-proline amino acids in synthetic proteins of varied sequences and lengths (Tables S2, S3). The proteins were decomposed into peptides with sequence coverage, noise, and overlap reminiscent of true HX/MS data (see Methods). We used FEATHER to calculate the exchange rates and PFs for each amino acid in these proteins (Fig. 3d, e). FEATHER performance was measured as the correlation coefficient of the true vs. calculated log(PF) (R) and root-mean-square error (RMSE) (Fig. 3f). FEATHER achieved R values near 1 for proteins of all sizes (Fig. 3g).

We found that peptide subtraction, PF swapping-sampling, and MS reconstruction contribute synergistically to FEATHER performance (Fig. 3g, Table S4); the benefit of including all three features exceeds the sum of their individual contributions (Table S4). Notably, FEATHER benefits more from increased data volume when using mass envelope data rather than centroids, highlighting further advantages of envelope fitting. Further, removing half of the timepoints, removing timepoints beyond 1e6 seconds, and noising the MS peaks did not significantly reduce accuracy (Tables S3, S5).

To compare FEATHER to other PF calculation methods[17–21], we also analyzed our optimal synthetic dataset using BayesianHDX[17], PyHDX[18], ExPfact[19], and HDsite[20] (Table S6). FEATHER considerably outperforms each method. Although FEATHER builds on its sampling method, BayesianHDX fits centroids and samples less efficiently, leading to diminished accuracy. HDsite was the first method to reconstruct full mass envelopes, but the iterative fitting approach and scoring function reduced its performance. Due to its incompatibility with dense data and weeks-long compute time, we could only test HDsite on one small dataset. Overall, FEATHER combines the best features of prior methods with sampling and architecture improvements for unprecedentedly accurate ΔG_{op} calculations.

PIGEON-FEATHER determines single-site ΔG_{op} values for ecDHFR

ecDHFR is a clinical target for treating infections.[22] After confirming its activity (Fig. S2), we tested the performance of PIGEON-FEATHER on ecDHFR (Fig. 4a-d). Comparing the fitted model with the experimental mass spectra and centroid-level D uptake demonstrates that FEATHER can accurately fit real HX/MS data (Fig. 4a) with low error (Fig. 4b). All D uptake plots and fitted models are in the SI files.

From 442 PIGEON-disambiguated peptides, we resolved 91% of positions at single-site resolution and 4% as dipeptides, totaling 95% of the protein sequence (Figs. 4c, S3, Tables S7, S8). The coverage and resolution exceeds previous HX/MS experiments for ecDHFR[23–25] due to using several acid proteases (Fig. S4), and the peptide-level exchange corroborates a recent study (Fig. S5).[25] The projection of ΔG_{op} on the structure highlights βB, βC, and βH as the regions most protected from exchange (Fig. 4c), with residues M42, I62, and F153 having the highest ΔG_{op}. Critically, PIGEON-FEATHER can resolve single-site ΔG_{op} in ecDHFR's catalytically important loops: the Met20 loop, the FG loop , and the GH loop. The Met20 loop residues exchange quickly (Fig. 4a, c, d) as expected given diffuse electron density in APO ecDHFR X-ray

crystal structures.[26,27] However, the FG and GH loops contain well-protected residues, like D127-Y128 (ΔG_{op}=40, 38 kJ/mol) and A143 (ΔG_{op}=70 kJ/mol).

The three PIGEON modes produce consistent FEATHER results with R values ranging from 0.84 to 0.90 (Fig. S6). Correlating ΔG_{op} from singlicate experiments against the pooled dataset (n=6) gives an average R value of 0.74, demonstrating the method's reproducibility (Figs. S7, S8a-g). In the absence of long timepoints (>4 hours), we observed a drop of 0.09 in R, similar to the synthetic data benchmark (Figs. S8h, S9). We found that three bootstrapping replicates are sufficient (Fig. S10), and that FEATHER converges to consistent results even without the structural and uptake priors (Fig. S11). FEATHER-derived PFs show a 0.48 correlation with a widely used phenomenological model that defines the PF as a function of the numbers of amide hydrogen bonds and local contacts (Fig. S12).[28] FEATHER-derived PFs also correlate with computationally calculated hydrogen bonds and solvent accessible surface area (SASA) using the macromolecular modeling suite Rosetta, suggesting a potential use of the method for structural modeling (Fig. S12).[29,30]

Single-site ΔG_{op} values for ecDHFR and human DHFR (hDHFR) enable comparison

Like protein sequences and structures, ΔG_{op} values may be evolutionarily conserved as features of the conformational ensemble.[31] Although the sequence identity between ecDHFR and hDHFR is only 26%[32], the X-ray crystal structures are almost identical (all-atom RMSD=1 Å, PDB IDs: 4M6J, 5DFR).[33,34] While peptide-level H-D exchange cannot be quantitatively compared for proteins with different sequences, absolute ΔG_{op} for structurally aligned regions of related proteins may be compared. We therefore applied HX/MS-PIGEON-FEATHER to determine ΔG_{op} values for hDHFR (Figs. 4e, S13, Tables S9, S10). Despite modest sequence identity, the residue-level ΔG_{op} values for the two DHFRs are moderately correlated (Fig. 4f, g).

Although the two orthologs are structurally similar, previous ^{15}N relaxation dispersion experiments showed that the ms-scale motions that dominate ecDHFR Met20 loop are absent in hDHFR, and mutations in ecDHFR that disrupt these motions slow $NADP^+$ dissociation.[35–37] It was hypothesized that lower intracellular $NADP^+$ and THF concentrations in vertebrates allow efficient cofactor exchange in DHFR, enabling a hinge-like open-close conformational change to control the ligand flux while maintaining the structure of the Met20 loop. Bacteria, which have ~1:1 NADPH:$NADP^+$, may instead have evolved a loop-mediated exchange mechanism to avoid end product inhibition.[38]

In support of this hypothesis, PIGEON-FEATHER yields the otherwise non-intuitive result that several positions in the central β sheet, a universally conserved DHFR motif, have much higher ΔG_{op} in ecDHFR than in hDHFR (Fig. 4g). We hypothesize that this stabilization compensates for the larger loop dynamics in ecDHFR and facilitates the open-close conformational change in hDHFR.[33] In line with this hypothesis, both hinge loops, which mediate the open-close conformational change in hDHFR and are much shorter in ecDHFR, are adjacent to β strands that are more stable in ecDHFR. Meanwhile, several peripheral residues in the M20 and FG loops and helix αC are stabilized by 9-14 kJ/mol in hDHFR compared to ecDHFR (Fig. 4f, g, S14). Even in the absence of ligands, hDHFR may have evolved to sample the hinge-closed conformational state with a higher probability than ecDHFR because this conformation has no specific functional role in ecDHFR. PIGEON-FEATHER thus introduces a strategy to classify the catalytic mechanisms for many other DHFR orthologs with varied-length functional loops and hinges towards developing pathogen-specific antibiotics.[33,39]

Two inhibitors differentially impact the ecDHFR ensemble

Trimethoprim (TMP) and methotrexate (MTX) are competitive inhibitors that bind ecDHFR with sub-nM affinities (Fig. 5a), but because TMP inhibits bacterial DHFRs at least four orders of magnitude more than vertebrate orthologs, it used as an antibiotic.[32,39,40] X-ray crystal structures for TMP- and MTX-ecDHFR have <0.5 Å RMSD (Fig. 5a, b).[41,42] NMR spectroscopy revealed near-identical μs-to-ms motions in the two inhibitor-

bound states.[43] Both states favor a conformation of the Met20 loop that blocks the active site.[41–44] As mounting drug resistance mutations[45] reduce inhibitor efficacy and motivate the search for new therapeutics, we revisited the question of how TMP and MTX impact the ecDHFR ensemble.

PIGEON-FEATHER-derived changes in ΔG_{op} ($\Delta\Delta G_{op}$) between the APO and inhibitor states revealed that TMP and MTX have distinct effects on ecDHFR (Fig. 5a-c, Figs. S14-16). Although both inhibitors stabilize the β sheet and binding pocket helix, they each also have unique energetic effects (Figs. 5a, b, S14, S17, S18). Compared to TMP, MTX more strongly impacts the helix αC (Figs. 5d, S18), GH loop (Figs. 5e, S18), Met20 loop (Figs. 5d, f, S18), and N-terminal region of αB (Fig. S18). Compared to the APO enzyme, αC is stabilized in MTX-ecDHFR via interactions with the Met20 loop, while TMP binding has no effect (Fig. 5a-d). In the αC-proximal loop, R57 forms bidentate hydrogen bonds with MTX that are not possible with TMP (Fig. S19). In αB, the difference is most pronounced in K32 ($\Delta\Delta G_{op,MTX-TMP}$=33 kJ/mol) due to its hydrophobic interaction with the methylene groups of the glutamate moiety of MTX (Fig. S19). Allosteric TMP-specific increases in ΔG_{op} are pervasive in the β sheet where TMP resistance mutations are enriched (Figs. 5a, b, S20). We highlight these effects to demonstrate the resolution and quantification afforded by PIGEON-FEATHER compared to conventional HX/MS analysis methods, and to emphasize results that are not obvious from structures.

