**Learning millisecond protein dynamics from what is missing in NMR spectra**

Hannah K. Wayment-Steele[1]*, Gina El Nesr[2]*, Ramith Hettiarachchi[3,4], Adedolapo M. Ojoawo[1], Hasindu Kariyawasam[5,6], Sergey Ovchinnikov[3], Dorothee Kern[1#]

[#] To whom correspondence should be addressed: dkern@scripps.edu
*These authors contributed equally.

[1]Department of Integrated Structural and Computational Biology, Scripps Research & Howard Hughes Medical Institute, La Jolla, CA
[2]Biophysics Program, Stanford University, Stanford, CA
[3]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA
[4]Current address: Ray and Stephanie Lane Computational Biology Department, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA
[5]Center for Advanced Imaging, Faculty of Arts and Sciences, Harvard University, Cambridge, MA
[6]Current address: Cornell Bowers College of Computing and Information Science, Gates Hall, Ithaca, NY

**Abstract**. Many proteins' biological functions rely on interconversions between multiple conformations occurring at micro- to millisecond (μs-ms) timescales. A lack of standardized, large-scale experimental data has hindered obtaining a more predictive understanding of these motions. After curating >100 Nuclear Magnetic Resonance (NMR) relaxation datasets, we realized an observable for μs-ms dynamics might be hiding in plain sight. Millisecond dynamics can cause NMR signals to broaden beyond detection, leaving some residues not assigned in the chemical shift datasets of ~10,000 proteins deposited in the Biological Magnetic Resonance Data Bank (BMRB)[1]. We made the bold assumption that residues missing assignments are exchange-broadened due to μs-ms motions and trained various deep learning models to predict missing assignments. Strikingly, these models also predict exchange measured via NMR relaxation experiments, indicative of μs-ms dynamics. The best of these models, which we named Dyna-1, leverages an intermediate layer of the multimodal language model ESM-3[2]. Notably, dynamics directly linked to biological function — including enzyme catalysis and ligand binding — are particularly well predicted by Dyna-1, which parallels our findings that residues experiencing μs-ms

exchange are more conserved. We anticipate the datasets and models presented here will be transformative in unlocking the common language of dynamics and function.

**Introduction**. The functions of proteins often depend on their ability to interconvert between multiple conformations[3]. Our ability to understand fundamental biological mechanisms would be significantly improved if we had greater ability to predict multiple conformations and the timescales at which they interconvert. Computationally modelling the motions of proteins has been a longstanding goal since the first molecular dynamics (MD) simulations of proteins[4,5]. However, our predictive power for the protein motions of proteins pales in comparison to our predictive power for single structures. AlphaFold-2 (AF2)[6] demonstrated that the experimental data available in the Protein Data Bank (PDB)[7], coupled with large-scale sequencing data and modern deep-learning architectures, could achieve unprecedented accuracy in structure prediction with notable impact in real-world tasks[8,9]. However, the task of predicting *protein dynamics* lacks what the task of predicting single structures had: large-scale, standardized benchmarks of experimental observables. Any computational method, be it deep-learning-based or simulation-based, currently faces a paucity of standardized experimental data available on dynamics to evaluate on, let alone train from (**Fig. 1a**).

We focus here on μs-ms dynamics, since it is well established that motions on this timescale are directly linked to biological function[10] and are often concerted in nature[11]. Where can we find experimental data on μs-ms dynamics? The primary source of structural data used in deep learning, the structure information in the PDB, lacks time information. Time-resolved X-ray diffraction and cryo-electron microscopy methods are gaining increasing traction, but processing these datasets requires nontrivially deconvolving signal to identify multiple conformations. In contrast, the last 50 years of methods development in NMR has led to experiments in which the

presence of μs-ms dynamics *creates* signal in the form the observable $R_{ex}$. $R_{ex}$ is an increase in $R_2$ relaxation times that unambiguously indicates that the monitored atom is experiencing changes in its chemical environment at the μs-ms timescale[12,13].

We wished to address the lack of experimental dynamics benchmarks by curating standardized datasets of $R_{ex}$ measurements on backbone amides. In analyzing the resulting "RelaxDB" dataset of 133 proteins, we found that residues with μs-ms exchange were more conserved than residues with no exchange, which in turn were more conserved than residues with pico-nanosecond motion. This suggested to us that with more experimental data on μs-ms dynamics, it might be feasible to use an evolutionary-guided deep learning approach, such as a protein language model, to predict μs-ms dynamics. This task—namely, can a model predict the presence of a concerted, millisecond-timescale process?—is distinct from other existing tasks: predicting backbone flexibility or order parameters[14,15] or predicting multiple conformations without information on timescales[16,17]. External conditions such as temperature, pH, and cofactors will change the rate and relative populations of such a process, but not the underlying process itself.

We were curious if the ~10,000 proteins with deposited NMR chemical shift information in the BMRB could provide a useful observable with which to train a model. We hypothesized that missing $^{15}N$ backbone assignments in these data are primarily caused by exchange-broadening due to μs-ms motions. We tested this key assumption by training deep learning models to predict which residues are missing assignments and consequently evaluating if these models also have predictive power for measured $R_{ex}$ for residues that are assigned. Our results demonstrate that signatures of millisecond-timescale dynamics are indeed learnable by modern deep learning approaches trained with these NMR data.

The developed model, Dyna-1, is able to predict biologically relevant μs-ms exchange as assessed on proteins characterized by many independent labs. Notably, our model led to

reinterpretation of published NMR relaxation data in several ways. While investigating "false negatives" in the RelaxDB dataset, we noticed that some proteins interact with phosphate ions as part of their biological function, yet the NMR relaxation experiments were conducted in phosphate buffer leading to artificially increased $R_2$ due to buffer binding. Dyna-1 did not predict exchange from such non-specific phosphate binding. Conversely, Dyna-1 did predict millisecond exchange that typical analysis of Carr-Purcell-Meiboom-Gill (CPMG) experiments missed, but more careful consideration of the existing NMR data validated. We anticipate Dyna-1 as well as the datasets used for training and evaluation will be transformative in better understanding how biomolecular dynamics leads to function.

**Results**

*Residues with micro-millisecond dynamics are more conserved than other dynamics categories.* In setting out to curate datasets of standardized $R_{ex}$ measurements, we started with the experiment where $R_{ex}$ is perhaps most directly measured: Carr-Purcell-Meiboom-Gill (CPMG) relaxation dispersion[12,13,18,19]. However, we could only find fewer than 10 sets of complete $^{15}N$ backbone CPMG dispersion data either publicly available or from private correspondences. We next turned to a more abundant type of experiments, which we will refer to as $R_1/R_2$/hetNOE ($R_1$ relaxation, $R_2$ relaxation, heteronuclear nuclear Overhauser enhancement) experiments, which provide information on backbone pico-nanosecond (ps-ns) motions and µs-ms exchange. We compiled data on 163 distinct protein domains (**Fig. 1b**), the earliest of which were published in 1990[20]. Roughly half of these were deposited in the BMRB, though many required further curation such as converting from incorrectly deposited units. The other half came from supplemental tables of published papers, plot digitization from figures in publications, and private correspondences, including a benchmark of 20 proteins compiled by Goodman et al. in 2000[21]. Because $R_{ex}$ reports on any process causing a

change in chemical environment, which can include monomer-dimer equilibrium or the binding-unbinding of a ligand, we only included datasets of apo proteins with no cofactors present and where the authors had verified the protein is monomeric under the experimental conditions used. We also excluded intrinsically disordered proteins.

Data collected at one magnetic field strength results in three values per residue: a $R_1$, $R_2$, and hetNOE value. $R_{ex}$ is the increase in $R_2$, the backbone amide's transverse relaxation (see Methods), caused by exchange between states with different chemical shifts on the μs-ms timescale. A common practice to estimate the presence of additional $R_{ex}$ contribution to $R_2$ is to look for high outliers in $R_2/R_1$ per residue. $R_1$ does not contain $R_{ex}$ and decreases if exchange is present at slower timescales. However, anisotropic tumbling can also alter the $R_2/R_1$ ratio. The model-free formalism[22-24] was developed to estimate $R_{ex}$ from these data; however, we were unable to robustly apply it at this scale of datasets since relaxation data from multiple field strengths—important for accurate estimations—were often not available. Therefore, we developed the following processing method to label residues with $R_{ex}$. We generated AF2 models for the 163 proteins and calculated their $R_2$ and $R_1$ values accounting for anisotropic rigid-tumbling only with HYDRONMR[25,26], fitting one constant per protein that represents the ratio between the experimental and the HYDRONMR overall tumbling time. We then designated a residue as having $R_{ex}$ present if the difference between experimental $R_2$ and rigid-tumbling $R_2$ ($dR_2$) was greater than the combined modelling and experimental uncertainty for that residue (see Methods). We also labeled residues as having ps-ns motion if their hetNOE was less than 0.65 [27-29].

**Fig. 1c** depicts the excellent degree to which HYDRONMR accounts for anisotropic effects for four example proteins from RelaxDB. Experimental data is in black, $R_2/R_1$ calculated from rigid tumbling is in grey, and labels for ps-ns motion and μs-ms exchange are in blue and orange, respectively. The protein GB3 is small yet has pronounced anisotropy in its $R_2/R_1$ arising from its

helix; it was a system first used to demonstrate that anisotropy in tumbling could lead to heightened $R_2/R_1$[30]. This "zigzag" pattern due to differing directions of backbone N-H amide bond vectors in a helix is also evident in the highly-anisotropic helical bundle of alpha spectrin[31]. Nevertheless, our fitting process discerns between anisotropic data features and increases in $R_2/R_1$ due to μs-ms exchange, exemplified in AcrIIA4 Cas9 inhibitor[32] and beta-lactamase[33]. The original 163 proteins were reduced to 133 as we realized some proteins had extensive missing data or ps-ns motion indicative of disordered regions and could not be reliably fit with this workflow (see Methods). All datasets are in **Extended Data Fig. 1**.

**Fig. 1d** depicts the distribution of all residues from these proteins along the two experimental observables defining the presence of different timescales of motion: namely, $dR_2/R_1$ (see Methods) for μs-ms exchange, and hetNOE for ps-ns motion. As expected, many residues with low hetNOE have a $dR_2/R_1$ less than zero, as motions at the ps-ns timescale lead to both a lower hetNOE and a reduced $R_2$ value. A few residues have both fast motions detected via hetNOE and elevated $dR_2/R_1$, indicative of μs-ms exchange. If anything, we anticipate that this method of looking at elevated $R_2/R_1$ is under-estimating exchange for residues where motions are present at more than one timescale, as the presence of μs-ms exchange in elevating $R_2$ and the presence of ps-ns motion decreasing $R_2$ would cancel.

Despite comprising only 133 proteins, the RelaxDB benchmark represents an unprecedented amount of experimental information on dynamics in one place. **Fig. 1e** depicts the overall distribution of different labels assigned per residue. With this curated dataset in hand, we were interested in seeing if there were any eminent relationships between sequence and dynamics. When calculating sequence conservation, we noticed a striking trend: the residues with μs-ms exchange were more conserved than residues with no designated label, which in turn were more conserved than residues with ps-ns motion. Overall, solvent-exposed residues are less conserved

than non-solvent-exposed residues (**Fig. 1f**). Residues involved in μs-ms processes are often involved in functions such as allosteric signaling, catalysis, or ligand binding. Nevertheless, we were surprised that this could be quantitatively detected in as few as 133 proteins. Further analysis revealed that the predicted local distance difference test (pLDDT) measure from AF2 is lowest for residues with ps-ns motion and shows little discrepancy between residues with no motion and μs-ms exchange (**Extended Data Fig. 2**).

Though we hoped RelaxDB would be useful for evaluating models, any model trained with only 133 proteins is unlikely to generalize. In looking at the data, we realized that our initial curation of the data lacked a key observable that could also give information on μs-ms dynamics. Some residues in the proteins of RelaxDB could not be assigned a label because they did not have measured $R_1$/$R_2$/hetNOE values. These residues lacking relaxation data fall into two fundamentally different categories. In the first category, a residue is assigned, but the authors did not report $R_1$/$R_2$/hetNOE values for a variety of reasons: possibly the peaks were too overlapped to fit exponentials reliably or an exponential did not fit well, for instance due to weak signal intensity. In the second category, a residue is missing $R_1$/$R_2$/hetNOE values because the residue was unable to be assigned in the first place. Crucially, residues can be missing from chemical shift assignments if they are "exchange-broadened", i.e., they are experiencing exchange between two states where the chemical shift difference of the two states is comparable to the frequency at which the residue is interconverting. Therefore, we identified the corresponding published assignments for each protein and created labels to distinguish whether residues without $R_1$/$R_2$/hetNOE data were assigned or not. We hypothesized that if the residues missing assignments were also reporting on residues with μs-ms exchange, they would show similar trends in conservation to residues with $R_{ex}$. Indeed, sequence conservation is statistically indistinguishable between residues with $R_{ex}$ and residues with missing assignments (**Fig. 1f**). Fundamentally, whether a residue is not assigned because it is

exchange-broadened or can be assigned but displays $R_{ex}$ is dependent on the chemical shift difference of its multiple states relative to their interconversion rate.

*Curating labels for μs-ms exchange across thousands of proteins*.

We realized that these missing assignments were present at orders-of-magnitude higher counts than in just the relaxation measurements (**Fig. 2a**); the Biological Magnetic Resonance Bank (BMRB) contains ~12,000 proteins with deposited chemical shifts. Could a model trained on all the proteins in the BMRB, where the label is "is this residue assigned or not?" generalize to predicting μs-ms exchange? The key assumption such a model would be leveraging is that backbone amide assignments are primarily missing due to exchange, as opposed to incomplete assignments or other reasons. Of course, residues that are completely exchange-broadened represent a subset of all residues with exchange. If we use residue counts in RelaxDB as a rough proxy for the relative abundances of detectable $R_{ex}$ versus missing assignments, we observe that 5% of residues have elevated $R_2/R_1$, whereas 4% have missing assignments (**Fig. 1e**). This rough estimate concludes that nearly half of the residues with exchange have missing assignments, which certainly is not all, but also not negligible.

We curated a new dataset: the mBMRB. In this dataset, we removed proteins that might have systematic, non-exchange reasons for assignments being missing. We began with all protein depositions in the BMRB as of May 2023 and removed entries with selective isotope labeling and methyl labelling. We also removed entries with deuterated labelling identified in the metadata, as incomplete back-exchange in $H_2O$ results in erroneously missing amide peaks in deuterated proteins. After further filtering to remove entries with outlier sets of missing assignments for other reasons (see Methods, **Extended Data Fig 3**), our dataset contained 9,381 proteins, almost two orders of magnitude more than in RelaxDB (**Fig. 2b**).

This dataset comprises “positives”, which are missing assignments curated as carefully as possible, but of course the “negatives" in this dataset are not unambiguously so: many residues are assignable while also experiencing exchange. Nevertheless, after applying these filters, we observed that a substantial fraction of proteins across the BMRB had significant numbers of non-proline and non-termini missing assignments (**Fig. 2c**). Disaggregating residues by identity and secondary structure type (calculated from structures predicted by ESMFold[34]) reveals that missing assignments span across both loops and structured elements for all amino acid types, though more frequently in loops (**Fig. 2d**). We note that while there will doubtlessly be noise in these labels arising from mis-assignment or other errors, there is well-established literature showing that in certain regimes of machine learning, noise can actually assist model convergence as a form of regularization[35,36].

We next compared the information in these “missing assignment” labels to sources of heterogeneity and uncertainty present in other forms of publicly available structure data to inform on what models trained on the PDB such as AF2 may or may not have already been exposed to. We used MMseqs[37] to align sequences in the mBMRB dataset (filtered to 80% sequence identity) against all proteins in the PDB. 95% of the mBMRB had one or more similar PDB structures (alignment bitscore > 50, sequence identity > 90%). 45% of these proteins (2680/5906) had no X-ray or EM structures but had one or more NMR structures (**Fig. 2e**). For proteins that were characterized by X-ray and/or EM and also assigned by NMR, we found that residues with missing assignments are distinct from residues unresolved in X-ray or EM structure models (**Fig. 2f,g**). Of the residues that are resolved, the distributions of B-factors are similar for both assigned and unassigned residues (**Fig. 2h**). In summary, our mBMRB dataset contains unique information on thousands of proteins in comparison to information present in X-ray and EM structures in the PDB.

*Deep learning models can predict missing assignments.*

With this data in hand, we tested if current deep learning models when trained with these NMR data have the capacity to predict which assignments are missing, and if such learning could transfer to predictive power for μs-ms exchange. We created a train/validation/test data split, initially holding out any sequences with >80% sequence identity to the RelaxDB or RelaxDB-CPMG datasets to use later as higher quality experimental evaluations with labels of exchange (**Fig. 3a**). To contextualize model performance on this new task, we began by calculating a set of naïve baselines. For instance, given that amino acids and secondary structures have different probabilities of having missing assignments in the training data (cf. **Fig. 2d**), we created a naïve classifier that classifies a residue as "missing" based on training set frequencies. For example, 14% of histidines in loops are missing in the training set, so this naïve classifier randomly assigns a histidine in loop to be missing 14% of the time. This classifier achieved an AUROC of 0.50 on the validation set, equivalent to a random predictor. We also created baseline models based on sequence, secondary structure, and solvent accessible surface area, sequence conservation, and elastic network models (**Fig. 3c, Extended Data Figure 6a**). Of these, the best performing model on the validation set used sequence, secondary structure, and SASA (see Methods for architecture details).

We were curious to explore the predictive power of existing deep learning models on this task, since AF2[6] and ESM-2[34] have shown predictive capabilities beyond the tasks for which they were trained. For example, the protein language model ESM-2 was trained in the self-supervised task of masked language modeling yet has predictive power for secondary structure, 3D structure, and functional labels. In this "transfer learning" paradigm, extracted representations from pretrained models are used to train new models with different datasets on related downstream tasks[38]. We designed architectures to test structure-focused representations from AF2[6] pair representations ("AF2, pair rep."), sequence-only representations in ESM-2[34], and interactions

between sequence and structure by using the multimodal language model ESM-3[2] (see Methods, **Fig. 3b, Extended Data Fig. 4**). We first compared these architectures by training with our curated NMR dataset (mBMRB) at 80% sequence identity cutoff; after assessing the models' predictive capability on this novel task given a majority of the available data, we then retrained models with a more stringently filtered training set (see below).

We observed that all three of these pretrained models out-performed other baseline models (**Fig. 3c**). The best-performing model was layer 22 of ESM-3 if provided both sequence and structure models as input, which achieved an AUROC of 0.77. Intriguingly though, the last layer of ESM-2 model achieved an AUROC of 0.75, and the AF2-pair model, which did not use MSAs, and rather only structure input, achieved 0.71. This suggests that both sequence and structure information provide utility in this new task. Performance from all ESM-2 and ESM-3 layers tested are depicted in Fig. 4c. AUPRC (Fig. 4c, right) mirrors AUROC values. We selected based on AUPRC since it is slightly more discriminative than AUROC. AUPRC visualizations for all subsequent per-protein comparisons discussed are included in **Extended Data Fig. 5**.

We then systematically tested the effect of sequence and structure similarity in the training data on performance by training with more stringent training splits: we tested 50% and 30% sequence identity cutoffs and structure-based similarity cutoffs of TM-score at 0.5 and 0.7. These significantly reduced the size of the training set, with the smallest including only ~2700 proteins. The most stringent training split decreased the AUROC of Dyna-1 on the validation set to 0.74 (**Fig. 3d**), and comparably on the remaining evaluation datasets in this work (**Extended Data Fig. 4**). We did not observe any individual proteins with significant jumps in AUROC between our most stringent training set (30% sequence identity, 0.5 TM-score cutoff) and our most lenient (80% sequence identity, 1.0 TM-score cutoff), which would have indicated model memorization (**Extended Data Fig. 4d**). We note that model performance does not improve significantly when

given almost three-fold more data (cf. **Extended Data Figs. 4-6**). The final model leverages the embeddings of layer 22 of ESM-3, uses both sequence and structure as inputs, and was trained with the most stringent data split (30% sequence identity, 0.5 TM-score cutoff).

**Fig. 3e** depicts AUROC per protein for those with one or more missing assignments in the mBMRB-Test dataset. **Fig. 3f** depicts the structures of some of the best-performing proteins in the test set with tube sized by the output of Dyna-1, i.e., the predicted probability each residue is missing, "p(missing)". The missing assignments in example (iv) are related to the function of this protein; in ribonuclease toxin MqsR, loop β2-β3 is known to have exchange, and its flexibility is reported to be related to its ability to bind mRNA[39]. The antitoxin MqsA acts on this protein not by directly inhibiting its active site, but rather by sequestering the β2-β3 loop and prohibiting its motion, which allows the ribonuclease to show broad non-specificity towards a variety of mRNA substrates[39].

