nature biotechnology

Article

https://doi.org/10.1038/s41587-026-03166-5

# Experiment-guided AlphaFold3 resolves measurement-consistent protein ensembles

Received: 11 October 2025

Accepted: 7 May 2026

Published online: 29 June 2026

Check for updates

Advaith Maddipatla[1], Nadav Sellam Bojan[1], Meital Bojan[1], Volodymyr Masalitin[1], Sanketh Vedula[2,3] ✉, Paul Schanda[1] ✉, Ailie Marx[4,5] ✉ & Alex M. Bronstein[1,6] ✉

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 nuclear magnetic resonance (NMR) spectroscopy, X-ray crystallography and cryogenic electron microscopy (cryo-EM) experiments and combinations thereof. Our approach can also incorporate data that explicitly report on dynamics, such as site-resolved order parameters. We demonstrate that this methodology generates compact structural ensembles whose ensemble-averaged observables agree with experiment, with fewer distance restraint violations than traditionally resolved NMR structures and with unmodeled alternate conformations uncovered in electron density. This methodology paves the way for experimentally aware predictive models that generate structural ensembles consistent with the measurements, potentially over multiple modalities, and that can be further refined toward thermodynamically grounded ensembles by incorporating energetics.

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

Building upon the fundamental discovery that amino acids in spatial proximity within a protein structure co-evolve[1,2], and training on the hundreds of thousands of experimentally determined X-ray crystal structures available today, deep-learning-based models such as AlphaFold[3,4] have enabled 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 generate protein structural ensembles that are consistent with 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 (for example, AlphaFold3) as sequence-conditioned structural priors and cast ensemble modeling as posterior inference of protein structures given experimental measurements. Recent studies have explored ways to incorporate experimental restraints into structure prediction.

[1]Institute of Science and Technology Austria (ISTA), Klosterneuburg, Austria. [2]Princeton University, Princeton, NJ, USA. [3]Eric and Wendy Schmidt Center, Broad Institute of MIT and Harvard, Cambridge, MA, USA. [4]Tel Hai – University of Kyriat Shmona in the Galilee, Upper Galilee, Israel. [5]MIGAL – Galilee Research Institute, Kiryat Shmona, Israel. [6]Technion – Israel Institute of Technology, Haifa, Israel. ✉e-mail: svedula@princeton.edu; paul.schanda@ist.ac.at; ailiem@migal.org.il; alexander.bronstein@ist.ac.at

https://doi.org/10.1038/s41587-026-03166-5

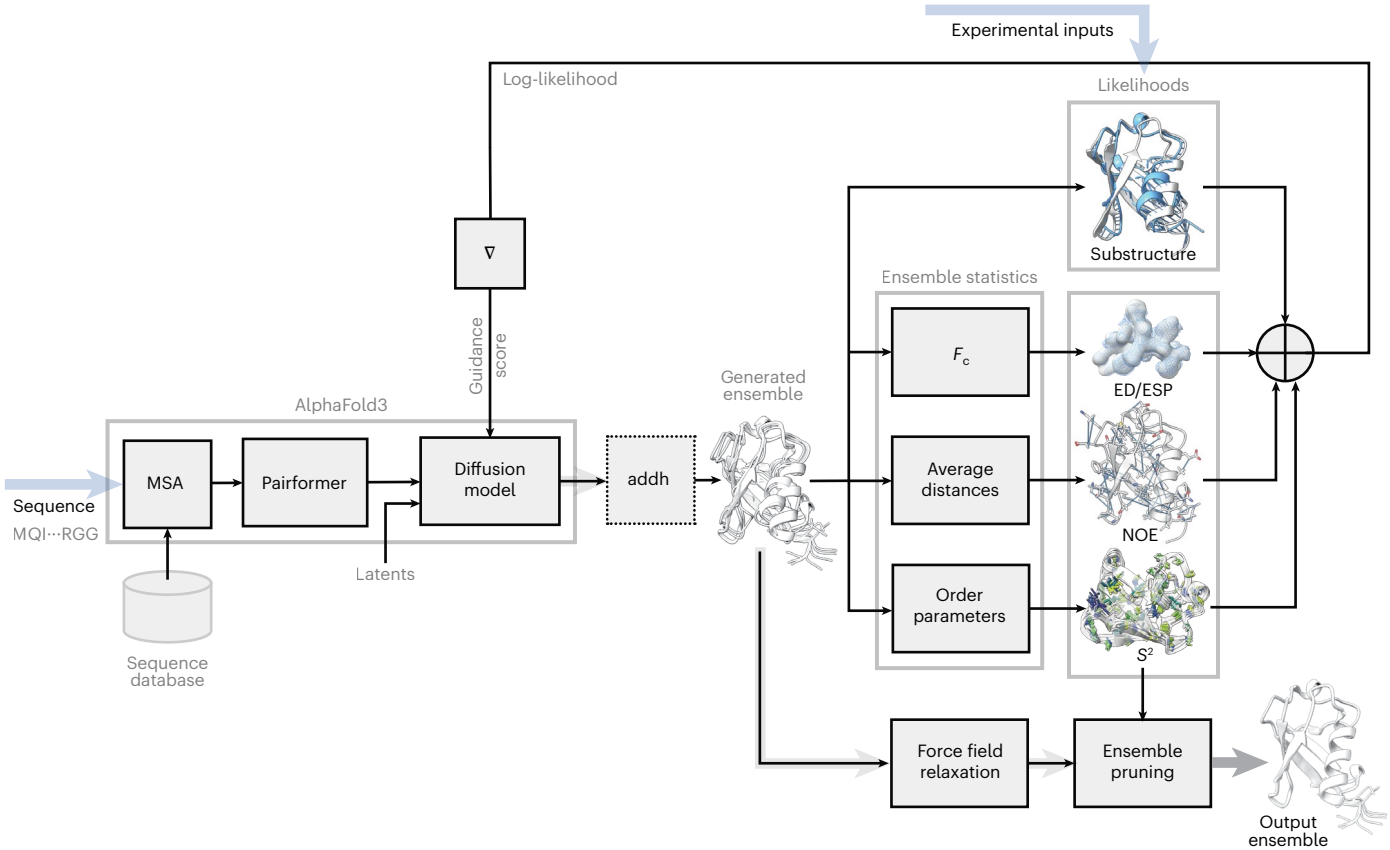



**Fig. 1 | Schematic depiction of the proposed method.** AlphaFold3 allows the sampling of protein structures given an amino acid sequence. To further condition the model by experimental observations, at each timestep of the AlphaFold3 diffusion model, an ensemble of structures is generated. Hydrogens are differentiably added to the heavy atoms (addh), and likelihoods of experimental observations are calculated given each individual ensemble member (for example, to enforce locations of certain atoms) and on ensemble averages (for example, calculated electron density, $F_c$, average interatomic 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. ED, electron density.

AlphaLink[5] integrates cross-linking mass spectrometry-derived distances as bias terms in AlphaFold's pair representation to guide single-structure inference. In the present study, we instead modify the generative process of AlphaFold3 itself: experimental likelihoods enter as guidance terms during sampling, enabling direct conditioning on any experimental modality to generate ensembles consistent with both sequence and experiment.

## 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. After 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 explains the experimental observations. See Fig. 1 for a schematic description of the proposed method and Sections 8.4 ('Protein structure inverse problems'), 8.5 ('X-ray crystallography models'), 8.6 ('NMR models') and 8.7 ('Cryo-EM models') for algorithmic and implementation details.

### Guiding AlphaFold3 with NMR distance restraints and order parameters efficiently generates experiment-consistent ensembles

NMR structure determination is largely based on the detection of interatomic proximities, which is a manifestation of the dipolar coupling between spins and detected in nuclear Overhauser effect (NOE) experiments in solution-state NMR[6]. Typically, the structure is then determined by applying these distance restraints in molecular dynamics simulations to obtain the set of lowest energy conformations[7,8]. 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[9,10], but it often requires highly accurate ('exact') NOE distances and is not yet common practice in NMR. Ensemble-based molecular dynamics approaches simulating multiple conformers simultaneously to satisfy experimental restraints remain computationally expensive, requiring days even for small systems like the 76-residue ubiquitin[11,12].

We use ubiquitin (Protein Bata Bank (PDB) 1D3Z; Biological Magnetic Resonance Bank (BMRB) 6457 and 5387) as an NMR structure determination benchmark. Guiding AlphaFold3 with NOE distance restraints generates ensembles with substantially fewer pairwise distance constraint violations. Figure 2a shows the cumulative distribution of violation magnitudes; a spatial visualization of the violated constraints is depicted in Fig. 2c. From Fig. 2b, it is evident that the

https://doi.org/10.1038/s41587-026-03166-5

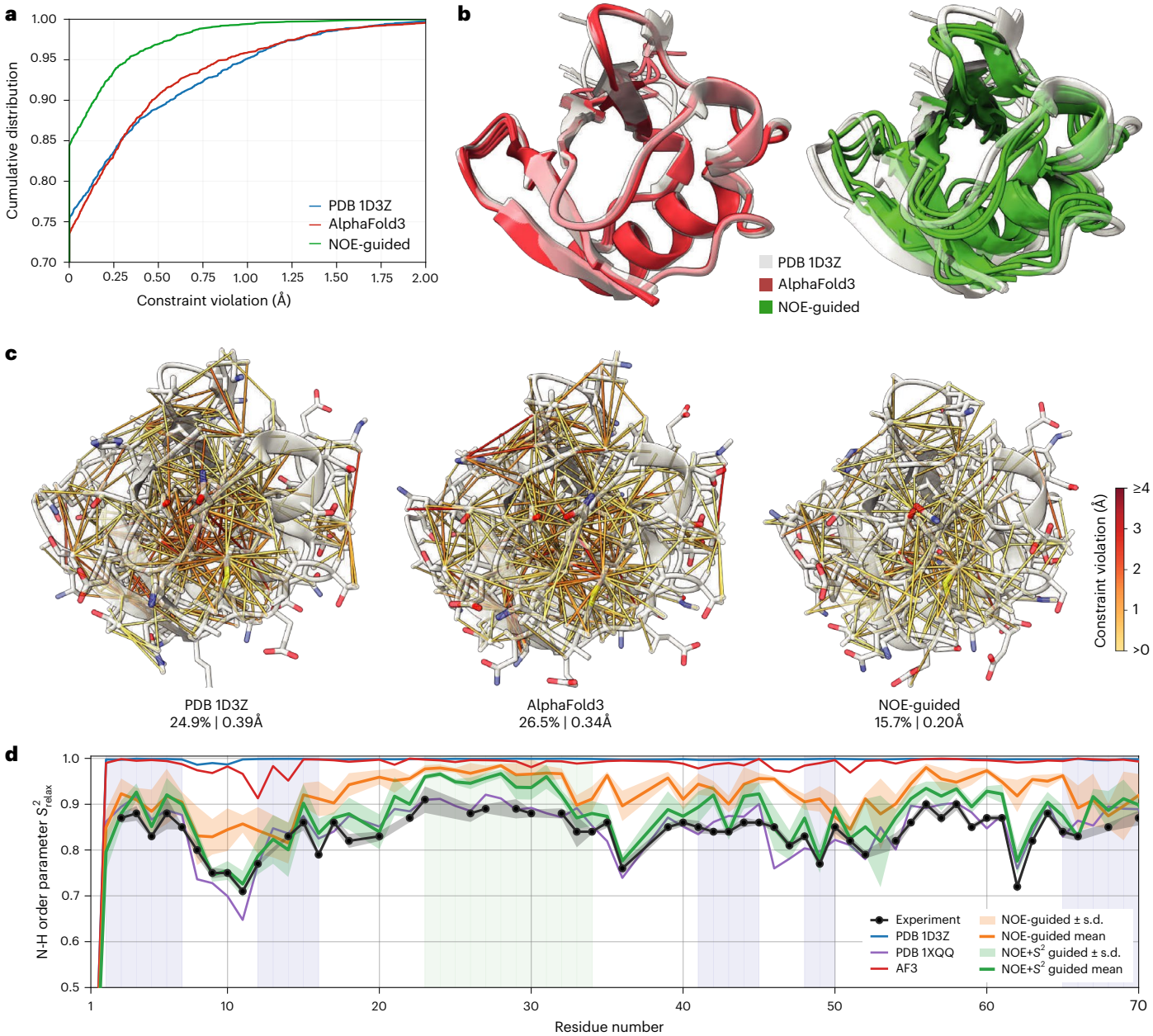



**Fig. 2 | NMR-guided AlphaFold3.** Guidance of ubiquitin structure by NOE pairwise distance bounds enforced as ensemble statistics produces substantially fewer constraint violations than the PDB structure ensemble 1D3Z. **a**, Cumulative distribution of constraint violations. **b**, Unguided (red) and 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 1,212) 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 ref. 14 (black) versus order parameter calculated from the ensembles of different methods. NOE-guided and NOE+$S^2$-guided curves are shown as the mean over five independent runs, with shaded bands denoting the standard deviation across runs. 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 similar to the PDB ensemble 1XQQ (purple) fitted to match structure and dynamics ($r = 0.87$, $q = 0.04$). For full results with relaxation and RDC data, see Supplementary Figs. 3 and 4. AF3, AlphaFold3.

ensemble produced by NOE-guided AlphaFold3 is more heterogeneous. Supplementary Figs. 1 and 2 evaluate 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 GPU 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 timescales shorter than a few nanoseconds[13]. We computed ensemble order parameters and compared them to experimental measurements reported in ref. 14. 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 Supplementary Fig. 3 shows that these ensembles exhibit only moderate agreement with the experimental $S^2$. By contrast, NOE-guided AlphaFold3 substantially improves agreement, better replicating the dynamic behavior observed in both flexible and structured regions.