Altogether, PIGEON-FEATHER shows how each inhibitor reshapes the ecDHFR ensemble (Fig. 5g). MTX interacts with the M20 loop and introduces a backbone-sidechain hydrogen bond between E17 in the Met20 loop and S49 in αC (Fig. 5d), further stabilizing the backbone-backbone hydrogen bond between N18 and G15. The FG loop is therefore released from its APO-state interactions with the Met20 loop to become relatively flexible, while the GH, FG, and Met20 loops at the edge of the sheet form a strong intramolecular hydrogen bond network (Fig. 5b-e, g). The destabilized FG loop disrupts backbone hydrogen bonds in βG by unfolding the small helix that connects them (Figs. 5g, S5). Because TMP binding does not bias the Met20 loop to the same conformation as MTX, the nucleotide-binding domain, FG loop, and βG strand remain relatively unperturbed (Figs. 5a, f, S14, S18).

These results combined with X-ray crystallography suggest that TMP's selective inhibition of bacterial DHFRs may be linked to its binding conformation and to allosteric effects on the βG and βH strands. Generally, the amino acid content and local structure is similar in the active site among bacterial and vertebrate orthologs (Fig. S21).[39,46] In ecDHFR and hDHFR complexes, the diaminopyrimidine ring of TMP is positioned similarly, but the trimethoxyphenyl group is positioned differently, with a 90° rotation in the dihedral angle for the pyrimidine-to-methylene bond (Fig. S22).[42,47] This rotameric change may destabilize αB to reduce TMP-hDHFR affinity. PIGEON-FEATHER shows that TMP binding also relies heavily on β-sheet stabilization, which may have a higher energetic cost in vertebrate DHFR orthologs where the N-terminal βG forms an extended loop with the adjoining FG loop (Figs. 4e, S22). PIGEON-FEATHER reveals how ligand-induced conformational reweighting in this region, which is far from the active site and least conserved among orthologs, is necessary for ligand binding and catalysis in both ecDHFR and hDHFR as determined previously by mutational studies.[48–51] TMP resistance mutations[52–54] in the β sheet of ecDHFR may limit TMP efficacy by similarly disrupting the wild-type pattern of ligand stabilization in this region (Fig. S20). By contrast, MTX binding destabilizes βG and likely energetically offsets this disruption through its interactions with αC and three functional loops (Fig. S19), resulting in the inhibition of both ecDHFR and hDHFR.

PIGEON-FEATHER clarifies functional switching in the lac *repressor (LacI)*

LacI is a model transcription factor from *E. coli*. Isopropyl β-D-1-thiogalactopyranoside (IPTG) binds the regulatory ligand-binding domain (LBD) of LacI to cause an order-disorder transition in the DNA-binding domain (DBD) more than 40 Å away, but X-ray crystal structures show <1.5 Å RMSD in among the apo, DNA-bound, and IPTG-bound LBDs.[55] We previously applied HX/MS to learn how IPTG and other ligands modulate the rigidity of secondary structures in the LBD.[56] The results were well-correlated with functional data from

mutational studies[57,58] and were validated by subsequent NMR[59] that confirmed the roles of the same regions in stabilized and truncated mutants. However, these studies did not show how the energetic changes in the LBD upon IPTG binding actually remodel the DBD ensemble. We applied PIGEON-FEATHER with fully replicated HX/MS experiments to probe the allostery of this large protein at the amino acid level, including the DBD (Fig. 6). The peptide-level trends were consistent with our previous study (Figs. 6a, b, S23-25), but PIGEON-FEATHER also revealed crucial biological insights about the IPTG to DNA ensemble transition.

We found that IPTG binding differentially reweights specific sets of interactions at the dimer interface over a 70 Å distance to disfavor LacI-DNA interactions and activate transcription (Fig. 6c, d). The C-terminal LBD subdomain is virtually identical by X-ray crystallography in the DNA- and IPTG-bound states.[55] Here, PIGEON-FEATHER shows that helices H17 and H15 are more protected in the DNA state due to better hydrophobic interactions among I283, Y282, and L251, while further stabilizing the D278-C281 and D278-Y282 backbone hydrogen bonds ($\Delta\Delta G_{op,\ DNA\text{-}IPTG}$ =48 kJ/mol) (Fig. 6e). Despite the 55 Å distance, this region is coupled to the DBD, where in the DNA state, a cross-interface hydrophobic interaction between L56 and V52 stabilizes the hinge helix and protects the R51-A116 hydrogen bond, further stabilizing helix H4 (Fig. 6f). This bond is a true molecular switch: in IPTG-LacI, the probability of this interaction is dramatically decreased to bias the ensemble to its active state.

Helix H1 in the N-terminal subdomain of the LBD is one of the most elusive helices in LacI. We previously observed polymorphism in its exchange behavior: the N-terminal half is more protected from H-D exchange in the IPTG state, while the C-terminal half is less protected (Fig. S24). Using PIGEON-FEATHER, we determined that the N-terminal protection at A75 and V79 originates from IPTG binding (alongside S193 and S197 in the pocket periphery), while the C-terminal deprotection is due to the weakening of the hydrogen bond between R86 and A82 ($\Delta\Delta G_{op,\ DNA\text{-}IPTG}$ =31 kJ/mol), caused by its hinge motion (Fig. 6g, h). The IPTG-triggered disruption to this region culminates in destabilization of the C-terminal region of helix H4 and the hinge helix (Figs. 6f, i). The interactions of the hinge helices with each other and with the minor groove of the DNA operator are therefore reduced[60], which linearizes the DNA and releases the helix-turn-helix motif of the DBD from the major groove via increased local unfolding (Fig. 6d).

Discussion

PIGEON-FEATHER determines high-resolution ΔG_{op} values as energetic features of protein ensembles from conventional HX/MS measurements. Detailed benchmarks show that it surpasses other methods and scales with protein size for nearly perfect accuracy, improving the utility and interpretability of HX/MS such that it may be seamlessly combined with cryo-EM and MD simulations. PIGEON-FEATHER includes several features that contribute to its high accuracy. After disambiguating peptides using tandem MS fragment identification, it fits their full isotopic mass envelopes, which circumvents the degeneracy of centroid fitting inherent in other methods by introducing additional constraints and minimizing calculation uncertainties. Inspired by the physics of the hydrogen exchange phenomenon, PIGEON-FEATHER's swapping-sampling mechanism significantly improves sampling efficiency on the rugged PFs space in the MCMC simulations. The peptide subtraction feature further improves the method's accuracy.

The power of PIGEON-FEATHER lies in its ability to reveal how proteins traverse free energy landscapes in response to various perturbations, including ligand binding, mutations, and environmental conditions (pH, temperature, pressure, solvent, and denaturant). The inclusion of ensemble energetic information adds an extra dimension to the sequence-structure-function paradigm for proteins. Integrated with recent advances in protein structure determination[61], structure prediction[7,62], structure search[63], and bioinformatic analysis[64,65], PIGEON-FEATHER enables a deeper understanding of proteins and protein families. Because the method permits the comparison of ΔG_{op} within fold families, biologists may directly observe changes in ensemble energies over the course of evolution. PIGEON-FEATHER reveals allosteric protein mechanisms at high resolution, which can aid the identification and design of new therapeutics.

PIGEON-FEATHER can also be integrated into protein engineering and computational structural modeling methods, for example to build residue-level scoring functions for enhanced docking and virtual epitope screening, integrating with structural modeling[66] and prediction[6,67]. With PIGEON-FEATHER, HX/MS datasets can be used to train machine learning models aimed at designing proteins with defined free energy landscapes. In principle, PIGEON-FEATHER can be extended to support multimodal distributions and HX-electron capture dissociation (ECD)/electron transfer dissociation (ETD)/MS. We will incorporate these capabilities in future developments.

References

1. Boehr, D. D., Nussinov, R. & Wright, P. E. The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol.* **5**, 789–796 (2009).
2. Cooper, A. Thermodynamic fluctuations in protein molecules. *Proc. Natl. Acad. Sci.* **73**, 2740–2741 (1976).
3. Hilser, V. J. An Ensemble View of Allostery. *Science* **327**, 653–654 (2010).
4. Kish, M. *et al.* Online Fully Automated System for Hydrogen/Deuterium-Exchange Mass Spectrometry with Millisecond Time Resolution. *Anal. Chem.* **95**, 5000–5008 (2023).
5. Espada, A. *et al.* A Decoupled Automation Platform for Hydrogen/Deuterium Exchange Mass Spectrometry Experiments. *J. Am. Soc. Mass Spectrom.* **30**, 2580–2583 (2019).
6. Krishna, R. *et al.* Generalized biomolecular modeling and design with RoseTTAFold All-Atom. *Science* **384**, eadl2528 (2024).
7. Abramson, J. *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024).
8. Konermann, L., Pan, J. & Liu, Y.-H. Hydrogen exchange mass spectrometry for studying protein structure and dynamics. *Chem. Soc. Rev.* **40**, 1224–1234 (2011).
9. Bai, Y., Sosnick, T. R., Mayne, L. & Englander, S. W. Protein Folding Intermediates: Native-State Hydrogen Exchange. *Science* **269**, 192–197 (1995).
10. Skinner, J. J., Lim, W. K., Bédard, S., Black, B. E. & Englander, S. W. Protein hydrogen exchange: Testing current models. *Protein Sci.* **21**, 987–995 (2012).
11. Bai, Y., Milne, J. S., Mayne, L. & Englander, S. W. Primary structure effects on peptide group hydrogen exchange. *Proteins Struct. Funct. Bioinforma.* **17**, 75–86 (1993).
12. Masson, G. R. *et al.* Recommendations for performing, interpreting and reporting hydrogen deuterium exchange mass spectrometry (HDX-MS) experiments. *Nat. Methods* **16**, 595–602 (2019).
13. Weis, D. D. Recommendations for the Propagation of Uncertainty in Hydrogen Exchange-Mass Spectrometric Measurements. *J. Am. Soc. Mass Spectrom.* **32**, 1610–1617 (2021).
14. Eggertson, M. J., Fadgen, K., Engen, J. R. & Wales, T. E. Considerations in the Analysis of Hydrogen Exchange Mass Spectrometry Data. in *Mass Spectrometry Data Analysis in Proteomics* (ed. Matthiesen, R.) 407–435 (Springer, New York, NY, 2020). doi:10.1007/978-1-4939-9744-2_18.
15. 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.* **91**, 7474–7481 (2019).
16. Guttman, M., Weis, D. D., Engen, J. R. & Lee, K. K. Analysis of Overlapped and Noisy Hydrogen/Deuterium Exchange Mass Spectra. *J. Am. Soc. Mass Spectrom.* **24**, 1906–1912 (2013).
17. Saltzberg, D. J. *et al.* A Residue-Resolved Bayesian Approach to Quantitative Interpretation of Hydrogen–Deuterium Exchange from Mass Spectrometry: Application to Characterizing Protein–Ligand Interactions. *J. Phys. Chem. B* **121**, 3493–3501 (2017).
18. Smit, J. H. *et al.* Probing Universal Protein Dynamics Using Hydrogen–Deuterium Exchange Mass Spectrometry-Derived Residue-Level Gibbs Free Energy. *Anal. Chem.* **93**, 12840–12847 (2021).
19. Stofella, M., Skinner, S. P., Sobott, F., Houwing-Duistermaat, J. & Paci, E. High-Resolution Hydrogen–Deuterium Protection Factors from Sparse Mass Spectrometry Data Validated by Nuclear Magnetic

