Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Experiment-guided AlphaFold3 resolves accurate protein ensembles

Alex Bronstein

alexbronst@gmail.com

ISTA https://orcid.org/0000-0001-9699-8730

Advaith Maddipatla

ISTA

Nadav Bojan Sellam

ISTA

Meital Bojan

ISTA

Sanketh Vedula

Princeton University

Ailie Marx

Migal

Paul Schanda

Kloserneuburg

Article

Keywords:

Posted Date: October 21st, 2025

DOI: https://doi.org/10.21203/rs.3.rs-7835261/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** Competing Interest.

Experiment-guided AlphaFold3 resolves accurate protein ensembles

Abstract

AlphaFold3 predicts highly accurate protein structures from sequence, but tends to collapse to a single dominant conformation, even when the underlying structure is inherently heterogeneous. Moreover, its predictions are oblivious to experimental conditions that can alter local sequence conformation. In this work we show that AlphaFold3 can be guided to match data obtained by NMR spectroscopy, X-ray crystallography and cryo-EM experiments. We demonstrate that this methodology can generate ensembles of conformations having less distance restraint violations than traditionally resolved NMR structures and uncover unmodelled alternate conformations detectable in electron density. We show that AlphaFold3 can also be guided by cryo-electron microscopy maps and that doing so in combination with NMR parameters improves model quality. This methodology paves the way for the development of experimentally aware predictive models that capture the ensemble nature of protein structures.

Introduction

Proteins are inherently dynamic entities, sampling variable conformational states in response to their surroundings and to fulfill their biological roles. While experimental techniques including X-ray crystallography, NMR spectroscopy and cryo-electron microscopy observe ensemble averages of these macromolecular structures, protein structural models often only report the dominant conformation, overlooking the underlying conformational heterogeneity.

Leveraging on the fundamental discovery that amino acids in spatial proximity within a protein structure coevolve (Göbel et al., 1994; Hopf et al., 2014), and training on the hundreds of thousands of experimentally determined X-ray crystal structures available today, deep learning-based models such as AlphaFold (Jumper et al., 2021; Abramson et al., 2024), have revolutionized structural biology by enabling protein structure

prediction approaching experimental accuracy. However, AlphaFold's training objective – to predict a single "most probable" structure – biases its output toward static snapshots, effectively marginalizing conformational heterogeneity encoded in its training data. This emphasizes the need for new models that can explicitly model protein ensembles that are faithful to experimental measurements.

To bridge this gap, we present a framework for building experiment-grounded protein structure generative models that infer conformational ensembles consistent with measured experimental data. The key idea is to treat state-of-the-art protein structure predictors (e.g., AlphaFold3) as sequence-conditioned structural priors, and cast ensemble modeling as posterior inference of protein structures given experimental measurements.

Results

We adapt AlphaFold3 to generate ensembles by modifying the reverse diffusion steps such that it includes a gradient-based guidance term derived from the experimental likelihood. This term introduces a data-dependent force that steers sampling toward structures compatible with measurements. A scaling hyperparameter controls the influence of this guidance, allowing interpolation between purely prior-driven and strongly data-conditioned outputs. Our framework supports multiple experimental modalities by defining appropriate differentiable likelihood functions. Following sampling, we perform energy minimization to correct geometric distortions and ensure chemical plausibility – a procedure that is used in many generative models including AlphaFold itself. We then apply ensemble selection to identify the minimal subset of structures that best explain the experimental observations. Refer to Figure 1 for a schematic description of the proposed method and to Supplementary Methods for algorithmic and implementation details.

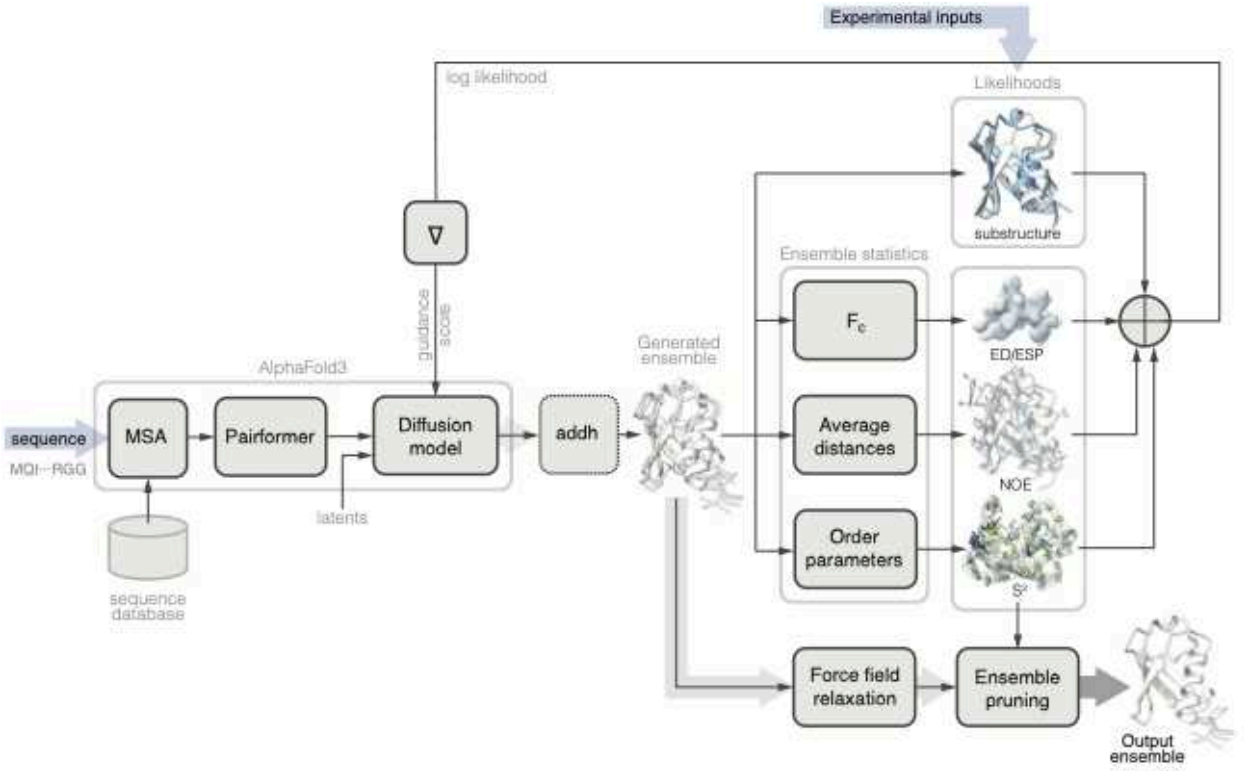


Figure 1 – Schematic depiction of the proposed method. AlphaFold3 allows the sampling of protein structures given an amino acid sequence. To further condition the model by experimental observations, at each time step of the AlphaFold3 diffusion model, an ensemble of structures is generated. Likelihoods of experimental observations are calculated given each individual ensemble member (e.g., to enforce locations of certain atoms) and on ensemble averages (e.g., calculated electron density Fc, average inter-atomic distances, or order parameters). The gradient of the combined log-likelihood terms is used as the guidance term. At the final diffusion step, the generated ensemble is refined by force field relaxation and pruned to select a minimal subset maximizing the likelihood terms.