Though these results appeared promising, missing assignments are an incomplete indicator of μs-ms dynamics, since residues can experience exchange broadening ($R_{ex}$) that does not necessarily preclude their assignment. Hence, "false positives" in this data cannot be held as ground truth. This is exemplified in (v), the telomerase protein Pof8[40]. Dyna-1 predicts high p(missing) for the α2-β4 loop, yet most of these residues are assigned. However, the authors described that assigning the α2-β4 loop and analyzing subsequent relaxation data for the loop was difficult "due to severe line broadening"[40], a clear indicator of μs-ms exchange. We therefore needed to evaluate Dyna-1's predictions using the more-informative relaxation data in RelaxDB to gain a better sense of the model's predictive capability for μs-ms dynamics. This next evaluation also tests if model learning can "transfer" beyond only predicting residues that are exchange-broadened to the extent that precludes their assignment to residues that are assignable, but have exchange detectable via relaxation.

*Dyna-1 can predict assigned residues with μs-ms exchange.*

We found that when evaluating only missing assignments (**Fig. 4ai**), representative models achieved comparable performance to the mBMRB validation and test sets. To test if the model indeed generalized beyond the data with which it was trained, we held out missing assignments and asked how well it could predict which assigned residues have exchange (**Fig. 4aii**). Performance does decrease across all models compared to the task of predicting missing assignments, but AF2-pair, ESM-2, and ESM-3-based models achieved AUROC well above the controls, indicating that the models can generalize learning from predicting missing assignments (e.g. p(missing)) to predicting exchange within assigned residues (e.g. p(exchange)). To evaluate the performance of the model to predict μs-ms exchange as manifested in both forms, we consider both missing assignments and residues with measured μs-ms exchange in RelaxDB (**Fig. 4aiii**). Out of the representative models, layer 22 of ESM-3 given both sequence and structure again performed the best in this evaluation, same as in the mBMRB validation set. We named this final model "Dyna-1".

**Fig. 4b** depicts AUROC for all proteins in RelaxDB vs. sequence length. We first noticed that there were numerous proteins in RelaxDB with an AUROC of <0.5, i.e. worse than a random predictor. By carefully reading the corresponding NMR papers, we learned that some of these were DNA- or RNA-binding proteins measured in phosphate buffer, and that the unpredicted but experimentally reported exchange often occurred at positively-charged surface residues (**Extended Data Fig. 7**). Several publications directly state this phenomena as an artefact about either their own or others' datasets being a limitation to interpreting measured exchange in the system they investigate.[41,42] We annotated these low-performing datasets further and found that out of the 22 proteins with AUROC ≤ 0.5, 10 of these (45%) are proteins that during their biological function would be expected to interact with phosphate groups (i.e. DNA- or RNA-binding proteins or

signaling domains), yet were experimentally characterized in phosphate buffer. Removing these proteins from RelaxDB increased Dyna-1's overall performance from an AUROC of 0.63 to 0.66.

**Figs. 4c,e** depict representative proteins where structures are sized by Dyna-1 predicted p(exchange) and colored by experimental data for comparison. In Bovine Pancreatic Trypsin Inhibitor (BPTI), Cys14, Lys15, Cys38, and Arg39 are known to undergo millisecond transitions from disulfide bond isomerization.[43,44] Promisingly for our automated labeling scheme, these four residues are also labeled as having μs-ms exchange from relaxation data in ref. [44] (depicted in **Fig. 4c**, top). Shaw et al. identified 5 kinetically distinct states on the micro-millisecond timescales via molecular dynamics simulations[45] (depicted in **Fig. 4c**, left). Strikingly, Dyna-1 predicts high p(exchange) for the two loops containing these residues (**Fig. 4c,** right), to achieve an AUROC of 0.92 for predicting millisecond exchange in BPTI.

*Thorough evaluation of reported NMR relaxation data and corresponding publications increases Dyna-1 performance*

While inspecting predictions for other proteins, we noticed that Dyna-1 predicted high p(exchange) for some residues with reported assignments but no relaxation data. This prompted us to revisit our initial curation of experimental data. For instance, in reduced Mercuric Transport protein (MerP)[46], Thr13 predicts high p(exchange), but has no relaxation data. Ref. [46] notes that the peak for Thr13 is very weak, which indicates exchange-broadening and we posit prohibited the authors from fitting $R_1$ and $R_2$. Though Ala15 was not labeled by our automated labeling scheme, it also has elevated exchange and a large error in $R_2/R_1$ (orange arrow, **Fig. 4c**). Adjusting the experimental label for both residues increases Dyna-1 AUROC from 0.98 to 1.

Another example is the Tudor domain of Survival of Motor Neuron protein (SMNTD)[47], some of the highest p(exchange) from Dyna-1 were in two residues that were assigned yet had no

relaxation data reported: Glu104 and Tyr130 (depicted in black, **Fig. 4c**). The amide hydrogen of Glu104 was assigned at 4.9 ppm, which is a highly unlikely value for an amide hydrogen. We posit that for this small domain and given the well-resolved spectra that the authors report, this residue is more likely exchange-broadened with an incorrect assignment reported. For Tyr130, no $R_1$ value was reported, yet the authors categorized it as fitting to "model 4" from the model-free formalism[47], which is a model containing $R_{ex}$ and requires an $R_1$ value. If we consider both residues as having evidence of exchange, Dyna-1's AUROC increases from 0.67 to 0.75.

In the NtrC receiver domain, Dyna-1 achieved an AUROC of 0.57. We revisited the 10 residues that were assigned yet had no relaxation data. We found that 2/10 had been erroneously assigned in chemical shift depositions corresponding to ref. [48] despite being exchange-broadened. 5/10 had very small peaks, indicative of exchange-broadening, and the remaining 3/10 were unanalyzable due to peak overlap. Taking these into account increases Dyna-1's AUROC to 0.60. Dyna-1 missed predicting exchange in Tyr101 posited to be from ring-flipping. This is exemplary of what we found more broadly across all of RelaxDB, that Dyna-1 is weakest at predicting exchange in beta-sheets (**Fig. 4d**).

In assessing what motifs Dyna-1 was able to predict in larger proteins (**Fig. 4e**), we were intrigued to see that many of the dynamic motifs it correctly predicts also have biological relevance. In flavodoxin YcqA, Dyna-1 predicts high exchange in the loop and helical regions that correspond to the FMN ligand-binding site. In Olfactory Marker Protein from humans, Dyna-1 predicts high p(exchange) in the omega-loop, which is described in ref. [49] to have substantial $R_{ex}$ and is highly conserved. More recent studies indicate that the omega loop serves as a nuclear export signal[50].[51] In APOBEC-2, Dyna-1 predicts high p(exchange) in a region that in structure homologues of APOBEC-2 is known to control the specificity of DNA substrates[52].

Another example is Tyrosine-phosphatase A from *M. tuberculosis* (MptpA)[53]. Dyna-1 has high p(exchange) in loops containing residues with differing orientations between the determined apo and holo structures (apo: 2LUO, holo: 1U2P) and which are postulated to be involved in regulating enzyme specificity[53]. Some of Dyna-1's high p(exchange) is in the P-loop of MptpA, which our curation indicated had been assigned (BMRB:18533) yet had no relaxation data (BMRB: 26513). However, the authors noted, "Due to exchange broadening the P-loop, residues $Cys^{11}$–$Gly^{13}$ and $Ile^{15}$–$Ser^{18}$ were not detectable"[53]. If we consider these labels, the AUROC for MptpA increases from 0.55 to 0.63.

To test our hypothesis that Dyna-1 was performing better in predicting conserved motions, we calculated AUROC across different sets of residues categorized by conservation score and core and surface designation. Indeed, we found that Dyna-1 had the most predictive power for more conserved residues in the core, supporting our observations on individual proteins (**Fig. 4f**).

Dyna-1 predictions in the MAP kinase p38 gamma (MK12) are indicative of room for future improvement. For MK12, we used labels derived from the authors' model-free fitting to multiple field strengths[54]. MK12 is a highly dynamic protein with ps-ns and µs-ms processes occurring throughout its structure, and intriguingly, the p(exchange) predicted by Dyna-1 was high across the whole protein. Improving discrimination between ps-ns motions and µs-ms exchange is the subject of future work. We identified numerous proteins where Dyna-1's highest p(exchange) is in regions with ps-ns motion (blue in **Fig. 4b**, structures in **Extended Data Fig. 7c**). One challenge in interpreting these is that residues can experience both ps-ns motions and µs-ms exchange, and a single field strength is insufficient to experimentally quantify motions at two timescales simultaneously. We expect that if dynamics is present at more than one timescale, because µs-ms exchange elevates $R_2$ and ps-ns motion decreases $R_2$, our current experimental labels for those

likely underestimate Rex, and better experimentally discerning both timescale motions in the future would potentially improve predictive power.

*Dyna-1 predicts exchange not typically considered*.

Due to these challenges in discerning different types of dynamics in single-field-strength $R_1$/$R_2$/hetNOE experiments, we lastly tested Dyna-1 on proteins with μs-ms exchange directly characterized by $^{15}$N CPMG relaxation dispersion (RelaxDB-CPMG dataset, Fig. 5). Unlike $R_1$/$R_2$/hetNOE experiments, CPMG experiments allow for directly measuring contribution from conformational exchange to $R_{2,eff}$[12,18,19]. In brief, data are collected with increasingly short gaps between spin-echo pulses, which increasingly suppress contributions from exchange between states with different chemical shifts ($R_{ex}$) to $R_2$[13,18,19]. In typical data processing (orange trace, **Fig. 5a**), $R_{2,eff}$ plateaus at a value that can be understood to be $R_2$ in the absence of chemical exchange ($R_{2,inf}$). The difference between $R_{2,eff}$ at the lowest $B_1$ field strength and $R_{2,inf}$ is the $R_{ex}$ contribution to $R_2$ that was suppressed. We note that because our RelaxDB-CPMG dataset is curated with $^{15}$N relaxation data, our evaluations on the experimental data do not consider cases where $R_{ex}$ is only seen for $^1$H caused by chemical shift differences only in $^1$H and not $^{15}$N, therefore underestimating other possible chemical exchange. Hence, Dyna-1 performance might be better than reported.

We first evaluated Dyna-1 by creating labels for residues with exchange following this data processing. **Fig. 5b** depicts Cyclophilin A (CypA) in a tube representation sized by Dyna-1 prediction probabilities, where residues with statistically significant $R_{ex}$ from ref. [55] are colored orange. Importantly, Dyna-1 predicted higher p(exchange) in the beta-sheet with a known concerted millisecond dynamics involving the sidechains of residues Ser99, Phe113, Met61, and Arg55[56] (shown in grey in **Fig. 5b**). In **Fig. 5b**, we also show the other beta-sheet on the back of CypA, which contains no exchange, for comparison. We were curious about residue Arg148: Dyna-1 predicted

high probability of exchange, but Arg148 did not register as having significant $R_{ex}$ as evaluated by the above method. When looking at the raw data, we found that Arg148 had high $R_{2,inf}$ compared to other residues. We propose that Arg148 has exchange in a process that is too fast to be suppressed by the $B_1$ fields used in the CPMG experiment for the given temperature. However, to confirm this requires a reasonable estimate for what $R_{2,inf}$ should be if $R_{2,eff}$ is never sufficiently suppressed by the CPMG field strength. As a solution to this problem, we sought a data-processing method that could be simply and uniformly applied across many proteins. We again used HYDRONMR to calculate $R_2$ from rigid tumbling and fit these values to best match the overall tumbling time. **Fig. 5c** depicts CPMG data for CypA collected at 25 ˚C and 600 MHz from a $B_1$ field strength of 33 (red) to 1000 Hz (blue), overlaid with predicted $R_2$ from rigid tumbling from HYDRONMR in black. CypA's tumbling is highly isotropic and deviations from $R_{2,inf}$ can be reasonably inferred. Indeed, several residues have unsuppressed $R_{ex}$ (in green) which conventional data processing would not have identified as having exchange.

We applied this new processing pipeline to 6 datasets where we had $^{15}N$ CPMG dispersion data for all non-overlapped residues. This included CypA[57], beta-lactamase[33], Adenylate Kinase from *Aquifex Aeolicus*[58], Biliverdin reductase B[59], K-Ras[60], and Dual-specificity phosphatase 3[61]. For all of these datasets, Dyna-1's performance improved when considering residues with unsuppressed $R_{ex}$, some quite dramatically; for instance, Dyna-1's AUROC on beta-lactamase from *M. tuberculosis* increases from 0.72 to 0.78 (**Fig. 5d**, structures in **Fig. 5e**). Dyna-1 also predicts substantial unsuppressed $R_{ex}$ in Biliverdin reductase B, a cellular redox regulator, where millisecond motions are implicated in efficient coenzyme engagement and catalysis[59]. Intriguingly, Dyna-1 also predicts concerted millisecond dynamics in the switch I and II regions of K-Ras that is well known to be crucial for its signaling activity[60]. All datasets are depicted in **Extended Data Fig. 8.** We additionally analyzed 4 proteins in the literature where CPMG had been performed but the raw data

was unavailable and found that Dyna-1 also had predictive power on all four after assigning labels to residues that had been reported to exchange. Structures sized by Dyna-1 probabilities with experimental labels are shown in **Extended Data Fig. 9**.

*Prospective experimental test of Dyna-1*.

To validate the predictive capabilities of Dyna-1, we conducted a prospective experimental test after public release of the model and initial review of the manuscript. We selected proteins from the "mBMRB test" set to characterize their dynamics, as these are proteins that are already assigned but many do not have published relaxation data. **Fig. 6a** depicts the mean p(exchange) from each protein in the mBMRB test set vs. sequence length, plotted by alpha helical / beta-strand content. This demonstrates that the predicted mean p(exchange) from Dyna-1 is broad for any given sequence length or helix / strand content. For proteins at the top and bottom range of mean p(exchange), we discarded proteins as candidates if they were fully intrinsically disordered, did not have published expression/purification methods sections, and already had some relaxation data publicly available. From the remaining proteins, we selected Chitinase 19 from *Bryum coronatum*[62] (BMRB: 11466), which had a high degree of p(exchange) (**Fig. 6b**), and hypothetical protein yjbJ[63] (BMRB entry: 5105), which had a low degree of p(exchange) (**Fig. 6c**). Performing $^{15}$N relaxation dispersion CPMG demonstrated that, resembling Dyna-1's predictions, the chitinase did have exchange in region that are crucial for biologic activity—in its active site and the loops known to bind chitin (Figure 6b,e)—while the yjbJ only had detectable exchange for residues 2 and 3 (Fig. 6c,d, supplemental dataset 1). Evaluated at a per-residue level, Dyna-1 obtained an AUROC of 0.62 for chitinase and an AUROC of 0.72 for yjbJ.

## Discussion

AlphaFold-2 (AF2) fundamentally altered the course of molecular biology and was made possible thanks to the clear task of protein structure prediction defined in the Critical Assessment for Structure Prediction (CASP) challenges and the large amount of standardized data deposited in the Protein Data Bank (PDB). Though widely acknowledged that understanding protein dynamics is essential for understanding protein function, the experimental measurements of dynamics that have uncovered foundational principles of allostery, specificity, catalysis, and many more biological functions have resisted being collected into standardized datasets. Given the predictive power displayed in Dyna-1, it is essential that the field shares and deposits experimental data on dynamics in macromolecules. One barrier to this in NMR has been the perceived diversity and complexity of experiments that measure dynamics. We have laid out in this work a clear task—classify residues with μs-ms exchange—and showed that we could train a model with predictive power across multiple experimental data types: missing assignments, $R_{ex}$ from $R_1$/$R_2$/hetNOE, and multiple exchange indicators from CPMG.

Further improvements to Dyna-1 doubtlessly exist for the task of classifying residue dynamics: for instance, ESM-3, the base of Dyna-1, was trained with C-alpha structure data, and not sidechain information. Incorporating sidechain-aware training will likely improve performance. Additionally, Dyna-1 is trained on proteins under 400 residues in length, possibly limiting the scope in which Dyna-1 can predict the dynamics involved in larger proteins. Dyna-1's sensitivity to within-family variations in dynamics and its ability to predict dynamics of *de novo* proteins is the subject of ongoing research. Our comparison of the Dyna-1 model architecture with a training split of 30% sequence identity, 0.5 TM-score cutoff to a larger and more lenient training split of 80% sequence identity, 1.0 TM-score cutoff showed only a marginal increase in performance. This indicates that future model improvements are likely to require innovations in model architectures, rather than simply increasing the dataset size alone.

We demonstrated that using one of the simplest possible NMR observables for dynamics—if a residue $^{15}N$ backbone assignment is missing—resulted in a model with predictive power for μs-ms exchange because that observable was available on a scale of ~10,000 proteins. We also showed that this observable is distinct from other measures of heterogeneity available in X-ray and EM structures. Our demonstrations were in spite of many issues with the current publicly-available data. The data that we used to obtain our labels, lists of atom assignments, are the end point of NMR data processing. The spectra themselves are not required to be deposited, and the lack of standardized requirements to deposit them is a major loss for science. In each spectra, the peak width is related to $R_2$[64]. Had the deposition of the raw data been required, we could have had roughly the information density of RelaxDB (i.e., $R_{ex}$ labels) for at least 10,000 proteins. Careful work[65] resulted in a curated set of 100 raw spectra, two orders of magnitude lower. There is also much room for improvement in metadata labeling in the BMRB. We filtered out entries that were indicated to be based on a deuterated sample, but while inspecting our test set, we found at least one entry (25080) where assignments were transferred from a deuterated sample, though it itself was not deuterated. Such curation is a useful direction for future data analysis and machine learning methods development. Looking forward, if one could prepare samples in higher throughput and diversity, in combination with emerging automated assignment algorithms[66], one could imagine collecting data with intentionally lower resolution to quickly collect information on exchange-broadened peaks and peak widths.

**Competing Interests.** D.K. is a co-founder of Relay Therapeutics and MOMA Therapeutics. H.K.W.-S. and D.K. are listed as inventors on a provisional patent application related to the methods described in this manuscript. The remaining authors declare no competing interests.

**Data availability.** The RelaxDB dataset is publicly available for noncommercial use at https://github.com/WaymentSteeleLab/Dyna-1 and at https://huggingface.co/spaces/gelnesr/Dyna-1. The mBMRB datasets will be made available upon publication.

**Code availability**. Code to run Dyna-1 is available for noncommercial use at https://github.com/WaymentSteeleLab/Dyna-1 and https://huggingface.co/spaces/gelnesr/Dyna-1.

**Author contributions**. H.K.W.-S., G.E.N., S.O., and D.K. designed the research. H.K.W.-S. curated the datasets. G.E.N., H.K.W.-S., R.H., H.K. contributed to model training and development. H. K.W.-S. and D.K. interpreted all existing NMR data during model validation. A.M.O. performed all NMR experiments for prospective experimental testing of Dyna-1. H.K.W.-S., G.E.N., D.K. wrote the manuscript. All authors reviewed the manuscript.

**Acknowledgments**. We thank Ricardo Padua, Hannes Ludewig, Lucy Colwell, Jeffrey Hoch, Art Palmer, Anthony Gitter, Sameer D'Costa, and Kern lab and Ovchinnikov lab members for useful discussions and advice. We thank Martin Stone for sharing the Indiana Dynamics Database data his group curated in 2000. We thank Katie Henzler-Wildman, Magnus Wolf-Watz, Elan Eisenmesser, J. Patrick Loria, Marcellus Ubbink, George Lisi, Sam Butcher, and Nicolas Doucet for sharing data. We thank Zeming Lin for assistance with ESM-3. We thank Possu Huang for financial and cluster support. Model training was performed on the Stanford Sherlock cluster. H.K.W-S. acknowledges funding from the Jane Coffin Childs fellowship. G. E. N. acknowledges funding from the NSF GRFP. This work was supported by the Howard Hughes Medical Institute (HHMI) to D.K.