Resonance Measurements. *J. Am. Soc. Mass Spectrom.* **33**, 813–822 (2022).
20. 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.* **110**, 16438–16443 (2013).
21. Zhang, Z., Zhang, A. & Xiao, G. Improved Protein Hydrogen/Deuterium Exchange Mass Spectrometry Platform with Fully Automated Data Processing. *Anal. Chem.* **84**, 4942–4949 (2012).
22. Huennekens, F. M. The methotrexate story: A paradigm for development of cancer chemotherapeutic agents. *Adv. Enzyme Regul.* **34**, 397–419 (1994).
23. Oyeyemi, O. A. *et al.* Comparative Hydrogen–Deuterium Exchange for a Mesophilic vs Thermophilic Dihydrofolate Reductase at 25 °C: Identification of a Single Active Site Region with Enhanced Flexibility in the Mesophilic Protein. *Biochemistry* **50**, 8251–8260 (2011).
24. Yamamoto, T. Mass Spectrometry on Segment-Specific Hydrogen Exchange of Dihydrofolate Reductase. *J. Biochem. (Tokyo)* **135**, 17–24 (2004).
25. Wales, T. E. *et al.* Resolving chaperone-assisted protein folding on the ribosome at the peptide level. *Nat. Struct. Mol. Biol.* 1–10 (2024) doi:10.1038/s41594-024-01355-x.
26. Schnell, J. R., Dyson, H. J. & Wright, P. E. Structure, Dynamics, and Catalytic Function of Dihydrofolate Reductase. *Annu. Rev. Biophys.* **33**, 119–140 (2004).
27. Wan, Q. *et al.* Toward resolving the catalytic mechanism of dihydrofolate reductase using neutron and ultrahigh-resolution X-ray crystallography. *Proc. Natl. Acad. Sci.* **111**, 18225–18230 (2014).
28. Vendruscolo, M., Paci, E., Dobson, C. M. & Karplus, M. Rare Fluctuations of Native Proteins Sampled by Equilibrium Hydrogen Exchange. *J. Am. Chem. Soc.* **125**, 15686–15687 (2003).
29. Alford, R. F. *et al.* The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *J. Chem. Theory Comput.* **13**, 3031–3048 (2017).
30. 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* **30**, 313-320.e3 (2022).
31. Hollien, J. & Marqusee, S. Structural distribution of stability in a thermophilic enzyme. *Proc. Natl. Acad. Sci.* **96**, 13674–13678 (1999).
32. Liu, C. T. *et al.* Functional significance of evolving protein sequence in dihydrofolate reductase from bacteria to humans. *Proc. Natl. Acad. Sci.* **110**, 10159–10164 (2013).
33. Bhabha, G. *et al.* Divergent evolution of protein conformational dynamics in dihydrofolate reductase. *Nat. Struct. Mol. Biol.* **20**, 1243–1249 (2013).
34. Bystroff, C. & Kraut, J. Crystal structure of unliganded Escherichia coli dihydrofolate reductase. Ligand-induced conformational changes and cooperativity in binding. *Biochemistry* **30**, 2227–2239 (1991).
35. Bhabha, G. *et al.* A Dynamic Knockout Reveals That Conformational Fluctuations Influence the Chemical Step of Enzyme Catalysis. *Science* **332**, 234–238 (2011).
36. Cameron, C. E. & Benkovic, S. J. Evidence for a functional role of the dynamics of glycine-121 of Escherichia coli dihydrofolate reductase obtained from kinetic analysis of a site-directed mutant. *Biochemistry* **36**, 15792–15800 (1997).
37. Miller, G. P. & Benkovic, S. J. Strength of an Interloop Hydrogen Bond Determines the Kinetic Pathway in Catalysis by Escherichia coli Dihydrofolate Reductase. *Biochemistry* **37**, 6336–6342 (1998).
38. Appleman, J. R. *et al.* Atypical transient state kinetics of recombinant human dihydrofolate reductase produced by hysteretic behavior. Comparison with dihydrofolate reductases from other sources. *J. Biol. Chem.* **264**, 2625–2633 (1989).
39. Blakley, R. L. Eukaryotic Dihydrofolate Reductase. in *Advances in Enzymology and Related Areas of Molecular Biology* 23–102 (John Wiley & Sons, Ltd, 1995). doi:10.1002/9780470123164.ch2.
40. Carroll, M. J. *et al.* Evidence for dynamics in proteins as a mechanism for ligand dissociation. *Nat. Chem. Biol.* **8**, 246–252 (2012).

41. Sawaya, M. R. & Kraut, J. Loop and Subdomain Movements in the Mechanism of Escherichia coli Dihydrofolate Reductase: Crystallographic Evidence,. *Biochemistry* **36**, 586–603 (1997).
42. Manna, M. S. *et al.* A trimethoprim derivative impedes antibiotic resistance evolution. *Nat. Commun.* **12**, 2949 (2021).
43. Mauldin, R. V., Carroll, M. J. & Lee, A. L. Dynamic Dysfunction in Dihydrofolate Reductase Results from Antifolate Drug Binding: Modulation of Dynamics within a Structural State. *Structure* **17**, 386–394 (2009).
44. Chen, J., Dima, R. I. & Thirumalai, D. Allosteric communication in dihydrofolate reductase: signaling network and pathways for closed to occluded transition and back. *J. Mol. Biol.* **374**, 250–266 (2007).
45. Krucinska, J. *et al.* Structure-guided functional studies of plasmid-encoded dihydrofolate reductases reveal a common mechanism of trimethoprim resistance in Gram-negative pathogens. *Commun. Biol.* **5**, 1–14 (2022).
46. Rodriguez, D. C. P., Weber, K. C., Sundberg, B. & Glasgow, A. MAGPIE: An interactive tool for visualizing and analyzing protein–ligand interactions. *Protein Sci.* **33**, e5027 (2024).
47. Bank, R. P. D. RCSB PDB - 2W3A: HUMAN DIHYDROFOLATE REDUCTASE COMPLEXED WITH NADPH AND TRIMETHOPRIM. https://www.rcsb.org/structure/2W3A.
48. Dion-Schultz, A. & Howell, E. E. Effects of insertions and deletions in a beta-bulge region of Escherichia coli dihydrofolate reductase. *Protein Eng. Des. Sel.* **10**, 263–272 (1997).
49. Bullerjahn, A. M. & Freisheim, J. H. Site-directed deletion mutants of a carboxyl-terminal region of human dihydrofolate reductase. *J. Biol. Chem.* **267**, 864–870 (1992).
50. Ratnam, M., Tan, X., Prendergast, N. J., Smith, P. L. & Freisheim, J. H. Ligand-induced structural constraints in human dihydrofolate reductase revealed by peptide-specific antibodies. *Biochemistry* **27**, 4800–4804 (1988).
51. Howell, E. E., Booth, C., Farnum, M., Kraut, J. & Warren, M. S. A second-site mutation at phenylalanine-137 that increases catalytic efficiency in the mutant aspartate-27 .fwdarw. serine Escherichia coli dihydrofolate reductase. *Biochemistry* **29**, 8561–8569 (1990).
52. Manna, M. S. *et al.* A trimethoprim derivative impedes antibiotic resistance evolution. *Nat. Commun.* **12**, 2949 (2021).
53. Watson, M., Liu, J.-W. & Ollis, D. Directed evolution of trimethoprim resistance in Escherichia coli. *FEBS J.* **274**, 2661–2671 (2007).
54. Toprak, E. Evolutionary paths to strong antibiotic resistance under dynamically sustained drug stress. *Nat. Genet.* (2011).
55. Daber, R., Stayrook, S., Rosenberg, A. & Lewis, M. Structural analysis of lac repressor bound to allosteric effectors. *J. Mol. Biol.* **370**, 609–619 (2007).
56. Glasgow, A. *et al.* Ligand-specific changes in conformational flexibility mediate long-range allostery in the lac repressor. *Nat. Commun.* **14**, 1179 (2023).
57. Markiewicz, P., Kleina, L. G., Cruz, C., Ehret, S. & Miller, J. H. Genetic studies of the lac repressor. XIV. Analysis of 4000 altered Escherichia coli lac repressors reveals essential and non-essential residues, as well as 'spacers' which do not require a specific sequence. *J. Mol. Biol.* **240**, 421–433 (1994).
58. Suckow, J. *et al.* Genetic Studies of the Lac Repressor XV: 4000 Single Amino Acid Substitutions and Analysis of the Resulting Phenotypes on the Basis of the Protein Structure. *J. Mol. Biol.* **261**, 509–523 (1996).
59. Romanuka, J. *et al.* Genetic switching by the Lac repressor is based on two-state Monod–Wyman–Changeux allostery. *Proc. Natl. Acad. Sci.* **120**, e2311240120 (2023).
60. Kalodimos, C. G., Folkers, G. E., Boelens, R. & Kaptein, R. Strong DNA binding by covalently linked dimeric Lac headpiece: Evidence for the crucial role of the hinge helices. *Proc. Natl. Acad. Sci.* **98**, 6039–6044 (2001).
61. Zhong, E. D., Bepler, T., Berger, B. & Davis, J. H. CryoDRGN: reconstruction of heterogeneous cryo-EM