To better match the N-H order parameter experimental data, we added an $S^2$ guidance term to the NOE-guided AlphaFold3. This

https://doi.org/10.1038/s41587-026-03166-5

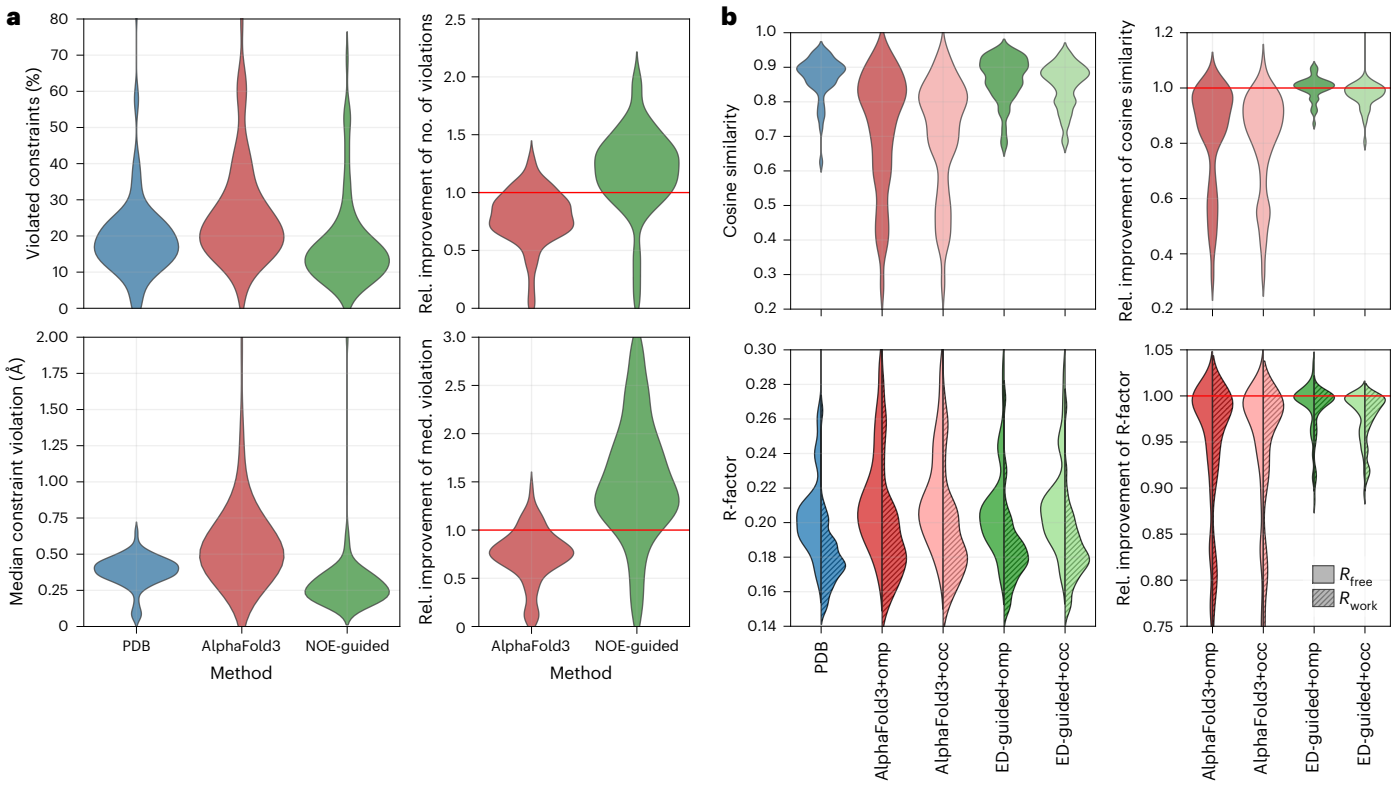



**Fig. 3 | Summary statistics on evaluated NMR and X-ray crystallographic benchmarks. a**, Distribution of percent of violated distance constraints (top left) and median violation (bottom left) for PDB, unguided and NOE-guided AlphaFold3 and the relative improvement of these metrics compared to the deposited 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. Additional summary statistics are shown in Supplementary Figs. 10 and 11.

resulted in a substantially higher agreement with the experimental $S^2$ measurements and similar to the result of computationally expensive NMR-guided ensemble molecular dynamics simulations (PDB 1XQQ from ref. 11) at a small fraction of the computational cost (Fig. 2d and Supplementary Fig. 3a–c). For the case of ubiquitin, numerous additional experimental data have been collected, including methyl axis order parameters from spin relaxation (reporting on ps–ns motions[15]) and order parameters from residual dipolar couplings (RDCs) for amide and methyl sites, which report on ps–ms motions[12,16,17]. Supplementary Figs. 3d–f and 4 show the result of guidance with these data. In all cases, guidance with NOE and $S^2$ terms shows high agreement with the experiment at no or little expense of pairwise distance constraint violation.

We further evaluated our method on two benchmarks: eight peptides shown to be mispredicted by AlphaFold3 (refs. 18,19) and a subset of 83 proteins from the recently compiled 100-protein NMR spectra database[20]. Supplementary Table 1 details quantitative performance indicators for individual structures; Fig. 3a 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 in only 15 cases (17%). NOE-guided AlphaFold3 outperforms its unguided counterpart in all cases. Additional comparison to inductive ensemble generative models, including AlphaFlow and ESMFlow[21], BioEmu[22] and AFCluster[23], is presented in Supplementary Table 6.

In all the reported cases, we used uniformly weighed ensembles. Supplementary Fig. 11 presents an additional evaluation when a force field predicted energy is used to weigh the different conformations either as a postprocessing step or by fully integrating the force field into the guidance term and producing ensembles that combine experimental guidance with force field energetic information, thereby moving the sampled distribution closer to a thermodynamically meaningful one. The latter weighing technique improves the distance constraint satisfaction in 78 cases (about 87%) with the median improvement of about 20%. For methodological details, see Section 8.13 ('Energy-reweighted guidance').

Figure 4 depicts examples of ensembles generated by NOE-guided AlphaFold3. We observe that AlphaFold3 tends to overpredict order (for example, 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, whereas the ensemble average guidance term 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 reduces the number of violations.

Finally, as an independent source of validation, we used the Accuracy of NMR Structures Using RCI and Rigidity (ANSURR) metrics introduced in ref. 24 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 root mean square deviation (RMSD) score in 92% of the cases and the correlation score in 54% of the cases when compared to the corresponding PDB ensembles. See Supplementary Fig. 5 for more information.

## Guiding AlphaFold3 with electron density captures previously unmodeled residues in X-ray crystallography data

AlphaFold3 is oblivious to environmental influences that may be present in experimental settings, such as interactions with ions, ligands and

https://doi.org/10.1038/s41587-026-03166-5

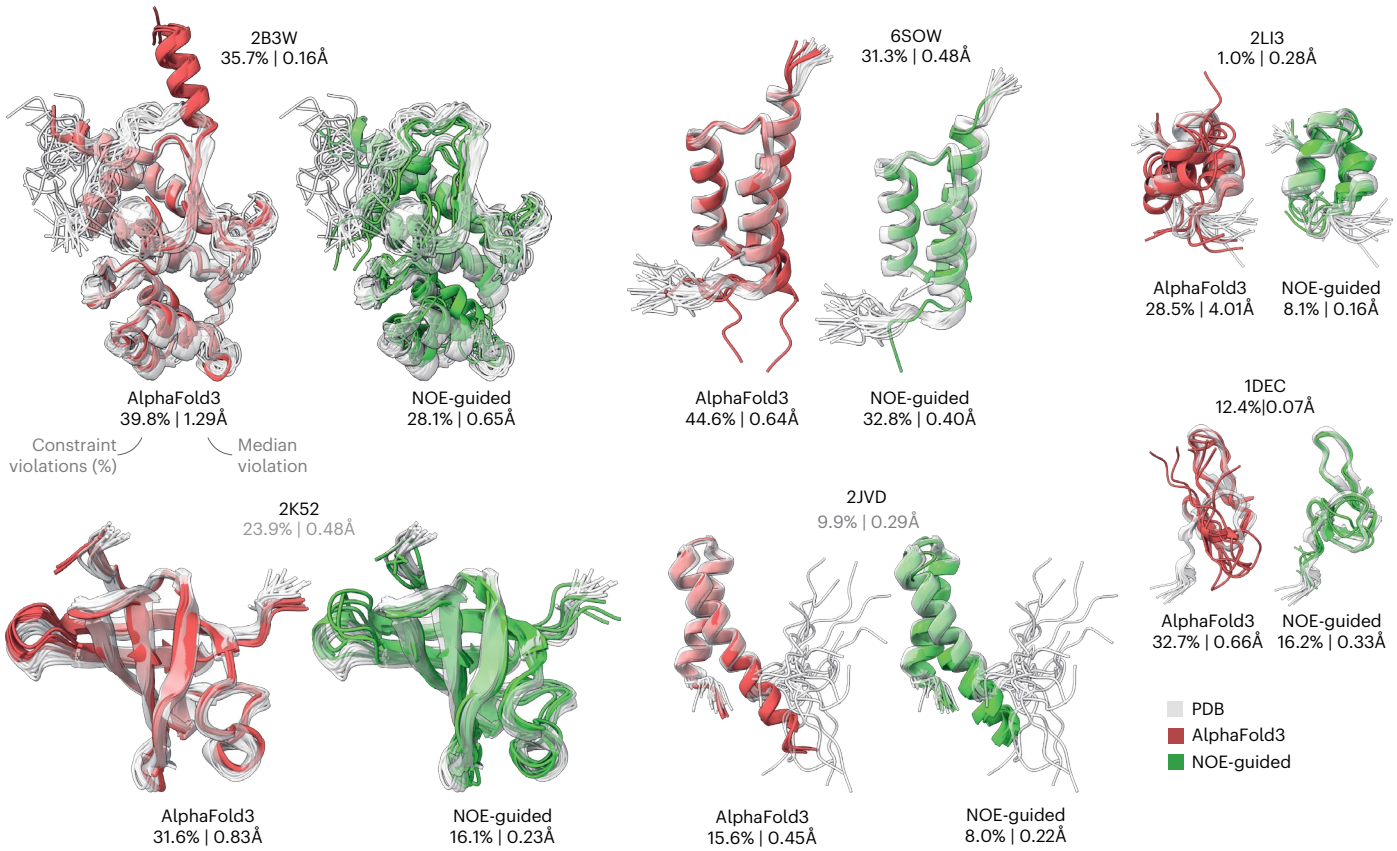



**Fig. 4 | Conformation ensembles generated for six proteins.** Structures obtained using unguided AlphaFold3 (red) and its NOE-guided version (green). Ensembles are overlaid on corresponding NMR structures solved from the same NOESY data shown in transparent white. PDB identifiers are indicated above each structure. The numbers below report the percentage of violated constraints and their median violation. ED, electron density.

other macromolecular partners or different crystal contacts, resulting in conformational variations. For example, HSP90α crystallizes 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 different conformation between the two chains, neither of which is accurately predicted by AlphaFold3. Figure 5a shows that the poor AlphaFold3 prediction is restored to experimental accuracy by electron density guidance. The method is similarly successful when applied across experiments. Supplementary 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 guidance 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 1.65 Å and PDB 9MZB at 2.8 Å) indicates the generalizability of our method. In another example, we consider a pair of myoglobin structures that differ in surface-exposed loop conformations due to altered crystal packing (PDB 1U7R and PDB 1U7S). Electron density-guided AlphaFold3 reproduces both conformations with density fits similar to deposited models. 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[25] and underscoring the biological relevance of training AlphaFold3 on ensembles detectable in crystal density, now an achievable task with our methodology.

Another region-specific phenomenon in X-ray crystal structures is unmodeled 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. 5b that density-guided AlphaFold3 has the potential to fill in missing regions to best explain the scarce density. The crystal structure of the 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. Figure 5b and Supplementary Fig. 7 show 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[26].

Another challenge in crystallography is modeling peptides. These short protein chains are often very poorly predicted by AlphaFold3 because they are highly adaptable to their binding partners. Figure 5c 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, because these ensembles were generated without multiple sequence alignment (MSA) conditioning. Additional examples in Supplementary Table 2 and the summary statistics in Fig. 3b highlight the method's ability to generalize across sequences of varying lengths. Comparison to ensemble generative models is presented in Supplementary Table 7. 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. 5d results from iterative improvements in real-space correlation (Supplementary Fig. 8). 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

https://doi.org/10.1038/s41587-026-03166-5

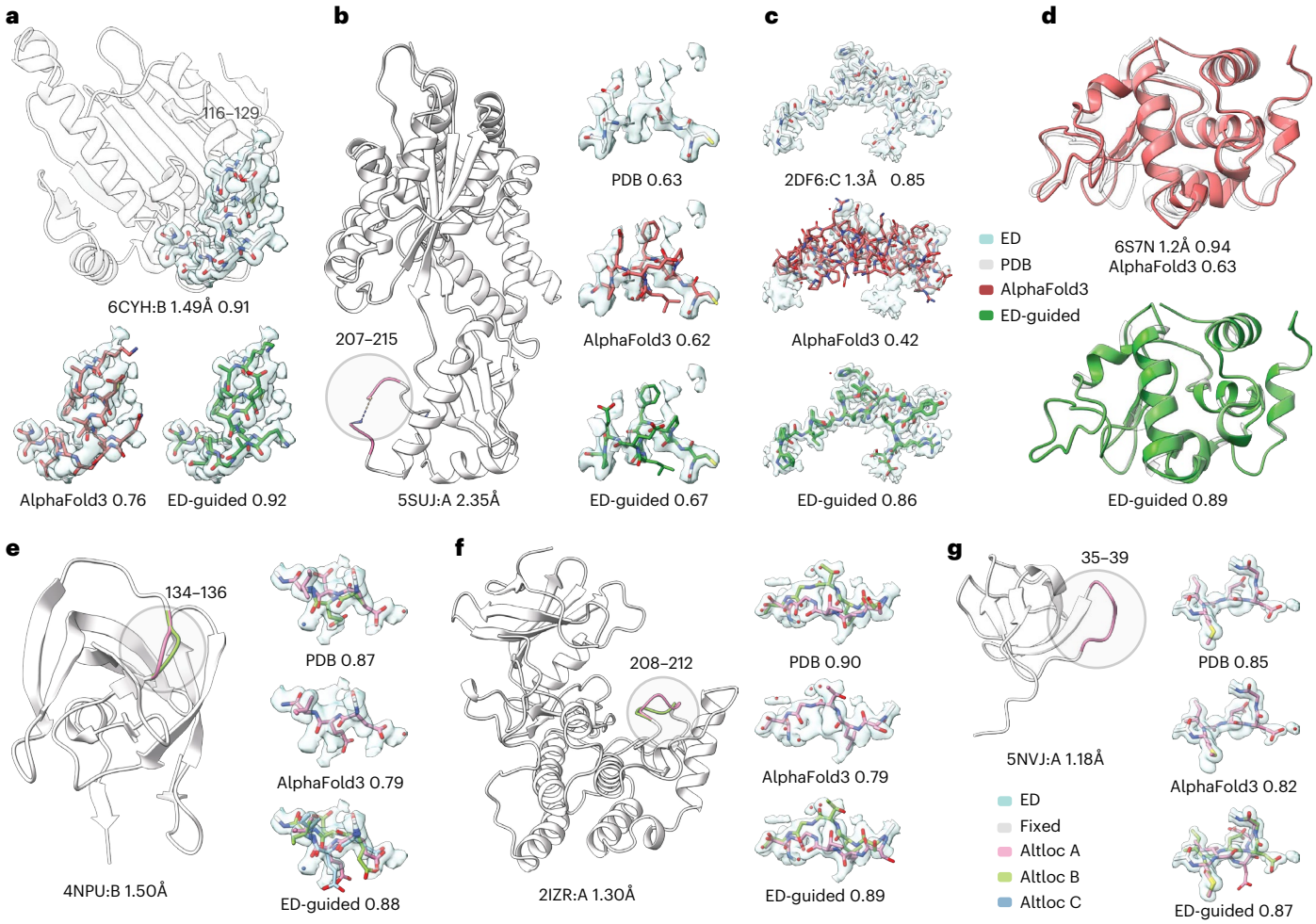



**Fig. 5 | X-ray electron density-guided AlphaFold produces structures and ensembles of equivalent or better accuracy to those deposited in the PDB and detects plausible conformations in unmodeled electron density. a**, 6CYH chain B is poorly predicted by AlphaFold3 in the region of amino acids 101–116. Guidance using the electron density maps restores the fit of the model. **b**, An unmodeled 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 $\beta$–PIX 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. Electron density guidance essentially reproduces the PDB structure. **e**,**f**, Electron density-guided AlphaFold3 produces conformational ensembles explaining the multimodal electron density on par with or better than the PDB model. **g**, Electron density-guided AlphaFold3 predicts unmodeled altlocs in 5NVJ, better explaining the electron density. Electron densities are visualized as 0.5 $e^{-}$ $Å^{-3}$ isosurfaces. Numbers below indicate cosine similarities between calculated and observed densities. ED, electron density; med., median; Rel., relative.

electron density map[27–30]. A recent study[31] compiled a comprehensive catalog of alternate locations (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[32]. Here, we demonstrate that electron density guidance on this same set of altlocs generates an ensemble that captures the backbone's bimodal nature, providing a better explanation of the density. This method is effective for different degrees of backbone separation spanning from 1 Å (Fig. 5e) to tens of angstroms (Fig. 5f). Supplementary 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, Supplementary 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 (Supplementary Fig. 9). Finally, we demonstrate the potential of this method for detecting previously unmodeled altloc conformations (Fig. 5g).

## Guiding AlphaFold3 with cryo-EM maps

Cryo-EM imaging allows to capture the conformations of intricate protein complexes in the form of electrostatic potential (ESP) maps. Because 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.

Figure 6a shows the reconstruction of insulin receptor IR-B in its symmetric apo state (PDB 8U4B; Electron Microscopy Data Bank (EMDB) 41877) and asymmetric conformation (PDB 8U4E; EMDB 41880) with three bound IGF2 factors. AlphaFold3 tends to oversymmetrize 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.