Guiding AlphaFold3 with NMR distance restraints and order parameters efficiently generates experimentally accurate ensembles

NMR structure determination is largely based on the detection of inter-atomic proximities, which is a manifestation of the dipolar coupling between spins, and detected in Nuclear Overhauser Effect (NOE) experiments in solution-state NMR (Shcherbakov et al., 2021). Typically the structure is then determined by applying these distance restraints in molecular dynamics (MD) simulations to obtain the set of lowest energy conformations (Wang et al., 2004; Schwieters et al., 2006; Güntert, 2004).

However, simulating conformers individually often leads to mode collapse, producing rigid ensembles that poorly capture true conformational dynamics. Also, oftentimes no single mode's pairwise Euclidean distances can fully explain the observed distance averages over multiple incongruent conformations. The explicit determination of multiple state-structural ensembles is becoming more accessible (Vögeli et al., 2016, Chi et al., 2015), but it often requires highly accurate ("exact") NOE distances, and is not yet common practice in NMR. Ensemble-based MD approaches simulating multiple conformers simultaneously to satisfy experimental restraints remain computationally expensive, requiring days even for small systems like the 76-residue ubiquitin (Lindorff-Larsen et al., 2005 and Lange et al., 2008).

We use ubiquitin (PDB 1D3Z, BMRB 6457, 5387) as an NMR structure determination benchmark. Guiding AlphaFold3 with NOE distance restraints generates ensembles with significantly less pairwise distance constraint violations. Fig. 2A shows the cumulative distribution of violation magnitudes; a spatial visualization of the violated constraints are depicted in Fig. 2C. From Fig. 2B, it is evident that the ensemble produced by NOE-guided AlphaFold3 is more heterogeneous. SI Fig. 1 shows the improvement of constraint satisfaction with the increase of the relative weight of the NOE guidance term. The runtime of the proposed method is several minutes, overcoming the computational bottleneck of traditional ensemble methods.

As an independent verification of ensemble fit to actual protein dynamics, we cross-validated our results using experimental spin relaxation measurements of the backbone amide bond order parameter (S^2) on time scales shorter than a few nanoseconds (Palmer, 2004). We computed ensemble order parameters and compared them to experimental measurements reported in (Lienin et al., 1998). As illustrated in Fig. 2D, the deposited ensemble derived from standard NMR protocols (PDB 1D3Z) as well as the unguided AlphaFold3 prediction are dominated by rigid conformations. Additional quantitative analysis in SI Fig. 2 shows that these ensembles exhibit only moderate agreement with the experimental S^2. In contrast, NOE-guided AlphaFold3 significantly improves agreement, better replicating the dynamic behavior observed in both flexible and structured regions.