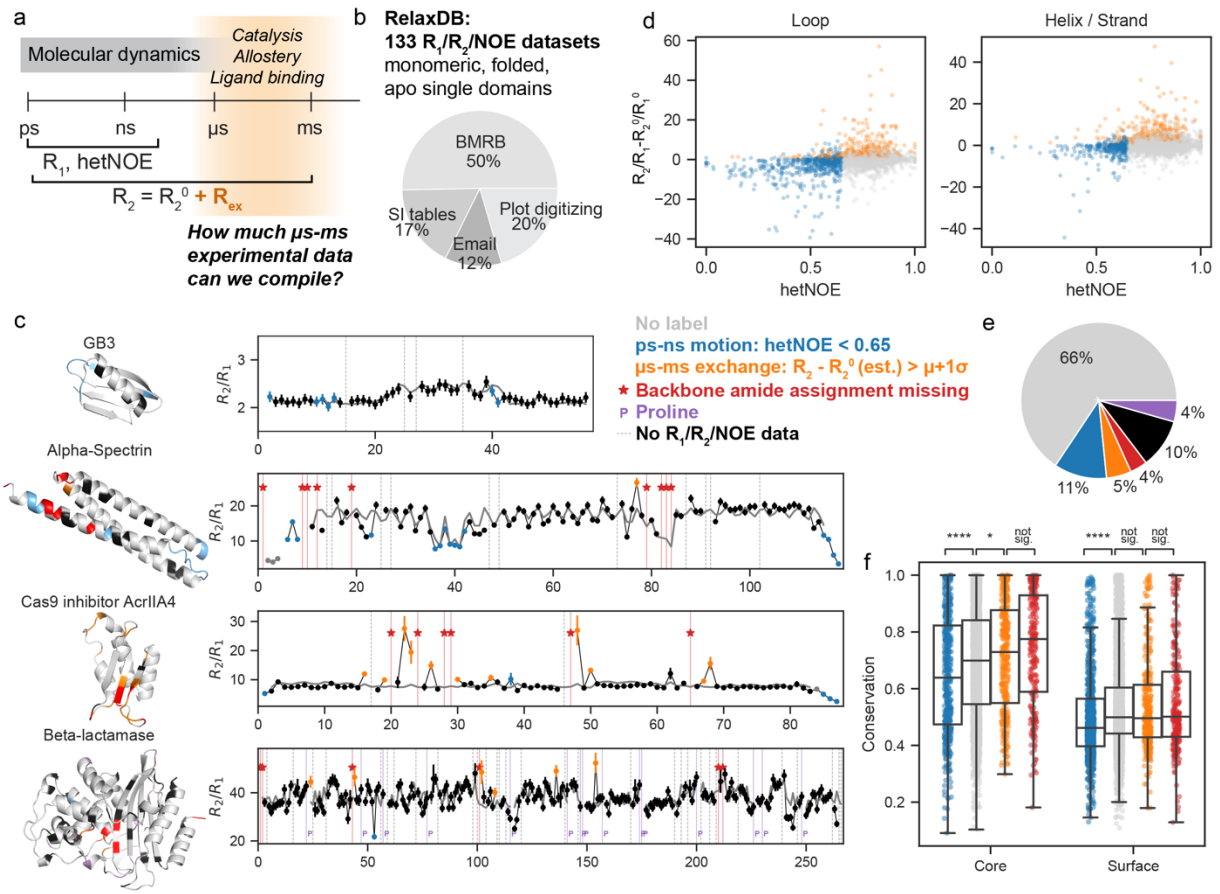



**Fig. 1. A benchmark of 133 protein NMR backbone relaxation measurements demonstrates residues with ps-ns motion are less conserved, and residues with µs-ms exchange are more conserved.** (a) Many biological processes occur at a µs-ms time regime, yet experimental data in this regime have not existed at a sufficient scale for machine learning. NMR $R_1$, $R_2$ and heteronuclear NOE (hetNOE) measure ps to low ns motion, while millisecond timescale exchange causes increased $R_2$ ($R_{ex}$). (b) The "RelaxDB" benchmark we curated consists of 133 single-domain $^{15}N$ protein backbone datasets with associated labels for the nature of motion present at each residue. Pie chart contains the origin of the data in RelaxDB. (c) Example raw $R_2$/ $R_1$ data (black) and comparison to calculation from rigid tumbling using HYDRONMR (grey line)[25,26] to obtain labels. Labels are shown in the different colors in legend and plotted onto the AF2 predicted structures. Error bars on $R_2/R_1$ data are calculated as the maximum of uncertainties reported by authors or

uncertainties estimated from the data (see Methods). (d) Distribution of hetNOE vs. calculated change in $R_2/R_1$ between experiment and rigid tumbling calculation. Blue: ps-ns motion. Orange: µs-ms exchange. Grey: no label. (e) Distribution of RelaxDB labels. (f) Calculating sequence conservation for these residues demonstrates that residues with µs-ms exchange are more conserved than residues with no assigned label, which in turn are more conserved than residues with ps-ns motion. Residues with missing backbone amide assignments are no different in conservation than residues with µs-ms exchange. Box plots depict median and 25/75% interquartile range, whiskers = 1.5 *interquartile range.

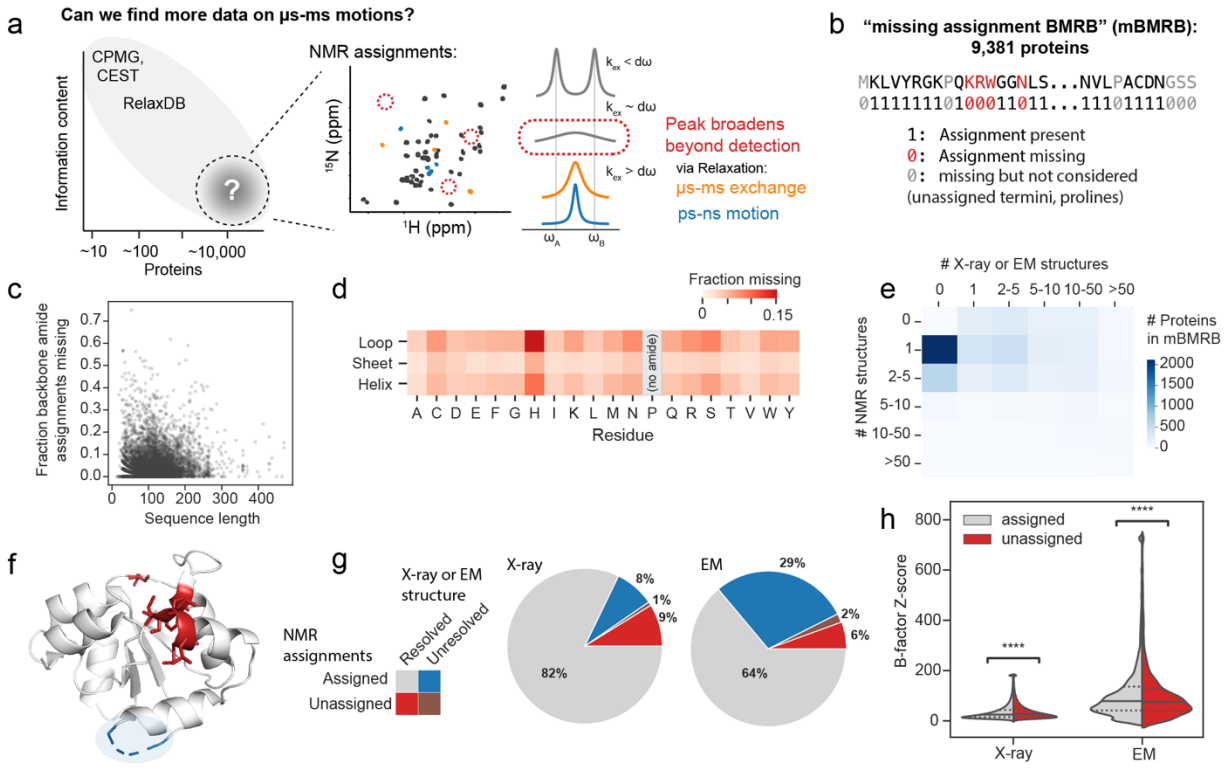



**Fig. 2. Missing NMR amide backbone assignments in BMRB used as label for µs-ms exchange.** (a) In seeking more experimental data for µs-ms dynamics, we realized that data on roughly 10,000 proteins exists in the form of missing backbone amide assignments in the BMRB. Theoretical NMR

line-shapes for 2-site exchange on different timescales shown. (b) Scheme for curating the "missing assignment BMRB" (mBMRB) dataset. (c) Fraction backbone amides missing vs. sequence length, not considering prolines or unassigned termini. (d) Fraction of backbone amides missing, disaggregated by amino acid and secondary structure type, in structures predicted by ESMFold[34]. (e) The majority of proteins in the BMRB have only one NMR structure and no X-ray or EM structure. (f) Residues with missing NMR peaks are distinct from residues missing in X-ray and EM structures, as exemplified by the MAP kinase binding domain of DUSP 16 (BMRB:19330, PDB:3TG3). Residues with missing assignments are in red, residues with missing X-ray density are in blue. (g) Fraction of residues assigned/unassigned in the mBMRB compared with resolved/unresolved in X-ray or EM structures shows little overlap. (h) B-factors of residues with present or missing NMR peaks have similar distributions. Solid lines: median, dotted lines: 25/75% percentile.

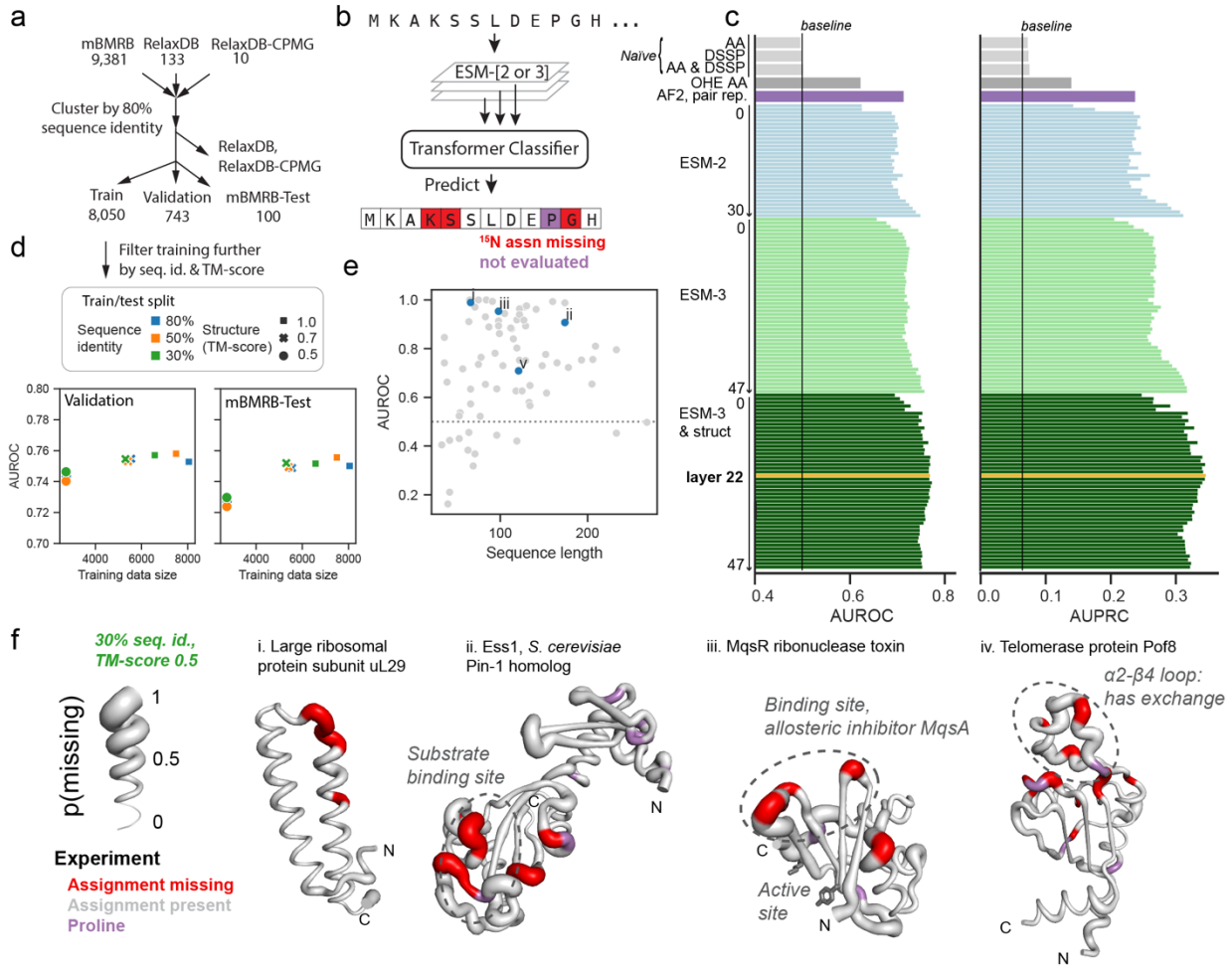



**Fig. 3. Deep learning has predictive power for the "missing assignment BMRB" (mBMRB).** (a) Train/validation/test split scheme to train from the mBMRB while holding out 100 proteins from mBMRB as well as all RelaxDB and RelaxDB-CPMG data as test sets. (b) Schematic of deep learning architectures tested. (c) AUROC and AUPRC on validation set testing layers from ESM-2[34] and ESM-3[2] models, AlphaFold2[6] pair representation, and baselines. AA: amino acid. DSSP: secondary structure assignment. OHE AA: one-hot encoded amino acid. Layer 22 was chosen as the best-performing model (Dyna-1). (d) Performance is not significantly impacted by removing sequence or structure homologues more stringently from training. (e) AUROC from Dyna-1, trained using the 30% sequence identity and 0.5 TM-score dataset split, on the mBMRB-Test set. Proteins in blue are shown in (f). (f) Examples of Dyna-1 predictions. Structures are generated with ESMFold

from mBMRB sequence (BMRB entries 5977, 50787, 51503, and 50002 for i-iv, respectively), the thickness of the tube is the probability missing as predicted by Dyna-1. Residues are colored for experimental data as follows: red: missing backbone assignment, grey: assignment present, purple: proline (no data).

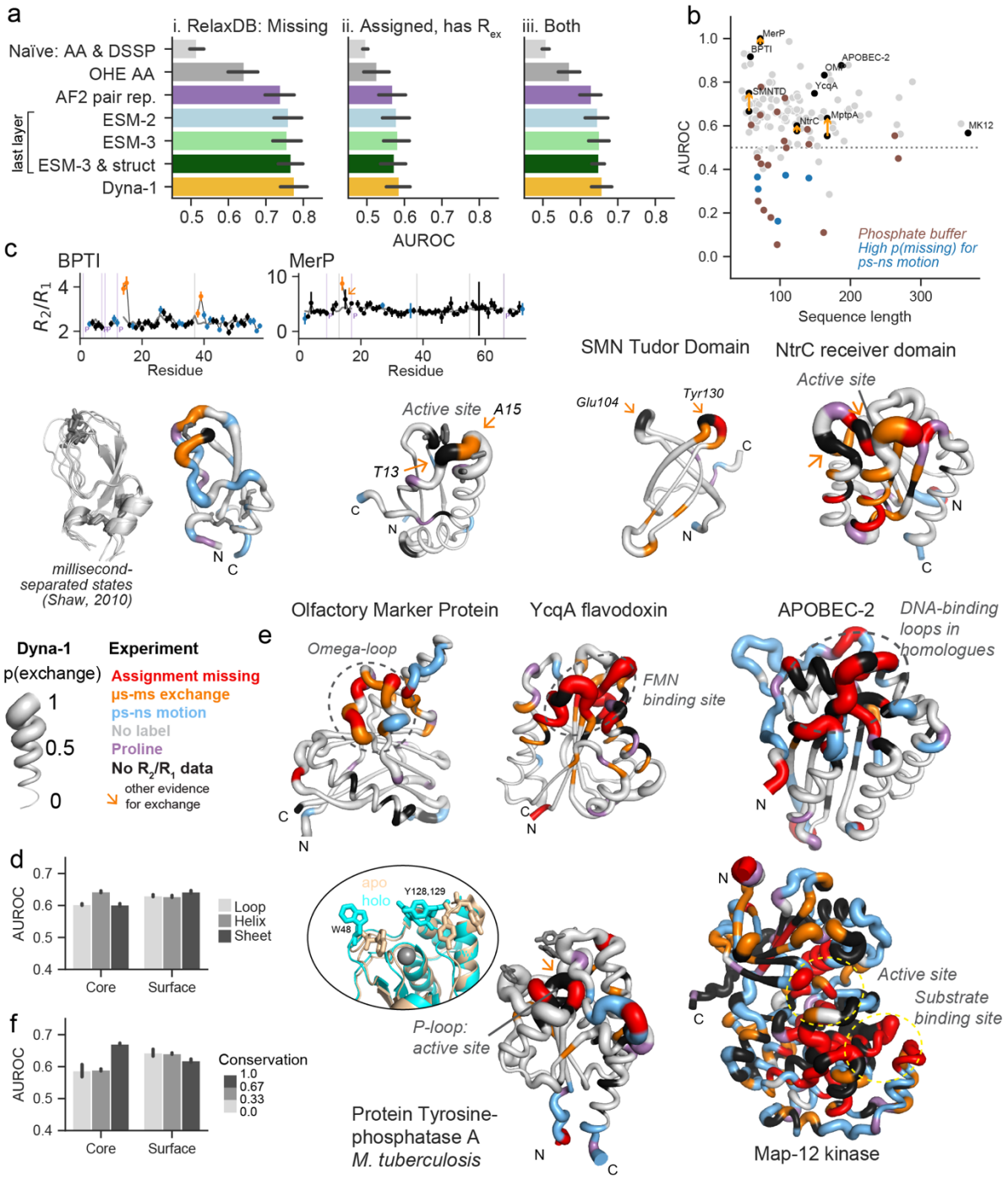

a
i. RelaxDB: Missing
ii. Assigned, has $R_{ex}$
iii. Both
Naïve: AA & DSSP
OHE AA
AF2 pair rep.
ESM-2
ESM-3
ESM-3 & struct
Dyna-1
last layer
AUROC
b
MerP
BPTI
APOBEC-2
OMP
YcqA
SMNTD
NtrC
MptpA
MK12
Phosphate buffer
High p(missing) for ps-ns motion
Sequence length
c
BPTI
MerP
$R_2/R_1$
Residue
millisecond-separated states (Shaw, 2010)
SMN Tudor Domain
Active site
A15
T13
Glu104
Tyr130
NtrC receiver domain
Olfactory Marker Protein
YcqA flavodoxin
APOBEC-2
DNA-binding loops in homologues
Omega-loop
FMN binding site
Dyna-1
p(exchange)
Experiment
Assignment missing
µs-ms exchange
ps-ns motion
No label
Proline
No $R_2/R_1$ data
other evidence for exchange
d
Loop
Helix
Sheet
Core
Surface
f
Conservation
1.0
0.67
0.33
0.0
e
apo
holo
Y128,129
W48
P-loop: active site
Protein Tyrosine-phosphatase A
M. tuberculosis
Active site
Substrate binding site
Map-12 kinase

**Fig. 4. Dyna-1 predicts biologically-relevant micro-millisecond dynamics in NMR relaxation data**. (a) Representative models evaluated on (i) missing assignments in RelaxDB, (ii) residues with assigned $R_{ex}$, and (iii) both missing and $R_{ex}$. Error bars represent 95% confidence interval for mean evaluated over proteins in RelaxDB. Normalized AUROC is calculated by subtracting the baseline AUROC from the AUROC for each protein, and the metric by which the final model was chosen. Error bars represent 95% confidence interval for mean normalized AUROC evaluated over proteins in RelaxDB. (b) AUROC per protein vs. sequence length for Dyna-1 in RelaxDB for both missing assignments and residues with $R_{ex}$. Only shown are proteins with 2 or more residues with missing assignment or exchange. Brown: proteins with exchange from phosphate buffer. Blue: proteins with high p(exchange) primarily in regions with ps-ns motions (cf. Extended Data Fig. 7). Orange arrows: proteins where follow-up on residues with no relaxation data resulted in increased AUROC (see c,e). (c) Dyna-1 predictions on smaller proteins offer an opportunity for detailed comparisons to experimental data. BPTI agrees with experimental $R_2/R_1$ data[44] (top) and heterogeneity among kinetically-distinct substates in molecular dynamics[45] (left). Mercuric Transport Protein, SMN Tudor Domain, and NtrC receiver domain offer prospective tests of residues with high p(exchange) for which no data was available (black), but other evidence for $R_{ex}$ (orange arrow). $R_2/R_1$ error bars are from the original curated datasets. (d) AUROC for residues binned by core and surface designation and secondary structure type indicates that Dyna-1 has the most predictive power in core helices and surface loops. (e) Representative Dyna-1 predictions for larger proteins. Biologically-relevant features are annotated in grey. The Dyna-1 prediction for MptpA, a phosphatase from *M. tuberculosis*, recapitulates measurements from NMR as well as structural heterogeneity in sidechains regulating specificity (W48, Y128, Y129) (apo: 2LUO, holo: 1UPA). (f) AUROC binned by core/surface location and sequence conservation indicate that Dyna-1 has the most predictive

power for more conserved residues. In (d,f), error bars represent standard error evaluated via bootstrapping.