Figure 6b shows the reconstruction of an amyloid-β fibril (PDB 9FH1; EMDB 50436) formed by a nearly planar homodimer that further

https://doi.org/10.1038/s41587-026-03166-5

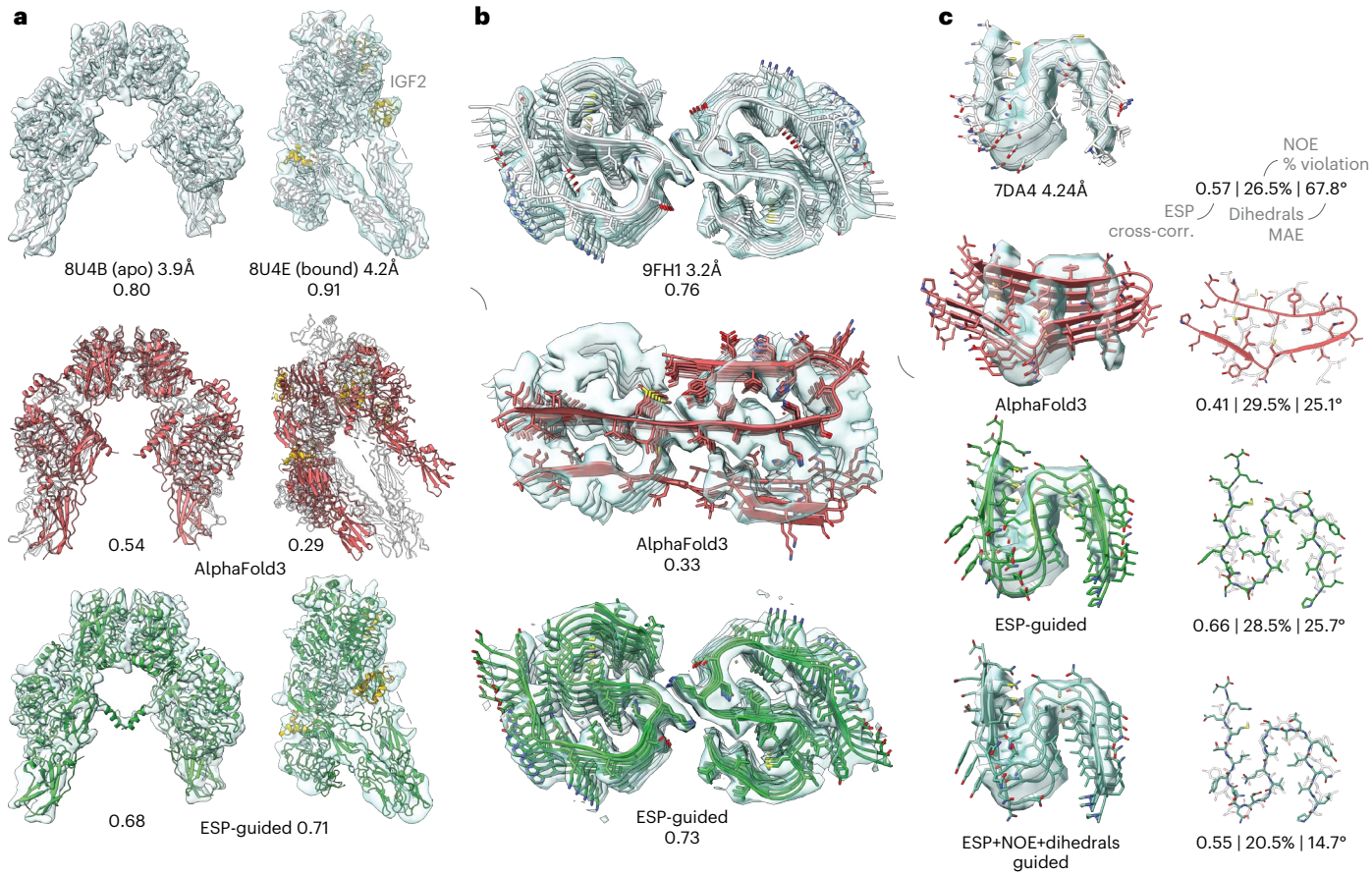



**Fig. 6 | Multimeric structure reconstruction using cryo-EM ESP map guidance.** **a**, Reconstruction of insulin receptor IR-B in its symmetric apo state (left) and asymmetric conformation (right) with three bound IGF2 factors (yellow). **b**, Reconstruction of an amyloid-β fibril. AlphaFold3 (red) 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 $\phi$ and $\psi$ (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. cross-corr., cross-correlation.

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.

### Guiding AlphaFold3 jointly with cryo-EM and NMR data

In many cases, structural biology approaches comprise data from several techniques. Our approach can integrate all the above sources of data and more. We exemplify such a case here for the combination of cryo-EM and NMR data, applied to amyloid fibrils. Figure 6c shows the reconstruction of a RIPK3 human amyloid fibril (PDB 7DA4; EMDB 30622) for which solid-state 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 improve the local structure accuracy at a reasonable tradeoff of density fit (see Supplementary Table 5 for quantitative analysis).

## 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. These ensembles should be interpreted as posterior structural hypotheses constrained by the supplied measurements and the model prior, not as fully calibrated Boltzmann equilibrium ensembles. In this sense, the method addresses a practically important inference problem in structural biology: generating compact sets of conformations whose ensemble-averaged observables explain heterogeneous experimental data. In practice, composable likelihoods from heterogeneous modalities—interatomic distances, dihedral angle order parameters from NMR, electron density from X-ray crystallography and ESP maps from cryo-EM—are injected as gradients during AlphaFold3's reverse-diffusion trajectory and followed by relaxation and sparsifying ensemble selection. 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 (for X-ray data) or ensembles of conformers from NMR, which, as we show here, do not reflect dynamic amplitudes faithfully.

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 reuses 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 (for example, MSA features) will determine how far inductive models can progress from predicting static structure to predicting relative populations of states. The comparison to existing inductive ensemble generators such as BioEmu and AlphaFlow suggests that, in the absence of experimental evidence, these models largely fail to produce ensembles in agreement with the experiment.

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 (for example, 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 reanalysis 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 AlphaFold3 prior; they are not calibrated thermodynamic populations. Agreement with a finite set of observables does not uniquely determine the underlying conformational distribution, and distinct ensembles may explain the same data comparably well. Accordingly, mixture weights should not be overinterpreted as equilibrium populations. Notably, the improvement observed when force field energy is incorporated indicates that energetic information contributes complementary constraints and can move the inferred ensembles closer to thermodynamically meaningful ones. A key direction for future work is, therefore, to integrate energetic priors and thermodynamic observables more tightly, for example, through maximum-entropy or evidence-based population estimation.

Guidance strength and gradient handling are important guardrails to avert models being pulled into noise, especially in low-resolution electron microscopy or poorly phased crystallographic maps. Although, 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, post-translational modifications, 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 nuclear Overhauser effect spectroscopy (NOESY) spectra and relaxation data without days of molecular dynamics 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.

## Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41587-026-03166-5.

## References


1. Göbel, U., Sander, C., Schneider, R. & Valencia, A. Correlated mutations and residue contacts in proteins. *Proteins* **18**, 309–317 (1994).
2. Hopf, T. A. et al. Sequence co-evolution gives 3D contacts and structures of protein complexes. *eLife* **3**, e03430 (2014).
3. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
4. Abramson, J. et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024).
5. Stahl, K., Graziadei, A., Dau, T., Brock, O. & Rappsilber, J. Protein structure prediction with in-cell photo-crosslinking mass spectrometry and deep learning. *Nat. Biotechnol.* **41**, 1810–1819 (2023).
6. Shcherbakov, A. A., Medeiros-Silva, J., Tran, N., Gelenter, M. D. & Hong, M. From angstroms to nanometers: measuring interatomic distances by solid-state NMR. *Chem. Rev.* **122**, 9848–9879 (2021).
7. Schwieters, C. D., Kuszewski, J. J. & Clore, G. M. Using XPLOR–NIH for NMR molecular structure determination. *Prog. Nucl. Magn. Reson. Spectrosc.* **48**, 47–62 (2006).
8. Güntert, P. Automated NMR structure calculation with CYANA. *Methods Mol. Biol.* **338**, 353–378 (2004).
9. Vögeli, B., Olsson, S., Güntert, P. & Riek, R. The exact NOE as an alternative in ensemble structure determination. *Biophys. J.* **110**, 113–126 (2016).
10. Chi, C. N. et al. A structural ensemble for the enzyme cyclophilin reveals an orchestrated mode of action at atomic resolution. *Angew. Chem. Int. Ed. Engl.* **54**, 11657–11661 (2015).
11. Lindorff-Larsen, K., Best, R. B., DePristo, M. A., Dobson, C. M. & Vendruscolo, M. Simultaneous determination of protein structure and dynamics. *Nature* **433**, 128–132 (2005).
12. Lange, O. F. et al. Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. *Science* **320**, 1471–1475 (2008).
13. Palmer III, A. G. NMR characterization of the dynamics of biomacromolecules. *Chem. Rev.* **104**, 3623–3640 (2004).
14. Lienin, S., Bremi, T., Brutscher, B., Brüschweiler, R. & Ernst, R. Anisotropic intramolecular backbone dynamics of ubiquitin characterized by NMR relaxation and MD computer simulation. *J. Am. Chem. Soc.* **120**, 9870–9879 (1998).
15. Liao, X., Long, D., Li, D.-W., Brüschweiler, R. & Tugarinov, V. Probing side-chain dynamics in proteins by the measurement of nine deuterium relaxation rates per methyl group. *J. Phys. Chem. B* **116**, 606–620 (2012).
16. Markwick, P. R. et al. Toward a unified representation of protein structural dynamics in solution. *J. Am. Chem. Soc.* **131**, 16968–16975 (2009).
17. Sabo, T. M. et al. ORIUM: optimized RDC-based iterative and unified model-free analysis. *J. Biomol. NMR* **58**, 287–301 (2014).
18. Vedula, S., Bronstein, A. M. & Marx, A. Improving prediction accuracy in chimeric proteins with windowed multiple sequence alignment. *Comput. Struct. Biotechnol. J.* **27**, 3292–3298 (2025).
19. McDonald, E. F., Jones, T., Plate, L., Meiler, J. & Gulsevin, A. Benchmarking AlphaFold2 on peptide structure prediction. *Structure* **31**, 111–119 (2023).
20. Klukowski, P. et al. The 100-protein NMR spectra dataset: a resource for biomolecular NMR data analysis. *Sci. Data* **11**, 30 (2024).

21. Jing, B., Berger, B. & Jaakkola, T. AlphaFold meets flow matching for generating protein ensembles. *ICML'24: Proceedings of the 41st International Conference on Machine Learning* (eds Salakhutdinov, R. et al.) 22277–22303 (JMLR.org, 2024).
22. Lewis, S. et al. Scalable emulation of protein equilibrium ensembles with generative deep learning. *Science* **389**, eadv9817 (2025).
23. Wayment-Steele, H. K. et al. Predicting multiple conformations via sequence clustering and AlphaFold2. *Nature* **625**, 832–839 (2024).
24. Fowler, N. J., Sljoka, A. & Williamson, M. P. A method for validating the accuracy of NMR protein structures. *Nat. Commun.* **11**, 6321 (2020).
25. Kondrashov, D. A., Zhang, W., Aranda IV, R., Stec, B. & Phillips Jr, G. N. Sampling of the native conformational ensemble of myoglobin via structures in different crystalline environments. *Proteins* **70**, 353–362 (2008).
26. Eastman, P. et al. OpenMM 7: rapid development of high performance algorithms for molecular dynamics. *PLoS Comput. Biol.* **13**, e1005659 (2017).
27. Smith, J. L., Hendrickson, W. A., Honzatko, R. B. & Sheriff, S. Structural heterogeneity in protein crystals. *Biochemistry* **25**, 5018–5027 (1986).
28. Furnham, N., Blundell, T. L., DePristo, M. A. & Terwilliger, T. C. Is one solution good enough? *Nat. Struct. Mol. Biol.* **13**, 184–185 (2006).
29. Van Den Bedem, H. & Fraser, J. S. Integrative, dynamic structural biology at atomic resolution—it's about time. *Nat. Methods* **12**, 307–318 (2015).
30. Wankowicz, S. A. et al. Automated multiconformer model building for X-ray crystallography and cryo-EM. *eLife* **12**, RP90606 (2024).
31. Rosenberg, A. A., Marx, A. & Bronstein, A. M. A dataset of alternately located segments in protein crystal structures. *Sci. Data* **11**, 783 (2024).
32. Rosenberg, A. A., Vedula, S., Bronstein, A. M. & Marx, A. Seeing double: molecular dynamics simulations reveal the stability of certain alternate protein conformations in crystal structures. Preprint at *bioRxiv* https://doi.org/10.1101/2024.08.31.610605 (2024).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2026

# Methods

## Models, resources and packages

For all experiments, we used Protenix[33], an open-source PyTorch reimplementation of AlphaFold3. For the AlphaFold3 baseline mentioned across experiments, we used the official AlphaFold3 weights and source code[4].

To implement the corresponding forward models, we used PyTorch (2.3.1 + cu121)[34]. However, to efficiently calculate the density for larger structures, we used PyKeOps[35] and PyTorch. All computations were performed on NVIDIA H100 and L40 GPUs running on Debian GNU/Linux 12 distribution.

Across all experiments, we retrieved the MSAs using a wrapper around the ColabFold MMseqs2 API[36]. The wrapper submits the query sequences to the remote MMseqs2 server via HTTP POST, polls for job completion and downloads and extracts the results. The output is provided in A3M format, which is compatible with AlphaFold3.

## Input data

**X-ray crystallography.** *Atomic models*. In the X-ray crystallography workflow, the input consists of a PDB ID, a chain identifier and an amino acid subsequence in the chain. For this modality, we restrict our modeling to single protein chains. The corresponding PDB structure and MTZ file are retrieved from PDB-Redo[37], and the specific chain is identified using Gemmi (version 0.6.5)[38]. From this chain, we extract only the standard amino acid residues explicitly modeled in the PDB file, while excluding structural waters, hydrogen atoms and all other non-standard amino acids, and save the resulting coordinates as a separate PDB file. In addition, we mutate selenium methionine to methionine by replacing selenium atoms with sulfur. We also mutate S-hydroxycysteine to cysteine by removing the oxygen from the hydroxyl group. If altlocs are present[32], the PDB is split into the corresponding altloc files. Because AlphaFold3 uses 1-based residue indexing, we renumber the residues in each split PDB to start from 1. In many cases, the retrieved PDB files can contain incomplete or unmodeled residues due to poor quality density. To address this, we use PDBFixer[39] (version 1.9.0), which employs standard residue templates from OpenMM[26] force fields, to model the missing atoms in all split PDBs. To avoid steric clashes after imputing the atoms with PDBFixer, we relax the structure using AMBER14 force field[40] relaxation. The isotropic B-factor for these imputed atoms is set to 100.00. For AlphaFold3, the input sequence is limited to the residues explicitly modeled in the reference chain.

To construct the atom mask for applying the substructure conditioner, we use the provided amino acid subsequence to identify residues that will primarily be optimized with the density-based loss versus those optimized using the substructure conditioner.

*Electron density maps*. Our proposed method generalizes across different types of real-space electron density maps. Here, we considered three map types.

- $2F_o - F_c$ maps: These maps are rendered using the PHENIX[41] `fft` command:

```
fft hklin mtz_path mapout map_output_path
LABIN F1= FWT PHI= PHWT SIG1= SIGFP VF000 volume f000
```

The unit cell volume (volume) and the total scattering factor at zero scattering angle F(0, 0, 0) (f000) are derived from the input atomic model. For a triclinic cell with edge lengths a, b, c (in Å) and angles α, β, γ (in radians), the unit cell volume is:

$$\texttt{volume} = a \cdot b \cdot c \cdot \sqrt{(1 - \cos^2\alpha - \cos^2\beta - \cos^2\gamma + 2\cos\alpha\cos\beta\cos\gamma)}$$

Unit cell parameters $(a, b, c, \alpha, \beta, \gamma)$ are parsed from the input atomic model. $F(0, 0, 0)$ is obtained using `phenix.f000`, which estimates the zero-angle scattering factor by summing electron counts from both the atomic model and the bulk solvent. The bulk solvent contribution is calculated using the default average solvent density of 0.35 $e^- Å^{-3}$. The atomic and solvent electron contributions are then combined to yield $F(0, 0, 0)$.