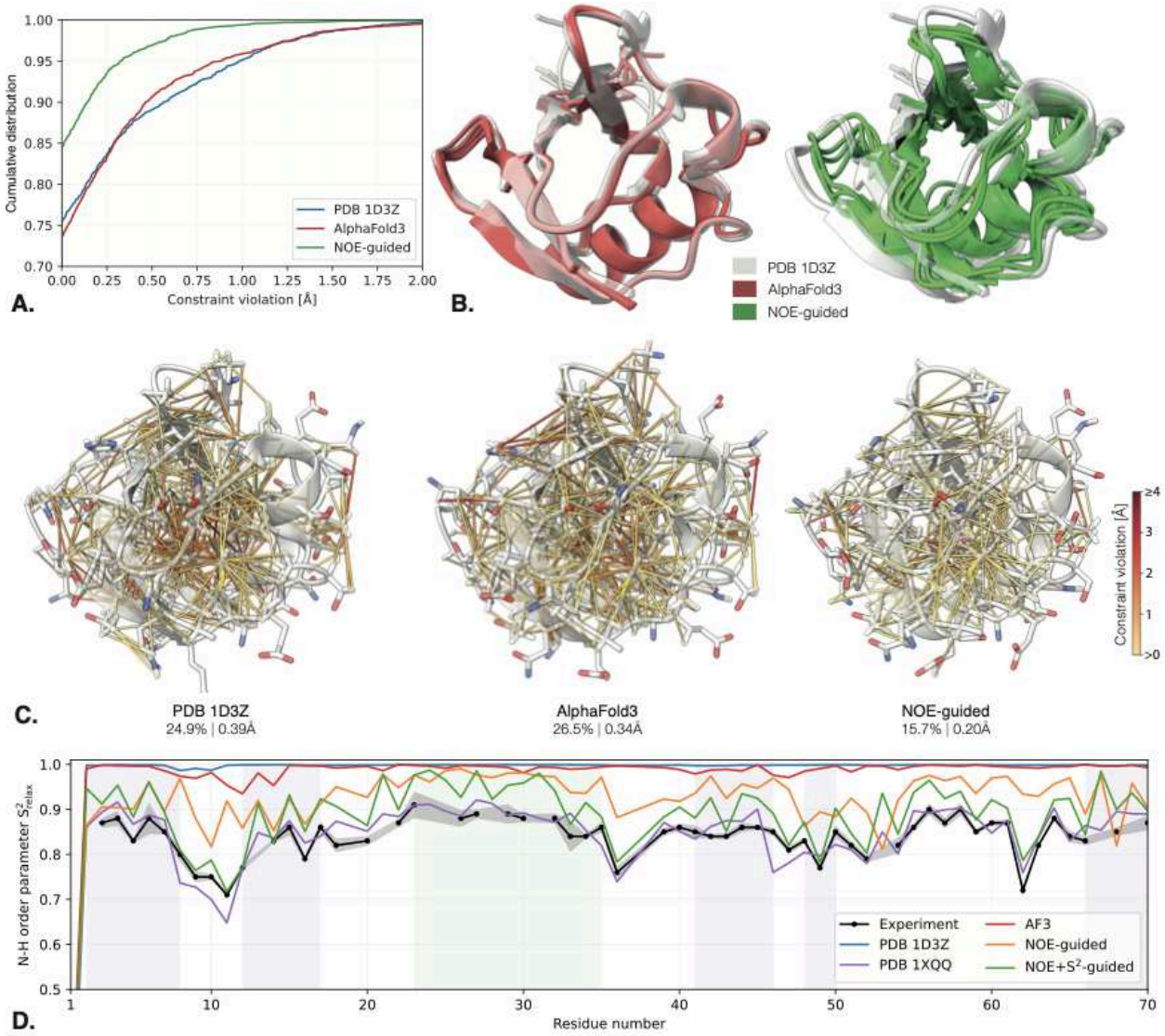


Figure 2 – **NMR-guided AlphaFold3.** Guidance of ubiquitin structure by NOE pairwise distance bounds enforced as ensemble statistics produces significantly less constraint violations than the PDB structure ensemble 1D3Z. (A) Cumulative distribution of constraint violations; (B) unguided (red) and NOE-guided (green) ensembles overlaid on the PDB structure (transparent white); (C) visualization of the pairwise distance constraint violations for the PDB ensemble, unguided and NOE-guided predictions. The fraction of the violated constraints (out of total 1212) and their median violation are reported below each structure. A single best-fitting structure from the ensemble is shown for clarity; (D) Relaxation amide order parameter S^2 from Lienin et al., 1998 (black) vs. order parameter calculated from the ensembles of different methods. Explicitly guiding the ensemble with NOE and ^{15}N-relaxation derived S^2 constraints (green) produces good agreement with the experiment (r=0.93, q=0.06) and is comparable to the PDB ensemble 1XQQ (purple) fitted to match structure and dynamics (r=0.87, q=0.04). For full results with relaxation and residual dipolar coupling (RDC) data, refer to Supplementary Figures S2, S3.

In order to better match the N-H order parameter experimental data, we added an S^2 guidance term to the NOE-guided AlphaFold3. This resulted in a significantly higher agreement with the experimental S^2 measurements and comparable to the result of computationally expensive NMR-guided ensemble MD simulations (PDB: 1XQQ from Lindorff-Larsen et al., 2005) at a small fraction of the computational cost (Fig. 2D and SI Fig. 2A-C). For the case of ubiquitin, a number of additional experimental data have been collected, including methyl axis order parameters from spin relaxation (reporting on ps-ns motions, Liao et al., 2012) and order parameters from residual dipolar couplings for amide and methyl sites, which report on ps-ms motions (Markwick et al., 2009; Sabo et al., 2014; Lange et al., 2008). SI Figs. 2D-F and 3 show the result of guidance with these data. In all cases guidance with NOE+S^2 terms shows remarkably high agreement with the experiment at no or little expense of pairwise distance constraint violation.

We further evaluated our method on two benchmarks: 8 peptides shown to be mispredicted by AlphaFold3 (Vedula et al. 2025, McDonald et al. 2023) and a subset of 83 proteins from the recently compiled 100 NMR spectra database (Klukowski et al., 2024). SI Table 1 details quantitative performance indicators for individual structures, while Fig. 5A presents summary statistics. We observe that compared to the ensembles deposited into the PDB, NOE-guided AlphaFold3 improves distance constraint satisfaction in 70 out of 91 cases (about 77%), compared to unguided AlphaFold3 which outperforms the PDB ensembles only in 15 cases (17%). NOE-guided AlphaFold3 outperforms its unguided counterpart in all cases.

Fig. 3 depicts examples of ensembles generated by NOE-guided AlphaFold3. We observe that AlphaFold3 tends to over-predict order (e.g., longer helical structures in 2B3W, 2JVD and 6SOW). In these cases, NOE guidance increases the ensemble heterogeneity, considerably improving constraint satisfaction and outperforming the corresponding PDB ensembles. In 2K52, no single conformer of the ensemble satisfies the distance constraints, while the ensemble average guidance term significantly improves constraint satisfaction, also compared to the PDB ensemble. In three other cases, AlphaFold3 completely mispredicts parts of the structure (2LI3, 2B3W and

1DEC) resulting in many distance constraint violations. NOE guidance significantly reduces the number of violations.

Finally, as an independent source of validation, we used the ANSURR metrics introduced in (Fowler et al., 2020) that compare the rigidity profile of the model to that inferred from NMR chemical shifts. We observe that ensembles generated by NOE-guided AlphaFold3 improve the RMSD score in 92% of the cases and the correlation score in 54% of the cases when compared to the corresponding PDB ensembles. Refer to SI Fig. 4 for more information.