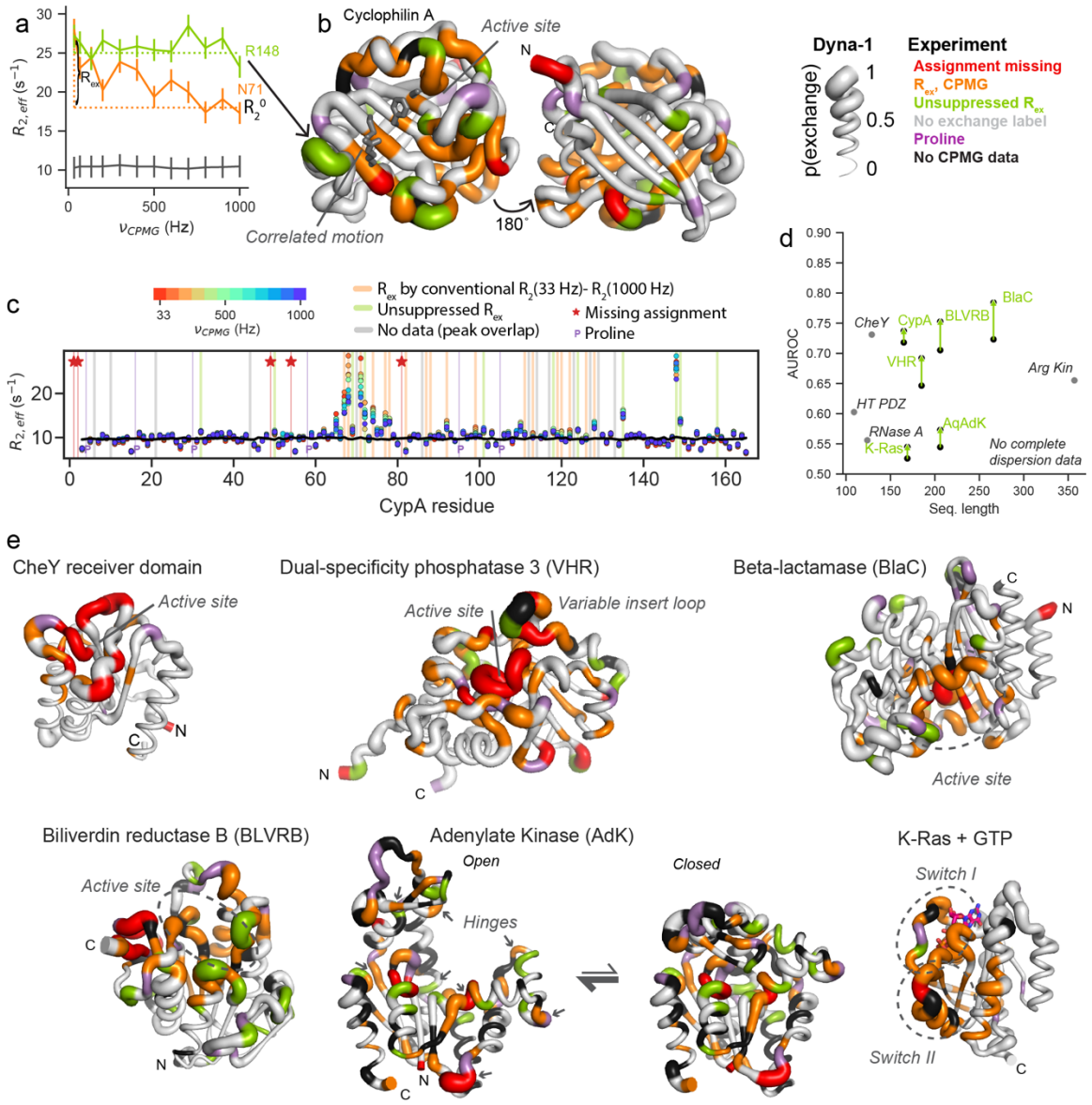



**Fig. 5. Dyna-1 even predicts exchange in residues that conventional CPMG data analysis misses**. (a) Example $^{15}$N CPMG data for Cyclophilin A (CypA)[57]. Error bars represent background noise in spectrum. $R_{ex}$ is typically estimated via difference in $R_{2,eff}$ at lowest field strength and $R_2^0$ value to which $R_{2,eff}$ plateaus. However, plotting $R_{2,eff}$ for Arg148 (green) demonstrates different behavior: $R_{2,eff}$ remains high over all $B_1$ field strengths used, indicative of exchange on the microsecond regime that is too fast to be suppressed with the CPMG field strengths used. (b) AF2

structure of CypA with tube thickness by Dyna-1 p(exchange). Legend coloring according to the experimental data is shown to the right. Dyna-1 predicts high exchange for Arg148. (c) We devised a data-processing method using $R_2$ predicted by HYDRONMR to systematically assign residues with $R_{ex}$ for residues where $R_{2,eff}$ get suppressed by CPMG pulse train (conventional $R_{ex}$, in orange) or residues with elevated, unsuppressed $R_{ex}$ (green). Spectra depicted with representative data for CypA was collected at 25 ˚C. (d) Incorporating labels for unsuppressed $R_{ex}$ increases reported AUROC for 5/6 proteins where complete dispersion data was available for reprocessing (shown by green arrows). Also shown is the AUROC for proteins where we could not find dispersion data for all residues, and instead curated residue labels based on descriptions in literature. (e) Representative experimental data and Dyna-1 predictions for a selection of proteins with dynamics characterized with $^{15}N$ CPMG. For Adenylate Kinase, structure predictions of the open and closed states from AF2 were used as input into Dyna-1.

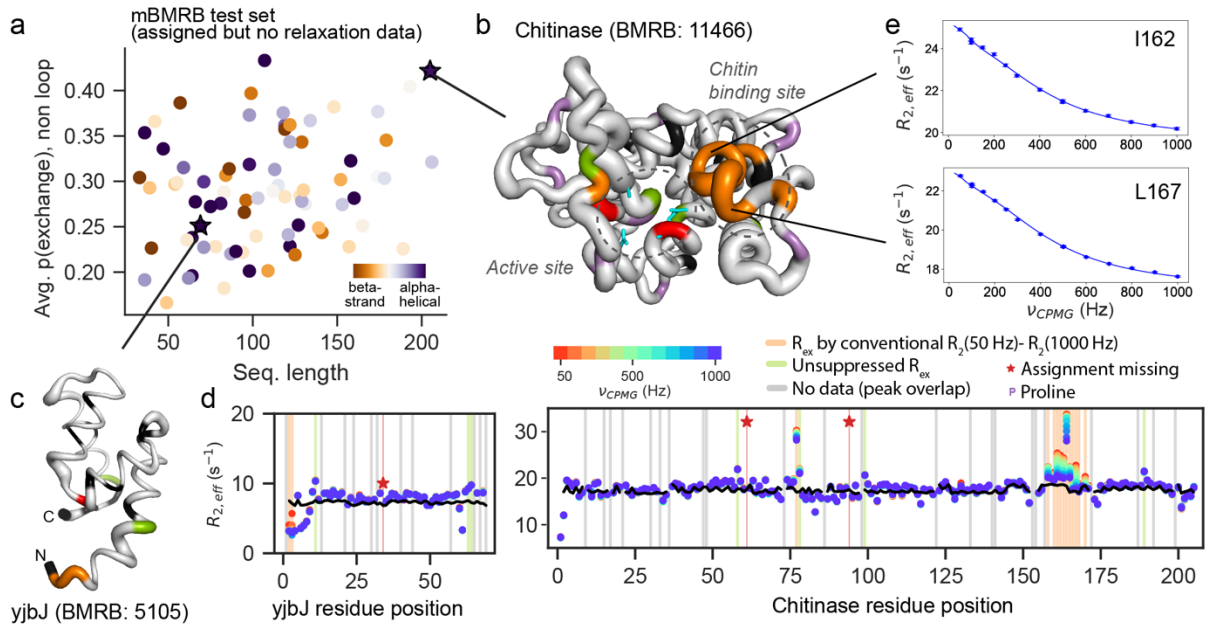



**Fig. 6: Dyna-1 has predictive power for enzymatic motions in prospective experimental test.** a) Dyna-1 proteins in mBMRB test set vary in net p(exchange) predicted across sequence length and helix/strand content. We selected two proteins to experimentally test Dyna-1 predictions: (b) Chitinase 19 from *Bryum coronatum*, and (c) hypothetical protein yjbJ. Residue coloring in (b,c,d) is same as in Fig. 5. (d) $^{15}$N CPMG relaxation dispersion only showed dispersion curves for two N-terminal residues in yjbJ, whereas chitinase has extensive exchange in the active site and chitin binding site (see also supplemental dataset 1). (e) Representative $^{15}$N CPMG dispersion data for I162 and L167 in the chitin binding site region.


References

1 Hoch, J. C. *et al.* Biological Magnetic Resonance Data Bank. *Nucleic Acids Res* **51**, D368-D376 (2023). https://doi.org:10.1093/nar/gkac1050

2 Hayes, T. *et al.* Simulating 500 million years of evolution with a language model. *Science*, eads0018 (2025). https://doi.org:10.1126/science.ads0018

3 Frauenfelder, H., Sligar, S. G. & Wolynes, P. G. The energy landscapes and motions of proteins. *Science* **254**, 1598-1603 (1991). https://doi.org:10.1126/science.1749933
4 McCammon, J. A., Gelin, B. R. & Karplus, M. Dynamics of folded proteins. *Nature* **267**, 585-590 (1977). https://doi.org:10.1038/267585a0
5 McCammon, J. A. & Karplus, M. Internal motions of antibody molecules. *Nature* **268**, 765-766 (1977). https://doi.org:10.1038/268765a0
6 Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 (2021). https://doi.org:10.1038/s41586-021-03819-2
7 Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res* **28**, 235-242 (2000). https://doi.org:10.1093/nar/28.1.235
8 Marks, D. S. *et al.* Protein 3D structure computed from evolutionary sequence variation. *PLoS One* **6**, e28766 (2011). https://doi.org:10.1371/journal.pone.0028766
9 Morcos, F. *et al.* Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *Proc Natl Acad Sci U S A* **108**, E1293-1301 (2011). https://doi.org:10.1073/pnas.1111471108
10 Henzler-Wildman, K. & Kern, D. Dynamic personalities of proteins. *Nature* **450**, 964-972 (2007). https://doi.org:10.1038/nature06522
11 van den Bedem, H. & Fraser, J. S. Integrative, dynamic structural biology at atomic resolution--it's about time. *Nat Methods* **12**, 307-318 (2015). https://doi.org:10.1038/nmeth.3324
12 Hansen, D. F., Vallurupalli, P. & Kay, L. E. Using relaxation dispersion NMR spectroscopy to determine structures of excited, invisible protein states. *J Biomol NMR* **41**, 113-120 (2008). https://doi.org:10.1007/s10858-008-9251-5
13 Tollinger, M., Skrynnikov, N. R., Mulder, F. A., Forman-Kay, J. D. & Kay, L. E. Slow dynamics in folded and unfolded states of an SH3 domain. *J Am Chem Soc* **123**, 11341-11352 (2001). https://doi.org:10.1021/ja011300z
14 Bruschweiler, R. New approaches to the dynamic interpretation and prediction of NMR relaxation data from proteins. *Curr Opin Struct Biol* **13**, 175-183 (2003). https://doi.org:10.1016/s0959-440x(03)00036-8
15 Cilia, E., Pancsa, R., Tompa, P., Lenaerts, T. & Vranken, W. F. From protein sequence to dynamics and disorder with DynaMine. *Nat Commun* **4**, 2741 (2013). https://doi.org:10.1038/ncomms3741
16 Wayment-Steele, H. K. *et al.* Predicting multiple conformations via sequence clustering and AlphaFold2. *Nature* **625**, 832-839 (2024). https://doi.org:10.1038/s41586-023-06832-9
17 Lewis, S. *et al.* Scalable emulation of protein equilibrium ensembles with generative deep learning. *Science* **389**, eadv9817 (2025). https://doi.org:10.1126/science.adv9817
18 Loria, J. P., Rance, M. & Palmer, A. G., 3rd. A TROSY CPMG sequence for characterizing chemical exchange in large proteins. *J Biomol NMR* **15**, 151-155 (1999). https://doi.org:10.1023/a:1008355631073

19 Vallurupalli, P., Bouvignies, G. & Kay, L. E. Increasing the exchange time-scale that can be probed by CPMG relaxation dispersion NMR. *J Phys Chem B* **115**, 14891-14900 (2011). https://doi.org:10.1021/jp209610v
20 Clore, G. M., Driscoll, P. C., Wingfield, P. T. & Gronenborn, A. M. Analysis of the backbone dynamics of interleukin-1 beta using two-dimensional inverse detected heteronuclear 15N-1H NMR spectroscopy. *Biochemistry* **29**, 7387-7401 (1990). https://doi.org:10.1021/bi00484a006
21 Goodman, J. L., Pagel, M. D. & Stone, M. J. Relationships between protein structure and dynamics from a database of NMR-derived backbone order parameters. *J Mol Biol* **295**, 963-978 (2000). https://doi.org:10.1006/jmbi.1999.3419
22 Mandel, A. M., Akke, M. & Palmer, A. G., 3rd. Backbone dynamics of Escherichia coli ribonuclease HI: correlations with structure and function in an active enzyme. *J Mol Biol* **246**, 144-163 (1995). https://doi.org:10.1006/jmbi.1994.0073
23 Lipari, G. & Szabo, A. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *Journal of the American Chemical Society* **104**, 4546-4559 (1982). https://doi.org:10.1021/ja00381a009
24 Cole, R. & Loria, J. P. FAST-Modelfree: a program for rapid automated analysis of solution NMR spin-relaxation data. *J Biomol NMR* **26**, 203-213 (2003). https://doi.org:10.1023/a:1023808801134
25 Bernado, P., Garcia de la Torre, J. & Pons, M. Interpretation of 15N NMR relaxation data of globular proteins using hydrodynamic calculations with HYDRONMR. *J Biomol NMR* **23**, 139-150 (2002). https://doi.org:10.1023/a:1016359412284
26 Garcia de la Torre, J., Huertas, M. L. & Carrasco, B. HYDRONMR: prediction of NMR relaxation of globular proteins from atomic-level structures and hydrodynamic calculations. *J Magn Reson* **147**, 138-146 (2000). https://doi.org:10.1006/jmre.2000.2170
27 Raza, T. *et al.* Insights into the NF-κB-DNA Interaction through NMR Spectroscopy. *ACS Omega* **6**, 12877-12886 (2021). https://doi.org:10.1021/acsomega.1c01299
28 Korn, S. M. *et al.* (1)H, (13)C, and (15)N backbone chemical shift assignments of the nucleic acid-binding domain of SARS-CoV-2 non-structural protein 3e. *Biomol NMR Assign* **14**, 329-333 (2020). https://doi.org:10.1007/s12104-020-09971-6
29 Johnson, E. C., Lazar, G. A., Desjarlais, J. R. & Handel, T. M. Solution structure and dynamics of a designed hydrophobic core variant of ubiquitin. *Structure* **7**, 967-976 (1999). https://doi.org:https://doi.org/10.1016/S0969-2126(99)80123-3
30 Hall, J. B. & Fushman, D. Characterization of the overall and local dynamics of a protein with intermediate rotational anisotropy: Differentiating between conformational exchange and anisotropic diffusion in the B3 domain of protein G. *J Biomol NMR* **27**, 261-275 (2003). https://doi.org:10.1023/a:1025467918856
31 Brenner, A. K., Kieffer, B., Trave, G., Froystein, N. A. & Raae, A. J. Thermal stability of chicken brain alpha-spectrin repeat 17: a spectroscopic study. *J Biomol NMR* **53**, 71-83 (2012). https://doi.org:10.1007/s10858-012-9620-y
32 Kim, I. *et al.* Solution structure and dynamics of anti-CRISPR AcrIIA4, the Cas9 inhibitor. *Sci Rep* **8**, 3883 (2018). https://doi.org:10.1038/s41598-018-22177-0

33 Elings, W. *et al*. Two beta-Lactamase Variants with Reduced Clavulanic Acid Inhibition Display Different Millisecond Dynamics. *Antimicrob Agents Chemother* **65**, e0262820 (2021). https://doi.org:10.1128/AAC.02628-20
34 Lin, Z. *et al*. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* **379**, 1123-1130 (2023). https://doi.org:10.1126/science.ade2574
35 Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research* **15**, 1929-1958 (2014).
36 Zhou, M. *et al*. in *International Conference on Machine Learning*. 7594-7602 (PMLR).
37 Hauser, M., Steinegger, M. & Soding, J. MMseqs software suite for fast and deep clustering and searching of large protein sequence sets. *Bioinformatics* **32**, 1323-1330 (2016). https://doi.org:10.1093/bioinformatics/btw006
38 Weiss, K., Khoshgoftaar, T. M. & Wang, D. A survey of transfer learning. *Journal of Big Data* **3**, 9 (2016). https://doi.org:10.1186/s40537-016-0043-6
39 Yu, V., Ronzone, E., Lord, D., Peti, W. & Page, R. MqsR is a noncanonical microbial RNase toxin that is inhibited by antitoxin MqsA via steric blockage of substrate binding. *J Biol Chem* **298**, 102535 (2022). https://doi.org:10.1016/j.jbc.2022.102535
40 Basu, R., Eichhorn, C. D., Cheng, R., Peterson, R. D. & Feigon, J. Structure of S. pombe telomerase protein Pof8 C-terminal domain is an xRRM conserved among LARP7 proteins. *RNA Biol* **18**, 1181-1192 (2021). https://doi.org:10.1080/15476286.2020.1836891
41 Sakhrani, V. V. *et al*. Backbone assignments and conformational dynamics in the S. typhimurium tryptophan synthase alpha-subunit from solution-state NMR. *J Biomol NMR* **74**, 341-354 (2020). https://doi.org:10.1007/s10858-020-00320-2
42 Morin, S. & Gagne, S. M. NMR dynamics of PSE-4 beta-lactamase: an interplay of ps-ns order and mus-ms motions in the active site. *Biophys J* **96**, 4681-4691 (2009). https://doi.org:10.1016/j.bpj.2009.02.068
43 Grey, M. J., Wang, C. & Palmer, A. G., 3rd. Disulfide bond isomerization in basic pancreatic trypsin inhibitor: multisite chemical exchange quantified by CPMG relaxation dispersion and chemical shift modeling. *J Am Chem Soc* **125**, 14324-14335 (2003). https://doi.org:10.1021/ja0367389
44 Beeser, S. A., Goldenberg, D. P. & Oas, T. G. Enhanced protein flexibility caused by a destabilizing amino acid replacement in BPTI. *J Mol Biol* **269**, 154-164 (1997). https://doi.org:10.1006/jmbi.1997.1031
45 Shaw, D. E. *et al*. Atomic-level characterization of the structural dynamics of proteins. *Science* **330**, 341-346 (2010). https://doi.org:10.1126/science.1187409
46 Steele, R. A. & Opella, S. J. Structures of the reduced and mercury-bound forms of MerP, the periplasmic protein from the bacterial mercury detoxification system. *Biochemistry* **36**, 6885-6895 (1997). https://doi.org:10.1021/bi9631632
47 Kawale, A. A. & Burmann, B. M. Inherent backbone dynamics fine-tune the functional plasticity of Tudor domains. *Structure* **29**, 1253-1265 e1254 (2021). https://doi.org:10.1016/j.str.2021.06.007

48 Volkman, B. F., Lipson, D., Wemmer, D. E. & Kern, D. Two-state allosteric behavior in a single-domain signaling protein. *Science* **291**, 2429-2433 (2001). https://doi.org:10.1126/science.291.5512.2429
49 Baldisseri, D. M., Margolis, J. W., Weber, D. J., Koo, J. H. & Margolis, F. L. Olfactory marker protein (OMP) exhibits a beta-clam fold in solution: implications for target peptide interaction and olfactory signal transduction. *J Mol Biol* **319**, 823-837 (2002). https://doi.org:10.1016/S0022-2836(02)00282-6
50 Nakashima, N., Nakashima, A., Nakashima, K. & Takano, M. Olfactory marker protein contains a leucine-rich domain in the Omega-loop important for nuclear export. *Mol Brain* **15**, 89 (2022). https://doi.org:10.1186/s13041-022-00973-0
51 Davis, J. H. & Agard, D. A. Relationship between enzyme specificity and the backbone dynamics of free and inhibited alpha-lytic protease. *Biochemistry* **37**, 7696-7707 (1998). https://doi.org:10.1021/bi972963p
52 Salter, J. D. & Smith, H. C. Modeling the Embrace of a Mutator: APOBEC Selection of Nucleic Acid Ligands. *Trends in Biochemical Sciences* **43**, 606-622 (2018). https://doi.org:10.1016/j.tibs.2018.04.013
53 Stehle, T. *et al.* The apo-structure of the low molecular weight protein-tyrosine phosphatase A (MptpA) from Mycobacterium tuberculosis allows for better target-specific drug development. *J Biol Chem* **287**, 34569-34582 (2012). https://doi.org:10.1074/jbc.M112.399261
54 Aoto, P. C., Stanfield, R. L., Wilson, I. A., Dyson, H. J. & Wright, P. E. A dynamic switch in inactive p38γ leads to an excited state on the pathway to an active kinase. *Biochemistry* **58**, 5160-5172 (2019).
55 Eisenmesser, E. Z. *et al.* Intrinsic dynamics of an enzyme underlies catalysis. *Nature* **438**, 117-121 (2005). https://doi.org:10.1038/nature04105
56 Fraser, J. S. *et al.* Hidden alternative structures of proline isomerase essential for catalysis. *Nature* **462**, 669-673 (2009). https://doi.org:10.1038/nature08615
57 Eisenmesser, E. Z., Bosco, D. A., Akke, M. & Kern, D. Enzyme dynamics during catalysis. *Science* **295**, 1520-1523 (2002). https://doi.org:10.1126/science.1066176
58 Henzler-Wildman, K. A. *et al.* Intrinsic motions along an enzymatic reaction trajectory. *Nature* **450**, 838-844 (2007). https://doi.org:10.1038/nature06410
59 Paukovich, N. *et al.* Biliverdin Reductase B Dynamics Are Coupled to Coenzyme Binding. *J Mol Biol* **430**, 3234-3250 (2018). https://doi.org:10.1016/j.jmb.2018.06.015
60 Hansen, A. L., Xiang, X., Yuan, C., Bruschweiler-Li, L. & Bruschweiler, R. Excited-state observation of active K-Ras reveals differential structural dynamics of wild-type versus oncogenic G12D and G12C mutants. *Nat Struct Mol Biol* **30**, 1446-1455 (2023). https://doi.org:10.1038/s41594-023-01070-z
61 Beaumont, V. A. *et al.* Allosteric Impact of the Variable Insert Loop in Vaccinia H1-Related (VHR) Phosphatase. *Biochemistry* **59**, 1896-1908 (2020). https://doi.org:10.1021/acs.biochem.0c00245
62 Ohnuma, T. *et al.* Chitin oligosaccharide binding to a family GH19 chitinase from the moss Bryum coronatum. *FEBS J* **278**, 3991-4001 (2011). https://doi.org:10.1111/j.1742-4658.2011.08301.x