structures using neural networks. *Nat. Methods* **18**, 176–185 (2021).
62. Lin, Z. *et al.* Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* **379**, 1123–1130 (2023).
63. van Kempen, M. *et al.* Fast and accurate protein structure search with Foldseek. *Nat. Biotechnol.* **42**, 243–246 (2024).
64. Halabi, N., Rivoire, O., Leibler, S. & Ranganathan, R. Protein Sectors: Evolutionary Units of Three-Dimensional Structure. *Cell* **138**, 774–786 (2009).
65. Steinegger, M. & Söding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* **35**, 1026–1028 (2017).
66. Leaver-Fay, A. *et al.* ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. in *Methods in enzymology* vol. 487 545–574 (Elsevier, 2011).
67. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
68. Skinner, J. J., Lim, W. K., Bédard, S., Black, B. E. & Englander, S. W. Protein dynamics viewed by hydrogen exchange. *Protein Sci.* **21**, 996–1005 (2012).
69. Smit, J. & roy. Jhsmit/HDXrate: HDXrate version 0.2.2. Zenodo https://doi.org/10.5281/zenodo.10022160 (2023).
70. Nguyen, D., Mayne, L., Phillips, M. C. & Walter Englander, S. Reference Parameters for Protein Hydrogen Exchange Rates. *J. Am. Soc. Mass Spectrom.* **29**, 1936–1939 (2018).
71. Röst, H. L., Schmitt, U., Aebersold, R. & Malmström, L. pyOpenMS: A Python-based interface to the OpenMS mass-spectrometry algorithm library. *PROTEOMICS* **14**, 74–77 (2014).
72. Bystroff, C. & Kraut, J. Crystal structure of unliganded Escherichia coli dihydrofolate reductase. Ligand-induced conformational changes and cooperativity in binding. *Biochemistry* **30**, 2227–2239 (1991).
73. Chen, E. A. & Porter, L. L. SSDraw: Software for generating comparative protein secondary structure diagrams. *Protein Sci.* **32**, e4836 (2023).
74. Zhang, C., Shine, M., Pyle, A. M. & Zhang, Y. US-align: universal structure alignments of proteins, nucleic acids, and macromolecular complexes. *Nat. Methods* **19**, 1109–1115 (2022).
75. Manna, M. S. *et al.* A trimethoprim derivative impedes antibiotic resistance evolution. *Nat. Commun.* **12**, 2949 (2021).
76. Bell, C. E. & Lewis, M. A closer view of the conformation of the Lac repressor bound to operator. *Nat. Struct. Biol.* **7**, 209–214 (2000).

Methods

PIGEON: detailed methods

Initial peptide identification and removal of systematic bias

PIGEON requires a list of peptide identifications for each MS/MS dataset. Peptide lists can be produced using any commercially available MS software that matches MS peaks by monoisotopic mass and retention time to sequence ranges and assigns a score based on *b*- and *y*- ion coverage (Fig. 2d). We collect 2-4 MS/MS datasets per HX/MS experiment and produce the peptide list using Bruker Biotools. To allow for interpreting datasets with a systematic *m/z* error from the MS calibration, we apply an initial high 30 ppm tolerance for *m*/*z* error in Biotools (Figs. 2e, i, f, S1a). To account for different protease specificities and heterogeneity in protease activities, our database includes every possible proteolytic fragment for the protein of interest up to a maximum length of 15 residues rather than only including peptide sequences resulting from pepsin cleavage (Fig. 2b). The 30 ppm tolerance introduces many false identifications, so we impose a lower tolerance threshold after correcting for calibration errors using PIGEON (Fig. 2e, ii). To do this, PIGEON first pools all identified peaks across MS2 datasets and identifies a subset of high-confidence peaks (with scores >150) from this pool (Fig. S1b), then fits a heuristic trend line to the subset using a fractional exponent function f(x) or polynomial function g(x):

$$f(x) = \frac{a}{\sqrt{x-b}} + c$$
$$g(x) = ax^2 + bx + c$$

We re-threshold the data to ±7 ppm to include only points around this trendline (Fig. S1c). Up to half of the peaks are discarded at this stage, with up to 100% increase in mean score.

Removal of redundant peptide matches
In HX/MS experiments, one to three charge states are typical per peptide. Each charge state for each peptide may have redundant matches that differ by small amounts in retention time or *m*/*z*, but correspond to the same peak. PIGEON uses the match with the highest fragment coverage for each charge state. In the approximately 30% of cases with equal fragment coverage among matches, PIGEON chooses the match with higher intensity. This leaves 100-1,000 peaks, with a 2-fold to 4-fold increase in mean score (Figs. 2e, ii, S1d). At this stage, the user may also impose a minimum score for matches, for use with higher quality datasets.

Removal of false duplicates
The most problematic case for HX/MS analysis is *m*/*z* degeneracy, where multiple peptides are redundantly assigned to a single peak in the MS because their theoretical *m*/*z* or masses are very close or identical (Fig. 2a-c). To check for these false duplicates, PIGEON compares each peak assignment to all other peak assignments in the dataset. If the monoisotopic *m*/*z* and the retention time for each sequence is within a threshold (default: 0.1 Da, 0.5 minutes) of another, PIGEON flags the pair. If the peptide in question is unsupported by fragment information (default: has score 0) it is dropped (Fig. 2e, iii, g). If multiple duplicate peptides have fragment support, this indicates that these peptides co-elute and are both present in an unknown ratio. Because the optimal treatment in this case depends on the details of the analysis (in particular on how high coverage is possible when excluding these co-eluting peptides, the resolution of information desired, and the extent of manual data refinement after rate fitting) several modes are provided in this case (Fig. 2e, iv). These are: to keep duplicates where both have fragment support ('KEEP'), to drop all ('DROP'), or to drop only those sets that do not mostly (default >80%) overlap in sequence ('SELECT').

FEATHER: detailed methods
The FEATHER method uses a Bayesian framework to reconstruct full isotopic mass envelopes to fit PFs given three inputs from the user: the protein sequence, a peptide list from PIGEON, and the HX/MS dataset.

Peptide subtraction
The experimental peptides from the HX/MS dataset for each functional state of the protein are categorized according to their overlaps. Subsequently, exhaustive combinations of peptides within each category are subjected to subtraction (Fig. 3a). Peptides are subtracted iteratively until no further new peptides can be produced. There is also an option to only perform one round of subtraction. We use a Monte Carlo (MC) simulation to subtract the isotopic mass envelope, aiming to minimize the error recalculated between the convolution of the two short peptides and the parent peptide. The initial guess for this process is obtained through the deconvolution of the parent peptides using a Fourier transform.

Isotopic mass envelope reconstruction
Following peptide subtraction, initial PF guesses for each individual amino acid are corrected using Markov Chain Monte Carlo (MCMC) sampling based on errors between the theoretical envelope and the measured

envelope along with the centroid value until the model converges (Fig. 3c). The theoretical envelopes are derived by convolving the natural isotopic mass envelope from the MS/MS experiment with the theoretical deuteration isotopic distribution at a specific time point. The deuteration isotopic distribution is calculated based on the single exponential kinetic curves of all residues in a peptide given the log(PF)s and the peptide's back exchange level. The peptide exchange does not take into account the first two residues and prolines. The forward model (f_{mod}), which calculates the isotopic mass envelope $E_{f,t}^{fmod}$ of peptide *f* at exchange time *t*, is defined as

$$E_{f,t}^{fmod} = Conv.\,(E_{f,t}^{nat,non-D}, E_{f,t}^{fmod,D})$$
$$E_{f,t}^{fmod,D} = f(\{\phi(1-\beta_f)(1-\partial_i \mathrm{e}^{-k_i t})\})$$

Where Conv. is the convolution function, f(x) is the function that calculates the probability distribution given the probability of each independent event, saturation level ϕ is the deuterium percent of the D_2O buffer, ∂_i is a delta function if residue $_i$ has an observable amide and zero otherwise, β_f is the back-exchange level of peptide f, and $E_{f,t}^{nat,non-D}$ is the natural isotopic mass envelope.