- Absolute-scale END maps: These density maps are generated using the END map workflow of ref. 42, which combines PHENIX and CCP4 programs to produce absolute-scale electron density maps ($e^- Å^{-3}$ units). We run

```
phenix.refine atomic_model_file mtz_file strategy= rigid_body
./END_RAPID.com pdb_refine_001.eff –norapid
```

The atomic model first undergoes rigid-body refinement to prevent artifacts from misaligned models. The END script then rescales the observed structure factors against the calculated ones and offsets the density by adding $F(0, 0, 0)$ to every voxel. This places the map on an absolute scale, where zero corresponds to vacuum.

- Composite omit maps: These density maps are computed in PHENIX using

```
phenix.composite_omit_map atomic_model.pdb reflection_data.mtz
```

These maps reduce model bias while minimizing phase degradation.

To align the real-space density maps with the corresponding atomic models from Section 8.2.1 ('X-ray crystallography'), we used `phenix.map_box` with `selection_radius = 10` to carve the map around the corresponding atomic model with sufficient padding.

Because our X-ray density guidance operates locally, we carved the input density map to include voxels within a radius of 5 Å of all atoms in the residue range (and chain) of interest from the reference PDBs. This preprocessing reduces VRAM memory consumption. During $F_c$ computation, the ensemble density is calculated at voxels within this 5-Å local neighborhood.

**NMR.** We used three sources of NMR observables for guidance: NOE-implied distance restraints, chemical-shift-implied backbone dihedral angles and per-residue order parameters.

*NOE-implied distance restraints*. NOE restraint lists were parsed from NMR-STAR files (via `pyNMR-STAR`), selecting only NOE-derived distance restraints; other restraint types such as dihedral angles, hydrogen bonds and RDC were excluded. 'Ambiguous NOEs'—cross-peaks whose atom pair assignment is uncertain—were retained and represented as disjunctive (`OR`) constraints: each ambiguity group defines multiple candidate atom pairs sharing the same bounds, and the restraint is satisfied if any one candidate meets those bounds (not necessarily all). Lower/upper bounds were read directly from the file; missing lower bounds were set to 0.0 Å. We note that, although NOE physics is intensity based, we adopt the common semiquantitative distance-bound representation here, as is standard practice in molecular dynamics-resolved NMR ensemble workflows; a future extension of our pipeline may incorporate intensity averaging within ambiguity groups prior to distance conversion. For each system, the corresponding NMR-STAR restraint file was retrieved from the PDB entry on the RCSB website.

*Chemical-shift-implied dihedral angles*. Backbone $\Phi^{tgt}/\Psi^{tgt}$ angles used in combined cryo-EM and NMR guidance were inferred from raw chemical shifts using TALOS-N[43], a tool that infers potential ranges of backbone dihedral angles from chemical shifts. Per target, the chemical shift lists were retrieved from the BMRB entry cross-referenced to the corresponding PDB ID.

*Order parameters*. Order parameter ($S^2$) values were taken on a per-residue basis from the respective publications (or associated

supplementary datasets) for each target. Across all observables, when required values or bounds were unavailable for a residue or atom pair, we masked that item and did not guide on it.

**Cryo-EM.** In the cryo-EM workflow, the inputs include a reference atomic model, amino acid sequences and the corresponding ESP map from the EMDB[44].

*Atomic models*. In cryo-EM, the objective is not to refine high-resolution reference structures, as in X-ray crystallography, but to fit amino acid sequences into experimentally resolved volumes. This distinction is because of the lower resolution of cryo-EM ESP maps, which makes them useful for studying larger protein structures. Here, reference PDB structures from RCSB[45] are primarily used to align chain labels, construct atom masks and serialize outputs. Unlike crystallographic preprocessing pipelines, missing residues or atoms are not imputed with modeling software (such as PDBFixer), and alternate conformations are not split into separate models. To ensure consistency with AlphaFold3, all residues are renumbered to be 1-indexed. Finally, because cryo-EM often targets large complexes, multimeric inputs along with multiple sequences can be provided. Relevant residue ranges and chain identifiers are specified alongside the sequences before guidance.

*ESP maps*. The input ESP volumes are obtained from the EMDB[44] in '.map' or '.mrc' format. To reduce VRAM usage without losing resolvable information, ESP maps with large side length $D$ are downsampled in Fourier space to a smaller side length $D' < D$. The reported resolution $r$ (in Å) is used to verify that the Nyquist frequency of the downsampled grid remains sufficient. This is checked using

$$D' \geq \frac{2D}{rd}$$

where $d$ is the original voxel size in Å. This relation follows from (1) conservation of the total physical side length $D \cdot d = D' \cdot d' \Rightarrow D' = \frac{Dd}{d'}$ and (2) the Nyquist criterion $r'_{\text{Nyquist}} = 2d' \leq r \Rightarrow \frac{1}{d'} \geq \frac{2}{r}$.

Lastly, atomic masks are generated using the reference protein to provide controlled guidance but are constructed in a way that maintains the generality of the approach:

- Global mask: A global mask is defined to include all voxels within $r_{\text{max}} = 2.1$ Å of the atoms in the reference protein. Empty regions of the masks are filled using convex hull implemented in `scipy.spatial.ConvexHull`[46].
- Per-chain mask: For multimeric structures (for example, amyloids), masks are generated per chain using a similar $r_{\text{max}}$ distance threshold. Because this is restricted to well-separable chains, user-defined masks could likewise be imported after manual generation. This is later used to obtain better-defined guidance directions.

## Force field

To map experimentally faithful ensembles onto Boltzmann-distributed populations, we evaluate per-structure energies $E(\mathbf{X})$ using ProteinEBM[47], a learned energy-based model derived from Boltz-2 (ref. 48). ProteinEBM is trained to assign scalar energies to protein conformations at 300 K by learning a Boltzmann-consistent score function over the conformational space. Although ProteinEBM does not constitute a classical force field, it defines an implicit, differentiable energy functional over protein conformations that serves as an effective structural prior for ensemble reweighting.

## Protein structure inverse problems

**Notation.** The amino acid sequence of a protein is denoted by $\mathbf{a}$, and the corresponding three-dimensional Cartesian coordinates of all atoms are denoted by $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_m)$, where $\mathbf{x}_i \in \mathbb{R}^3$ is the position of the $i$-th atom.

**Problem setting.** Given a protein's amino acid sequence $\mathbf{a}$ and its corresponding experimental observation $\mathbf{y}$ (from crystallography, NMR or cryo-EM), our goal is to sample an experimentally faithful ensemble of $n$ structures $\mathcal{X} = (\mathbf{X}^1, \ldots \mathbf{X}^n)$ from the posterior distribution $p(\mathcal{X}|\mathbf{a}, \mathbf{y})$.

The posterior distribution can be factorized as $p(\mathcal{X}|\mathbf{a}, \mathbf{y}) \propto p(\mathbf{y}|\mathcal{X}, \mathbf{a}) \cdot p(\mathcal{X}|\mathbf{a})$ using Bayes' Rule. The prior term is captured by $p(\mathcal{X}|\mathbf{a})$, which is the likelihood of an ensemble conditioned on the amino acid sequence $\mathbf{a}$. Using probabilistic laws, individual terms can be factorized into a product of independent samples in the prior term:

$$p(\mathcal{X} = (\mathbf{X}^1 \ldots \mathbf{X}^n)|\mathbf{a}) = \prod_{i=1}^{n} p(\mathbf{X}^i|\mathbf{a}) \tag{1}$$

In our case, the prior $p(\mathcal{X}|\mathbf{a})$ is implicitly modeled by AlphaFold3 (ref. 4).

The data term $p(\mathbf{y}|\mathcal{X}, \mathbf{a})$ represents the probability of the experimental data $\mathbf{y}$ given the protein structure ensemble $\mathcal{X}$ and the amino acid sequence $\mathbf{a}$. Because the data term is conditioned on the entire ensemble, we cannot separate the likelihood into separate structures. In essence, the data term models the underlying physics of the experiment. In this paper, we model the data term for three experimental modalities:

1. X-ray crystallography: where we use real-space electron density maps
2. NMR: where we use NOE pairwise distance restraints, amide (N-H) and methyl (C-C) order parameter obtained using relaxation (sensitive to ps–ns motion timescale) and RDC (sensitive to µs–ms motion timescale) and backbone dihedral angles $\phi$ and $\psi$ estimated from NMR chemical shifts using TALOS-N[43]
3. Cryo-EM: where we use ESP maps

Additionally, in all modalities we use the following terms:

1. 'Substructure' conditioning allowing to fix parts of the structure to follow an existing model
2. 'Validity' ensuring physically plausible bond lengths, valence angles and lack of steric intersections

In the sequel, we describe in detail the construction of these likelihoods.

*Guided diffusion and non-i.i.d. sampling*. AlphaFold3 generates structures using a diffusion-based generative model[49] over all-atom coordinates. In its original formulation, a single structure $\mathbf{X}$ is produced by integrating a variance-preserving stochastic differential equation (SDE)[50]:

$$d\mathbf{X} = -\left(\frac{1}{2}\mathbf{X} + \nabla_{\mathbf{X}} \log p_t(\mathbf{X}|\mathbf{a})\right)\beta_t\, dt + \sqrt{\beta_t}\,\mathbf{N} \tag{2}$$

where $\beta_t$ is the noise schedule; $\mathbf{N} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ is sampled from an isotropic normal distribution; and $\nabla_{\mathbf{X}} \log p_t(\mathbf{X}|\mathbf{a})$ is the learned score function modeled by AlphaFold3 (ref. 3). To predict a single structure, AlphaFold3 samples from the prior distribution $\mathbf{X}_T \sim \mathcal{N}(\mathbf{0}, \beta_0 \cdot \mathbf{I})$ and integrates the SDE in equation (2) from $t = T$ to $t = 0$ to iteratively denoise the diffusion variable $\mathbf{X}_T$ to a noiseless protein structure $\mathbf{X}_0$.

In our setting, we expand the current SDE to model ensembles instead of single structures. Formally, we sample an ensemble $\mathcal{X} = (\mathbf{X}^1, \mathbf{X}^2, \ldots, \mathbf{X}^n)$ of $n$ structures. Thus, we can generalize the SDE in equation (2) to sample ensembles instead of a single structure:

$$d\begin{bmatrix}\mathbf{X}^1\\ \vdots \\ \mathbf{X}^n\end{bmatrix} = -\left(\frac{1}{2}\begin{bmatrix}\mathbf{X}^1\\ \vdots \\ \mathbf{X}^n\end{bmatrix} + \begin{bmatrix}\nabla_{\mathbf{X}^1} \log p_t(\mathbf{X}^1|\mathbf{a})\\ \vdots \\ \nabla_{\mathbf{X}^n} \log p_t(\mathbf{X}^n|\mathbf{a})\end{bmatrix}\right)\beta dt + \sqrt{\beta_t}\begin{bmatrix}\mathbf{N}^1\\ \vdots \\ \mathbf{N}^n\end{bmatrix} \tag{3}$$

where $\mathbf{N}^k \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and the unconditional score term $\nabla_{\mathbf{X}^k} \log p_t(\mathbf{X}^k|\mathbf{a})$ can be separated into independent structures like in equation (1). To sample a non-i.i.d. ensemble from the posterior distribution, we plug in the guidance score to equation (3), obtaining:

$$d\begin{bmatrix}\mathbf{X}^1\\ \vdots\\ \mathbf{X}^n\end{bmatrix} = -\left(\frac{1}{2}\begin{bmatrix}\mathbf{X}^1\\ \vdots\\ \mathbf{X}^n\end{bmatrix} + \begin{bmatrix}\nabla_{\mathbf{X}^1}\log p_t(\mathbf{X}^1|\mathbf{a})\\ \vdots\\ \nabla_{\mathbf{X}^n}\log p_t(\mathbf{X}^n|\mathbf{a})\end{bmatrix}\right)\beta dt - \boldsymbol{\eta}\nabla_{\mathscr{X}}\log p\,(\mathbf{y}|\mathscr{X},\mathbf{a})\,\beta dt + \sqrt{\beta_t}\begin{bmatrix}\mathbf{N}^1\\ \vdots\\ \mathbf{N}^n\end{bmatrix} \tag{4}$$

Unlike the unconditional score term, the guidance score term equation (4) is not separable because it is conditioned on ensemble $\mathscr{X}$. Also, the hyperparameter $\boldsymbol{\eta}$ can be used to scale the strength of the guidance score and direct the flow toward higher posterior likelihood regions. The pseudocode for guided AlphaFold3 is provided in Supplementary Algorithm 1.

**Parameter settings.** For X-ray crystallography, we consistently set $\boldsymbol{\eta} = 0.1$. For NOE constraints under NMR, we experimented with values $\boldsymbol{\eta} \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$ and selected the best-performing result. For order parameters under NMR, we fixed $\boldsymbol{\eta} = 0.3$. To ensure numerical stability during the diffusion process, we apply gradient clipping to clip the guidance score. For cryo-EM, the used values were $\eta \in 0.1, 0.15, 0.2$ where the best values per structures were chosen.

## X-ray crystallography models

Let $\mathscr{X} = \{\mathbf{X}^k\}_{k=1}^n$ be an ensemble of $n$ conformers and conformer $\mathbf{X}^k$ have atomic coordinates $\{\mathbf{x}_i^k\}$ and isotropic B-factors $B_i^k$. The observed electron density is denoted as $F_o : \mathbb{R}^3 \to \mathbb{R}$. To restrict map comparisons around a region of interest, we use a spatial mask.

**Forward model.** To compute the crystallographic likelihood term, we first calculate the theoretical electron density map $F_c : \mathbb{R}^3 \to \mathbb{R}$ for all structures in ensemble $\mathscr{X}$. $F_c$ can be calculated at any spatial coordinate location $\boldsymbol{\xi} \in \mathbb{R}^3$. For a single conformer $\mathbf{X}^k = (\mathbf{x}_1^k \ldots \mathbf{x}_m^k)$, $F_c$ is a sum over kernel density estimates built from six spherical Gaussians centered at the symmetry operation[51] of every atom:

$$F_c(\boldsymbol{\xi}; \mathbf{X}^k, \mathbf{a}) = \sum_{q=1}^{N_s}\sum_{i=1}^{m}\sum_{j=1}^{6} a_{i,j}\cdot\left(\frac{4\pi}{b_{i,j}+B_i^k}\right)^{3/2}\exp\left(\frac{-4\pi^2}{b_{i,j}+B_i^k}\cdot\|(\mathbf{R}_q\mathbf{x}_i^k+\mathbf{t}_q)-\boldsymbol{\xi}\|_2^2\right) \tag{5}$$

Here, $N_s$ is the number of symmetry operations[51]; $m$ is the number of atoms in the asymmetric unit of conformer $\mathbf{X}^k$; $\mathbf{R}_q \in SO(3)$ is the rotation matrix of symmetry operation $q$; $\mathbf{t}_q \in \mathbb{R}^3$ is the translation vector of the symmetry operation $q$; $a_{i,j}$ and $b_{i,j}$ are tabulated atomic form factor coefficients for each heavy atom $\mathbf{x}_i^k$ (ref. 52); and $B_i^k$ is the B-factor that represents the isotropic displacement parameter[53]. Here, we consider the B-factor to be a bandwidth parameter.

We can extend the density calculation at spatial coordinates $\boldsymbol{\xi}$ to ensembles:

$$F_c\left(\boldsymbol{\xi}; \{\mathbf{X}^k\}_{k=1}^n, \mathbf{a}\right) = \frac{1}{n}\sum_{k=1}^{n} F_c(\boldsymbol{\xi}; \mathbf{X}^k, \mathbf{a}) \tag{6}$$

$F_c(\boldsymbol{\xi}; \mathbf{X}^k, \mathbf{a})$ is computed using equation (5). For our experiments, we assume that each atom has a uniform B-factor that is inversely related to the size of the ensemble, $B_i^k = \frac{4}{n}\forall i, k$. Additionally, we use an ensemble size of $n = 16$ for all our experiments. Our guidance framework differentiably incorporates all symmetry mates of the non-hydrogen atoms belonging to the chain of interest. In subsequent postprocessing steps, we further append structural waters, non-standard residues, ligands, ions and atoms from other chains of the proteins together with their respective symmetry mates.