63 Yee, A. *et al*. An NMR approach to structural proteomics. *Proc Natl Acad Sci U S A* **99**, 1825-1830 (2002). https://doi.org:10.1073/pnas.042684599
64 Dudley, J. A., Park, S., MacDonald, M. E., Fetene, E. & Smith, C. A. Resolving overlapped signals with automated FitNMR analytical peak modeling. *J Magn Reson* **318**, 106773 (2020). https://doi.org:10.1016/j.jmr.2020.106773
65 Klukowski, P. *et al*. The 100-protein NMR spectra dataset: A resource for biomolecular NMR data analysis. *Sci Data* **11**, 30 (2024). https://doi.org:10.1038/s41597-023-02879-5
66 Klukowski, P., Riek, R. & Guntert, P. Rapid protein assignments and structures from raw NMR spectra with the deep learning technique ARTINA. *Nat Commun* **13**, 6151 (2022). https://doi.org:10.1038/s41467-022-33879-5
67 Kneller, J. M., Lu, M. & Bracken, C. An effective method for the discrimination of motional anisotropy and chemical exchange. *J Am Chem Soc* **124**, 1852-1853 (2002). https://doi.org:10.1021/ja017461k
68 Pei, J. & Grishin, N. V. AL2CO: calculation of positional conservation in a protein sequence alignment. *Bioinformatics* **17**, 700-712 (2001). https://doi.org:10.1093/bioinformatics/17.8.700
69 Shrake, A. & Rupley, J. A. Environment and exposure to solvent of protein atoms. Lysozyme and insulin. *J Mol Biol* **79**, 351-371 (1973). https://doi.org:10.1016/0022-2836(73)90011-9
70 McGibbon, R. T. *et al*. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys J* **109**, 1528-1532 (2015). https://doi.org:10.1016/j.bpj.2015.08.015
71 Chen, H. & Zhou, H. X. Prediction of solvent accessibility and sites of deleterious mutations from protein sequence. *Nucleic Acids Res* **33**, 3193-3199 (2005). https://doi.org:10.1093/nar/gki633
72 Beaumont, V. A. *Functional Implications of Conformational Dynamics of Protein Tyrosine Phosphatases*, Yale University, (2019).
73 McDonald, L. R., Boyer, J. A. & Lee, A. L. Segmental motions, not a two-state concerted switch, underlie allostery in CheY. *Structure* **20**, 1363-1373 (2012). https://doi.org:10.1016/j.str.2012.05.008
74 Aspholm, E. E., Lidman, J. & Burmann, B. M. Structural basis of substrate recognition and allosteric activation of the proapoptotic mitochondrial HtrA2 protease. *Nat Commun* **15**, 4592 (2024). https://doi.org:10.1038/s41467-024-48997-5
75 Beach, H., Cole, R., Gill, M. L. & Loria, J. P. Conservation of mus-ms enzyme motions in the apo- and substrate-mimicked state. *J Am Chem Soc* **127**, 9167-9176 (2005). https://doi.org:10.1021/ja0514949
76 Davulcu, O., Flynn, P. F., Chapman, M. S. & Skalicky, J. J. Intrinsic domain and loop dynamics commensurate with catalytic turnover in an induced-fit enzyme. *Structure* **17**, 1356-1367 (2009). https://doi.org:10.1016/j.str.2009.08.014
77 Zhang, Y. & Skolnick, J. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res* **33**, 2302-2309 (2005). https://doi.org:10.1093/nar/gki524

78 Case, D. A. Normal mode analysis of protein dynamics. *Current Opinion in Structural Biology* **4**, 285-290 (1994). https://doi.org:https://doi.org/10.1016/S0959-440X(94)90321-2
79 Zhang, S. *et al.* ProDy 2.0: increased scale and scope after 10 years of protein dynamics modelling with Python. *Bioinformatics* **37**, 3657-3659 (2021). https://doi.org:10.1093/bioinformatics/btab187
80 Loshchilov, I. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101* (2017).
81 Menon, A. K. *et al.* Long-tail learning via logit adjustment. *arXiv preprint arXiv:2007.07314* (2020).
82 Waskom, M. L. Seaborn: statistical data visualization. *Journal of Open Source Software* **6**, 3021 (2021).
83 Taira, T. *et al.* Cloning and characterization of a small family 19 chitinase from moss (Bryum coronatum). *Glycobiology* **21**, 644-654 (2011). https://doi.org:10.1093/glycob/cwq212
84 Hansen, D. F., Vallurupalli, P. & Kay, L. E. An improved 15N relaxation dispersion experiment for the measurement of millisecond time-scale dynamics in proteins. *J Phys Chem B* **112**, 5898-5904 (2008). https://doi.org:10.1021/jp074793o
85 Jiang, B., Yu, B., Zhang, X., Liu, M. & Yang, D. A (15)N CPMG relaxation dispersion experiment more resistant to resonance offset and pulse imperfection. *J Magn Reson* **257**, 1-7 (2015). https://doi.org:10.1016/j.jmr.2015.05.003
86 Ahlner, A., Carlsson, M., Jonsson, B. H. & Lundstrom, P. PINT: a software for integration of peak volumes and extraction of relaxation rates. *J Biomol NMR* **56**, 191-202 (2013). https://doi.org:10.1007/s10858-013-9737-7
87 Vallurupalli, P., Bouvignies, G. & Kay, L. E. Studying "invisible" excited protein states in slow exchange with a major state conformation. *J Am Chem Soc* **134**, 8148-8161 (2012). https://doi.org:10.1021/ja3001419

## Methods

### Curating the "RelaxDB" dataset.

*Theoretical basis for relaxation measurements*. The following is condensed from ref. [67]. The longitudinal ($R_1$) and transverse ($R_2$) relaxation rates of a backbone nitrogen atom attached to a single proton may be written in a simplified form as

$$R_1 = \frac{A}{10} J(\omega_N) \,,$$

$$R_2 = \frac{A}{15} [J(0) + \frac{3}{4} J(\omega_N)] + R_{ex},$$

where $J(\omega) = S^2\tau_C/(1+\omega^2{\tau_C}^2)$. $J(\omega)$ is the spectral density function describing the intensity of exchange occurring at frequency $\omega$. These expressions do not include terms on the order of $J(\omega_H)$, which are small on the timescale of protein tumbling ($\omega_N\tau_C \gg 1$). In the above, $A = 3d^2 + 4c^2$, where d=$\frac{\mu_0\, h\, \gamma_H\, \gamma_N}{r_{NH} 8\pi^2}$ and $c = \frac{\gamma_N B_0 \Delta\sigma}{\sqrt{3}}$. $\mu_0$ is the permeability of free space, *h* is Planck's constant, $\gamma_X$ is the gyromagnetic ratio of atom X (nitrogen N or hydrogen H), $r_{NH}$is the bond length of the N-H bond, $B_0$ is the static magnetic field strength, and $\Delta\sigma$ is the chemical shift anisotropy of nitrogen.

Note that the ratio $R_2/R_1$ for residues that do not demonstrate intramolecular dynamics can be used to estimate the global tumbling time $\tau_C$ via the following relationship:

$$\frac{R_2}{R_1} = \frac{2}{3}\frac{J(0)}{J(\omega_N)} + \frac{1}{2} = \frac{2}{3}{\omega_N}^2{\tau_C}^2 + \frac{7}{6}.$$

*BMRB entries.*

To get entries from the BMRB that contained $R_1$, $R_2$ and hetNOE data, we used https://bmrb.io/software/query/ to download a list of all entries with one entity of polymer type polypeptide(L) that contained $R_2$ relaxation data (referred to as $T_2$ in the user interface). This resulted in 344 entries when queried in January 2023. We manually went through the list and identified entries that were monomers, had no cofactors present, were not in micelles/bicelles/nanodisks, were not in denaturing conditions, and had at least one set of $R_1/R_2$/hetNOE data all at the same field strength. This resulted in 91 entries. We included data collected at multiple field strengths when available. Several of the BMRB entries had incorrect units, which we fixed by visually inspecting data and comparing to figures of plotted data in the corresponding publications. For three entries, we realized that they contained two or more domains, but that these two or more domains could be modelled as tumbling independently. We therefore separated these into two separate entries.

*Data from literature.*

To find data from literature, we searched in PubMed and Google Scholar for keywords “NMR”, “backbone”, and “relaxation”. Upon finding a dataset of an apo, monomeric, stably folded protein and if the data was only available in figure form, we used PlotDigitizer (https://plotdigitizer.com/) to compile values for $R_1$/$R_2$/hetNOE. We did not include datasets if the figures were too low-resolution to allow for accurately using PlotDigitizer. We only used datasets where we could identify a corresponding set of assignments either cited in the same paper or clearly referencing another set of assignments.

**Fitting data / assigning labels.**

*Generating MSAs.*

We generated MSAs using the `run_mmseqs2` command in ColabFold with `filter=True`. We used these MSAs to predict AF2 models. Prior to calculating conservation metrics (below), we removed sequences with more than 10% gaps and used MMseqs easy-cluster to further filter with options `-c 0.9 --cov-mode 1 --min-seq-id 0.9`.

*Generating structure models.*

We used AF2 as implemented in ColabDesign (https://github.com/sokrypton/ColabDesign.git@gamma) to generate the models used for HYDRONMR calculations. We ran AF2 for 3 recycles, seed=0. We used model 1 for each protein. For the Adenylate Kinase datasets (both from *E. coli* and *A. Aquifex*), AF2 predicts the closed state though it is known to occupy the open state predominantly in solution. Therefore, we templated AF2 using structure model PDB ID 4AKE of the open state.

We used these structure models as inputs into HYDRONMR[25,26] with default settings to estimate R2/R1 from rigid tumbling, denoted as $\widehat{R_2^0/R_1^0}$. We trimmed unstructured termini of the 161 proteins by removing residues from the N or C termini until the pLDDT of the next residue was over 70, or over 80 for CBFB and PCF11. Residues corresponding to unstructured termini are not considered during evaluation.

*Fitting $R_2/R_1$ to rigid tumbling estimates.*

We needed to account for a constant offset in the overall tumbling time between each experimental dataset and the corresponding prediction. To do so, we fit $\widehat{R_2^0/R_1^0}$ to the experimental data $R_2/R_1$ as follows: for each dataset, we wish to minimize the squared difference between $R_2/R_1$ and $C\ \widehat{R_2^0/R_1^0}$ where C is a scalar to be fit, i.e.:

$$\min_C \frac{1}{n} \sum_{i=1}^{N} \left[\frac{R_2}{R_1}(i) - C\,\frac{\widehat{R_2^0}}{R_1^0}(i)\right]^2 .$$

This has a closed-form solution for C:

$$C = \frac{\sum_i \frac{R_2}{R_1}(i) \times \frac{\widehat{R_2^0}}{R_1^0}(i)}{\sum_i \frac{\widehat{R_2^0}}{R_1^0}(i)^2}.$$

To avoid potential influence of outliers, we first calculate C excluding residues with hetNOE <0.65 as used in refs. [27-29] to designate residues with ps-ns motion. We use this C to calculate residuals (i.e., $\frac{R_2}{R_1}(i) - C\,\frac{\widehat{R_2^0}}{R_1^0}(i)$), then calculated C again excluding residues with a residual greater than the interquartile range.

Residues exhibiting $R_2/R_1$ greater than that predicted by rigid tumbling, i.e. $R_{ex}$, were identified as those with $dR_2/R_1$ > experimental error + fitting error. Fitting error for each protein was estimated as the interquartile range of the residuals, as this is more robust to outliers than a standard deviation. Experimental error for each residue was taken as the maximum of 5% of the

reported $R_2/R_1$ value or the reported error in $R_2/R_1$, whichever value was larger. We also assigned a label for residues with ps-ns motion as residues with hetNOE < 0.65 as used in ref. [27-29] to designate residues with ps-ns motion.

All final datasets and corresponding HYDRONMR fits are depicted in **Extended Data Fig. 1**, as well as the normalized RMSE over all $R_2/R_1$ values for residues that were not assigned to have low hetNOE or elevated $R_2/R_1$. RMSE alone would be larger for larger proteins, so we report RMSE that is normalized to the mean $R_2/R_1$ value for each protein. Distributions of norm. RMSE for the final RelaxDB evaluation set, datasets which were initially curated but then excluded, and datasets of proteins removed for containing phosphate buffer and phosphate-related biological function are depicted in **Extended Data Fig. 2a**. In the final dataset, 65/111 datasets had a norm. RMSE of less than 0.1, and 107/111 had a norm. RMSE of less than 0.2 (**Extended Data Fig. 2b**). A norm. RMSE of 0.1 can be interpreted as the experimental data for residues that did not have dynamics assigned were on average within 10% of the predicted HYDRONMR value for $R_2/R_1$.

*Filtering datasets resistant to fitting protocol.*

After curating 163 datasets initially, we realized in the process of finding a standardized way to fit all of them that some would not be well fit for a variety of reasons. For instance, datasets ISDHN3, MECP2, CBP6C, 5330, and MH35LIF were missing many $R_2$ and $R_1$ values, datasets 19356, CBP6N, and CBP6C all had hetNOEs < 0.65, and datasets 18773, 6758, and 5762 had relatively high Norm. RMSE values (see **Extended Data Fig. 2a** for comparison of all). We show the raw data for all the excluded datasets in **Extended Data Fig. 1**.

*Per-residue analysis.*

For the RelaxDB proteins and the analysis shown in **Fig. 1e**, we calculated conservation scores following the method described in ref. [68]. To determine buried/exposed residues, we calculated solvent accessible surface area (SASA) using the Shrake-Rupley method[69] implemented in MDtraj[70] and normalized values by the maximum value per secondary structure type (loop, helix, sheet) over the whole dataset, following the practice in ref. [71]. We assigned residues with a normalized SASA ≤ 0.2 as core and > 0.2 as surface residues, following ref. [71].

**The "mBMRB" dataset**

*Curating data*.

We downloaded a list of all entries consisting of polymer type polypeptide(L) from the BMRB[1] using https://bmrb.io/software/query/. We processed this to retrieve all entities within each entry which were type polypeptide(L). On May 31, 2023, this resulted in 12102 entities.

We removed entities with isotope labels corresponding to deuterated samples, as a deuterated sample that has not completely back-exchanged with $H_2O$ would lead to missing assignments. We also removed methyl-labeled samples or samples otherwise partially labeled based on keywords in the isotope metadata. Additionally, we removed any sequences shorter than 30 residues. After this filtering, we had 10123 entities.

We removed any entities with "X" in the sequence (242 removed). Some sequences deposited included multiple copies if the entity consisted of a multimer. We detected and edited these sequences to contain only one copy of the monomeric subunit to match the deposited data. We realized that 1834 sequences contained His-tags, many of which were unassigned. We therefore masked the N- or C-termini of sequences if 6 or more contiguous histidines were present, including

everything before the His tag if it was an N-terminal His tag, or everything after if it was a C-terminal His tag. We further filtered the remaining entities to include only entities with 10 or more assignments, to remove entries that included only a few backbone chemical shifts (removed 273).

For each residue, we assigned labels corresponding to whether the $^{15}N$ backbone chemical shift is missing or not based on if that residue had a chemical shift entry corresponding to an "Atom_ID" of N.

We further removed entities that had 15 or more contiguous missing assignment labels between any assigned residues. This was to filter entities where we noticed that entire domains were unassigned. We selected the cutoff of 15 contiguous residues because the most residues missing in a row in the RelaxDB dataset was 14. This removed 227 entities.

Of the remaining 9,381 entries, 5% did not have isotope information present. However, we elected to keep these, as we saw that they had on average fewer missing assignments and shorter sequences than the 95% of other remaining sequences which had 1H labeling (**Extended Data Figure 3a,b**). Roughly 60% of these are older entries (entry_ID < 10000).

**Extended Data Fig. 3d** contains average fraction missing across the mBMRB divided by pH, temperature, and sequence length. From this we observe that there are certain well-known factors that play a role in more missing assignments—for instance, pH greater than 8, or sequence length greater than 250—but these entries also represent small fractions of the data in total. Understanding how to more systematically account for such systematic outliers will be the topic of future research.

Structure models for each protein in the mBMRB dataset were generated using ESMFold v1[34].

*Comparing to structure models in PDB*

We used MMseqs[37] to align all the mBMRB sequences (9,381) against all protein chains in the PDB (curated based on PDB data in NMRbox, October 2024). We kept all hits with a bitscore greater than 50 and sequence identity greater than 90%. We kept one chain per protein sequence per model.

**RelaxDB-CPMG curation**

*Complete dispersion data available*

AdK: $R_{2,eff}$ vs $\nu_{CPMG}$ for apo AdK from *Aquifex aeolicus* at 10˚C collected at 800 MHz, presented in ref. [58], as well as a corresponding peak list, was found in the Kern lab data archive, with thanks to K. Henzler-Wildman.

BlaC: $R_{2,eff}$ vs $\nu_{CPMG}$ of apo BlaC collected at 600 MHz and 850 MHz, presented in ref. [33], was kindly provided by M. Ubbink. A peak list to identify missing assignments was taken from BMRB entry 27888 (same as in RelaxDB).

BLVRB: $R_{2,eff}$ vs $\nu_{CPMG}$ of apo BLVRB, presented in ref. [59], was kindly provided by E. Eisenmesser. A peak list to identify missing assignments was taken from BMRB entry 27462.

CypA: $R_{2,eff}$ vs $\nu_{CPMG}$ of apo CypA, collected at 10˚C and 25˚C at 600 MHz in ref. [57] , as well as peak lists to identify missing assignments at both temperatures, were found in the Kern lab data archive. Because different processes are more prominent at 10˚C or 25˚C, the final labels created represent whether a residue experienced exchange at either temperature.

VHR: $R_{2,eff}$ vs $v_{CPMG}$ was taken from 800 MHz data provided in Table 7.10.3 of ref. [72]. Missing assignments were taken from assignments in Table 7.10. Note that these differ from assignments by the authors in BMRB 27950; the assignments in the dissertation table are more complete.

K-Ras: $R_{2,eff}$ vs $v_{CPMG}$ of K-Ras with GTP bound, presented in ref. [60], had been publicly deposited at https://doi.org/10.5061/dryad.j6q573nm0.

*Complete dispersion data unavailable*

CheY: Residues with missing assignments were taken as the black residues in Fig. 2 from ref. [73]. Residues were labeled as having exchange when CPMG could individually be fit (all residues in Table 2 in ref. [73]). This is CheY in the presence of 1 mM EDTA to remove any $Mg^{2+}$, which alters the dynamics.

HtrA2 PDZ domain: Values for $R_{ex}$ were taken from source data corresponding to Fig. 2a in ref. [74]. Assignments were taken from BMRB:51320.

RNase: Residues were labeled as having exchange when $R_{ex}$ >3 Hz at 20 ˚C in either 500MHz or 600 MHz data. This recreates Fig. 4B in ref. [75]. The $R_{2,eff}$ vs $v_{CPMG}$ data for these residues was kindly provided by P. Loria, but the dispersion data for all residues was not available.

Arginine kinase: Labels for residues with exchange were taken from Table 1 in ref. [76], which summarized all residues for which CPMG data could be fit individually. Table S1 in ref. [76] indicated if residues were missing due to exchange-broadening or overlap, and we used those labels.

**Identifying unsuppressed $R_{ex}$ in CPMG experiments**

We first compared model predictions to labels for exchange where we assigned labels using a conventional assignment from the literature: any residue where $R_{ex}$ is significantly apparent, calculated as $R_{2,eff}$ (first field strength) - $R_{2,eff}$ (last field strength) is greater than background noise.

However, we noticed that in some datasets there were residues with noticeably high $R_{2,eff}$'s across all field strengths but $R_{2,eff}$(first) - $R_{2,eff}$(last) was not substantial enough to be greater than 2-3 Hz. These correspond to residues that have exchange at the microsecond timescale: at the fast end of the micro-millisecond regime that CPMG can detect.

However, assigning these residues with confidence requires determining what residues have elevated $R_{2,eff}$ when there is no suppression, the equivalent of measuring $R_{ex}$ as is conventionally done. Given our success using HYDRONMR to quantitatively fit $R_2$/ $R_1$ ratios in RelaxDB, we used HYDRONMR to calculate $R_2$values and fit a scalar to find the closest fit between $R_{2,inf}$ and $R_{2,rigid}$.

The $R_{2,eff}$ data for all residues, corresponding $R_{2,rigid}$ values, and labels are in **Extended Data Fig. 8**.

**Data splitting**

We clustered sequences using MMseqs[37] `easy-cluster` with arguments `--cluster-mode 1 --min-seq-id 0.8 -k 5`. We used the resulting clusters to remove all sequences similar to those in RelaxDB or RelaxDB-CPMG. From the remaining sequences, we randomly separated out 500 clusters for the validation set and 100 clusters for the test set.