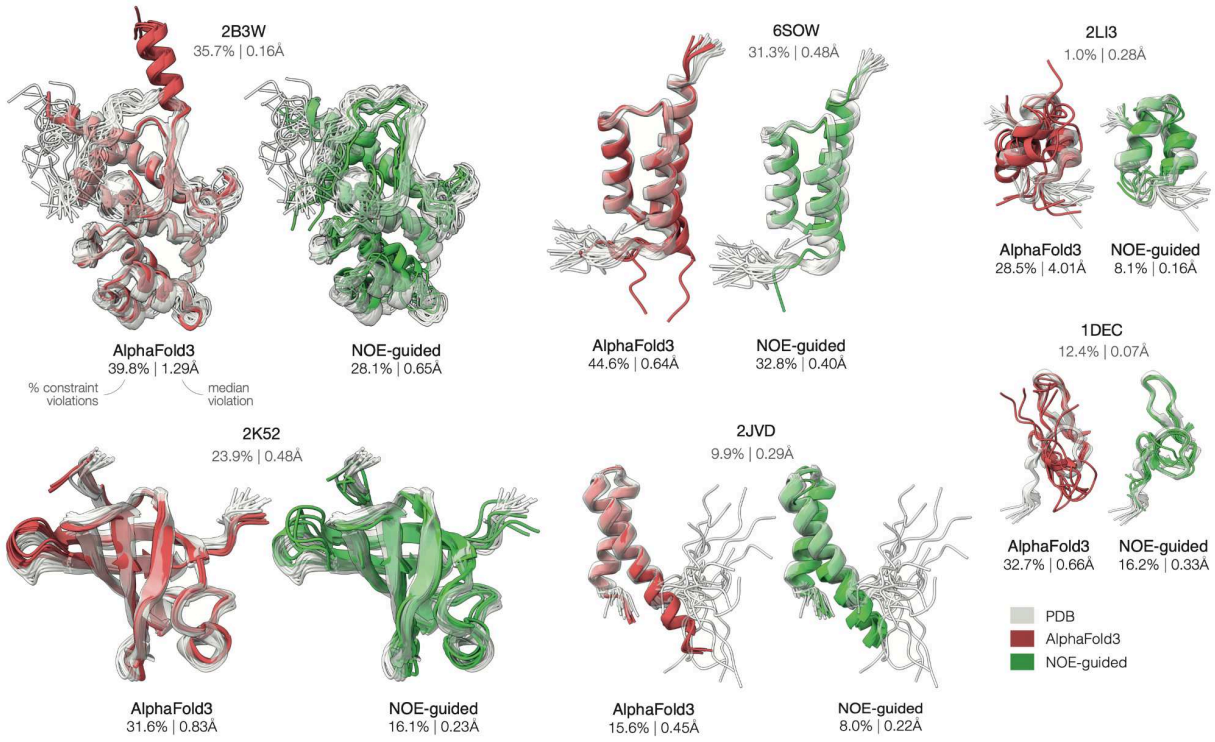


Figure 3 – **Conformation ensembles generated for six proteins** using unguided AlphaFold3 (red) and its NOE-guided version (green). Ensembles are overlaid on corresponding NMR structures solved from the same NOESY data shown in transparent white. PDB identifiers are indicated above each structure. The numbers below report the percentage of violated constraints and their median violation.

Guiding AlphaFold3 with electron density captures previously unmodelled residues in X-ray crystallography data.

AlphaFold3 is oblivious to environmental influences such as interactions with ions, ligands and other macromolecular partners or different crystal contacts resulting in conformational variations. For example, heat shock protein 90-alpha crystallises as a dimer in the asymmetric unit cell with only chain A, and not chain B, captured in the ligand bound state (PDB 6CYH). The loop adjacent to this ligand adopts a significantly different conformation between the two chains, neither of which is accurately predicted by AlphaFold3. Fig. 4A shows that the poor AlphaFold3 prediction is restored to experimental accuracy by electron density guidance. The method is similarly successful when applied across experiments. SI Table 2 shows that a pair of sequence identical constructs for the Orf9b protein from SARS-CoV-2 exhibit varying conformations at several protein chain locations (PDB 9N55 and 9MZB). Electron density guidances successfully restores the AlphaFold3 model generation to match the PDB deposited structures. The success of fitting on electron density maps having very different resolutions (PDB 9N55 at 1.65Å and 9MZB at 2.8Å) pays tribute to the generalizability of our method. In yet another example, we consider a pair of myoglobin structures that differ in surface-exposed loop conformations due to altered crystal packing (PDB 1U7R and 1U7S). Electron density guided AlphaFold reproduces both conformations with density fits comparable to deposited models. Interestingly, it has been noted that these alternate conformations closely resemble those observed in solution by NMR (PDB 1MYF), suggesting that the lattice selects from intrinsically adopted conformations (Kondrashov et al., 2008) and underscores the biological relevance of training AlphaFold on ensembles detectable in crystal density, now an achievable task with our methodology.

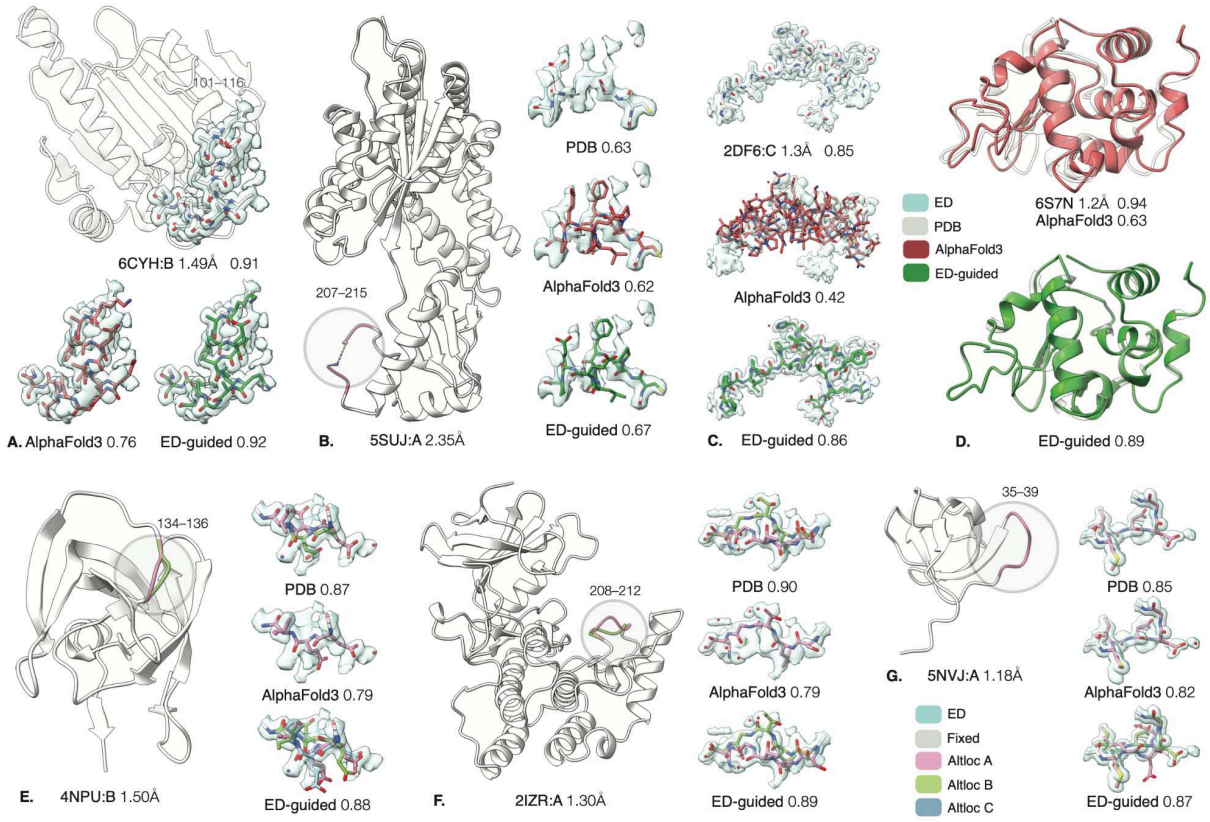


Figure 4 – X-ray electron density (ED)-guided Alphafold produces structures and ensembles of equivalent or better accuracy to those deposited in the PDB and detects plausible conformations in unmodelled electron density. (A) 6CYH chain B is poorly predicted by AlphaFold3 in the region of amino acids 101–116. Guidance using the electron density maps restores the fit of the model. (B) An unmodelled region in 5SUJ has different conformations at two locations due to the presence of ligands in one chain. (C) The structure of an 18 amino acid peptide from PAK2 bound to the SH3 Domain of betaPIX is poorly predicted by AlphaFold3 but this prediction can be guided by the electron density to recapture the experimental structure. (D) AlphaFold3 poorly predicts multiple regions of 6S7N. ED guidance essentially reproduces the PDB structure. (E-F) ED-guided AlphaFold3 produces conformational ensembles explaining the multi-modal electron density on par or better than the PDB model. (G) ED-guided AlphaFold3 predicts unmodelled altlocs in 5NVJ better explaining the electron density. Electron densities are visualized as $0.5e^{-}/Å^{3}$-isosurfaces. Numbers below indicate cosine similarities between calculated and observed densities.