**X-ray log-likelihood.** To quantify the agreement between observed and calculated electron density maps, we measure the $L_1$ norm of the difference between the calculated ensemble-averaged density $F_c$ (from equation (6)) and the observed density $F_o$:

$$\log p(\mathbf{y} = F_o|\mathscr{X}, \mathbf{a}) = -\left\|F_o(\cdot) - F_c(\cdot; \{\mathbf{X}^k\}_{k=1}^n, \mathbf{a})\right\|_1 \tag{7}$$

The log-likelihood computation is summarized in Supplementary Algorithm 2.

## NMR models

**Differentiable hydrogen placement.** AlphaFold3 does not explicitly model hydrogen atoms, yet their positions are essential for accurately modeling NMR observables. To address this, at each sampling step we first reconstruct hydrogen atoms of each atom in an amino acid in a differentiable manner using a PyTorch port of Hydride's[54] hydrogen placement algorithm. Each hydrogen-bearing center is identified by its local bonding environment and assigned a matching reference fragment. The fragment's non-hydrogen atoms are then aligned to the target by a single rigid-body superposition, after which the same transformation is applied to its hydrogens. In this way, proton positions adjust smoothly with the heavy-atom coordinates, yielding accurate and fully differentiable NMR calculations.

**NOE distance restraints.** We used solution-state NOE restraints as ensemble-averaged distance constraints. Conceptually, NOEs arise from through-space dipolar interactions between protons with observable effects typically for internuclear separations less than approximately 6 Å; intensities in NOESY spectra reflect an (approximate) $r^{-6}$ dependence and are commonly used semiquantitatively to bound distances. For an ensemble $\mathscr{X} = \{\mathbf{X}^k\}_{k=1}^n$, we modeled the ensemble-average distance

$$d_{ij}(\mathscr{X}) = \frac{1}{n}\sum_{k=1}^{n}\left\|\mathbf{x}_i^k - \mathbf{x}_j^k\right\|_2 \tag{8}$$

and enforced lower/upper bounds $(\underline{d}_{ij}, \bar{d}_{ij})$ via the log-likelihood

$$\log p(D|\mathscr{X}; \mathbf{a}) = -\sum_{(i,j)\in D}\left(\left[\underline{d}_{ij} - d_{ij}(\mathscr{X})\right]_+^2 + \left[d_{ij}(\mathscr{X}) - \bar{d}_{ij}\right]_+^2\right)$$

where $[x]_+ = \max(x, 0)$. Additionally, ambiguous NOE assignments were retained as OR groups; for evaluation, the group-level violation equals the minimum violation among its members (that is, the group is satisfied if any member lies within bounds). For restraints involving methyl groups, we modeled rapid three-fold internal rotation by averaging over the three methyl proton positions generated by differentiable hydrogen placement. The log-likelihood computation is summarized in Supplementary Algorithm 3.

We guided the diffusion sampler with the NOE likelihood using non-i.i.d. ensemble sampling, which couples conformers through the ensemble-level likelihood. Substructure conditioning was not used for NOE.

**Order parameters.** Order parameters $S^2$ for amide and methyl bonds represent ensemble flexibility (lower values imply higher disorder). At each sampling step, we rigidly align all conformers $\mathbf{X}^k \in \mathscr{X}$ in the ensemble to a reference $\mathbf{X}^{\text{ref}}$ to remove global rotation and translation. For each bond $p$ in ensemble member $\mathbf{X}^k$, we compute a unit bond vector $\mathbf{d}_p^k$ (amide uses the N → H vector, and methyl uses the C → C vector pointing to a methyl carbon; some amino acids, such as leucine, isoleucine

and valine, have two methyl groups and, hence, two vectors per residue). Given these normalized vectors, the calculated bond order $S^2_{c,p}$ of the bond $p$ is given by

$$S^2_{c,p}(\mathcal{X};\mathbf{a}) = \frac{1}{|\mathcal{X}|(|\mathcal{X}|-1)} \sum_{i \neq j} P_2\left(\mathbf{d}^i_p \cdot \mathbf{d}^j_p\right)$$

where $P_2(x) = \frac{1}{2}(3x^2 - 1)$ is the Legendre polynomial of order 2. The log-likelihood is computed according to

$$\log p(S^2|\mathcal{X},\mathbf{a}) = -\sum_p \left(S^2_{c,p}(\mathcal{X};\mathbf{a}) - S^2_p\right)^2$$

where the summations are performed over the collection of bonds on which the experimental data $S^2 = \{S^2_p\}$ have been collected. The log-likelihood computation is summarized in Supplementary Algorithm 4.

**Dihedrals likelihood.** Backbone dihedral restraints assess the agreement between predicted Ramachandran angles ($\phi^{\text{pred}}$, $\psi^{\text{pred}}$) from ensemble $\mathcal{X}$ and target angles ($\Phi^{\text{tgt}} = \{\phi^{\text{tgt}}_i, \delta_{\phi,i}\}_{i=1}^M$ and $\Psi^{\text{tgt}} = \{\psi^{\text{tgt}}_i, \delta_{\psi,i}\}_{i=1}^M$) defined over $M$ residues. The target angles and their associated uncertainty half-widths are obtained from TALOS-N[43] predictions. The log-likelihood is computed using a $2\pi$-periodic (wrapped) angular discrepancy function that respects the circular nature of angles. This term encourages locally accurate backbone geometry and regular secondary structure, complementing global guidance terms derived using NOE distance restraints or cryo-EM ESP maps. The dihedral log-likelihood is defined as

$$\log p\left(\Phi^{\text{tgt}}, \Psi^{\text{tgt}}|\mathcal{X},\mathbf{a}\right) = -\frac{1}{2M}\sum_{r=1}^{M}\left[\frac{\Delta\left(\phi^{\text{pred}}_r(\mathcal{X}), \phi^{\text{tgt}}_r\right)^2}{\delta^2_{\phi,r}} + \frac{\Delta\left(\psi^{\text{pred}}_r(\mathcal{X}), \psi^{\text{tgt}}_r\right)^2}{\delta^2_{\psi,r}}\right]$$

where $\Delta(\alpha,\beta)$ denotes the wrapped angular difference (implemented as `angle_diff`). This log-likelihood is averaged over all chains of structures in the ensemble.

**Overall NMR likelihood.** When guiding with order parameters, NOE observations and dihedrals, the log-likelihoods are combined as:

$$\begin{aligned}\log p(S^2, D, \Phi^{\text{tgt}}, \Psi^{\text{tgt}}|\mathcal{X},\mathbf{a}) = & -\lambda_S \sum_p \left(S^2_{c,p}(\mathcal{X};\mathbf{a}) - S^2_p\right)^2 \\ & -\lambda_D \sum_{(i,j)\in D}\left(\left[\underline{d}_{ij} - d_{ij}(\mathcal{X})\right]^2_+ + \left[d_{ij}(\mathcal{X}) - \overline{d}_{ij}\right]^2_+\right) \\ & +\lambda_T \log p(\Phi^{\text{tgt}}, \Psi^{\text{tgt}}|\mathcal{X},\mathbf{a})\end{aligned} \tag{9}$$

We guide the model either without order parameter information or using one of two types: relaxation-derived order parameters (ps–ns motions) or RDC-derived order parameters (µs–ms motions). Dihedral restraints are incorporated only in experiments that involve cryo-EM ESP maps. The scaling parameters $(\lambda_S, \lambda_D, \lambda_T)$ are set as follows:

$$(\lambda_S,\lambda_D,\lambda_T) = \begin{cases} (0,\ \lambda_D \in \{0.1,\ldots,0.5\},0), & \text{NOE}, \\ (0.30, 0.45, 0), & \text{Methyl relaxation \& NOE}, \\ (0.45, 0.45, 0), & \text{Amide relaxation \& NOE}, \\ (0.45, 0.40, 0), & \text{Methyl } S^2 \text{ from RDC \& NOE}, \\ (0.30, 0.40, 0), & \text{Amide } S^2 \text{ from RDC \& NOE}, \\ (0.0, 0.40, 0.30), & \text{NOE, Cryo ESP, \& Dihedrals}. \end{cases} \tag{10}$$

When guiding with NOEs alone, we selected the best-performing value of $\lambda_D$ from the tested set.

## Cryo-EM models

**ESP forward model.** As discussed in ref. 30, the forward model for cryo-EM ESP maps is approximated by atomic scattering potentials, following equations (5) and (6). Unlike X-ray crystallography, cryo-EM maps typically represent a single particle, so the model reduces to

$$F_c(\boldsymbol{\xi};\mathbf{X}^k,\mathbf{a}) = \sum_{i=1}^{m}\sum_{j=1}^{5} a_{i,j}\cdot\left(\frac{4\pi}{b_{i,j}+B^k_i}\right)^{3/2}\exp\left(\frac{-4\pi^2}{b_{i,j}+B^k_i}\cdot\left\|\left(\mathbf{R}^k\mathbf{x}^k_i + \mathbf{t}^k\right) - \boldsymbol{\xi}\right\|^2_2\right), \tag{11}$$

where $\phi = (\mathbf{R}^k, \mathbf{t}^k)$ represents the pose of conformation $\mathbf{X}^k$ instead of a crystallographic symmetry operator. In the presence of significant blurring, the ensemble contribution can be evaluated as an average using equation (6).

**Strategies for aligning proteins with ESP maps.** AlphaFold3's noisy time marginals $\mathbf{X}_t$ (atom coordinates at timestep $t$ of the diffusion process) and the denoised predictions $\hat{\mathbf{X}}_0(\mathbf{X}_t, t)$ evolve on the learned diffusion manifold and not in a fixed frame. As a result, accurately predicted structures will generally be misaligned with the ERP map's voxel grid. Additionally, the $\hat{\mathbf{X}}_0(\mathbf{X}_t, t)$ predictions have a random pose at every iteration. In X-ray crystallography, the Kabsch algorithm[55] is used to solve the alignment problem. On the other hand, in cryo-EM ab initio protein density fitting, no reference protein structure is available, making this alignment method inapplicable. Instead, a pure protein density alignment strategy has to be implemented. Depending on whether the diffusion is performed over the whole ESP map or over well-separable chains, two different alignment strategies have been implemented:

- Full ESP map rigid alignment: global pose determined by solving

$$\phi^k = (\mathbf{R}^k, \mathbf{t}^k) = \arg\min_{\phi^k} d\left(\mathbf{R}^k\mathbf{X}^k + \mathbf{t}^k, F_o\right) \tag{12}$$

  where $d(\mathbf{X}, F_o)$ is the error objective defined in the following sections. The resulting pose is applied as a rigid transformation, $\tilde{\mathbf{X}}^k = \mathbf{R}^k\mathbf{X}^k + \mathbf{t}^k$, and all subsequent guidance scores are evaluated with respect to the aligned structure.
- Per-chain non-rigid ESP alignment: In this case, the alignment is performed independently for each chain. A set of per-chain poses $\{\phi^{k,c}\}_{c=1}^{N_c} = \{(\mathbf{R}^{k,c}, \mathbf{t}^{k,c})\}_{c=1}^{N_c}$ is estimated and applied to obtain the aligned structure $\tilde{\mathbf{X}}^k = \cup_c\left(\mathbf{R}^{k,c}\mathbf{X}^{k,c} + \mathbf{t}^{k,c}\right)$, where $N_c$ is the number of chains.

In AlphaFold3, a structure $\mathbf{X}_t$ is represented as a tensor of shape $(N_a, 3)$ where $N_a = \sum_{c=1}^{N_c} N_{a_c}$ denotes the total number of atoms and $N_{a_c}$ denotes the number of atoms in a chain $c$. After concatenation of $\tilde{\mathbf{X}}^k$, the application of the per-chain poses is a non-rigid transformation with respect to the full structure $\mathbf{X}^k$.

The chosen objective $d$ is the trilinear interpolation $-F_o(\tilde{\mathbf{X}}^k)$, and the maximization of $F_o$ was implemented as a gradient descent over the pose $\phi^k$. For this, hyperparameters of number of iterations $N_{\text{iter}}$, $N_{\text{init}}$ initial random rotation, learning rates $\alpha_R$, $\alpha_t$, the side length $T$ of the box around the weighted centroid of $F_o$ and the positive B-factor $B$ to be applied to the volume to ensure non-zero gradients for initial poor alignments can be defined in the configuration file. For cases where predictions $\hat{\mathbf{X}}_0(\mathbf{X}_t, t)$ have a starkly different tertiary structure from the intended $\mathbf{X}^k$ and this creates numerous unstable local minima (for example, like what we observed in the insulin receptor structures 8U4B and 8U4E), a momentum decay $\alpha_m$ can be specified. This adds decaying Kabsch-aligned anchoring to the pose from the previous diffusion iteration to prevent large fluctuations of poses between diffusion timesteps.

**After-alignment diffusion manifold correction of $\mathbf{X}_t$ and $\widehat{\mathbf{X}}_0(\mathbf{X}_t, t)$.** Because in the diffusion process $\mathbf{X}_{t-1}$ is computed as a function of the noisy structure and the $t = 0$ prediction $\mathbf{X}_{t-1} = f(\mathbf{X}_t, \widehat{\mathbf{X}}_0(\mathbf{X}_t, t))$, in case a transformation was applied to the aligned $\widehat{\mathbf{X}}_0$, for the correct $\mathbf{X}_{t-1}$ evaluation, an equivalent change must be applied to $\mathbf{X}_t$ in its relative frame.

- Rigid alignment correction: Here the correction can be avoided if the direction gradient is evaluated using autograd as $\frac{\partial}{\partial \mathbf{X}_t}\widetilde{\mathbf{X}}_0(\mathbf{X}_t, t)$ and the aligned $\widetilde{\mathbf{X}}_0$ is aligned back through $\mathbf{T}^{-1}(\widetilde{\mathbf{X}}) = \mathbf{R}^T(\widetilde{\mathbf{X}} - \mathbf{t})$.
- Non-rigid alignment correction: For correction and preservation of the manifold, non-rigid alignment is applied in two steps: (1) a rigid alignment $\mathbf{T}_1(\mathbf{X}) = \mathbf{R}_1\mathbf{X} + \mathbf{t}_1$ that aligns the complete $\mathbf{X}$ with the complete density $F_o$ and (2) the non-rigid per-chain $c$ transformations $\mathbf{T}_2^c(\mathbf{R}_2^c\mathbf{X}^c + \mathbf{t}_2^c)$ computed using gradient ascent. In this way, movements equivalent to the additional non-rigid fine detail alignments $\mathbf{T}_2$ of $\widetilde{\mathbf{X}}_0$ can be defined in the relative frame of $\mathbf{X}_t$ and $\widehat{\mathbf{X}}_0$ as $\mathbf{T}_2^*$, where the corrected structures we will denote as $\mathbf{X}_t^*$ and $\widehat{\mathbf{X}}_0^*$. In other words, $\mathbf{T}_2 \circ \mathbf{T}_1 = \mathbf{T}_1 \circ \mathbf{T}_2^*$. Thus, $\widehat{\mathbf{X}}_{0,c}^* = \mathbf{T}_1^{-1} \circ \mathbf{T}_2 \circ \mathbf{T}_1(\widehat{\mathbf{X}}_{0,c})$ and $\mathbf{X}_t^* = \mathbf{T}_1^{-1} \circ \mathbf{T}_2 \circ \mathbf{T}_1(\mathbf{X}_t)$.

**ESP $L_1$ likelihood.** The objective $L_1$ is defined similarly to equation (7) and is computed as $-\|F_c(\mathbf{X}) - F_o\|_1$. This likelihood is best suited for homogeneous proteins or for heterogeneous ones with modest flexibility.