For training with a 50% or 30% sequence identity cutoff, we used –min-seq-id 0.5 [or 0.3] to cluster but kept the same proteins in the validation and test set. To create training splits with structure homologues removed at TM-score = 0.5 or 0.7, we calculated all pairwise TM-scores using TMalign[77] between the mBMRB-Train set and mBMRB-Val, mBMRB-Test, RelaxDB, and RelaxDB-CPMG sets. We used TM-score normalized by the average length of the query and target protein, as recommended in ref. [77].

The model sweep depicted in **Fig. 3c** was performed with a different dataset split that considered all 163 sequences initially curated for RelaxDB and used a split of 80% sequence identity, 1.0 TM-score cutoff. The splits used for the model sweep in **Fig. 3c** can be found in the `oct2024` data repository. Subsequent model training was reperformed using a data split considering only the later filtered 133 sequences in RelaxDB and used the most stringent training data split (30% sequence identity, 0.5 TM-score cutoff). These subsequent models are depicted in **Figs. 3d** and onwards.

**Baseline training**

Our baseline training aimed to test simpler representations of protein sequences and/or structure. We use the mBMRB data to train these methods. The results of these baseline models on the mBMRB-Validation, mBMRB-Test, and RelaxDB datasets are depicted in **Extended Data Fig. 6a**. None outperform the models based on pre-trained models.

<u>Naïve baselines:</u> The simple, naïve baseline uses amino acid frequencies in the mBMRB training data as weighting of a random classifier.

- The simplest is the AA baseline, which computes the frequency of each amino acid being missing from the mBMRB training data. We then predict for our validation and test set whether a residue is missing at random, weighted by that relative frequency.
- We evaluate similar random classifier using frequency of secondary structure (loop, sheet, helix) as calculated by DSSP, and a combination termed the "AA & DSSP" baseline using the frequency of each amino acid type given its secondary structures (loop, sheet, helix).

Simple classifiers: We use a similar Transformer classifier as that described for the final Dyna-1 architecture trained on simpler representations of the input mBMRB training data. Each representation is described.

- "OHE AA": we input each protein as a one-hot encoded sequence, where each residue is represented by a unique combination of 0s and 1s, into the transformer classifier.
- "SASA" we evaluate solvent-accessible surface area using the Shrake-Rupley method[69] implemented in MDTraj[70]. These values are encoded by a simple one-dimensional MLP of hidden size 128 and then input into the transformer classifier.
- The "SASA & DSSP", "SASA & AA", and "SASA & AA & DSSP" concatenate the outputs of the MLP with a one-hot encoding of the secondary structure, sequence, and secondary structure and sequence, respectively. This concatenated embedding is subsequently used as input into a transformer and trained accordingly.

Other representations:

- Normal mode analysis (NMA) is a computational technique that has long been of interest for modelling dynamic processes in proteins[78]. To evaluate NMA, we use ProDy[79] to compute the top 20 normal modes of the models generated using ESMFold (for the mBMRB data) or AF2 (for the RelaxDB data) using the Anisotropic Network Model (ANM) and Gaussian Network Model (GNM) and build the Hessian matrix and Kirchhoff matrix,

respectively. We then calculate the mean square fluctuation and evaluate the AUROC and AUPRC using these values. No model is trained.

- To use sequence conservation as a baseline, we computed sequence conservation scores and coverage index per protein as described in ref. [68] and evaluated the AUROC and AUPRC using these values as logits. No model is trained.

**Deep learning training**

We trained Dyna-1 on the mBMRB-Train dataset, sampling using a sequence-cluster-based approach. One protein per sequence cluster is chosen randomly during each epoch. We used the validation set to compute the validation loss and validation metrics. The final model weights for each training run for all models are chosen using max AUROC on the validation set. We excluded unassigned termini and prolines (which have no backbone amide and therefore are always missing) from model evaluations.

We used the ESM2-t30-150M model for ESM-2 embeddings, which has 31 layers and 150M parameters. Initial attempts using ESM2-t6-8M and ESM2-t12-35M indicated that smaller models did not perform as well, and ESM2-t33-650M performed marginally similar. We used the outputs of AF2 model 1 to construct the AF-pair representation. The weights from ESM and AF2 were fixed (frozen) during training.

Models were trained with the AdamW optimizer[80] with a learning rate of 1e-6. The AF2-pair model was trained with batch size of 4 with an accumulation step every 4 epochs, while the ESM-2 and ESM-3 based models were trained with a batch size of 16 and accumulation each epoch. Training is completed within 500 epochs, and we use early stopping when validation loss begins to overfit. We

choose our best performing model by comparing normalized AUPRC on the validation set, as we saw this had slightly more discrimination than AUROC. Normalized AUPRC is calculated by subtracting the baseline AUPRC from the AUPRC for each protein.

The transformer architecture used for the ESM-2 embeddings has 8 heads and 12 layers. The transformer for the ESM-3 embeddings has 6 heads and 12 layers, and the transformer for AF2-pair representation has 4 heads and 4 layers. We note that increase in number of heads or number of layers had negligible improvements for ESM-based models, and we were limited by GPU memory for AF2-pair. The hidden dimension for the ESM-based models is equivalent to the hidden size of the embedding, while for the AF2-pair model is 128. We apply a 0.1 dropout rate. All models were trained on a single NVIDIA A40 GPU, which takes between 12 hours and 2 days to complete training.

**Adjusting logits to account for class imbalance.**

When converting from logits to probabilities, we adjusted the raw logit outputs predicted by the model to account for the significant class imbalance in the problem following the practice in ref. [81] (cf. **Extended Data Fig. 10**). This practice seeks probabilities that minimizes the "balanced error", the prediction error averaged over all classes. In brief, for data *x* labelled with classes *y*, the native posterior probability *P(y | x)* is proportional to the prior probability of class y, $\pi(y)$:

$$P(y|x) \propto \pi(y)P(x|y).$$

In seeking to minimize the balanced error, the balanced class-probability function instead implicitly becomes

$$P^{bal}(y|x) \propto P(x|y),$$

where *L* is the number of classes (L=2 in our binary classification context).

Menon et al. describe how $P^{bal}(y|x)$ is a Bayes-optimal scorer for minimizing the balanced error. For a Bayes-optimal scorer, the accuracy of any given class is not proportional to its degree of representation in the training data, which is desirable in our context.

Consider our model to be a scoring function $\hat{s}$, with learned class probabilities $\hat{P}(y|x) \propto e^{\hat{s}(x)}$. Via the definition of $P^{bal}$, we can approximate the Bayes-optimal scorer using our given scorer by using an adjusted scorer $s'$:

$$P^{bal}(y|x) \approx e^{s'(x)} = \frac{\hat{P}(y|x)}{\pi(y)} = \frac{e^{\hat{s}(x)}}{\pi(y)}.$$

Accordingly, the adjusted logits may be written as $s'(x) = s'(x) - \log \pi(y)$.

In the context of our binary classification problem, let $\pi^m$ be the fraction missing in the training data set and let *l(x)* be the raw logit output given input *x*. Dyna-1 is structured to just predict one logit, the logit of the "missing" class, hence the logit predicted for the "not missing" class is always zero. We calculate our probability P(missing) as

$$P(missing|x) = \frac{p(missing|x)}{p(present|x)+p(missing|x)} = \frac{e^{l(x)-\log \pi^m}}{e^{-\log(1-\pi^m)}+e^{l(x)-\log \pi^m}} = \frac{e^{l(x)}}{\frac{\pi^m}{1-\pi^m}+e^{l(x)}}.$$

This can be rewritten in terms of the sigmoid function $\sigma(x) = (1 + e^{-x})^{-1}$, which is typically used to convert logits to probabilities:

$$P(missing|x) = \sigma(x - \log \frac{\pi_m}{1-\pi_m}).$$

**Uncertainty estimation.**

We estimated uncertainty for comparing models in the RelaxDB dataset (**Fig. 4a, Extended Data Fig. 5d,e**) using a 95% confidence interval calculated in Seaborn v0.11.2 [82].

We estimated uncertainty in binned residue analyses (**Fig. 4e, 4g**) using bootstrapping. We resampled residue subsets with replacement to obtain 100 bootstrap iterations that contained at least one of each class (required for calculating AUROC).

All depicted p-values (**Fig. 1e, Fig. 2h**) are for a 2-sided hypothesis test whose null hypothesis is that two sets of data are uncorrelated, performed using statannotations v0.2.3 (https://pypi.org/project/statannotations/).

We curated uncertainties for the RelaxDB datasets in the following way: if authors reported $R_1$/$R_2$/hetNOE uncertainties, we used those. In the BMRB, we found a few with a deposited uncertainty of zero. We set these to be the smallest nonzero uncertainty value given for that data type in the dataset. If no uncertainties were given, we set the uncertainty for $R_1$ as 10% the measured value, and 5% for $R_2$ and hetNOEs. If the data was plot digitized and had error bars, we digitized the error bars as uncertainties.

**Preparation of recombinant proteins**

The hypothetical protein yjbJ (YjbJ_ecoli) (NCBI: WP_058905752.1) and wild-type *Bryum coronatum (*BcChi-A) constructs were ordered from GenScript (Supplementary Table 1). The codon-optimized YjbJ_ecoli sequence was cloned into a pET-15b expression vector with an N-terminal 6xHis tag followed by a TEV protease cleavage site. The wild-type BcChi-A (without any affinity tag) was

cloned into a pET-22b vector between the NdeI (5') and BamHI (3') restriction sites, replacing the pelB leader sequence of the vector.

The individual plasmids encoding each protein were transformed into *E. coli BL21(DE3)* cells (New England Biolabs). Uniformly $^{13}C$, $^{15}N$-labeled proteins were expressed in *E. coli* grown in 1L M9 minimal medium supplemented with 2.0 g/L U-$^{13}C$ as the main carbon source and 1.0 g/L of $^{15}NH_4Cl$ (Cambridge Isotope Laboratories). Cells were grown to an $OD_{600nm}$ of 0.6 before induction with 1 M isopropyl thiogalactoside (IPTG). Then grown overnight at 25°C for the BcChi-A construct or 15 °C for the YjbJ construct for ~18 h.

Both constructs were purified using a similar method as previously described [63,83]. Briefly, YjbJ cell pellets were resuspended in 20 mM Tris, pH 8, 500mM NaCl, 30mM Imidazole (Buffer A), DNAse 1 (Sigma-Aldrich), and lysozyme (Sigma-Aldrich). Lysate was sonicated on ice for 10 min (20 sec on, 30 sec off, output power of 40 W), followed by clarification via centrifugation at 19000 rpm for 45 min at 4°C. A 5 ml His-trap $Ni^{2+}$ affinity column (Cytiva) was pre-equilibrated in buffer A. The column was washed with 10 column volumes of buffer A after loading the supernatant, followed by elution of the his-tagged YjbJ protein with buffer containing 500 mM imidazole. The His tag was cleaved off with TEV protease and removed by loading the cleavage reaction back onto the His-trap $Ni^{2+}$ affinity column. YjbJ with no tag was collected in the flow-through with 5 column volumes of buffer A containing no Imidazole. The $^{13}C$,$^{15}N$ labeled YjbJ was further purified on a Superdex 75 26/600 pg column in 25 mM $Na_2PO4$, pH 6.5. BcChi-A construct was purified by resuspending the pellet in 20mM Tris-HCl, pH 7.5, and sonicated on ice. The lysate was spun down at 18500 rpm for 45 min, followed by dialysis of the supernatant against 10 mM sodium acetate, pH 4 and additional clarification by spinning down at 18500 rpm for 20 min. The supernatant was then dialyzed in 10

mM sodium acetate, pH 5, before loading on a 5 ml HiTrap Q HP column (Cytiva) pre-equilibrated in the same dialysis buffer. The protein was eluted with a linear gradient of NaCl from 0 to 150 mM NaCl in the same buffer. The $^{13}C$, $^{15}N$ labeled BcChi-A was further purified on a Superdex 75 26/600 pg column in 50 mM Acetate, pH 5.

**NMR spectroscopy**

$^{15}N$- Carr-Purcell-Meiboom-Gill (CPMG) data were carried out on a 900 MHz Bruker spectrometer equipped with a cryogenic triple resonance probe. Measurements for BcChi-A were performed at 300 K on a 1 mM $^{13}C$, $^{15}N$ labelled sample in 50 mM Acetate, pH 5.0, while YjbJ experiments were conducted at 298 K on a 1 mM sample of $^{13}C$, $^{15}N$ labelled in 25 mM $Na_2PO_4$, pH 6.5. $^{15}N$- CPMG[84,85] data was acquired as pseudo 3Ds with a fixed total CPMG relaxation period ($T_{CPMG}$) of 40 ms, and 16 $n_{cyc}$ values ranging from 1 to 40. A reference spectrum was collected with no CPMG relaxation delay ($n_{cyc}$=0), and duplicate measurements at $n_{cyc}$ of 4 and 20 were collected for error estimation. The effective transverse relaxation rate $R_{2,eff}$ was calculated from these datasets, and errors were taken as the larger of noise-based versus duplicate-based error estimation. Each dataset was collected with 32 scans per FID for the BcChi-A or 8 scans per FID for the YjbJ protein with 2 sec repetitive delay between scans. All CPMG traces for both proteins are in Supplemental Data 1.

Backbone assignments for the BcChi-A and YjbJ were obtained from BMRB accession codes 11441 and 5105, respectively.

**NMR data analysis and fitting**

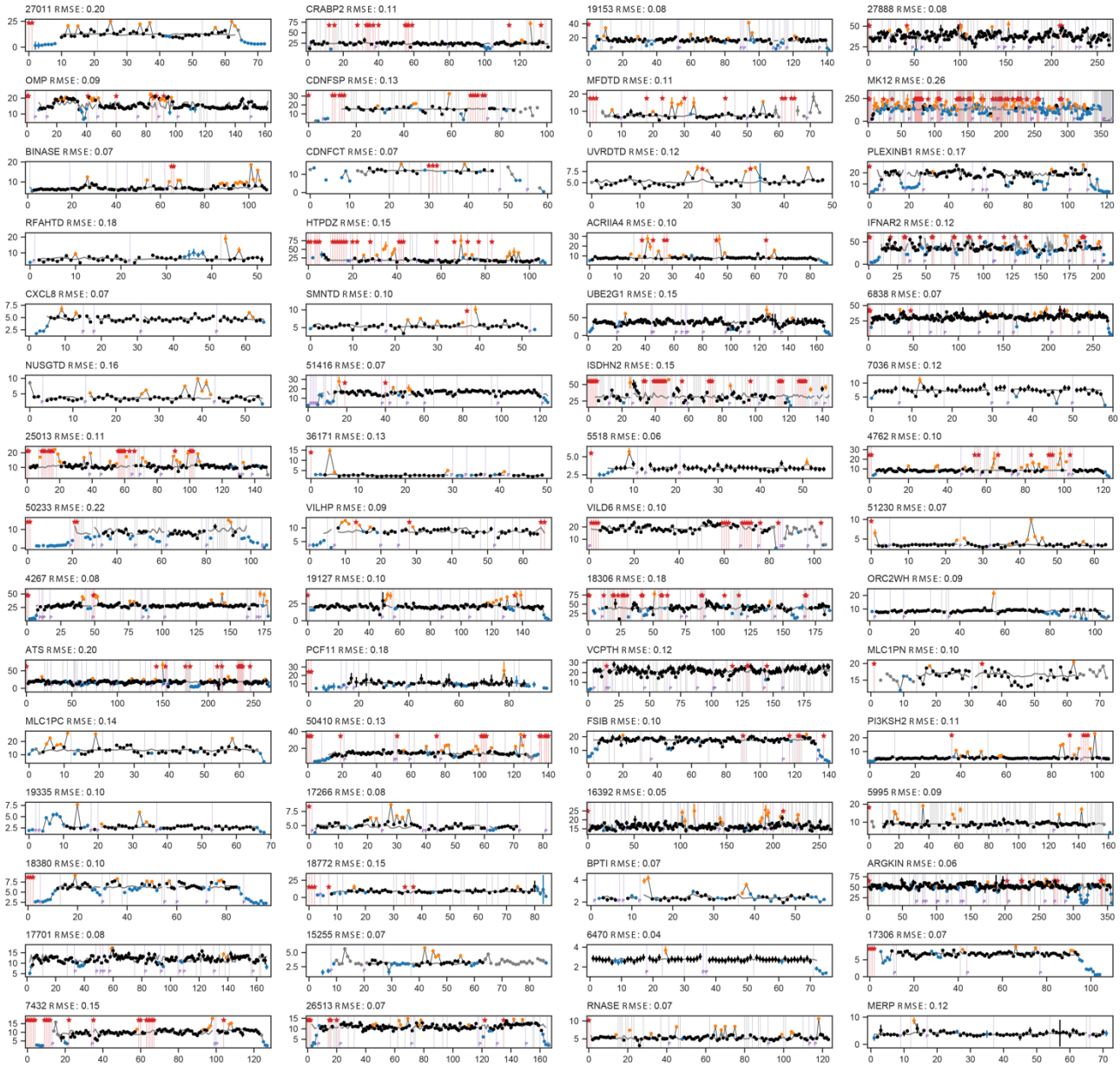
Peak heights for individual residues were collected from each 2D $^{15}$N-HSQC in the pseudo 3D $^{15}$N-CPMG by lineshape fitting using PINT[86] software. The $R_{2,eff}$ at a given CPMG field strength ($v_{cpmg}$) was calculated using:

$$R_{2,eff} = \frac{1}{T_{CPMG}} \ln \frac{V(0)}{V(v_{cpmg})}$$

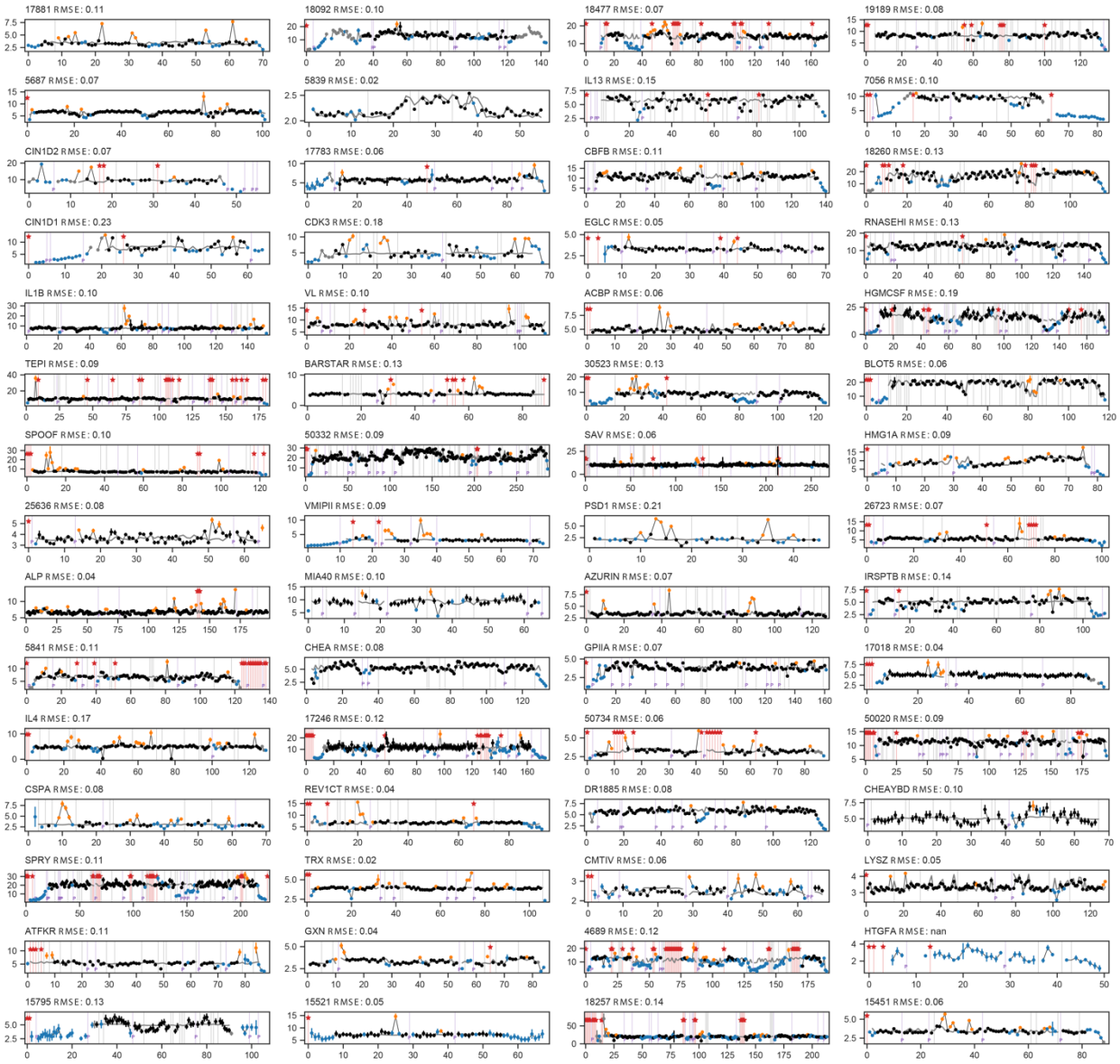
where $T_{cpmg}$ is the constant CPMG relaxation period, V(0) is the peak volume and $V(v_{cpmg})$ is the peak volume at each CPMG field strength.