Another region-specific phenomenon in X-ray crystal structures are unmodelled regions. Local flexibility results in a smeared electron density map (measured by the B factor), to the point where the density is too blurred to readily infer structure. We show in Fig. 4B that density guided AlphaFold3 has the potential to fill in missing regions to best explain

the scarce density. The crystal structure of an uncharacterized protein LPG2148 from Legionella pneumophila (PDB 5SUJ) crystallizes as a dimer in the asymmetric unit cell and both chains have a missing segment in the model at the same surface exposed loop region. Fig. 4B and SI Fig. 6 shows that the density guided version fills the missing space in both chains with better local real space correlation to the sparse density than the AlpaFold3 prediction or the commonly used PDBfixer (Eastman *et al.*, 2017).

Another challenge in crystallography is modelling peptides. These short protein chains are often very poorly predicted by AlphaFold3 since they are highly adaptable to their binding partners. Fig. 4C shows that our density-guided method restores the fit of the peptide to the electron density from a very poor AlphaFold3 prediction. This example leverages AlphaFold3 as a simple structural prior, since these ensembles were generated without MSA conditioning. Further examples in SI Table 2 and the summary statistics in Fig. 5B highlights the method's ability to generalize across sequences of varying lengths.

Going beyond segment based guidance we automatically guide an entire protein chain through an iterative process of detection and improvement of regions poorly fit to the electron density and phase refinement. The final fitted structure shown in Fig. 4D results from iterative improvements in real space correlation (SI Fig. 7). We hypothesize that the results appear slightly inferior to those deposited in the PDB mainly due to the manual placement of ions and structural waters that AlphaFold does not generate.

Although early crystallographic refinements prioritized single-conformer models, often protein crystal structures can also capture ensembles as two or more distinct conformations within the same electron density map (Smith et al., 1986; Furnham et al., 2006; van den Bedem & Fraser, 2015; Wankowicz et al., 2024). A recent study by Rosenberg et al. (2024a) compiled a comprehensive catalog of altlocs from PDB structures and also showed that even for regions with well-separated and stable altlocs, structural ensemble predictors such as AlphaFold3 fail to reproduce the experimentally determined distributions or capture the bimodal nature of backbone conformations (Rosenberg et al., 2024b). Here, we demonstrate that electron density guidance on this

same set of altlocs generates an ensemble that captures the backbone's bi-modal nature, providing a better explanation of the density. This method is effective for different degrees of backbone separation spanning from 1 Å (Fig. 4E) to tens of Ångstroms (Fig. 4F). SI Table 2 further demonstrates that alternate conformations are successfully detected across different sequence types and lengths and within electron density maps having different resolutions and qualities. Moreover, SI Table 4 shows that these conformations are consistently reproduced across independent runs using different types of density maps, with the consistency further improving as batch size increases (SI Fig. 8). Finally, we demonstrate the potential of this method for detecting previously unmodelled altloc conformations (Fig. 4G).

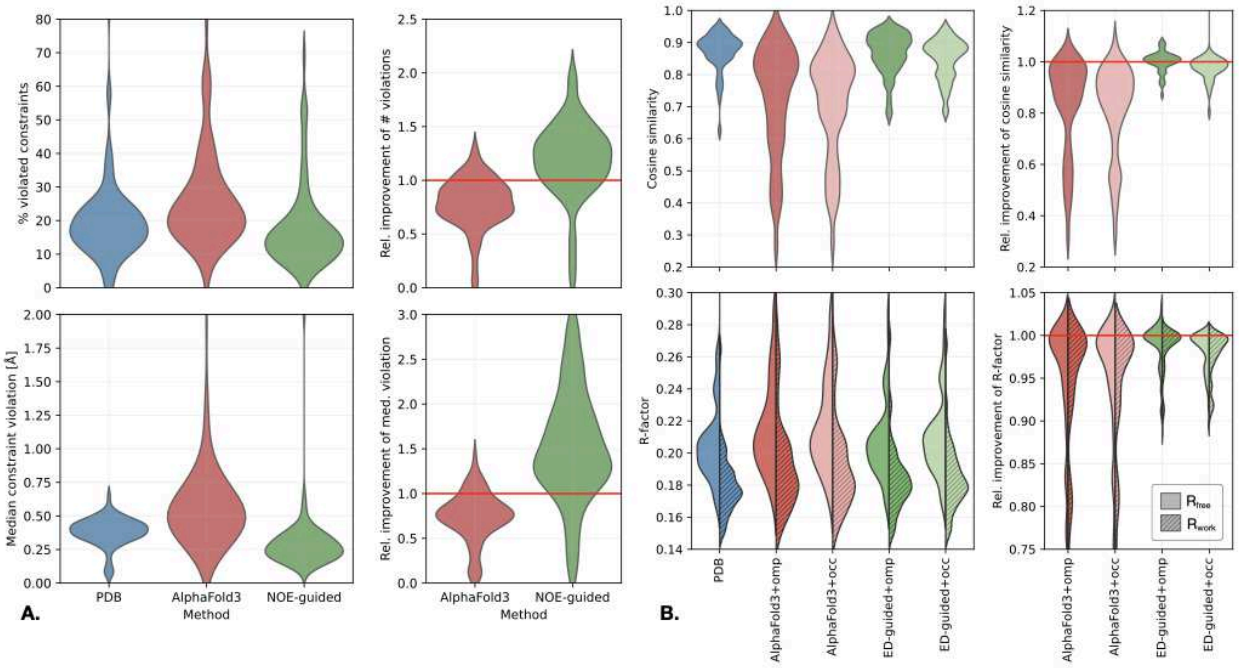


Figure 5 – Summary statistics on evaluated NMR and X-ray crystallographic benchmarks. (A) Distribution of % of violated distance constraints (top left) and median violation (bottom left) for PDB, unguided and NOE-guided AlphaFold3 and the relative improvement of these metrics compared to the deposited PDB structures. (B) Distribution of F_o and F_c cosine similarity (top left) and R-factors (top right, split between R_{free} and R_{work}), with the relative improvement shown on the right.