Working from the BayesianHDX[17] score function, we replaced the original noise model with a Gaussian function of the summed squared fitting error for individual isotopic mass envelope peaks, $E_{f,t,n,p}^{exp}$, for peptide *f* at time *t*, replicate *n,* and peak *p*. P_E represents the probability of observing the envelope E.

$$P_E(E_{f,t,n}^{\exp} \mid M, \sigma_{E,f,t}) = \frac{1}{\sqrt{2\pi\sigma_{E,f,t}^2}} \exp\left(-\frac{1}{2}\sum_{p=1}^{N} \frac{(E_{f,t,n,p}^{\exp} - E_{f,t,p}^{f\mathrm{mod}})^2}{\sigma_{E,f,t}^2}\right)$$

Here, the model M = ({k_i}), is of the set of residue-resolved exchange rates, {k_i}. N is the total number of peptides, $\sigma_{E,f,t}$ is the average mass envelope peak error estimate for peptide *f* at time *t*.

We also incorporate the agreement of the fits with the centroid-level observations. The probability P_C of experimentally observed centroid value of the mass spectrum C^{exp} for peptide *f* at time *t*, replicate *n,* is given by

$$P_C(C_{f,t,n}^{\exp} \mid M, \sigma_{C,f,t}) = \frac{1}{\sqrt{2\pi\sigma_{C,f,t}^2}} \exp\left(-\frac{1}{2}\frac{(C_{f,t,n}^{\exp} - C_{f,t}^{f\mathrm{mod}})^2}{\sigma_{C,f,t}^2}\right)$$

$\sigma_{C,f,t}$ is the average centroid value error estimate for peptide *f* at time *t*.

The joint likelihood of isotopic mass envelope and centroid value of the mass spectrum is:

$$P(E_{f,t,n}^{\exp}, C_{f,t,n}^{\exp} \mid M, \sigma_{E,f,t}, \sigma_{C,f,t}) = P_E(E_{f,t,n}^{\exp} \mid M, \sigma_{E,f,t})^{w_C} \times P_C(C_{f,t,n}^{\exp} \mid M, \sigma_{C,f,t})^{w_E}$$

Where w_C and w_E are the weights of the centroid and envelope contributions, respectively. Their default values are both set to 1.

The likelihood function is the joint likelihood of all the peptides, timepoints and replicates.

$$P(\{E^{\text{exp}}_{f,t,n}\}, \{C^{\text{exp}}_{f,t,n}\} \mid M, \{\sigma_{E,f,t}\}, \{\sigma_{C,f,t}\}) = \prod_{f,t,n} P(E^{\text{exp}}_{f,t,n}, C^{\text{exp}}_{f,t,n} \mid M, \sigma_{E,f,t}, \sigma_{C,f,t})$$

The PFs are derived by combining results from multiple posterior analyses, each of which employs bootstrapping techniques where 90% of the peptides are randomly resampled from the entire dataset. In a typical bootstrapping replicate fitting process, simulated annealing is applied with the temperature decreasing from 50 to 0.1. At temperatures of 2 and 0.01, 1500 models are saved. This process is repeated five times, resulting in the collection of 7500 models. Finally, the top models (n=50 by default) from each bootstrap replicate are collected and pooled. K-means clustering is then applied to obtain the PFs within each resolution segment, where the number of clusters corresponds to the size of the resolution segment (1 for a single resolved residue). The standard deviation of PFs within the cluster, when combined across multiple bootstrapping runs, is used as an error measure.

Log(PF) swapping

During each sampling step, 20% of residues are randomly selected. Each of the residues in this subset may swap its log(PF) value with that of an adjacent residue. The probability that this swap is accepted is contingent upon the Metropolis criterion based on the score function at a temperature (T) of 0.

Priors

We also implemented two optional Bayesian priors to enhance FEATHER performance. The *structural prior* leverages computationally predicted or experimentally solved protein structures to bias the sampling based on probable hydrogen bonding patterns within the protein ensemble. This prior is justified by seminal work that demonstrated how backbone amide groups involved in hydrogen bonds with other protein atoms, ligands, or ordered water molecules limit H-D exchange.[10,68] The *uptake prior* compares the FEATHER-determined exchange rates for each amino acid with the D uptake plots to ensure consistency with empirical observations in the calculated exchange rates. During the sampling process, two broad Gaussian biases are applied to the score function for positions with low and high PFs, based on their categorization according to D uptake.

In the structural and uptake priors, observable residues are categorized into three groups: high PF, low PF, and regular PF residues. This classification is based on an analysis of backbone hydrogen bonds in the protein structure and peptide uptake plots. For the structural prior, a residue is considered a high PF residue if it forms hydrogen bonds with other protein atoms, binding partners, or water molecules. Residues that lack backbone hydrogen bonds are classified as low PF residues. For the uptake prior, residues covered by slow-exchange peptides, where the maximum deuterium uptake is less than 20%, are considered high PF residues. Conversely, residues covered by fast-exchange peptides, where the minimum deuterium uptake is more than 100% (exceeding 100% due to full D sample error), are classified as low PF residues. Residues covered by both slow and fast exchange peptides are categorized as regular PF residues. To account for these categorizations in the sampling process, two soft biases are applied to the score function. For high PF residues, a prior Gaussian norm(6, 4.0) is added, and for low PF residues, a bias norm(1.5, 2.0) is included based on seminal NMR studies.[10,68] The scales of the structural and uptake priors are set to 0.7 and 0.5, respectively. The PFs of regular residues are sampled evenly.

Simulated data creation

An amino acid sequence and a corresponding array of PFs of identical length and within a specified range (*e.g.*, 2–10) were randomly generated using the python built-in *random* module. The intrinsic exchange rates (k_{int}) were calculated using the HDXrate library[69] given the protein sequence and the pH and temperature of the experiment (pH 7, 293.0 K)[11,70]. Subsequently, the deuterium uptake kinetics of each residue *i* were determined using a single exponential function that incorporates PF and k_{int}.

$$D_i = 1 - \exp\left(-\frac{k_{int,i}}{PF_i}t\right)$$

Peptides of lengths ranging from 5 to 12 residues were then randomly and evenly generated along the sequence range. The theoretical isotopic mass of a peptide was obtained using the forward model described above. We employed pyOpenMS[71] to calculate the natural isotopic mass envelope of a simulated peptide. The peptide deuterium incorporation is a sum of deuterium uptake of all the residues within it. The simulated data generation is optimized to achieve coverage comparable to that of the experimental dataset by ensuring a similar number of subtracted peptides after one round of subtraction.

Data availability

All data needed to reproduce this work are available at the PRIDE database (dataset identifier PXD054550).

Code availability

The source code for the software is available at https://github.com/glasgowlab/PIGEON-FEATHER, along with the synthetic datasets used for benchmarking, an example HX/MS dataset, and a tutorial.

Acknowledgements

We gratefully acknowledge Glasgow Lab members for requesting, adding, discussing, testing, and troubleshooting various features of the software. PIGEON-FEATHER also benefited from testing by Danielle Swingle at the Advanced Science Research Center (ASRC) at the City University of New York (CUNY). We thank Dr. Andrea Piserchio, Dr. Kevin Gardner, and Dr. Rinat Abzalimov from the ASRC, as well as Dr. Shawn Costello, Sophie Shoemaker, and Dr. Naomi Latorraca for feedback on the method. We also thank Dr. Abzalimov in his role as the Mass Spectrometry Facility Manager at the ASRC, where the HX/MS experiments were performed. We are grateful to Dr. Supriya Pratihar, Dr. Hashim Al-Hashimi, and Dr. Art Palmer at Columbia University Medical Center for valuable discussions, and to Dr. Costello, Dr. Al-Hashimi, Dr. Abzalimov, Dr. Roksana Azad, Dr. Jeff Glasgow, Dr. Tanja Kortemme, and Dr. Stavros Lomvardas for critical feedback on this manuscript.