Relaxation dispersion curves, expressed as $R_{2,\,eff}$ versus CPMG field strength, were fit per residue using ChemEx software (https://github.com/gbouvignies/chemex) [87].

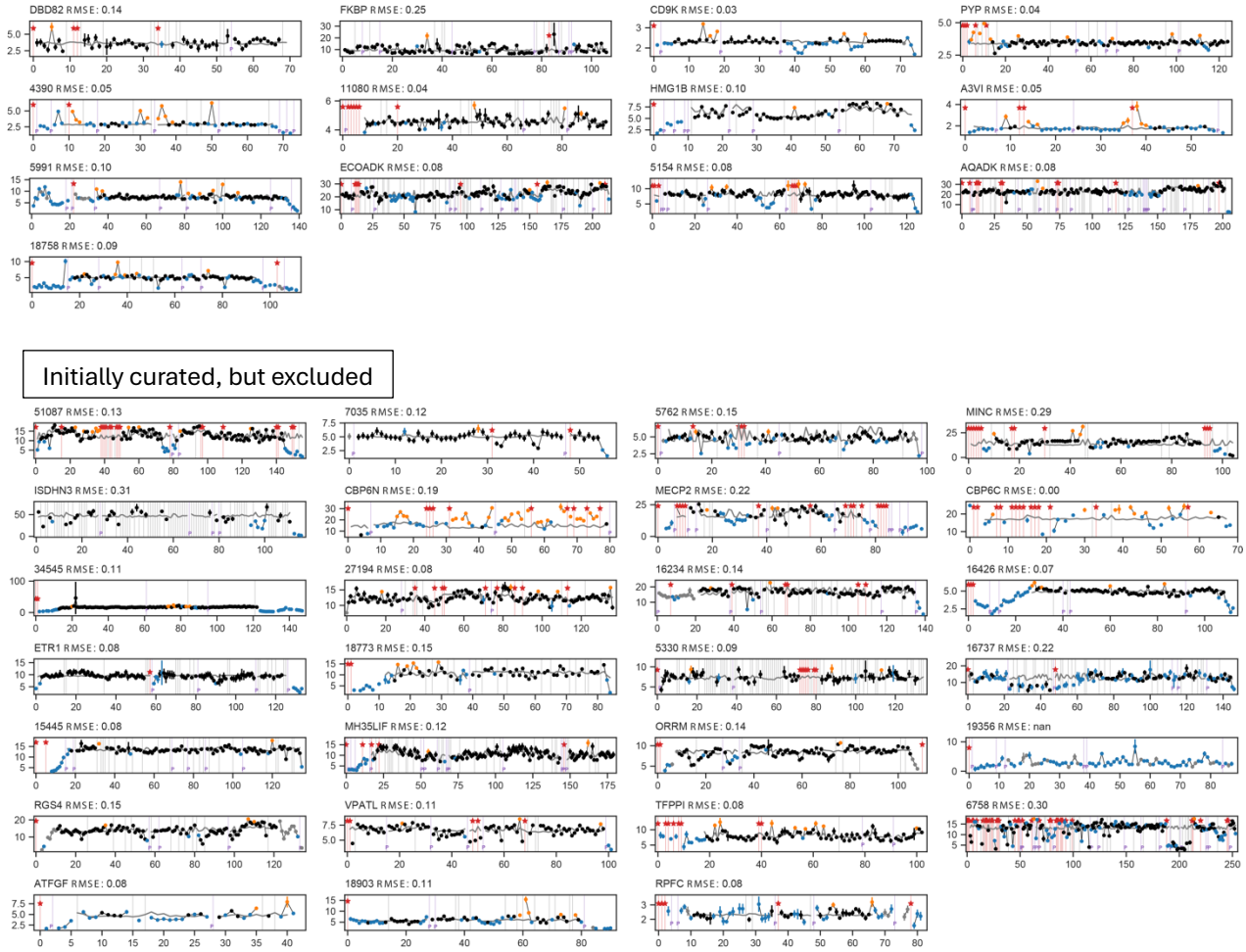
## Extended Data



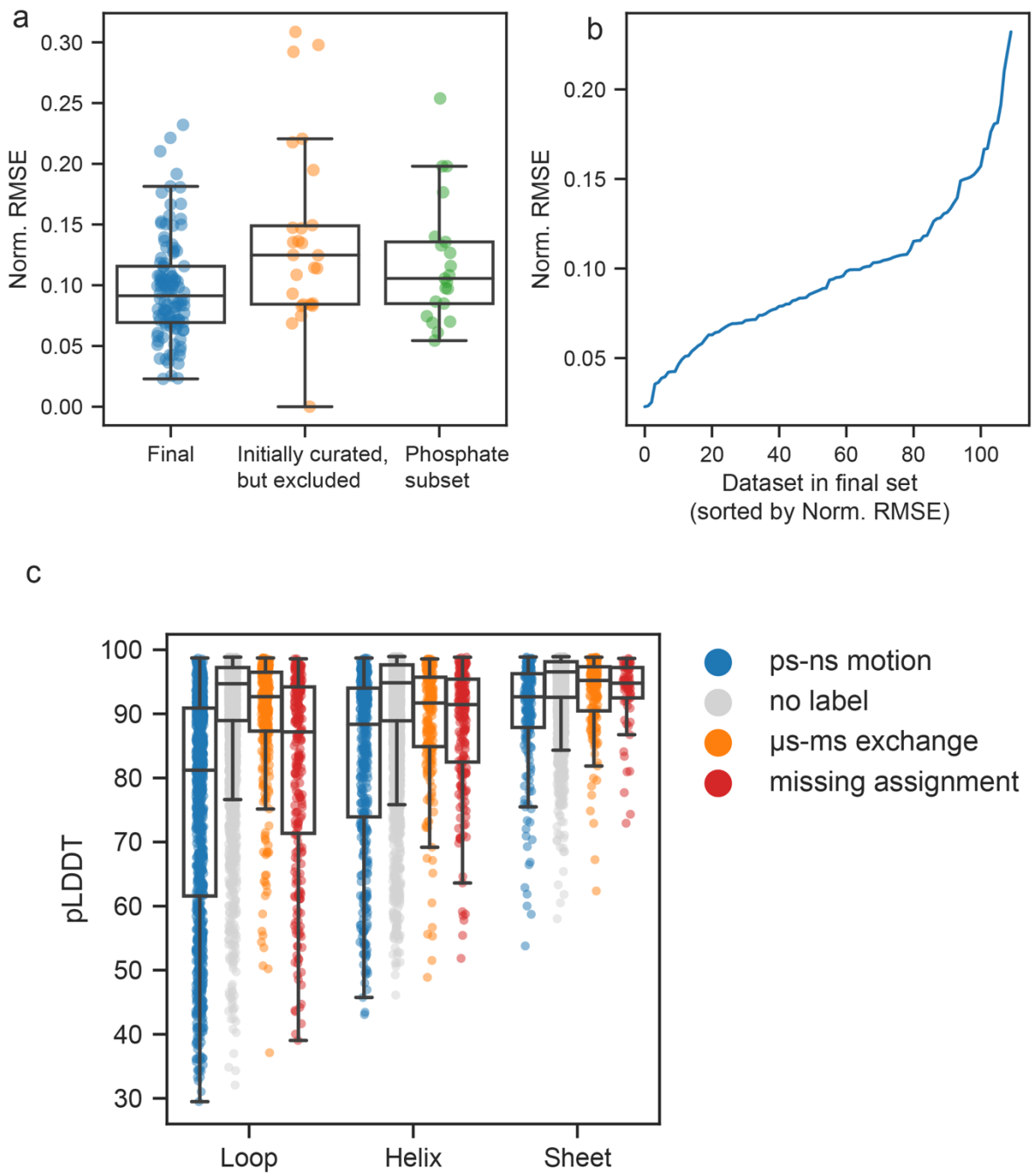

**Extended Data Figure 1**: (page 1/3) $^{15}N$ $R_2/R_1$ data and labels for all 133 proteins in RelaxDB and the 30 which were initially curated but excluded from final evaluation set for various reasons (see Methods). Calculated $R_2/R_1$ from HYDRONMR[25,26] for rigid tumbling is depicted by grey line. Colors used are the same as in Fig. 1. Title for each is RelaxDB entry ID followed by normalized RMSD.



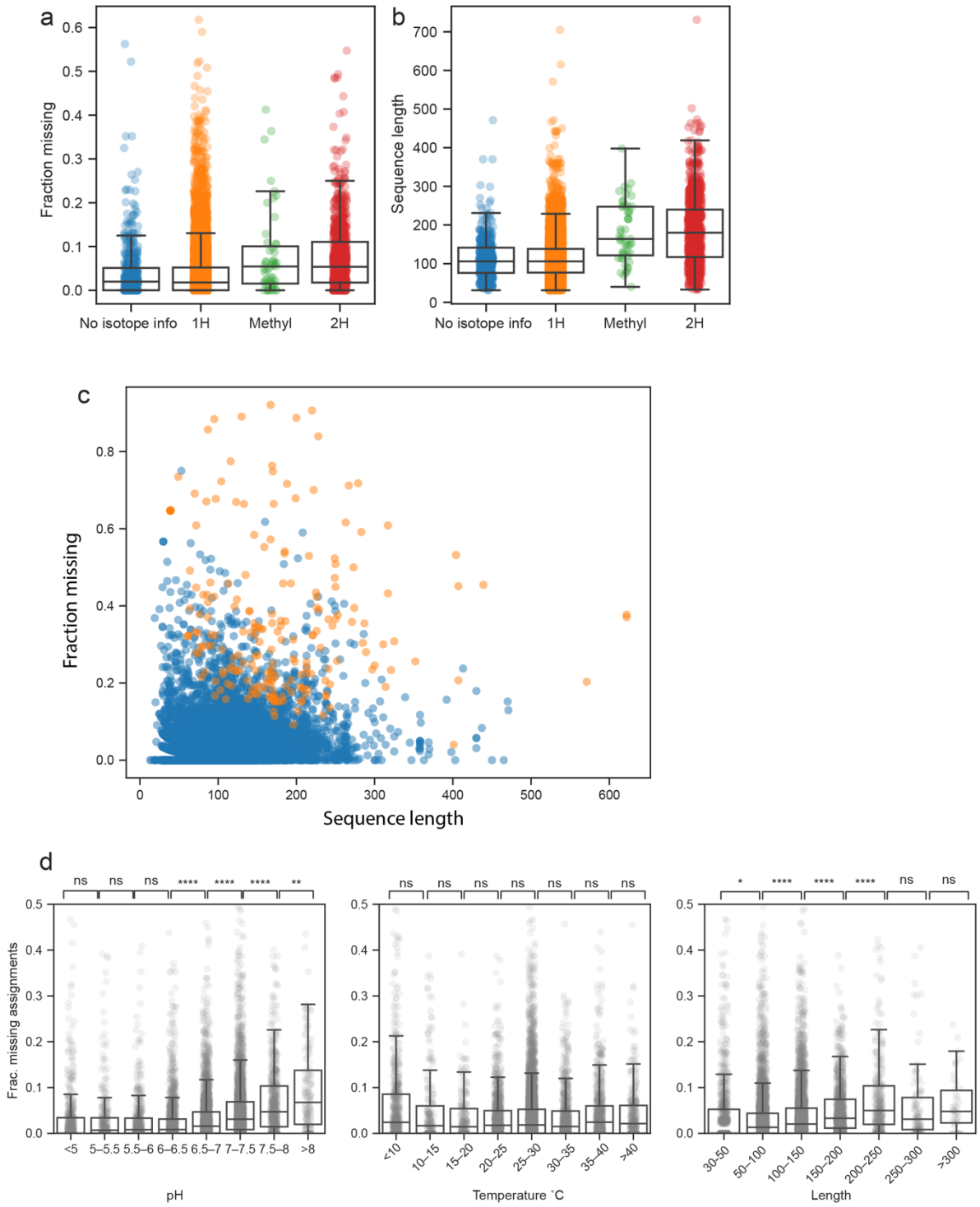

**Extended Data Figure 1**: (page 2/3)




**Extended Data Figure 1**: (page 3/3)

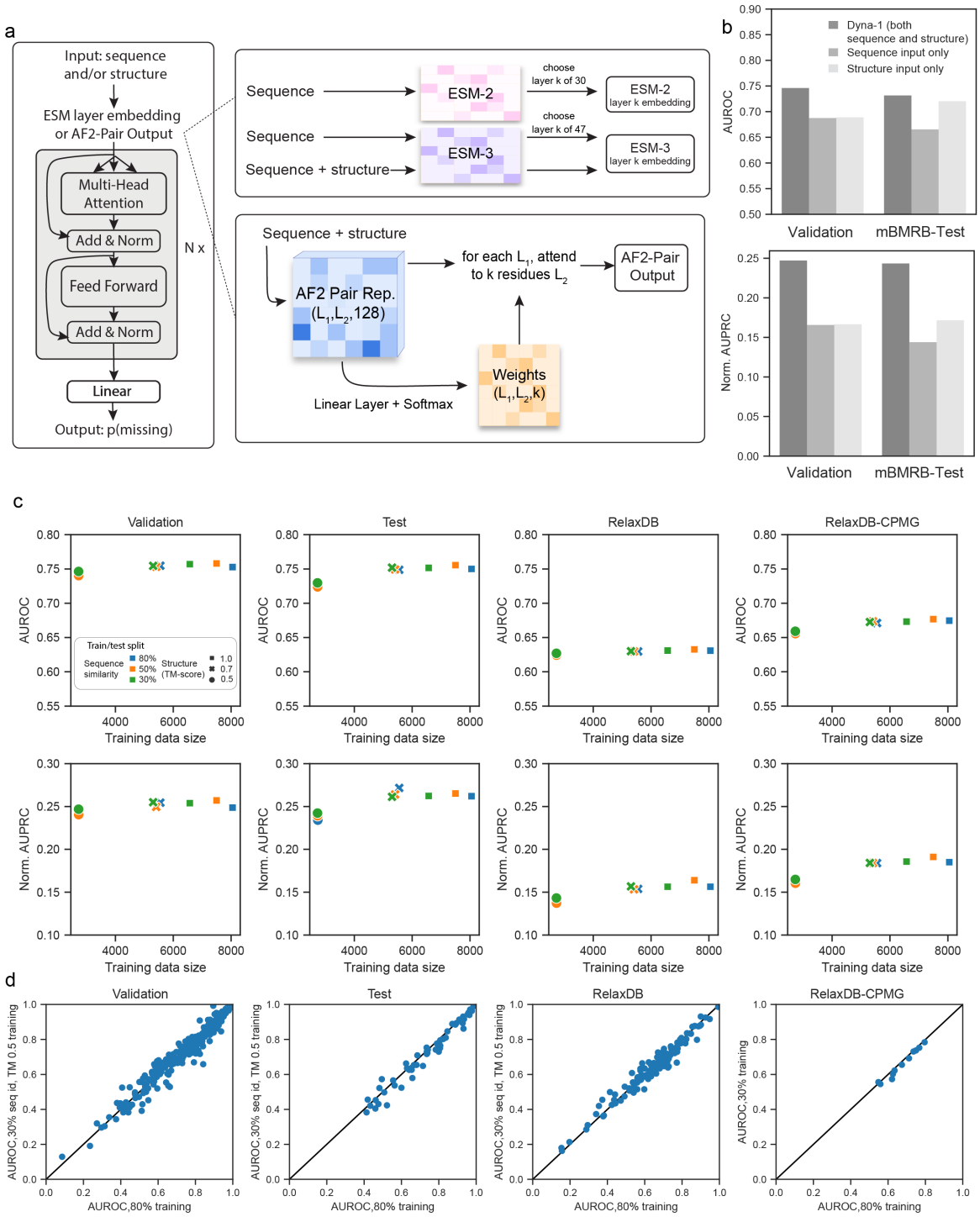



**Extended Data Figure 2**: (a) RMSE normalized to median R2/R1 per protein for the final RelaxDB dataset, protein datasets that were initially curated but excluded, and proteins with biological function involving phosphate binding, but measured in phosphate buffer. (b) Norm. RMSE distribution of final RelaxDB dataset used for evaluation. (c) pLDDT (predicted local distance difference test) from AF2 by secondary structure and dynamics type. Low AF2 pLDDT is most indicative of ps-ns motion in RelaxDB.



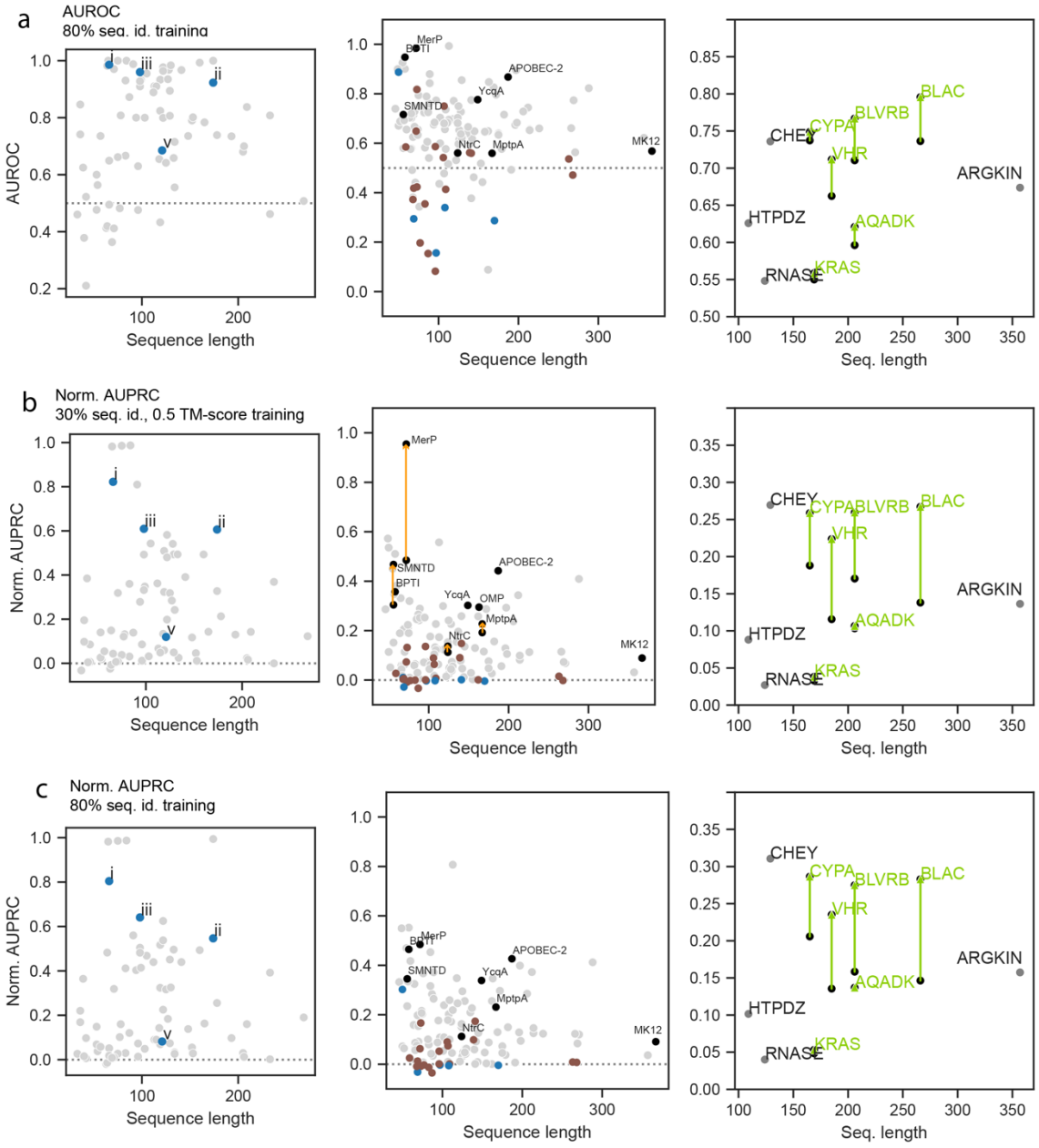

**Extended Data Figure 3**: Statistics from curating the "missing assignment BMRB" (mBMRB). We removed proteins with deuteration or methyl or specific labelling, based on metadata. Some proteins had no isotope information, yet we elected to keep them, as they had fewer missing

assignments compared to proteins with $^{1}H$ isotope labels and were overall shorter than other isotope categories. (a) Fraction missing split by isotope category. (b) Sequence length by isotope category. (c) We removed entries with 15 or more consecutive missing assignments (orange). (d) Missing assignments as a function of pH, Temperature, and sequence length. In (a,b,d): Box plots depict median and 25/75% interquartile range, whiskers = 1.5 times the interquartile range.