Guiding AlphaFold3 with cryoEM maps in combination with NMR parameters improves model quality.

CryoEM imaging allows to capture the conformations of intricate protein complexes in the form of electrostatic potential (ESP) maps. Since the standard reconstruction pipelines rely on rigid body consensus sets, individual ESP maps do not capture the true thermodynamic heterogeneity of the particle. At the same time, the achievable resolution – especially in flexible regions – is often insufficient for precise placement of individual atoms. We demonstrate on several examples that AlphaFold3 guidance with ESP maps achieves better *ab initio* modeling accuracy.

Fig. 6A shows the reconstruction of insulin receptor IR-B in its symmetric apo state (PDB 8U4B, EMDB 41877) and asymmetric conformation (PDB 8U4E, EMDB 41880) with three bound IGF2 factors. AlphaFold3 tends to over-symmetrize the predicted homodimers, mispredicting large portions of the asymmetric conformation but also inaccurately predicting the symmetric parts. ESP map guidance produces conformations that agree with the observed densities.

Fig. 6B shows the reconstruction of an amyloid-beta fibril (PDB 9FH1, EMDB 50436) formed by a nearly planar homodimer that further oligomerizes in the perpendicular direction forming a very long fiber structure. AlphaFold3 completely mispredicts the dimerization interface and the structure of the dimeric unit. ESP guidance produces accurate agreement with the density.

Fig. 6C shows the reconstruction of a RIPK3 human amyloid fibril (PDB 7DA4, EMDB 30622) for which NMR chemical shift-derived dihedral angles are also available (PDB 7DAC, BMRB 36392). In this case, AlphaFold3 predicts the global structure of the multimer relatively well within the resolution of the ESP map, and ESP-only guidance improves the density fit quality while producing a relatively poor fit of the backbone dihedrals and NOE constraints. Guiding jointly by ESP, NOE and dihedral angles significantly improves the local structure accuracy at a reasonable tradeoff of density fit (see SI Table 5 for quantitative analysis).

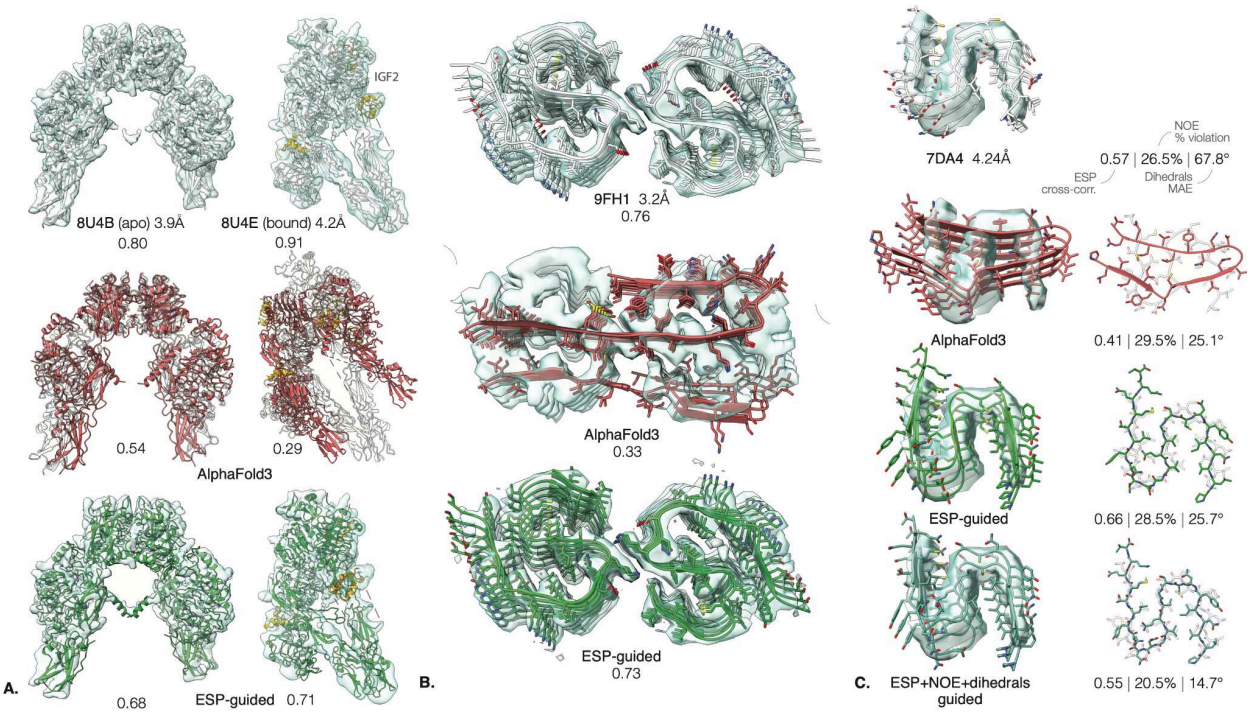


Figure 6 – Multimeric structure reconstruction using cryoEM electrostatic potential (ESP) map guidance. (A) Reconstruction of insulin receptor IR-B in its symmetric apo state (left) and asymmetric conformation (right) with three bound IGF2 factors (yellow); and (B) reconstruction of an amyloid-beta fibril. AlphaFold3 (red) and ESP-guided (green) structures are shown alongside the deposited PDB models. Density cross-correlation values are indicated below each structure. (C) Reconstruction of a human RIPK3 amyloid fibril using unguided AlphaFold3 (red), ESP-guided AlphaFold3 (green) and ESP-guided AlphaFold3 with the addition of NMR chemical shift-inferred backbone dihedral angles ϕ, ψ (cyan). The addition of the dihedrals further improves agreement with both experimental observations. Cross-correlations, NOE constraint violations and dihedral errors are reported below each structure.

Discussion

This work reframes AlphaFold3 as a powerful sequence-conditioned prior that can be steered by experimental evidence to yield small, testable conformational ensembles rather than a single consensus model. Casting structure determination as posterior inference makes it straightforward to compose likelihoods from heterogeneous modalities—inter-atomic distances, dihedral angles order parameters from NMR, electron density from X-ray crystallography, and electrostatic-potential maps from

cryo-EM—and to direct sampling toward structures that better explain the data. In practice, guidance is injected as gradients during AF3's reverse-diffusion trajectory and followed by relaxation and sparsifying ensemble selection, producing a minimal set of conformers that is both chemically plausible and data-faithful. The method's ability to sample from the joint distribution of all atoms unlocks a richer picture of the protein's conformational landscape that is partially lost in the traditional PDB representation of its structure as a mixture of Gaussian marginals of individual atoms.