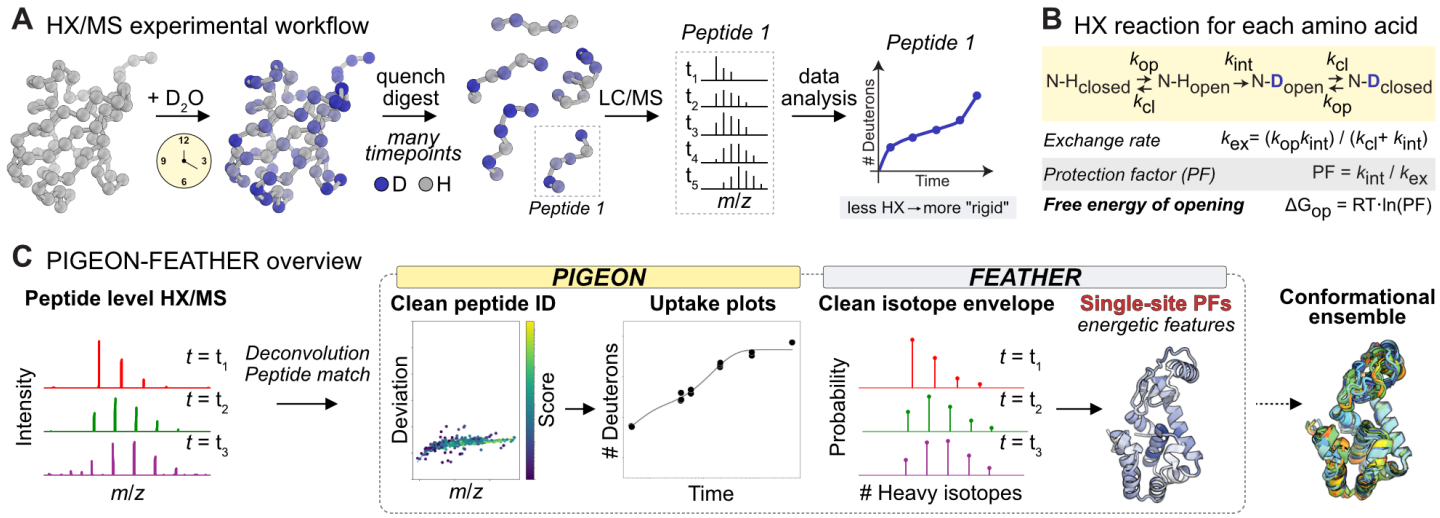


Figure 1. PIGEON-FEATHER overview. A. HX/MS workflow. Increased H-D exchange over the timecourse for a peptide in one functional state compared to another state indicates less involvement in backbone amide hydrogen bond interactions. **B.** Derivation of free energies of opening for protein ensembles from protection factors (PFs), calculated from H-D exchange rates. **C.** Summary of the PIGEON-FEATHER method to disambiguate peptides from HX/MS data and calculate single-site ΔG_{op} values as features of the protein ensemble.

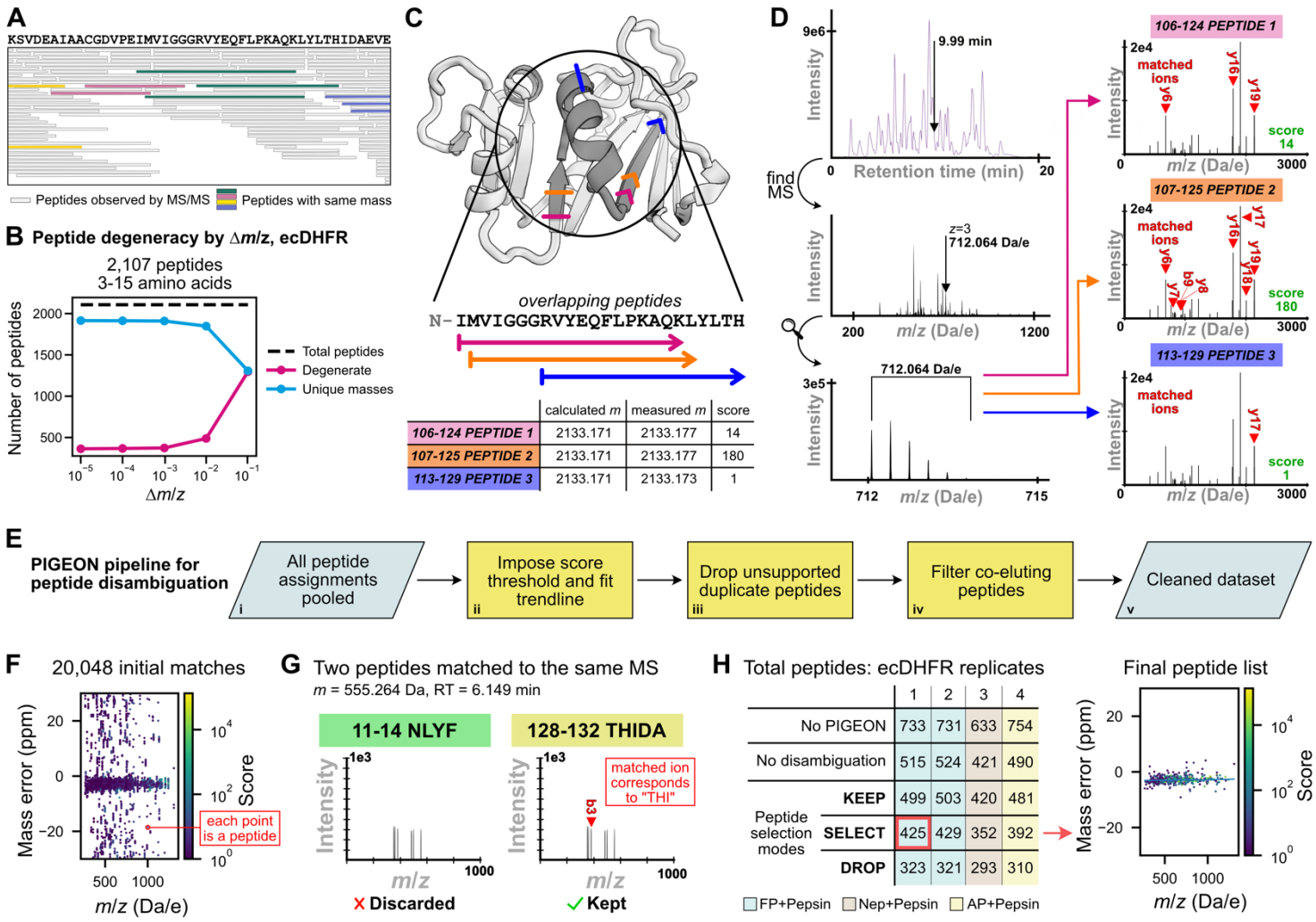
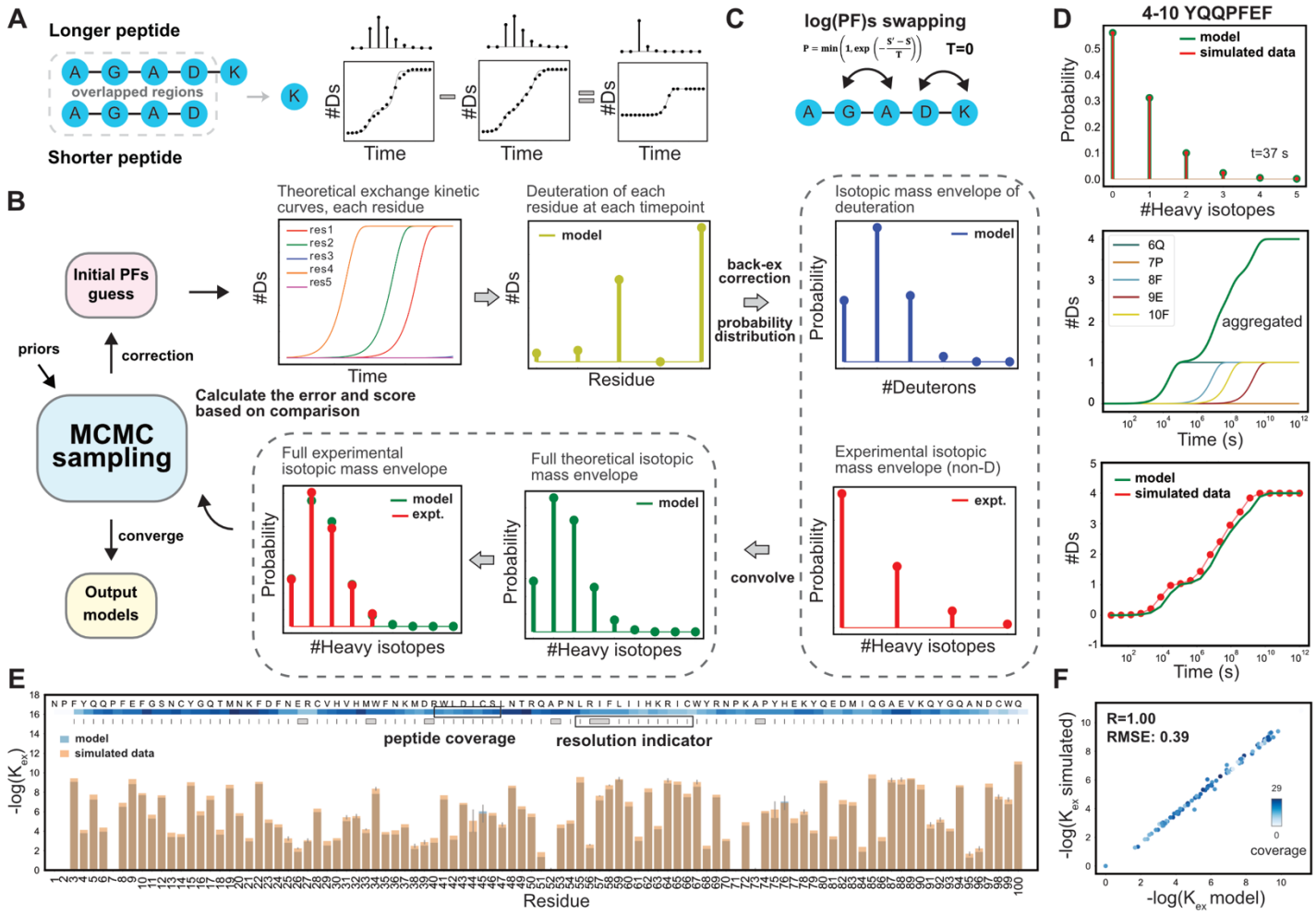