**Optimal transport likelihood.** Different cryo-EM ESP maps for a single sequence or protein assembly typically represent distinct states of a heterogeneous protein. Depending on the range of movement between the conformations of such proteins and the original predictions of AlphaFold3, the $L_1$ density objective gradients are often unable to guide the AlphaFold3 prediction to the map. Instead, due to floating point limitations and the absence of gradients for corresponding atoms that are far apart, this guidance strategy tends to push the noisy structures $\mathbf{X}_t$ and the corresponding $\widehat{\mathbf{X}}_0$ predictions off the diffusion manifold. Therefore, a better approach for evaluating guidance gradients is to use a likelihood that directly points toward the density mass—namely, optimal transport.

We model the protein and ESP map as discrete measures $\mu = \sum_{i=1}^{N} \alpha_i \, \delta_{\mathbf{x}_i}$ and $\nu = \sum_{j=1}^{M} \beta_j \, \delta_{\mathbf{y}_j}$, where $\mathbf{x}_i \in \mathbb{R}^3$ are all the atoms of a protein and $\mathbf{y}_j \in \mathbb{R}^3$ are all the voxel centers inside the global mask (or, equivalently, per-chain masks and per-chain coordinates). $\alpha_i$ is proportional to the atomic number $Z_i$, and $\beta_j$ is proportional to the observed value in the center of the voxel of $F_o$. The total mass of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ is normalized as $\sum_{i=1}^{N} \alpha_i = \sum_{j=1}^{M} \beta_j = 1$. The ground cost matrix is defined as $\mathbf{C} \in \mathbb{R}^{N \times M}$ with $C_{ij} = c(\mathbf{x}_i, \mathbf{y}_j)$ proportional to the 1-norm or 2-norm between these two vectors. The 'entropy optimal transport objective' is then defined as the 'primal problem'

$$\mathrm{OT}_\varepsilon(\boldsymbol{\alpha}, \boldsymbol{\beta}; \mathbf{C}) = \min_{\substack{\mathbf{P} \in \mathbb{R}_{\geq 0}^{N \times M} \\ \mathbf{P}\mathbf{1}_M = \boldsymbol{\alpha},\ \mathbf{P}^\top \mathbf{1}_N = \boldsymbol{\beta}}} \{\langle \mathbf{P}, \mathbf{C} \rangle - \varepsilon H(\mathbf{P})\}, \; H(\mathbf{P}) = -\sum_{i,j} P_{ij} \log P_{ij}, \tag{13}$$

with $\langle \mathbf{P}, \mathbf{C} \rangle = \sum_{i,j} P_{ij} C_{ij}$ and $\mathbf{1}_K$ the $K$-vector of 1s. Here, $\mathbf{P}$ is a transport plan, where $P_{ij}$ specifies the proportion of mass from source atom $\mathbf{x}_i$ assigned to voxel $\mathbf{y}_j$. The entropy term $\varepsilon H(\mathbf{P})$ promotes smooth, stable couplings and well-behaved gradients via discretizing assignments that are too sparse or sharp (contain a lot of zero entries) and makes the problem strictly convex.

In practice, to reduce computational cost, we use the 'Sinkhorn entropy optimal transport loss' in its dual representation with atom/voxel potentials $\boldsymbol{f} \in \mathbb{R}^N$ and $\boldsymbol{g} \in \mathbb{R}^M$:

$$\begin{aligned} \mathcal{L}_{\mathrm{Sinkhorn},\varepsilon}(\boldsymbol{\alpha}, \boldsymbol{\beta}; \mathbf{C}) \; &:= \; \mathrm{OT}_\varepsilon(\boldsymbol{\alpha}, \boldsymbol{\beta}; \mathbf{C}), \\ &= \max_{\boldsymbol{f}, \boldsymbol{g}} \boldsymbol{\alpha}^\top \boldsymbol{f} + \boldsymbol{\beta}^\top \boldsymbol{g} - \varepsilon \sum_{i,j} \alpha_i \beta_j \left( \exp\left( \tfrac{f_i + g_j - C_{ij}}{\varepsilon} \right) - 1 \right). \end{aligned} \tag{14}$$

We chose a powerful optimized implementation of optimal transport from `GeomLoss`[56], particularly to avoid memory bottlenecks. The hyperparameters of choice were `p=2, blur=`$\epsilon$`=0.005` and the binary `reach` flag, depending on whether mass needed to be dropped for a structure.

In cases where ESP maps contain separable chains, the optimal transport objective was also applied per chain. The rationale was to separate the guidance directions and to prevent unwanted misaligned stable local minima. This was required for structures 6W0O, 7DA4 and 9FH1. A typical issue with those structures was the discrepancy in AlphaFold3's tensor representation and the spatial arrangement of the subunits, where those were not ordered correspondingly. To solve that, we tested both the Hungarian algorithm for the linear assignment problem (LAP), defined by chainwise distances, and brute force approaches. However, the most optimal solution was the per-chain optimal transport objective: such guidance naturally ordered the AlphaFold'3 tensor representation to the spatial chains and, thereby, optimally solved for the correspondence implicitly.

**Recycling and extended optimization of proteins.** In cases where AlphaFold3 early-time or unguided predictions poorly match the protein of interest, the initial protein optimization required to reach sufficient quality for ESP map alignment may consume the majority of available timesteps in the diffusion process. Such a trend was observed for structures 6W0O and 9FH1. This, in turn, can leave an insufficient number of steps in the diffusion process for fine detail protein-ESP map fitting or make it completely impossible due to the noise schedule being insensitive to any change after the noise level $\sigma_t$ reaches its lowest values (in practice, no significant change is observed after 160 steps). Additionally, the aforementioned proteins suffer from strong, stable local minima in the optimal transport likelihood—for example, 180° rotations about any axis lying on the plane perpendicular to the first principal component of 6W0O. This significantly increases the average time it takes for the protein to align to the intended spot during the diffusion process.

To allow additional time for finding these optimal protein-ESP map arrangements, we implement recycling: once the noisy structure reaches a specified timestep $t$, instead of proceeding with the backwards diffusion process to obtain $\mathbf{X}_{t-1} = f_\theta(\mathbf{X}_t, \widehat{\mathbf{X}}_0)$, noise is added as in the forward diffusion process to advance from timestep $t$ to $t + k$:

$$\mathbf{X}_{t+k} = \mathbf{X}_t + \sqrt{\sigma_{t+k} - \sigma_t}\,\epsilon, \; \epsilon \sim \mathcal{N}(0, I).$$

For example, if protein-ESP alignment of 6W0O converged to all the intended rotations for each chain instead of the 180° opposite symmetry poses after 160 timesteps, it would not be further optimized. Adding an additional small level of noise, typically equivalent to 25 timesteps of the forward process, does not break the correctly found tertiary structure while giving extra time for fine detail fitting.

**Objective normalization.** To balance multiple objectives with the ESP guidance, we employ a specific normalization strategy: one objective, typically the ESP $L_1$ or optimal transport likelihood, is chosen as the reference, and all other likelihood terms are rescaled at each iteration so that their contributions remain a fixed proportion $\lambda_i \in [0, 1]$ of this main objective. This prevents magnitude imbalances that are otherwise observed when combining losses from very different modalities and ensures stable relative weighting throughout the diffusion process.

**Overall cryo-EM likelihood.** The overall (pure) cryo-EM guidance objective is defined as a weighted combination of the density-based $L_1$ and the optimal transport likelihood

$$\log p(\mathbf{y} = F_o | \mathcal{X}, \mathbf{a}) = -\lambda_1(t) \left\| F_o(\cdot) - F_c\left(\cdot; \{\mathbf{X}^k\}_{k=1}^n, \mathbf{a}\right) \right\|_1 - \lambda_2(t) \mathrm{OT}(\mathcal{X}, F_o, \mathbf{C}), \tag{15}$$

where $\lambda_1(t)$ and $\lambda_2(t)$ are time-dependent weights normalized as described above: one of these weights is set to $\lambda_i = 1$ as the main objective, whereas the other is some portion of it normalized as per the previous section. For the pure cryo-EM loss, we typically simply use $\lambda_1 = 0$ and $\lambda_2 = 1$ until the structure roughly aligns to the density mass and then switch to pure density loss to optimize fine detail features and optimize B-factors—that is, $\lambda_1 = 1$ and $\lambda_2 = 0$. Some more complicated mixtures were explored; however, they showed better results for combined guidance with other modalities. For example, in the NMR setting, the likelihood is scaled to 0.15 of the active objective, first relative to the optimal transport loss (when it is turned on) and later relative to the $L_1$ density loss (once it becomes active).

### Additional data terms

**Substructure conditioner.** Specifically for crystallographic structures, we aim to optimize a certain region of the protein while the remainder of the structure is stabilized by bootstrapping the diffusion process to a set of reference atomic coordinates. This is similar to `SubstructureConditioner` in Chroma[57].

As input, consider a list of reference atom locations $Y = \{\mathbf{y}_i : i \in A\}$ for atom indices $A \subseteq \{1 \ldots m\}$, where $m$ is the number of atoms in conformer $\mathbf{X}^k$. The log-likelihood is as a quadratic penalty on the deviation from reference atom locations:

$$\log p(Y|\mathcal{X}, \mathbf{a}) = -\lambda_{\text{sub}} \frac{1}{n} \sum_{k=1}^{n} \sum_{i \in A} \| \mathbf{x}_i^k - \mathbf{y}_i \|_2^2 \tag{16}$$

Before computing the log-likelihood, we align all structures in the ensemble to the reference atom locations (limited to indices in $A$) using the Kabsch algorithm[55]. For crystallography, $\lambda_{\text{sub}} = 0.1$, whereas, for NMR and cryo-EM, $\lambda_{\text{sub}} = 0.0$.

**Validity likelihood.** To prevent ensembles with elongated bonds and steric clashes, we incorporate a validity log-likelihood as regularization akin to AlphaFold2's violation loss[3]. Consider an ensemble of $n$ protein conformations $\mathcal{X} = \{\mathbf{X}^k\}_{k=1}^n$, where each structure $\mathbf{X}^k \in \mathbb{R}^{m \times 3}$. We define a binary bond matrix $\mathbf{B} \in \{0, 1\}^{m \times m}$ such that $B_{ij} = 1$ if atoms $i$ and $j$ are covalently bonded and 0 otherwise. The bond length loss for conformation $\mathbf{X}^k$ over bonded atom pairs ($B_{ij} = 1$) is given as

$$\mathcal{L}_{\text{bond}}(\mathbf{X}^k; \mathbf{a}) = \sum_{i=1}^{m} \sum_{j=i+1}^{m} B_{ij} \cdot \left(\max\left(0, |d_{ij}^k - d_{ij}^{\text{ideal}}| - \delta_{\text{bond}}\right)\right)^2$$

where $d_{ij}^{\text{ideal}}$ is the ideal bond length approximated as the sum of covalent radii of atoms $i$ and $j$; $d_{ij}^k = \| \mathbf{x}_i^k - \mathbf{x}_j^k \|_2$ is the between-atoms $i$ and $j$ in conformation $k$; and $\delta_{\text{bond}} = 0.2$ Å is a tolerance margin. For non-bonded pairs (between and within residues), steric clashes are penalized using a soft collision loss, defined as the maximum violation over neighbors of each atom:

$$\mathcal{L}_{\text{collision}}(\mathbf{X}^k; \mathbf{a}) = \sum_{i=1}^{m} \max_{j \neq i, B_{ij}=0} \left(\max\left(0, (d_{ij}^{\text{ideal}} + p^{\text{collision}}) - d_{ij}^k\right)\right)$$

Here, we use a padding distance $p^{\text{collision}} = 0.4$ (in Å) to prevent over-penalization of near-contact atoms. Lastly, we define a bond angle violation loss as

$$\mathcal{L}_{\text{angle}}(\mathbf{X}^k; \mathbf{a}) = \sum_{j=1}^{m} \sum_{i \neq j} \sum_{k > i} B_{ij} B_{jk} \cdot \left(\max\left(0, |\theta_{ijk} - \theta_{ijk}^{\text{ideal}}| - \delta_{\text{angle}}\right)\right)$$

where $\theta_{ijk}$ (in degrees) is calculated using the dot product of the bond vectors from central atom $j$ to atoms $i$ and $k$; $\theta_{ijk}^{\text{ideal}}$ (in degrees) is retrieved from the valence shell electron pair repulsion (VSEPR) theory[58]; and $\delta_{\text{angle}} = 12°$ is a tolerance margin. The resulting validity log-likelihood across the ensemble is given as

$$\log(\mathbf{B}|\mathcal{X}, \mathbf{a}) = -\sum_{k=1}^{n} \lambda_{\text{bond}} \mathcal{L}_{\text{bond}}(\mathbf{X}^k; \mathbf{a}) + \lambda_{\text{collision}} \mathcal{L}_{\text{collision}}(\mathbf{X}^k; \mathbf{a}) + \lambda_{angle} \mathcal{L}(\mathbf{X}^k; \mathbf{a})$$

where $\lambda_{\text{bond}}, \lambda_{\text{collision}}$ and $\lambda_{\text{angle}}$ are scaling factors that control the contribution of bond, collision and bond angle terms. For ensembles guided using crystallographic density maps, we use $\lambda_{\text{bond}} = \lambda_{\text{collision}} = \lambda_{\text{angle}} = 0.075$. For NMR-guided ensembles, the scaling factors ($\lambda_{\text{bond}}, \lambda_{\text{collision}}, \lambda_{\text{angle}}$) depend on the choice of ($\lambda_S, \lambda_D, \lambda_T$) as shown in Supplementary Table 8. (For NOE-guided experiments, we report the best-performing configuration out of two settings.)

### Force field relaxation

Despite applying validity likelihood, a select few samples have broken bonds or steric clashes. Due to the nonphysical nature of the ensemble, we filter out these samples from the ensemble. We consider a sample to have broken bonds if the distance between any pair of bonded atoms exceeds $\tau_{\text{bond}}$. Likewise, we consider a sample to have steric clashes if the distance between any two atoms is less than $\tau_{\text{clash}}$. Across all experiments, $\tau_{\text{bond}} = 2.1$ Å and $\tau_{\text{clash}} = 1.1$ Å.

Despite this, certain geometries can go undetected. To resolve this while maximizing the log-likelihood of the experimental observations, we relax the remaining samples in the ensemble using an off-the-shelf harmonic force field. Specifically, we use OpenMM's[26] implementation of the AMBER14 (ref. 40) force field parameters for energy minimization within ColabFold[36]. In ColabFold, the energy is minimized for a maximum of 2,000 iterations with energy tolerance threshold of 2.39 kcal mol$^{-1}$ and stiffness of 100.0 kcal mol$^{-1}$ Å$^{-2}$ for NMR and 10.0 kcal mol$^{-1}$ Å$^{-2}$for the other modalities, thereby controlling the strength of positional restraints applied to the atoms.

### Ensemble pruning

Note that, here, 'ensemble' refers to the set of samples that remain after relaxation, from which we prune a nonredundant subset. After relaxation, we aim to report a nonredundant subset of samples $\mathcal{X}_\mathcal{I} = \{\mathbf{X}^k : k \in \mathcal{I}\}$ that best explains experimental observation $\mathbf{y}$. $\mathcal{I}$ denotes the indices of selected samples from the full ensemble. We propose two complementary strategies for populating $\mathcal{I}$.

**Orthogonal Matching Pursuit.** Orthogonal Matching Pursuit (OMP)[59] is a greedy selection algorithm that begins with $\mathcal{I} = \emptyset$. At each iteration, the candidate $k \notin \mathcal{I}$ that maximizes $\log p(\mathbf{y}|\mathcal{X}_{\mathcal{I} \cup \{k\}}, \mathbf{a})$ is added to the support set $\mathcal{I}$ until the log-likelihood no longer increases. After termination, each sample in $\mathcal{X}_\mathcal{I}$ is assigned a uniform occupancy of $\frac{1}{|\mathcal{I}|}$.

To avoid overfitting to noise in $\mathbf{y}$ and reduce redundancy in the ensemble, we introduce two early-stopping criteria. (1) The ensemble size is capped at $n_{\max} = 5$. Once $|\mathcal{I}| = 5$, the algorithm terminates. (2) Before including a candidate, we evaluate the ensemble's validation loss $f_{\text{val}}$, and the algorithm terminates if the loss increases. The OMP selection procedure is applied only to ensembles derived from X-ray crystallography.