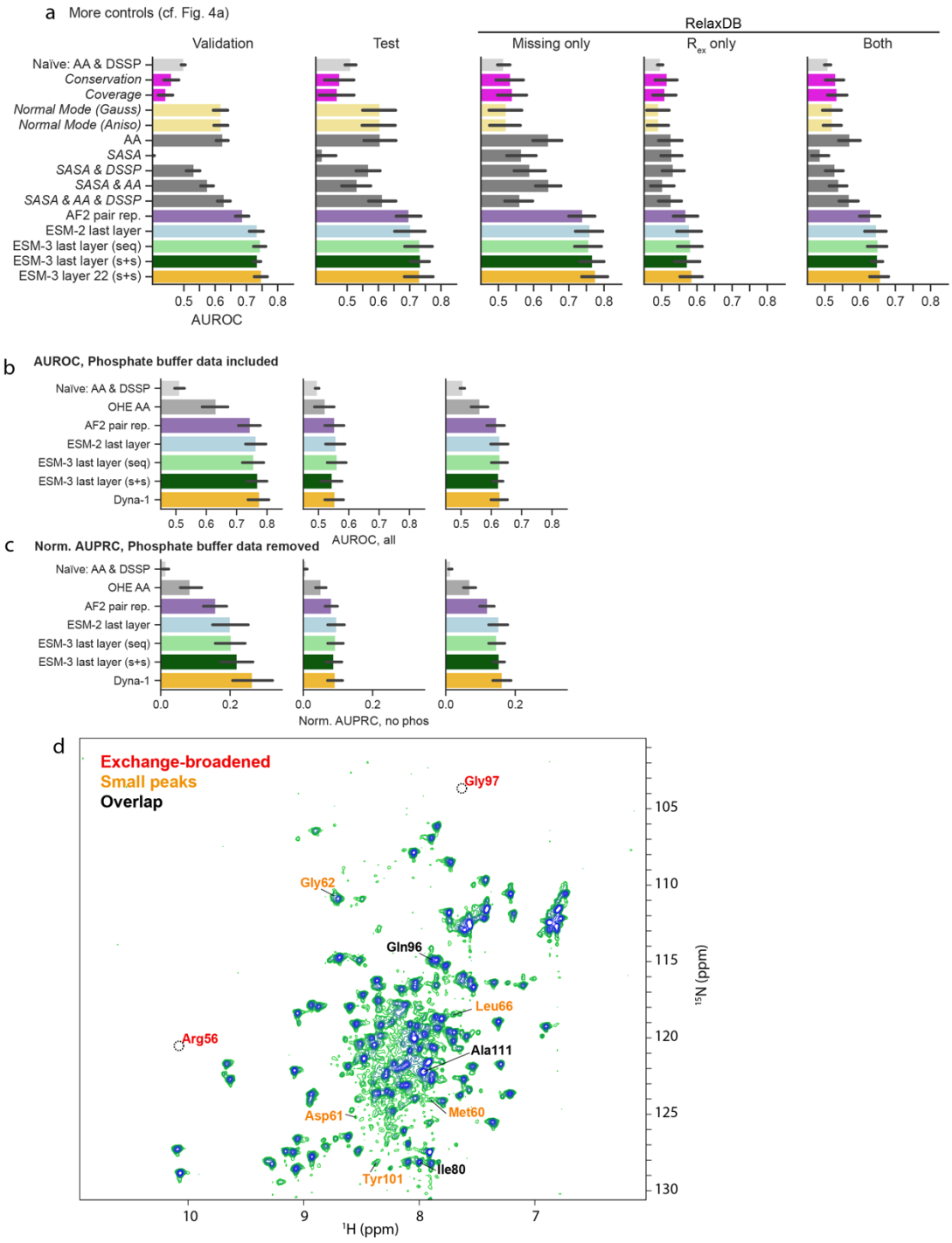



**Extended Data Figure 4**: Details of training Dyna-1. (a) Architectures of tested ESM and AF2 inputs, outlining the model architecture. At training time, either a sequence-structure pair is passed through AF2 or ESM-3, or a sequence is passed through ESM-2 or ESM-3. The corresponding embedding from the $k^{th}$ layer from ESM or AF2-pair output is fixed ("frozen") and used as input into a one-dimensional Transformer architecture. Output is one value per residue, the logit for that

residue being missing. (b) Comparing Dyna-1 performance given both sequence and structure (blue) vs. sequence only (orange) or structure only (green). Performance is best when given both sequence and structure. (c) Dyna-1 performance given different training split cutoffs based on sequence identity and structure similarity (cf. Fig. 3d) across all datasets. (d) AUROC per protein for either training with the 80% sequence identity and 1.0 TM-score split or 30% sequence identity and 0.5 TM-score split. Performance on individual proteins in the validation, test, RelaxDB, and RelaxDB-CPMG does not significantly change between more stringent sequence similarity cutoffs.

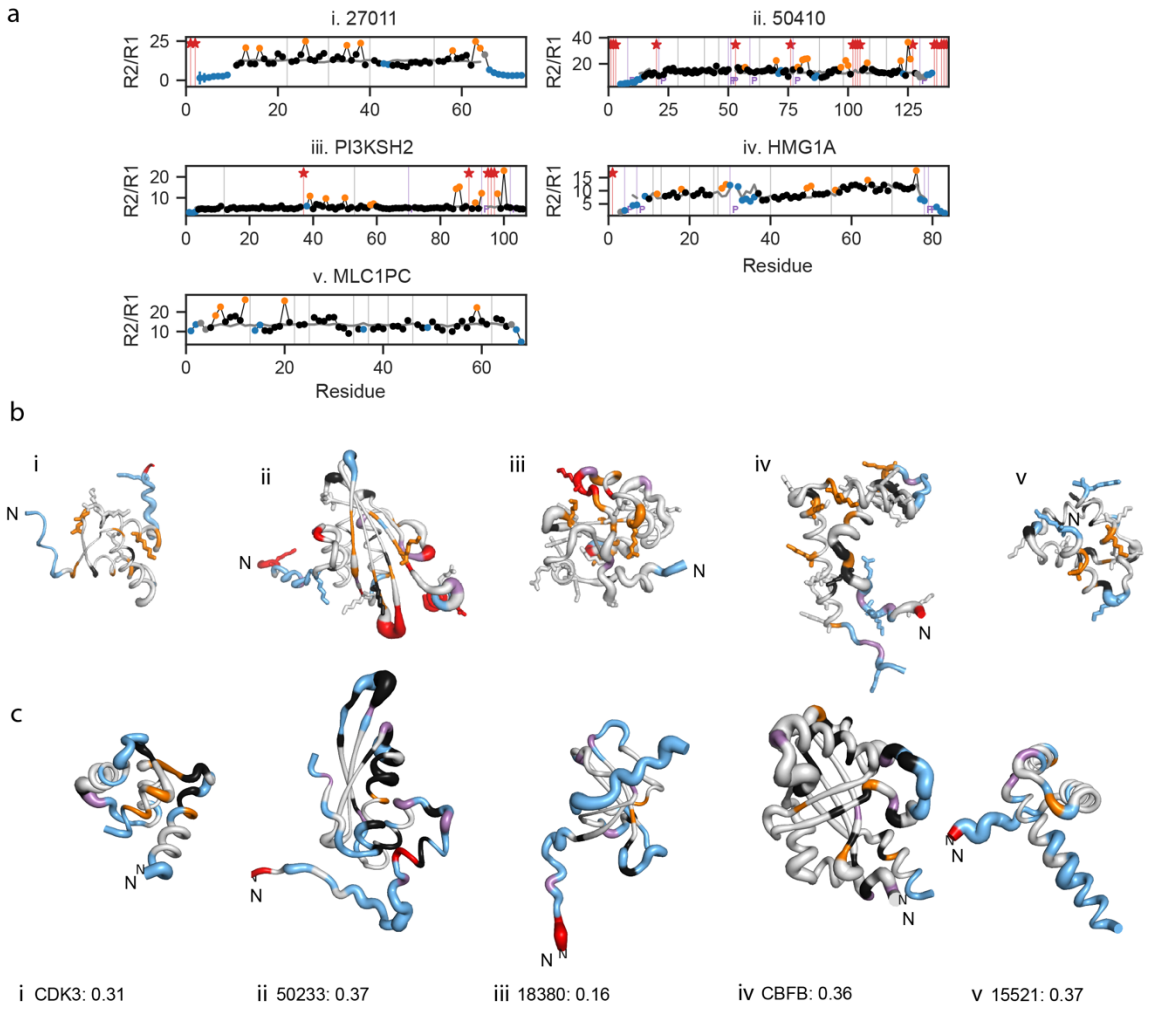



**Extended Data Figure 5.** Dyna-1 trained with 80% sequence identity cutoff is comparable to results presented in main text for 30% sequence identity and 0.5 TM-score cutoff. (a) AUROC for mBMRB-Test (cf. Fig. 3e), RelaxDB (cf. Fig. 4b), RelaxDB-CPMG (cf. Fig. 5d). Coloring for each follows respective main-text figures. (b) Normalized AUPRC per protein for Dyna-1, trained with 30% sequence identity and 0.5 TM-score cutoff. (c) Normalized AUPRC per protein for Dyna-1, trained with 80% sequence identity cutoff.

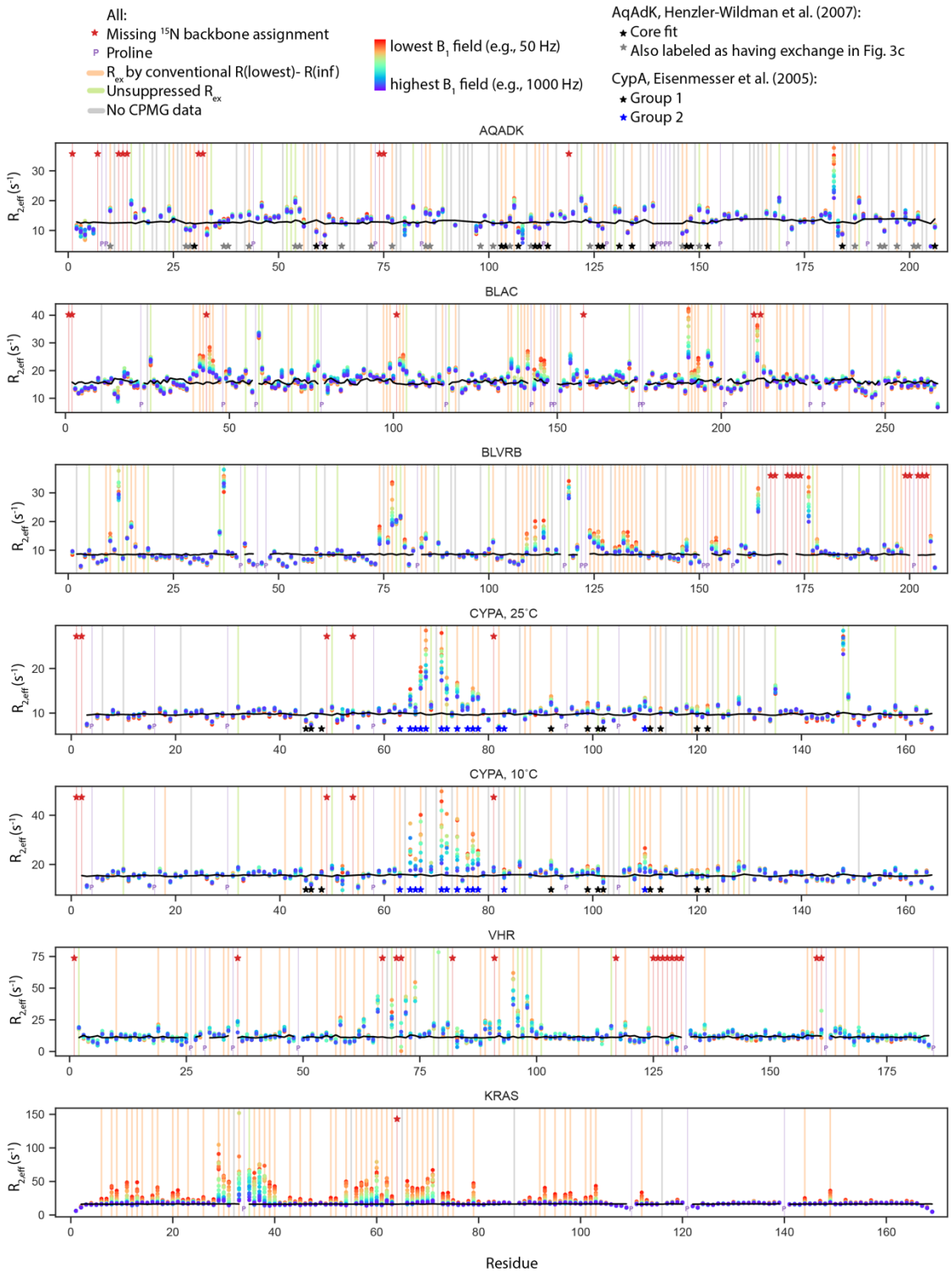



**Extended Data Figure 6.** (a) Validation, test, and RelaxDB datasets comparing further benchmarks at training cutoff of 30% sequence identity, TMscore 0.5. Benchmark models in italics are not in Fig. 4a. (b) RelaxDB evaluated by AUROC across different models tested, using all 133 proteins including those with exchange from phosphate binding (cf. **Fig. 4a**). (c) RelaxDB evaluated across different models evaluated by normalized AUPRC, with phosphate binding proteins removed. In (a,b,c), error bars represent 95% confidence interval evaluated over proteins in RelaxDB. (d) $^{1}$H-$^{15}$N HSQC of NtrC at 25˚C, annotated with residues for which assignments exist in BMRB entry 4527

(non-phosphorylated NtrC), but relaxation data is missing from BMRB entry 4762. The Arg56 and Gly97 assignments in entry 4762 (shown in red here) are presumed to be erroneously copied from entry 4528, which are assignments corresponding to phosphorylated NtrC. All residues labeled orange are very weak peaks indicative of $R_{ex}$. For these, we concluded exchange broadening was the reason for no reported relaxation data. Black residues are overlapped peaks.

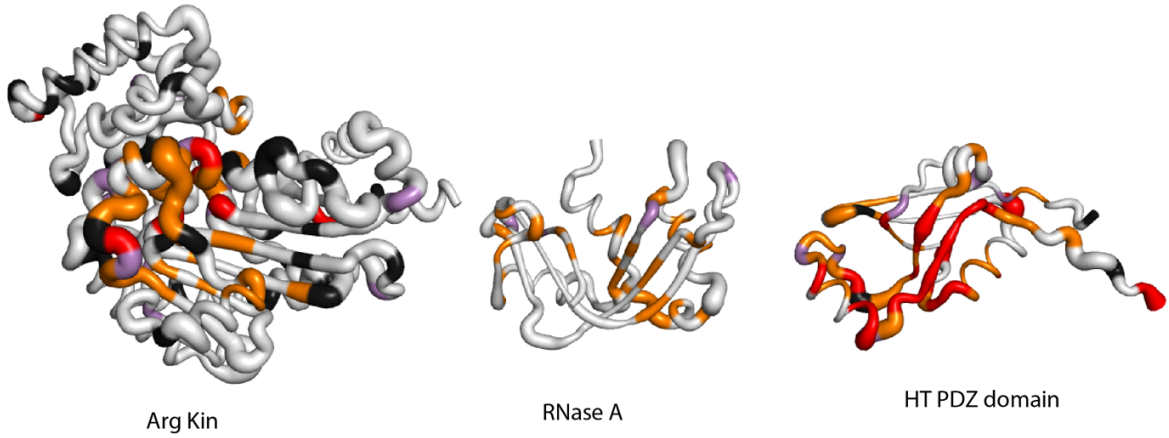



**Extended Data Figure 7:** Some proteins with poorly predicted exchange in RelaxDB reflect phosphate-binding exchange or high probability for ps-ns motion. 14/29 of proteins in RelaxDB predicted with an AUROC < 0.5 are these proteins that bind phosphate ion moieties during their biological function, yet the NMR experiments were measured in phosphate buffer. (a) $R_2/R_1$ data for 6 of these proteins which bind RNA or DNA. (b) All positively-charged sidechains (arginine, lysine) are depicted in stick form. Many of the residues with experimental exchange (orange) that are not predicted by Dyna-1 are on the surface of these proteins and can be explained by phosphate binding. (c) Proteins with low AUROC and high p(exchange) predicted in disordered N- or C- termini, or areas with characterized ps-ns motion, labeled in blue in Fig. 4b and here. In (b,c): structures are sized by Dyna-1 p(exchange) and colored by experimental labels as in Fig. 4.

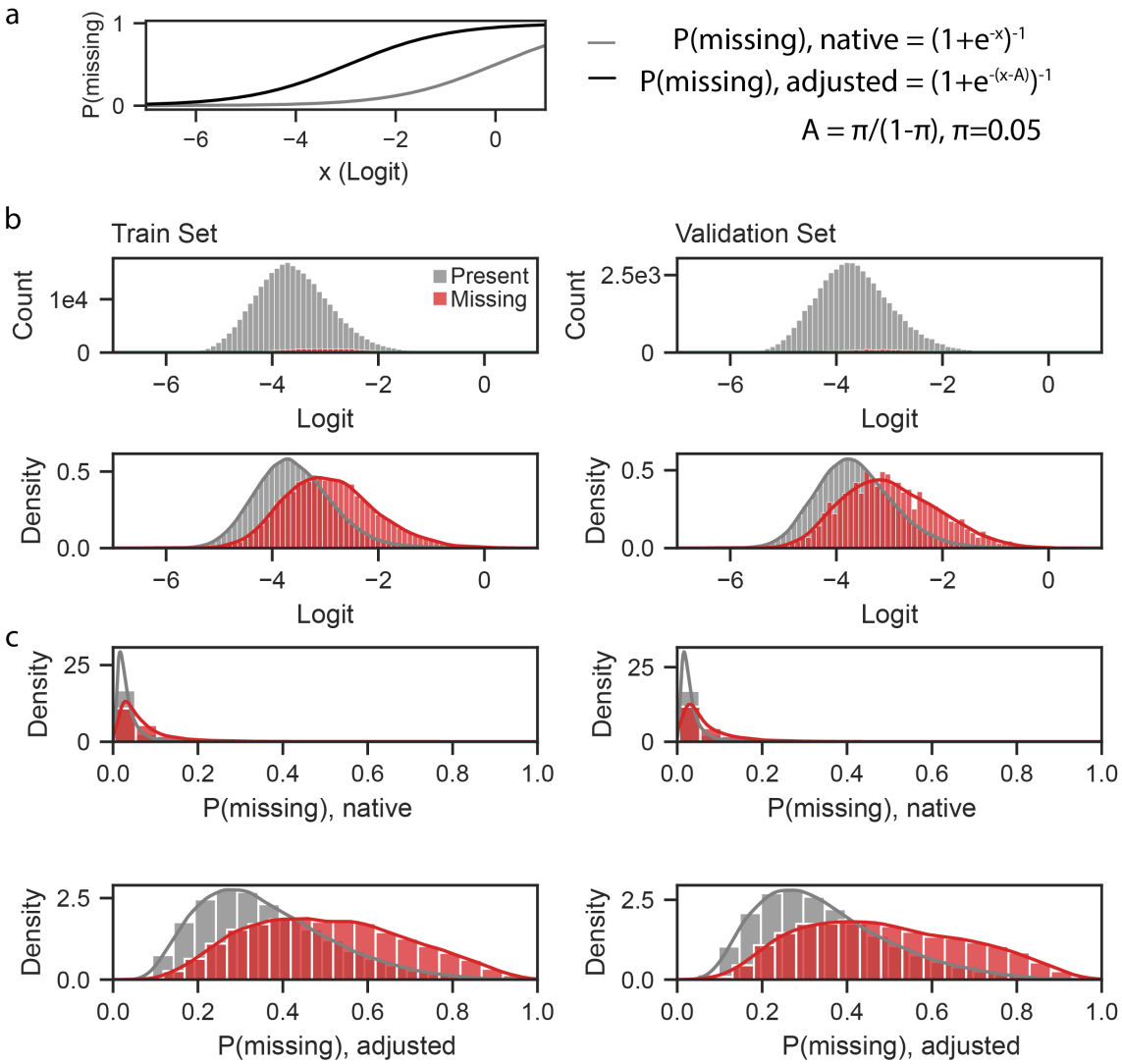



**Extended Data Figure 8**: RelaxDB-CPMG datasets where dispersion data was available for all residues. Legend: Missing assignments (red star), prolines (purple "P"), exchange via $R_{ex}$ calculated

conventionally (light orange), residues with unsuppressed $R_{ex}$ (light green), residues with no CPMG data due to peak overlap (gray). The color gradient represents $B_1$ field strength from lowest (red, e.g., 50 Hz) to highest (purple, e.g., 1000 Hz). The AqAdk and CypA datasets are annotated with residue groups reported in refs. [58] and [55], respectively.




**Extended Data Figure 9**: Additional $^{15}N$ CPMG datasets with no dispersion data published for all residues (cf. Figure 5e).




**Extended Data Figure 10**: Adjusting native logits from Dyna-1 to account for class imbalance. (a) We follow the practice described in ref. [81] to adjust probabilities calculated from logits to maximize the "balanced accuracy", i.e. accuracy averaged across both classes, rather than accuracy averaged over all datapoints. We set our prior for the missing class based on the training data, π, as 5%. Comparing the adjusted curve (black) to the probability densities in (b) demonstrates how this intuitively matches where a discriminator between the two classes should be (at x ~ -3). (b) Histograms and probability densities of logits predicted for residues labeled as "present" and "missing" over train and validation sets. (c) Densities of resulting p(missing) if calculated using native logits or adjusted logits.