Figure 2. Peptide disambiguation with PIGEON. A. HX/MS peptides are frequently degenerate in mass. **B.** The number of degenerate peptides in *E. coli* DHFR (ecDHFR) given measurement uncertainty (Δ*m*/*z* ~1e-3 within experiments, ~1e-2 between experiments). **C.** ecDHFR peptides are duplicatively matched to the same peak. **D.** Disambiguation of peptides in panel C using tandem MS fragment information. Summed data for the *m*/*z* range of 200–1200 *m*/*z* for the marked portion of the base peak intensity chromatogram are zoomed on a single natural isotopic distribution, where specific fragment ions from MS/MS are matched. **E. i.** Peak assignments are pooled for all charge states for all replicates. **ii.** Systematic mass errors are corrected by fitting a trendline to the highest scoring matches and re-thresholding around the trendline. All but the best peak for each charge state are discarded. **iii.** Each charge state for which there exists an identical match and which is not supported by MS fragments is dropped. **iv.** Each charge state for which there exists an identical match (as in iii.) but which has fragment support is either kept ('KEEP'), discarded if all matches overlap by more than some threshold in sequence (default 80% of exchanged residues) ('SELECT'), or discarded regardless of score ('DROP'). These peptides co-elute and both contribute to the signal. **v.** The final cleaned dataset contains peptides with scores distributed over the full range. **F.** Initial score and *m*/*z* error distribution for a representative pooled dataset (as in E, i). **G.** A duplicate match where the fragment ions support one peptide but not the other. **H.** Impact of each PIGEON step (systematic error correction and disambiguation) and peptide selection mode for four ecDHFR HX/MS datasets (Tables S1, S7). The use of multiple proteases increases peptide coverage: fungal protease (FP), pepsin, nepenthesin (np), and alanyl aminopeptidase (AP). The *m*/*z* and score distribution for one resulting peptide pool (red box) are shown.



G **Testing PIGEON-FEATHER features:** effects of removing log(PF) swapping, peptide subtraction, or isotope envelope reconstruction
Format: **R**/RMSE

# Amino acids	100	100	100	200	200	200	300	300	300	300	500	500
# Peptides	50	100	150	100	200	300	100	200	300	500	300	500
FEATHER	**0.99** / 0.49	**0.99** / 0.51	**1.00** / 0.39	**0.96** / 0.79	**0.98** / 0.57	**0.98** / 0.66	**0.89** / 1.24	**0.92** / 1.04	**0.97** / 0.75	**1.00** / 0.40	**0.91** / 1.10	**0.96** / 0.78
- swapping	**0.92** / 1.11	**0.89** / 1.23	**0.88** / 1.29	**0.90** / 1.19	**0.91** / 1.07	**0.88** / 1.24	**0.87** / 1.35	**0.85** / 1.42	**0.88** / 1.28	**0.88** / 1.29	**0.87** / 1.31	**0.90** / 1.17
- subtraction	**0.96** / 0.82	**0.94** / 0.96	**0.98** / 0.57	**0.94** / 0.89	**0.97** / 0.69	**0.99** / 0.49	**0.90** / 1.20	**0.90** / 1.21	**0.95** / 0.91	**0.96** / 0.82	**0.93** / 1.00	**0.94** / 0.96
- envelope	**0.97** / 0.73	**0.90** / 1.19	**0.93** / 1.03	**0.94** / 0.94	**0.95** / 0.88	**0.90** / 1.15	**0.87** / 1.33	**0.86** / 1.37	**0.95** / 0.86	**0.95** / 0.86	**0.88** / 1.24	**0.91** / 1.10

Figure 3. MS envelope fitting with FEATHER. A. Peptide subtraction to generate data at a higher resolution. **B.** Isotopic mass envelope reconstruction incorporates a Markov Chain Monte Carlo (MCMC) sampling framework. **C.** During MCMC sampling, each residue has a chance to swap its log(PF) with a neighboring residue. **D.** Representative fitting results on a simulated dataset: the reconstructed envelope vs. the simulated counterpart, alongside FEATHER-calculated D uptake kinetic curves for individual residues. These are aggregated to produce a peptide uptake plot and juxtaposed with centroid-level data. **E.** Overlay of the model-derived exchange rates with simulated data. The standard deviation of PFs/exchange rates within a cluster aggregated across multiple bootstrapping runs is represented as an error bar (see Methods). **F.** Correlation between the modeled and simulated kinetic exchange rates. Points represent individual residues colored by peptide coverage. **G.** FEATHER benchmarks on synthetic data.

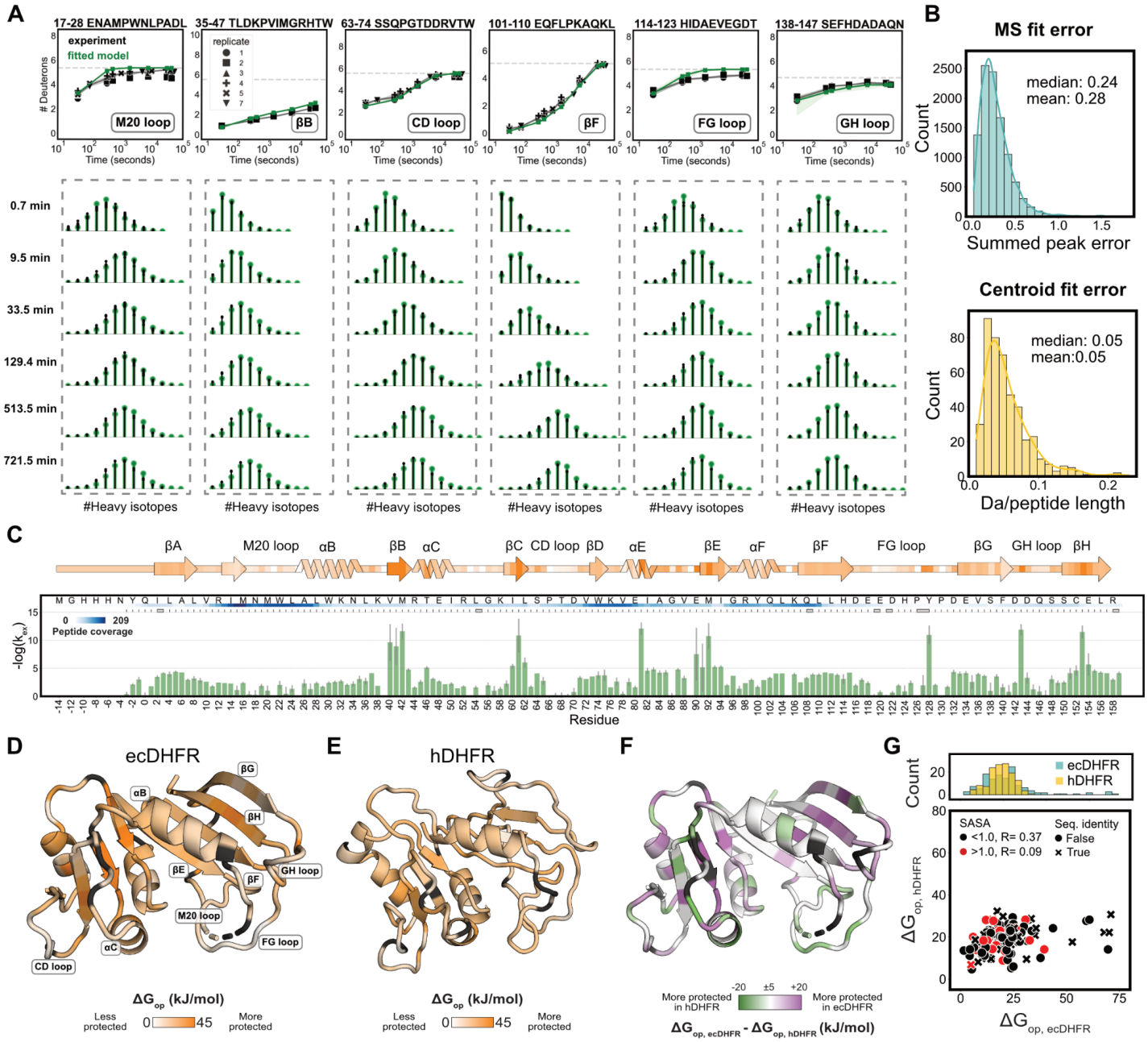


Figure 4. PIGEON-FEATHER reveals site-resolved local energetics of ecDHFR. A. Representative fitted centroid uptake plots with corresponding isotopic mass envelope fitting. Dotted line represents experimental full D level. The fitted models are shown in green, and the shadows on the uptake plots are calculated based on the standard deviation of PFs within the cluster. Peptide positions on the protein are annotated in the lower right corner. **B.** Distribution of mass spectrum fit errors and centroid fit errors. **C.** FEATHER-derived exchange rates, formatted similarly to Fig. 3E. **D.** Absolute ΔG_{op} of ecDHFR projected onto the crystal structure in the APO state (PDB: 5DFR[72]) and the secondary structure of ecDHFR using SSDraw[73] (C, top). Darker hues indicate amino acids with backbone amide groups that are more protected from HX. **E.** Absolute ΔG_{op} of hDHFR projected onto the crystal structure in the APO state (PDB: 4M6J). **F.** Difference between hDHFR and ecDHFR ΔG_{op} values projected onto the crystal structure of ecDHFR based on structural alignment with USalign[74]. Proline and residues with insufficient or noisy data coverage are colored black in panels D-F. **G.** Correlation of ΔG_{op} between ecDHFR and hDHFR. ecDHFR contains several comparatively highly protected residues ($-\log(k_{ex})>8$ in panel C).