The selection of the best-performing sample at each iteration is based on equation (6). The log-likelihood is computed over all voxels $\boldsymbol{\xi}$ in $F_o$ that are within 2.5 Å of atoms in the relevant residue range of the reference PDB files (Section 8.2.1 'X-ray crystallography'). Before a candidate is added to $\mathcal{X}_\mathcal{I}$, a scalar B-factor $B$ (applied uniformly to all atoms in the structure) is optimized to maximize the log-likelihood. This effectively tunes the bandwidth of $F_c(\boldsymbol{\xi}; \mathcal{X}_\mathcal{I}, \mathbf{a})$ to best match the observed density $F_o(\boldsymbol{\xi})$. Optimization of $B$ is carried out using Adam[60] for 100 iterations with a learning rate of 1.0. As a validation loss, we use $R_{\text{free}}$ (Section 8.11 'Metrics'), computed with PHENIX's `phenix.model_vs_data`. A detailed pseudocode is provided in Supplementary Algorithm 5.

**Occupancy optimization.** In crystallography, we propose an alternative pruning strategy because OMP does not always yield the minimal subset explaining $F_o$. Instead, we optimize conformer occupancies in the relaxed ensemble $\mathcal{X}_\mathcal{J}$ to best fit $F_o$. Each candidate is initially assigned a uniform weight $p_k = \frac{1}{n}$ ($\sum_k p_k = 1$) with $n$ samples in the ensemble $\mathcal{X} = (\mathbf{X}^1 \ldots \mathbf{X}^n)$.

In this algorithm, given $\mathcal{X}$ and amino acid sequence $\mathbf{a}$, the 'weighted average' calculated electron density is defined as

$$F_c(\boldsymbol{\xi}; \mathcal{X}, \mathbf{a}) = \sum_k p_k \cdot F_c(\boldsymbol{\xi}; \mathbf{X}^k, \mathbf{a})$$

where $F_c(\boldsymbol{\xi}; \mathbf{X}^k, \mathbf{a})$ is the calculated density of conformer $\mathbf{X}^k$ weighted by occupancy $p_k$. The occupancy weights are optimized by minimizing the following objective:

$$\mathcal{L} = \left\| F_o(\boldsymbol{\xi}) - F_c(\boldsymbol{\xi}; \mathcal{X}, \mathbf{a}) \right\|_1 + \lambda \sum_k p_k \log p_k,$$

The first term maximizes the log-likelihood $\log p(F_o | \mathcal{X}, \mathbf{a})$; the second term serves as an entropy minimizer that promotes selection of a small subset of conformers. The objective is minimized using Adam optimizer for $T_{max} = 300$ iterations with a learning rate of $\alpha = 10^{-3}$ and $\lambda = 0.7$. We observe that most occupancies decay to zero, thereby best explaining $F_o$ with minimal redundancy. After convergence, the pruned ensemble is defined as:

$$\mathcal{J} = \{k : p_k > 0.1\}, \mathcal{X}_\mathcal{J} = \{\mathbf{X}^k : k \in \mathcal{J}\}$$

A detailed pseudocode is provided in Supplementary Algorithm 6. In practice, we apply both algorithms and ultimately select the minimal ensemble that provides the best fit to $F_o$.

**Refinement.** In crystallography, after the termination of the ensemble pruning algorithms, the selected ensemble $\mathcal{X}_\mathcal{J}$ is merged into a single PDB file. In this merged structure, atoms included in the set $A$ (outside residue range of interest) are retained from the corresponding chain of the PDB-Redo entry[39], whereas atoms not included in $A$ (in residue range of interest) are replaced with those modeled in $\mathcal{X}_\mathcal{J}$. Each conformer in $\mathcal{X}_\mathcal{J}$ is assigned a unique altloc identifier (`A`, `B`, etc.). Altloc occupancies are then set according to the probabilities estimated by the selection algorithms (Supplementary Algorithms 5 and 6). Other structural attributes, such as isotropic and anisotropic B-factors, are retained from the relaxed conformers.

The merged structure is then refined against the experimental diffraction data using `REFMAC5`, part of the CCP4 library[61]. `REFMAC5` refines a structural model by minimizing the difference between the observed structure factors (provided in the MTZ file from PDB-Redo) and the calculated structure factors (computed from the input model). The refinement program adjusts atomic coordinates and atomic displacement parameters (B-factors) to better fit the observed structure factors. During optimization, `REFMAC5` applies stereochemical restraints to maintain chemically plausible geometry. Because the coordinate updates are typically small, the refinement step preserves the conformational diversity introduced by our guidance framework while ensuring statistical consistency with the diffraction data. In practice, we perform five cycles.

## Metrics

**X-ray crystallography.** To evaluate how well our density-guided structures agree with experimental X-ray crystallographic data, we report three metrics: *R*-values, real space correlation coefficient (RSCC) and cosine similarity.

*R-factors.* The *R*-factor (or crystallographic *R*-value) is a standard metric for assessing how well a crystal structure's model explains experimental X-ray diffraction data. These data are typically stored in MTZ files. The metric measures the alignment between the observed structure factor amplitudes $|F_{obs}|$ and the calculated structure factor amplitudes $|F_{calc}|$. Formally,

$$R = \frac{\sum_{\mathbf{h}} | \, |F_{obs}(\mathbf{h})| - |F_{calc}(\mathbf{h})| \, |}{\sum_{\mathbf{h}} |F_{obs}(\mathbf{h})|}$$

Here, $\mathbf{h}$ denotes a Miller index triple $(h, k, l)$, which denotes a discrete reflection in the reciprocal (Fourier) space. To avoid overfitting, $R_{free}$ value is calculated in a similar manner but using a small subset $T$ of reflections that are excluded from the refinement pipeline and consequently serve as a validation set. Hence,

$$R_{free} = \frac{\sum_{\mathbf{h} \in T} | \, |F_{obs}(\mathbf{h})| - |F_{calc}(\mathbf{h})| \, |}{\sum_{\mathbf{h} \in T} |F_{obs}(\mathbf{h})|}$$

It is important to note that $|F_{obs}|$ and $|F_{calc}|$ are structure factor amplitudes in reciprocal space. On the other hand, $F_o$ and $F_c$ used in our density guidance framework denote the real-space three-dimensional volumetric electron density grids. Electron density maps are obtained by estimating phase to the structure factor amplitudes followed by an inverse Fourier transform into real space. For each experiment, we report both the $R$ and $R_{free}$ values as calculated by `REFMAC5` after refinement.

*RSCC.* The RSCC is a local validation metric that measures the alignment between a structural model and the observed experimental electron density map. Unlike *R*-values, RSCC is calculated directly in real space by comparing observed electron density $F_o(\boldsymbol{\xi})$ with calculated electron density $F_c(\boldsymbol{\xi})$ over an amino acid. Formally, RSCC for a grid of voxels $\mathcal{G}$ around a residue $a \in \mathbf{a}$ is given by

$$\mathrm{RSCC}_a = \frac{\sum_{\boldsymbol{\xi} \in \mathcal{G}} (F_o(\boldsymbol{\xi}) - \overline{F_o})(F_c(\boldsymbol{\xi}) - \overline{F_c})}{\sqrt{\sum_{\boldsymbol{\xi} \in \mathcal{G}} (F_o(\boldsymbol{\xi}) - \overline{F_o})^2} \sqrt{\sum_{\boldsymbol{\xi} \in \mathcal{G}} (F_c(\boldsymbol{\xi}) - \overline{F_c})^2}}$$

where $\overline{F_o} = \mathbb{E}_\mathcal{G}[F_o]$ and $\overline{F_c} = \mathbb{E}_\mathcal{G}[F_c]$. RSCC ranges from −1 to 1, with values closer to 1 indicating a residue that is well supported by the experimental density. For our use case, we used `phenix.real_space_correlation` to measure the RSCC for each residue (and altloc) in our protein.

*Cosine similarity.* In addition to the RSCC, which provides a residue-level measure of model-to-map alignment in real space, we report in a single global score over the entire residue range of interest. For this purpose, we use cosine similarity to quantify the overall alignment between the observed density $F_o$ and the calculated density $F_c$. To compute this metric, we first identify the set of spatial coordinates $\boldsymbol{\xi} \in \mathbb{R}^3$ by selecting all voxels from $F_o$ that lie within 2.5 Å of atoms in the relevant residue range of the reference PDB structures. Formally,

$$\text{Cosine Similarity} = \frac{\sum_{\boldsymbol{\xi}} F_o(\boldsymbol{\xi}) F_c(\boldsymbol{\xi})}{\sqrt{\sum_{\boldsymbol{\xi}} F_o^2(\boldsymbol{\xi})} \cdot \sqrt{\sum_{\boldsymbol{\xi}} F_c^2(\boldsymbol{\xi})}}$$

where $F_o(\boldsymbol{\xi})$ and $F_c(\boldsymbol{\xi})$ denote vectors of density values at the selected voxel coordinates. As with RSCC, values closer to 1 indicate strong agreement between calculated and observed densities, signaling a good overall structural fit to the experimental map.

**NMR.** *Percentage of violated NOE constraints.* Given a set of NOE distance-bound constraints grouped into ambiguous OR groups, $G_m = \{\mathbf{r} = (i, j, \underline{d}_{ij}, \overline{d}_{ij})\}$, define the ensemble-averaged interproton distance

$$d_{ij}(\mathcal{X}; \mathbf{a}) = \frac{1}{|\mathcal{X}|} \sum_{k=1}^{|\mathcal{X}|} \| \mathbf{x}_i^k - \mathbf{x}_j^k \|_2,$$

and the per-restraint violation magnitude

$$v_{ij} = \max\left\{\underline{d}_{ij} - d_{ij}(\mathcal{X};\mathbf{a}), d_{ij}(\mathcal{X};\mathbf{a}) - \bar{d}_{ij}, 0\right\}.$$

The group-level violation is $v_{G_m} = \sum_{\mathbf{r}\in G_m} v_{ij}$, so an OR group is satisfied if any member falls within bounds. The percentage of violated NOE constraints is

$$100 \times \frac{|\{m : v_{G_m} > 0\}|}{M}$$

where $M$ is the number of OR groups. Ambiguity and methyl restraints are treated as in the NOE likelihood (ensemble-averaged distances; methyls averaged over the three proton positions).

*Median violation magnitude*. Given the group-level violations $\{v_{G_m}\}_{m=1}^{M}$ defined above, the median violation magnitude is

$$\text{median}\left(\{v_{G_m} : v_{G_m} > 0\}\right) \ [\text{Å}]$$

—that is, the median taken over violated groups only.

*Order parameter correlation (r-factor)*. Given a set of experimentally determined order parameters, $\{S_p^2\}$ for some collection of bonds $p \in P$ and a corresponding set of order parameters $\{S_{c,p}^2\}$ estimated from the generated ensemble, the correlation coefficient is defined as

$$r = \frac{\sum_p \left(S_{c,p}^2 - \overline{S_c^2}\right)\left(S_p^2 - \overline{S^2}\right)}{\sqrt{\sum_p \left(S_{c,p}^2 - \overline{S_c^2}\right)^2} \cdot \sqrt{\sum_p \left(S_p^2 - \overline{S^2}\right)^2}}$$

where $\overline{S^2} = \frac{1}{|P|}\sum_p S_p^2$ and $\overline{S_c^2} = \frac{1}{|P|}\sum_p S_{c,p}^2$ are the averages of the experimental and estimated order parameters, respectively. The coefficient ranges from −1 to 1, with higher values indicating better agreement with the experiment.

*Order parameter normalized fitting error (q-factor)*. Similary to the correlation coefficient, the normalized fitting error is defined as:

$$q = \frac{\sqrt{\sum_p \left(S_{c,p}^2 - S_p^2\right)^2}}{\sqrt{\sum_p S_{c,p}^2}}.$$

Lower values of $q$ indicate a better agreement with the experiment. Although $r$ is insensitive to a constant offset in the estimated values, $q$ is sensitive both to the error standard deviation and bias.

*ANSURR metrics*. ANSURR[24] validates a protein model by comparing per-residue flexibility inferred from backbone chemical shifts with rigidity predicted from the three-dimensional structure. Flexibility from NMR data is estimated via the random coil index (RCI), whereas structural rigidity is computed from the network of covalent and hydrogen-bond constraints using rigidity theory (for example, Floppy Inclusions and Rigid Substructure Topography (FIRST)). ANSURR reports two centile-ranked scores (0–100): a 'correlation score' that reflects agreement in the pattern of flexible versus rigid regions and an 'RMSD' score that reflects agreement in the magnitude of rigidity/flexibility. For structural ensembles, ANSURR is computed separately for each ensemble member, and the median of RMSD and correlation scores across the ensemble is used as a representative metric.

**Cryo-EM.** For cryo-EM, the same metrics as for crystallography were taken, mainly the cosine similarity and the per-residue model map cross-correlation scores from PHENIX from Section 8.11.1 ('X-ray crystallography').

## Minimal end-to-end code invocation

**X-ray crystallography.** To guide AlphaFold3 predictions using real-space crystallographic electron density maps, run the following command in your terminal:

```
python3 run_xray.py <pdb_id> <chain_id> <region_subseq>\
--ccp4_setup_sh <CCP4 SETUP PATH>\
--phenix_setup_sh <PHENIX SETUP PATH>
```

Here, `<pdb_id>` is the four-character PDB identifier of the protein; `<chain_id>` specifies the target chain; and `<region_subseq>` is the subsequence of the amino acid sequence to be guided. The `--ccp4_setup_sh` and `--phenix_setup_sh` arguments point to the respective CCP4 and PHENIX installation scripts.

**NMR.** To guide AlphaFold3 predictions using NOE distance restraints, run the following command in your terminal:

```
python3 run_nmr.py <pdb_id>
```

Here, `<pdb_id>` is the four-character PDB identifier of the protein. Optionally, users may also supply methyl RDC and relaxation files (`--methyl_relax_file` and `--methyl_relax_file`), as well as amide RDC and relaxation files (`--amide_rdc_file` and `--amide_relax_file`), to provide additional experimental guidance during structure prediction.

**Cryo-EM.** To guide AlphaFold3 predictions using cryo-EM ESP maps, run the following command in your terminal:

```
python3 run_em.py <pdb_id> <emdb_id> <renumbered_pdb_file>
<assembly_identifier> --phenix_setup_sh <PHENIX SETUP PATH>
--dihedrals_file <TALOS dihedrals> --noe_restraints_file
<NOE restraints> --noe_pdb_file <renumbered NOE file path> --sequences
<space separated sequence> --counts <#chains for each sequence>
```

Here, `<pdb_id>` is the four-character PDB identifier of the protein; `<emdb_id>` is the EMDB[44] identifier of the cryo-EM ESP map; and `<renumbered_pdb_file>` is the path to a PDB file whose residues have been renumbered to match the 1-indexed amino acid sequences. In other words, if residue A in the input sequence is at index 5, the corresponding residue in the PDB must also be renumbered to index 5. Missing or unresolved atoms are permitted. Because the modeling is performing ab initio, the PDB is not used during guidance. Instead, it is used exclusively for evaluation and for constructing the initial volume mask when the default density-based alignment is selected (as opposed to RMSD alignment to atomic coordinates). Multichain compositions are specified with `--sequences` (space-separated amino acid sequences for each unique chain), and `--counts` (the corresponding copy numbers, in order). `<assembly_identifier>` specifies the biological assembly or asymmetric unit to model. The `-phenix_setup_sh` argument points to the PHENIX installation script. Our framework also supports joint guidance using ESP maps. These optional restraints can be provided via `--dihedrals_file` (for example, TALOS-N dihedrals), `--noe_restraints_file` (NOE distance restraints) and `--noe_pdb_file`, in the similar fashion to the renumbered PDB file; all the residues in these files should be reordered to match the 1-indices in the sequences given to our guidance model.