Currently, the field lacks human-intelligible visualization and machine-interpretable representations of the joint probability distributions that govern proteins and complexes. An intriguing question is whether Nature re-uses a restricted vocabulary of local distributional motifs—analogous to secondary structure—but defined over distributions of backbone/side-chain states, and whether there exist non-local distributional motifs of correlated motion and allostery, analogous to structural domains. The degree to which such motifs are imprinted in evolutionary signals (e.g., MSA features) will determine how far inductive models can progress from predicting static structure to predicting relative populations of states.

Methodologically, three features are central in this study. First, ensembles are jointly guided during diffusion, which preserves diversity where the data admit multiple explanations while avoiding trivial replication of a single conformation; subsequent relaxation ensures chemical plausibility; and a matching-pursuit–like pruning step yields a compact set that still explains the observations. Second, the framework naturally supports "local-to-global" strategies—targeting problematic segments (e.g., loops) and then iterating to the whole chain—mirroring how crystallographers and NMR spectroscopists often work. Third, the computational footprint is small (minutes per target in typical cases, which is likely to be further optimizable), making the method practical as an interactive assistant within existing pipelines rather than a replacement for refinement or simulation. For the same reason, the tool enables large-scale re-analysis of legacy datasets ("PDB redo") across single modalities and their combinations.

Several caveats warrant emphasis. The reported ensembles are explanatory with respect to the provided data and the AF3 prior; they are not calibrated thermodynamic populations. Without additional constraints, mixture weights should not be over-interpreted as equilibrium populations. Future work could link pruning weights to maximum-entropy or evidence-based population estimates, or incorporate sparse thermodynamic measurements (e.g., ΔΔG, KD) to at least partially calibrate populations. Guidance strength and gradient handling are important guardrails to avert models being pulled into noise, especially in low-resolution EM or poorly phased crystallographic maps. While in practice ensemble pruning and cross-validation techniques mitigate overfitting, a more principled statistically robust procedure is needed to disentangle true thermodynamic entropy from epistemic uncertainty. Finally, the present implementation focuses primarily on proteins; treatment of ligands, metals, PTMs, structural waters, and complex symmetries remains limited when the only evidence is a density-like signal. A more faithful modeling of ligands will also allow disentangling between native thermodynamic and compositional conformational heterogeneity.

These limitations notwithstanding, the posterior-sampling view opens practical routes to improve mainstream structural workflows. In crystallography, experiment-guided AlphaFold3 can propose altlocs and occupancies during model building, repair loops with weak electron density, and accelerate convergence by offering a small set of candidate conformers that already satisfy the map. In NMR, it enables rapid, ensemble-aware interpretation of NOESY spectra and relaxation data without days of MD, and can suggest mutations or ligands that stabilize minor predicted states for validation. In cryo-EM, it provides a principled way to inject atomic detail into flexible regions and to couple map-based guidance with dihedral or secondary-structure information where local resolution is limited. We anticipate that experiment-guided generative models will make ensemble-centric modeling routine in crystallography,

NMR, and cryo-EM, with near-term impact on ligand discovery, variant interpretation, and the design of experiments that stabilize or reveal otherwise cryptic states.

Data and Code Availability

All structures and metrics reported in this paper are available at https://doi.org/10.7910/DVN/PLYUHN

The code is available at https://doi.org/10.5281/zenodo.17307005

References

1. Abramson, J. et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024). https://doi.org/10.1038/s41586-024-07487-w.
2. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D* **66**, 213–221 (2010).
3. Baek, M. et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871–876 (2021).
4. Baek, M. et al. Accurate prediction of protein–nucleic acid complexes using RoseTTAFoldNA. *Nature Methods* **21**, 117–121 (2024).
5. Burley, S. K. et al. Protein Data Bank (PDB): the single global macromolecular structure archive. *Protein Crystallography: Methods and Protocols* 627–641 (2017).
6. Chen, X. et al. Protenix—advancing structure prediction through a comprehensive AlphaFold3 reproduction. *bioRxiv* (2025). Preprint
7. Chi, C. N., Vögeli, B., Bibow, S., Strotz, D., Orts, J., Güntert, P. & Riek, R. A structural ensemble for the enzyme cyclophilin reveals an orchestrated mode of action at atomic resolution. *Angewandte Chemie International Edition* **54**, 11657–11661 (2015). https://doi.org/10.1002/anie.201503698.
8. Eastman, P. *et al.* OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comput. Biol.* **13**, e1005659 (2017).
9. Fowler, N. J., Sljoka, A. & Williamson, M. P. A method for validating the accuracy of protein NMR structures. *Nature Communications* **11**, 6321 (2020). https://doi.org/10.1038/s41467-020-20065-1.

10. Furnham, N., Blundell, T. L., DePristo, M. A. & Terwilliger, T. C. Is one solution good enough?. *Nature Structural & Molecular Biology* **13**, 184–185 (2006).
11. Göbel, U., Sander, C., Schneider, R. & Valencia, A. Correlated mutations and residue contacts in proteins. *Proteins* **18**, 309–317 (1994). https://doi.org/10.1002/prot.340180402.
12. Güntert, P. Automated NMR structure calculation with CYANA. *Protein NMR Techniques* 353–378 (2004).
13. Gutermuth, T., Sieg, J., Stohn, T. & Rarey, M. Modeling with alternate locations in X-ray protein structures. *Journal of Chemical Information and Modeling* **63**, 2573–2585 (2023).
14. Ho, J., Jain, A. & Abbeel, P. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems* **33**, 6840–6851 (2020).
15. Hopf, T. A. et al. Sequence co-evolution gives 3D contacts and structures of protein complexes. *eLife* **3**, e03430 (2014).
16. Ingraham, J. B. et al. Illuminating protein space with a programmable generative model. *Nature* **623**, 1070–1078 (2023). https://doi.org/10.1038/s41586-023-06728-8.
17. Jing, B., Berger, B. & Jaakkola, T. AlphaFold meets flow matching for generating protein ensembles. *arXiv* (2024). arXiv:2402.04845
18. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
19. Kondrashov, D. A., Zhang, W., Aranda, R., 4th, Stec, B., & Phillips, G. N., Jr (2008). Sampling of the native conformational ensemble of myoglobin via structures in different crystalline environments. Proteins, **70**(2), 353–362. https://doi.org/10.1002/prot.21499
20. Klukowski, P. et al. The 100-protein NMR spectra dataset: A resource for biomolecular NMR data analysis. *Scientific Data* **11**, 30 (2024).
21. Lang, P. T., Holton, J. M., Fraser, J. S. & Alber, T. Protein structural ensembles are revealed by redefining X-ray electron density noise. *Proceedings of the National Academy of Sciences USA* **111**, 237–242 (2014).
22. Lange, O. F. et al. Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. *Science* **320**, 1471–1475 (2008).
23. Lange, O. F., Lakomek, N.-A., Fares, C., Schroder, G. F., Walter, K. F. A., Becker, S., Meiler, J., Grubmuller, H., Griesinger, C., & de Groot, B. L. Recognition Dynamics Up to Microseconds Revealed from an RDC-Derived Ubiquitin Ensemble in Solution. *Science*, **320**(5882), 1471–1475 (2008). https://doi.org/10.1126/science.1157092