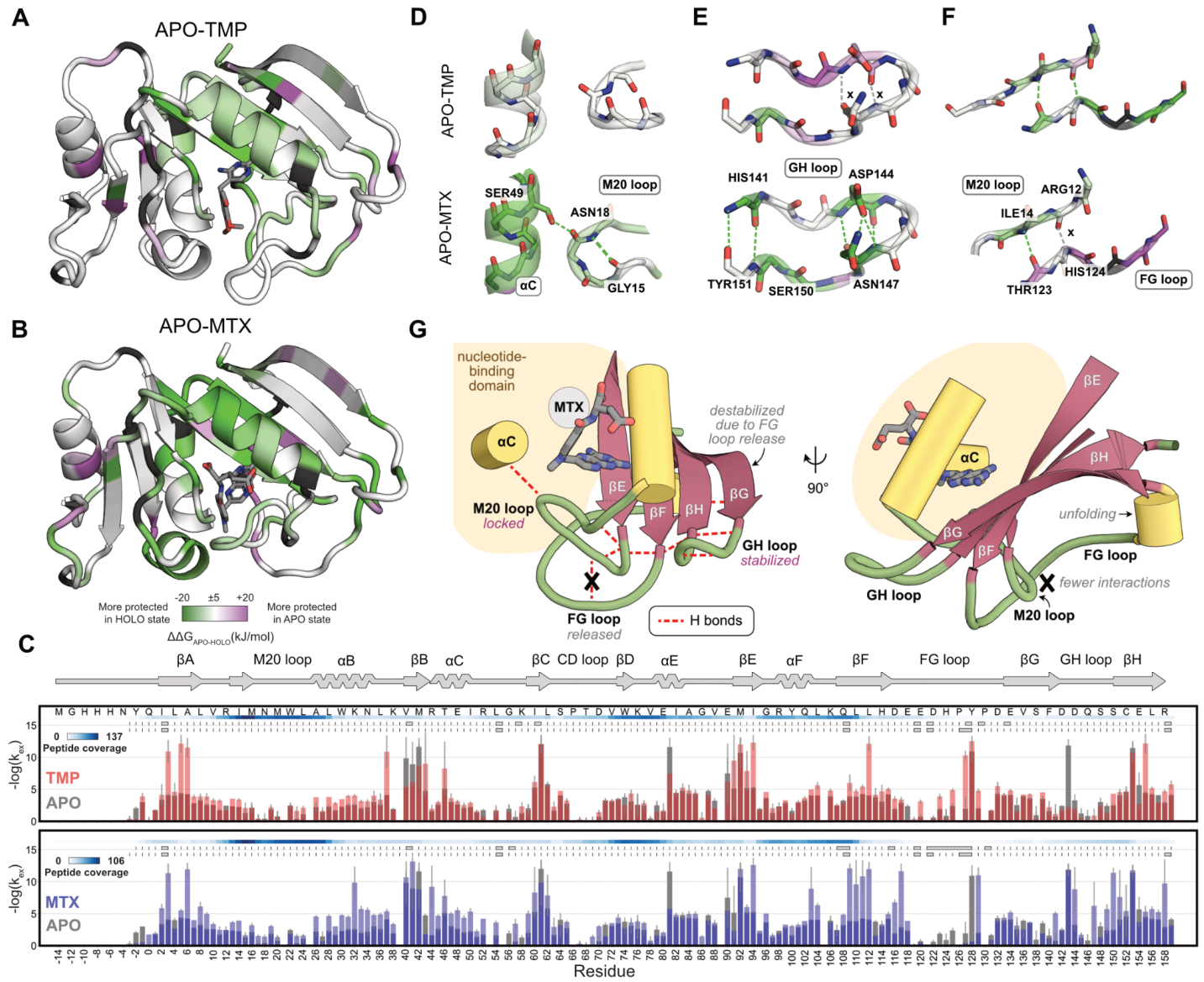


Figure 5. Effects of antifolate therapeutics on the ecDHFR ensemble. A. $\Delta\Delta G_{op\ (TMP\text{-}APO)}$ mapped onto the crystal structure of ecDHFR (PDB: 6XG5[75]). **B.** $\Delta\Delta G_{op\ (MTX\text{-}APO)}$ mapped onto the crystal structure of ecDHFR (PDB: 1RG7). **C.** Overlay of FEATHER-derived exchange rates for TMP and MTX with the APO state, formatted similarly to Fig. 3e. **D-F.** Close view of the Met20 loop, GH loop, and FG loop, showing molecular details of hydrogen bonds that differentiate the $\Delta\Delta G_{op\ (TMP\text{-}APO)}$ and $\Delta\Delta G_{op\ (MTX\text{-}APO)}$. **G.** Conformational changes involved in MTX inhibition.

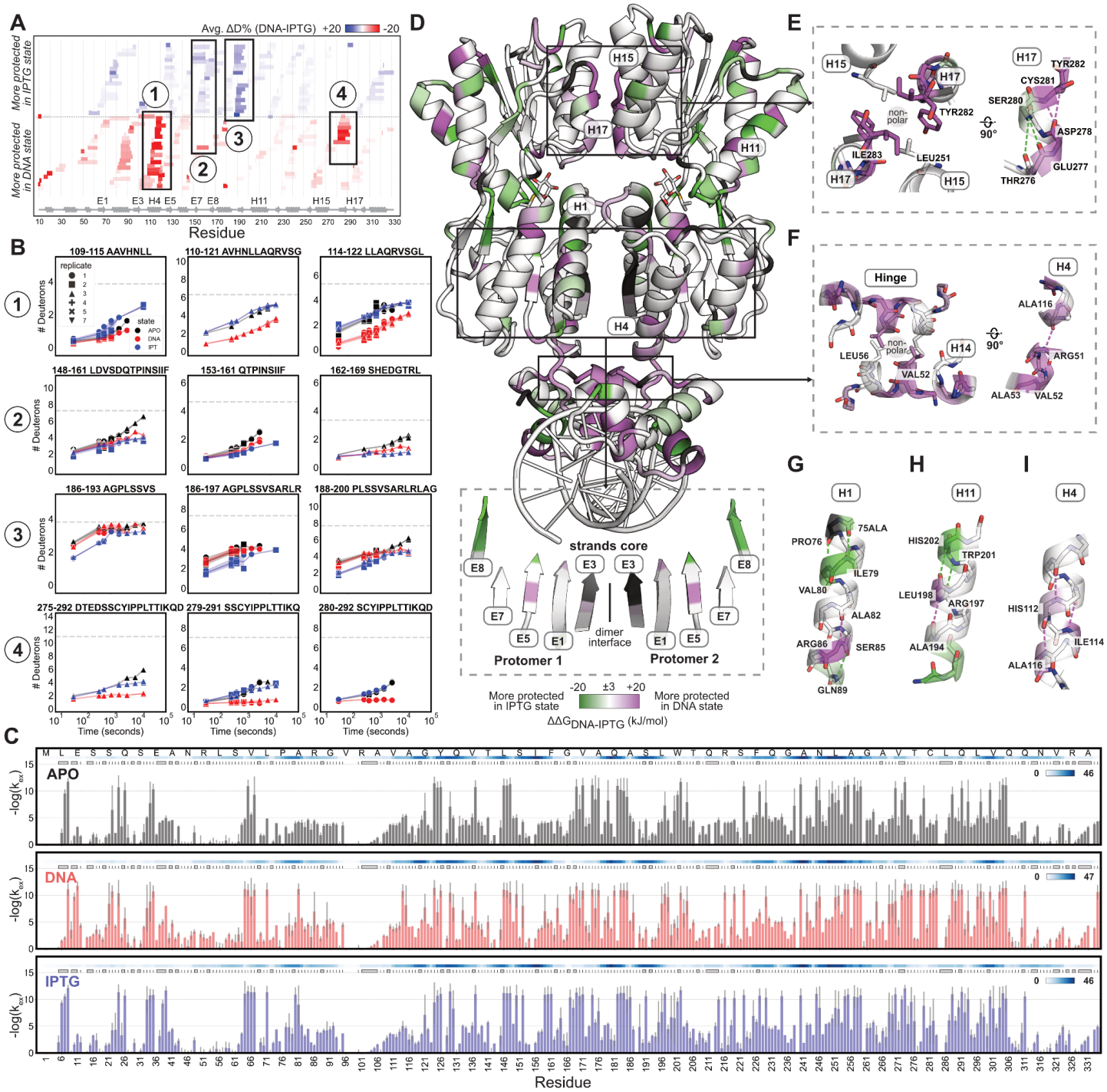


Figure 6. PIGEON-FEATHER reveals differences in the local energetics of LacI functional states. A. Woods plot showing the deuterium uptake difference between the DNA and IPTG states. **B.** Representative uptake plots of peptides in the regions that differentiate the DNA and IPTG states, boxed in panel A. **C.** FEATHER-derived exchange rates, formatted similarly to Fig. 3e. **D.** $\Delta\Delta G_{op\ (IPTG\text{-}DNA)}$ mapped onto the crystal structure of ecDHFR (PDB: 1EFA[76]), with IPTG from PDB: 2P9H[55] aligned to show the binding pocket. **E-I.** Close-up views of the dimer interface at helix H15/H17, hinge helix and helix H4, helix H1, H11, and H4, showing molecular details of hydrogen bonds with $\Delta\Delta G_{op\ (IPTG\text{-}DNA)}$.