24. Liao X, Long D, Li DW, Brüschweiler R, Tugarinov V. Probing side-chain dynamics in proteins by the measurement of nine deuterium relaxation rates per methyl group. *The Journal of Physical Chemistry B*. **116**(1):606-20 (2012).
25. Lienin, S. et al. Anisotropic intramolecular backbone dynamics of ubiquitin characterized by NMR relaxation and MD. *Journal of the American Chemical Society* **120**, 9870–9879 (1998).
26. Lindorff-Larsen, K. et al. Simultaneous determination of protein structure and dynamics. *Nature* **433**, 128–132 (2005).
27. Mallat, S. G. & Zhang, Z. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing* **41**, 3397–3415 (1993).
28. Markwick, P. R. L., Bouvignies, G., Salmon, L., McCammon, J. A., Nilges, M., & Blackledge, M. (2009). Toward a unified representation of protein structural dynamics in solution. *Journal of the American Chemical Society*, **131**(46), 16968–16975. https://doi.org/10.1021/ja907476w
29. McDonald, E. F. et al. Benchmarking AlphaFold2 on peptide structure prediction. *Structure* **31**, 111–119 (2023).
30. Mirdita, M. et al. ColabFold: making protein folding accessible to all. *Nature Methods* **19**, 679–682 (2022).
31. Palmer, A. G. NMR characterization of the dynamics of biomacromolecules. *Chemical Reviews* **104**, 3623–3640 (2004). https://doi.org/10.1021/cr030413t.
32. Pettersen, E. F. et al. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Science* **30**, 70–82 (2021).
33. Riley, B. T. et al. qFit 3: Protein and ligand multiconformer modeling for X-ray and cryo-EM maps. *Protein Science* **30**, 270–285 (2021).
34. Rosenberg, A. A., Marx, A. & Bronstein, A. M. A dataset of alternately located segments in protein crystal structures. *Scientific Data***11**, 783 (2024). https://doi.org/10.1038/s41597-024-03595-4.
35. Rosenberg, A. A., Vedula, S., Bronstein, A. M. & Marx, A. Seeing double: MD simulations reveal the stability of certain alternate protein conformations in crystal structures. *bioRxiv* (2024). https://doi.org/10.1101/2024.08.31.610605. Preprint
36. Sabo, T. M., Smith, C. A., Ban, D., Mazur, A., Lee, D., & Griesinger, C. ORIUM: Optimized RDC-based Iterative and Unified Model-free analysis. *Journal of Biomolecular NMR*, **58**(4), 287–301 (2014). https://doi.org/10.1007/s10858-013-9775-1
37. Schwieters, C. D., Kuszewski, J. J. & Clore, G. M. Using XPLOR–NIH for NMR molecular structure determination. *Progress in Nuclear Magnetic Resonance Spectroscopy* **48**, 47–62 (2006).

38. Shapovalov, M. V. & Dunbrack, R. L. A smoothed backbone-dependent rotamer library for proteins. *Structure* **19**, 844–858 (2011).
39. Shcherbakov AA, Medeiros-Silva J, Tran N, Gelenter MD, Hong M. From angstroms to nanometers: measuring interatomic distances by solid-state NMR. *Chemical reviews* **122**, 9848-9879 (2021).
40. Smith, J. L., Hendrickson, W. A., Honzatko, R. B. & Sheriff, S. Structural heterogeneity in protein crystals. *Biochemistry* **25**, 5018–5027 (1986).
41. Smyth, M. & Martin, J. X-ray crystallography. *Molecular Pathology* **53**, 8 (2000).
42. Terrell, G. R. & Scott, D. W. Variable kernel density estimation. *The Annals of Statistics* **20**, 1236–1265 (1992).
43. van den Bedem, H. & Fraser, J. S. Integrative, dynamic structural biology at atomic resolution—it's about time. *Nature Methods* **12**, 307–318 (2015). https://doi.org/10.1038/nmeth.3324.
44. van Zundert, G. C. et al. qFit-ligand reveals widespread conformational heterogeneity of drug-like molecules in X-ray electron density maps. *Journal of Medicinal Chemistry* **61**, 11183–11198 (2018).
45. Vögeli, B., Olsson, S., Güntert, P. & Riek, R. The exact NOE as an alternative in ensemble structure determination. *Biophysical Journal* **110**, 113–126 (2016). https://doi.org/10.1016/j.bpj.2015.11.031.
46. Vögeli, B. The nuclear Overhauser effect from a quantitative perspective. *Progress in Nuclear Magnetic Resonance Spectroscopy* **78**, 1–46 (2014).
47. Wankowicz, S. A. et al. Automated multiconformer model building for X-ray crystallography and cryo-EM. *eLife* **12**, RP90606 (2024).
48. Weiss, T., Mayo Yanes, E., Chakraborty, S., Cosmo, L., Bronstein, A. M. & Gershoni-Poranne, R. Guided diffusion for inverse molecular design. *Nature Computational Science* **3**, 873–882 (2023). https://doi.org/10.1038/s43588-023-00532-0.
49. Wu, X. et al. The structure of a minimum amyloid fibril core formed by necroptosis-mediating RHIM of human RIPK3. *Proceedings of the National Academy of Sciences USA* **118**, e2022933118 (2021). https://doi.org/10.1073/pnas.2022933118.
50. Vedula, S., Bronstein, A. M. & Marx, A. Improving prediction accuracy in chimeric proteins with windowed multiple sequence alignment. *Computational and Structural Biotechnology Journal* **27**, 3292–3298 (2025).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- natbiotechSI6.pdf