## Energy-reweighted guidance

Although the ensembles generated using the above method are experimentally faithful, they do not reflect equilibrium population weights implied by an underlying physical energy model. For this, we introduce

an additional reweighting step that maps experimentally faithful ensembles onto approximately Boltzmann-distributed populations. We assume access to an energy function $E(\mathbf{X})$, derived from a force field model that induces a Boltzmann density:

$$\pi(\mathbf{X}) \propto \exp(-\tau \cdot E(\mathbf{X}))$$

where $\tau = (k_B T)^{-1}$; $k_B = 0.001987$ kcal mol$^{-1}$ K$^{-1}$ is the Boltzmann constant; and $T$ is the temperature in Kelvin (K). Using this Boltzmann density, we modify the ensemble statistic in equation (8) to define an energy-weighted ensemble statistic $d_{ij}^{w}$,

$$d_{ij}^{w}(\mathcal{X}) = \sum_{k=1}^{n} w_k(\mathbf{X}^k) \cdot \left\| \mathbf{x}_i^k - \mathbf{x}_j^k \right\|_2 \tag{17}$$

where the normalized Boltzmann weights are given by $w_k(\mathbf{X}^k) = \frac{1}{Z}\exp(-\tau \cdot E(\mathbf{X}^k))$ with $Z = \sum_{i=1}^{n} \exp(-\tau \cdot E(\mathbf{X}^i))$ as the normalization constant. Substituting this weighted ensemble statistic into the NMR log-likelihood yields

$$\log p(D|\mathcal{X}; \mathbf{a}) = -\sum_{(i,j)\in D} \left( \left[ \underline{d}_{ij} - d_{ij}^{w}(\mathcal{X}) \right]_+^2 + \left[ d_{ij}^{w}(\mathcal{X}) - \overline{d}_{ij} \right]_+^2 \right)$$

Guiding with this log-likelihood biases the ensemble toward structures that are both experimentally consistent and energetically favorable.

To balance exploration and energetic refinement during the diffusion process, we anneal $\tau$ over time $t$ according to

$$\tau(t) = \begin{cases} \tau_{\text{high}}, & t \le t_1, \\ \tau_{\text{high}}(1-g) + \tau_{\text{low}} \cdot g, & t_1 \le t < t_2, \\ \tau_{\text{low}}, & t \ge t_2, \end{cases} \tag{18}$$

where

$$g = 3\alpha^2 - 2\alpha^3,\ \alpha = \frac{t - t_1}{t_2 - t_1}\ \tau_{\text{high}}, \tau_{\text{low}}, t_1, t_2 = 35.0, 0.6, 100, 180$$

We set $\tau_{\text{high}} = 35$, $\tau_{\text{low}} = 0.6$, $t_1 = 100$ and $t_2 = 180$. This schedule enables early-stage exploration under a flattened energy landscape, followed by progressive sharpening toward low-energy configurations.

Finally, to mitigate noise and short-range artifacts in the force field energy estimates during late-stage refinement, we apply an exponential moving average (EMA) to the energies after 160 diffusion steps:

$$E\left(\mathbf{X}_t^k\right)^{\text{EMA}} = \kappa \cdot E(\mathbf{X}_t^k) + (1-\kappa) \cdot E\left(\mathbf{X}_{t-1}^k\right)^{\text{EMA}}$$

with smoothing parameter $\kappa = 0.3$. This improves the robustness of energy-based guidance without suppressing structural diversity.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

## Data availability

All structures and metrics reported in this paper are openly available on Harvard Dataverse[62].

## Code availability

All code is openly available on GitHub (https://github.com/sai-advaith/guided_alphafold); the version used for this paper (version 0.9.1) is permanently archived on Zenodo[63].

## References


33. ByteDance AML AI4Science Team et al. Protenix - advancing structure prediction through a comprehensive AlphaFold3 reproduction. Preprint at *bioRxiv* https://doi.org/10.1101/2025.01.08.631967 (2025).
34. Ansel, J. PyTorch 2: faster machine learning through dynamic Python bytecode transformation and graph compilation. 29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems. Vol. 2 (2024).
35. Charlier, B., Feydy, J., Glaunès, J. A., Collin, F.-D. & Durif, G. Kernel operations on the GPU, with Autodiff, without memory overflows. *J. Mach. Learn. Res.* **22**, 1–6 (2021).
36. Mirdita, M. et al. ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).
37. Joosten, R. P., Long, F., Murshudov, G. N. & Perrakis, A. The PDB_REDO server for macromolecular structure model optimization. *IUCrJ* **1**, 213–220 (2014).
38. Wojdyr, M. et al. Gemmi - library for structural biology. *GitHub*. https://github.com/project-gemmi/gemmi (2021).
39. Eastman, P. et al. PDBFixer. *GitHub*. https://github.com/openmm/pdbfixer (2013).
40. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **25**, 1157–1174 (2004).
41. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
42. Lang, P. T., Holton, J. M., Fraser, J. S. & Alber, T. Protein structural ensembles are revealed by redefining X-ray electron density noise. *Proc. Natl Acad. Sci. USA* **111**, 237–242 (2014).
43. Shen, Y. & Bax, A. Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J. Biomol. NMR* **56**, 227–241 (2013).
44. Turner, J. et al. EMDB—the Electron Microscopy Data Bank. *Nucleic Acids Res.* **52**, D456–D465 (2024).
45. Rose, P. W. et al. The RCSB Protein Data Bank: redesigned web site and web services. *Nucleic Acids Res.* **39**, D392–D401 (2010).
46. Virtanen, P. et al. Scipy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).
47. Roney, J. P., Ou, C. & Ovchinnikov, S. Protein diffusion models as statistical potentials. Preprint at *bioRxiv* https://doi.org/10.64898/2025.12.09.69307 (2025).
48. Passaro, S. et al. Boltz-2: towards accurate and efficient binding affinity prediction. Preprint at *bioRxiv* https://doi.org/10.1101/2025.06.14.659707 (2025).
49. Ho, J., Jain, A. & Abbeel, P. Denoising diffusion probabilistic models. *NIPS'20: Proceedings of the 34th International Conference on Neural Information Processing Systems* (eds Larochelle, H., Ranzato, M.,Hadsell, R., Balcan, M. F. & Lin, H.) 33, 6840–6851 (2020). Curran Associates,.
50. Song, Y. & Ermon, S. Generative modeling by estimating gradients of the data distribution. *Proceedings of the 33rd International Conference on Neural Information Processing Systems* (eds Wallach, H. M., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F. & Fox, E. B.) 11918–11930 (Curran Associates, 2019).
51. *International Tables for Crystallography Volume A: Space-group Symmetry* (ed. Hahn, T.) (D. Reidel Publishing Company, 1983).
52. *International Tables for Crystallography, Volume C: Mathematical, Physical and Chemical Tables* 3rd edn (ed Prince, E.) (Springer, 2004).
53. Trueblood, K. et al. Atomic dispacement parameter nomenclature. report of a subcommittee on atomic displacement parameter nomenclature. *Acta Cryst.* **52**, 770–781 (1996).
54. Kunzmann, P., Anter, J. M. & Hamacher, K. Adding hydrogen atoms to molecular models via fragment superimposition. *Algorithms Mol. Biol.* **17**, 7 (2022).
55. Kabsch, W. A solution for the best rotation to relate two sets of vectors. *Acta Cryst.* **32**, 922–923 (1976).

56. Feydy, J. et al. Interpolating between optimal transport and MMD using Sinkhorn divergences. *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics* (eds Chaudhuri, K. & Sugiyama, M.) 2681–2690 (PMLR, 2019).
57. Ingraham, J. B. et al. Illuminating protein space with a programmable generative model. *Nature* **623**, 1070–1078 (2023).
58. Gillespie, R. J. The VSEPR model revisited. *Chem. Soc. Rev.* **21**, 59–69 (1992).
59. Mallat, S. G. & Zhang, Z. Matching pursuits with time-frequency dictionaries. *IEEE Trans. Signal Process.* **41**, 3397–3415 (1993).
60. Kingma, D. P. & Ba, J. Adam: a method for stochastic optimization. Preprint at *arXiv* https://doi.org/10.48550/arXiv.1412.6980 (2014).
61. Agirre, J. et al. The CCP4 suite: integrative software for macromolecular crystallography. *Acta Crystallogr. D Struct. Biol.* **79**, 449–461 (2023).
62. Bronstein, A. et al. Experiment-guided AlphaFold3 resolves accurate protein ensembles. Harvard Dataverse. https://doi.org/10.7910/DVN/PLYUHN (2025).
63. sai-advaith/guided_alphafold: experiment-guided AlphaFold (v0.9.1). *Zenodo*. https://doi.org/10.5281/zenodo.17307005 (2025).

## Acknowledgements

A. Marx acknowledges the financial support of the Helmsley Fellowships Program for Sustainability and Health. A.M.B. and P.S. are supported by the Institute of Science and Technology Austria Internal Project Call grant Generative Protein NMR. S.V. was supported in part by funding from the Eric and Wendy Schmidt Center at the Broad Institute of MIT and Harvard.

## Author contributions

A.M.B., A. Marx, S.V. and A. Maddipatla conceived the study, designed the algorithmic framework and shaped the overall scientific direction. A. Maddipatla led the implementation of the open-source codebase, with contributions from N.S.B., M.B. and V.M. A. Maddipatla and N.S.B. conducted the X-ray crystallography case studies. M.B. and S.V. conducted the NMR case studies. V.M. conducted the cryo-EM case studies. A. Marx selected the X-ray case studies, P.S. selected the NMR case studies and A.M.B., together with A. Marx, selected the cryo-EM case studies. A.M.B., A. Marx, P.S. and S.V. supervised the project. A. Maddipatla led the writing of the Methods section. All authors contributed to writing the main text. All authors reviewed and approved the final paper.

## Funding

Open access funding provided by Institute of Science and Technology (IST Austria).

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41587-026-03166-5.

**Correspondence and requests for materials** should be addressed to Sanketh Vedula, Paul Schanda, Ailie Marx or Alex M. Bronstein.

**Peer review information** *Nature Biotechnology* thanks Alexandre Bisson, Jianping Fu and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

**Reprints and permissions information** is available at www.nature.com/reprints.

nature portfolio

Corresponding author(s): Alex M. Bronstein, Ailie Marx, Paul Schanda, Sanketh Vedula

Last updated by author(s): May 4, 2026

# Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

## Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

| n/a | Confirmed |
|---|---|
| ☐ | ☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☒ | ☐ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☒ | ☐ The statistical test(s) used AND whether they are one- or two-sided<br>*Only common tests should be described solely by name; describe more complex techniques in the Methods section.* |
| ☒ | ☐ A description of all covariates tested |
| ☒ | ☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐ | ☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☒ | ☐ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted<br>*Give P values as exact values whenever suitable.* |
| ☒ | ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☒ | ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☒ | ☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated |

*Our web collection on statistics for biologists contains articles on many of the points above.*

## Software and code

Policy information about availability of computer code

**Data collection**

No new experimental data were collected for this study; all input data are obtained from public databases and pre-existing publications.
Public-database queries and parsing were performed using:
1. ColabFold v1.5.5 MMseqs2 API (custom HTTP wrapper) for multiple sequence alignment retrieval
2. PDB-Redo, RCSB PDB, BMRB, and EMDB online interfaces for reference atomic models, NMR restraints, chemical shifts, and cryo-EM electrostatic potential maps
3. Gemmi v0.6.5 for chain identification and structure parsing
4. pynmrstar for parsing NMR-STAR restraint files
5. TALOS-N for inferring backbone dihedral angles from chemical shifts

**Data analysis**

Modelling and ensemble generation:
1. Protenix v0.2.0 (open-source PyTorch re-implementation of AlphaFold3) and AlphaFold3 (official weights and source code) for sequence-conditioned st v0.1.0ructure prediction
2. PyTorch v2.3.1+cu121 (with CUDA 12.1) for forward models, gradients, and guidance
3. PyKeOps v2.3 for memory-efficient density computation on large structures

Structure preparation and force-field relaxation:
1. PDBFixer v1.9.0 (with OpenMM force-field templates) for missing-atom imputation
2. AMBER-14 force field for energy minimisation

Crystallographic tools:
1. Phenix 1.21.2

2. CCP4 8.0

Energy / ensemble reweighting:
1. ProteinEBM v0.1.0, derived from Boltz-2, for per-structure energy estimation

Numerical utilities:
1. SciPy v1.15.0 (scipy.spatial.ConvexHull) for ESP-map mask construction

Custom code:
guided_alphafold, openly available at https://github.com/sai-advaith/guided_alphafold; the version used for the results in this paper (v0.9.1) is permanently archived on Zenodo (https://doi.org/10.5281/zenodo.17307005).

All computations were performed on NVIDIA H100 and L40 GPUs running Debian GNU/Linux 12.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

## Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

All data underlying this study are publicly available; no datasets are restricted or available only on request.

Generated data. All structural ensembles and evaluation metrics produced for this study are openly deposited on Harvard Dataverse: https://doi.org/10.7910/DVN/PLYUHN.

Public-repository inputs. All experimental ground-truth data are obtained from established public repositories. The complete lists of accession codes used as inputs are provided in SI Tables 2 (NMR benchmark) and 3 (X-ray benchmark); the cryo-EM systems are listed below.

1. Protein Data Bank (RCSB PDB / PDB-Redo) - atomic models and crystallographic reflection data. Specific entries used are listed in SI Tables 2 and 3.
2. Biological Magnetic Resonance Data Bank (BMRB) - chemical shifts and NMR-STAR restraint files. Entries are cross-referenced to the corresponding PDB IDs in SI Table 2.
3. Electron Microscopy Data Bank (EMDB) - cryo-EM electrostatic potential maps. Entries used: EMD-30622, EMD-41877, EMD-41880, and EMD-50436.

Benchmark datasets. Drawn from prior published work and used as released by the original authors:

1. 100-NMR-spectra database (Klukowski et al. 2024) - 83-protein subset.
2. 8 AlphaFold3-mispredicted peptides (Vedula et al. 2025; McDonald et al. 2023).
3. Altloc catalogue (Rosenberg et al. 2024).

Human, clinical, or third-party data. None used.

Restrictions. None. All deposited data are released under open terms, and all queried public repositories are freely accessible.

## Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

| | |
|---|---|
| Reporting on sex and gender | N/A - No human participants |
| Reporting on race, ethnicity, or other socially relevant groupings | N/A - No human participants |
| Population characteristics | N/A - No human participants |
| Recruitment | N/A - No recruitment performed |
| Ethics oversight | N/A - no human-subjects research; no ethics-board approval needed |

Note that full information on the approval of the study protocol must also be provided in the manuscript.

# Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

# Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| | |
|---|---|
| Sample size | Sample size was not statistically pre-determined; benchmarks were assembled to match published evaluation sets used in prior NMR / X-ray / cryo-EM modelling work. The NMR benchmark contains n = 91 proteins (83 from the 100-NMR-spectra subset of Klukowski et al. 2024 plus 8 AlphaFold3-mispredicted peptides from Vedula et al. 2025 and McDonald et al. 2023). The X-ray benchmark consists of the structures listed in SI Table 3. The cryo-EM evaluation comprises three systems: insulin receptor IR-B in the apo state and IR-B with three bound IGF2 factors, an amyloid-β fibril, and a human RIPK3 amyloid fibril. |
| Data exclusions | No structures were excluded from the benchmarks. |
| Replication | All findings are reported on the full benchmark sets. Reproducibility of altloc and ensemble generation across independent random initialisations is quantified in SI Table 5, and SI Fig. 9 shows convergence as a function of batch size. |
| Randomization | N/A. There are no experimental groups to allocate or randomise. |
| Blinding | N/A. Evaluation is fully computational. |

# Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

## Materials & experimental systems

| n/a | Involved in the study |
|---|---|
| ☒ | ☐ Antibodies |
| ☒ | ☐ Eukaryotic cell lines |
| ☒ | ☐ Palaeontology and archaeology |
| ☒ | ☐ Animals and other organisms |
| ☒ | ☐ Clinical data |
| ☒ | ☐ Dual use research of concern |
| ☒ | ☐ Plants |

## Methods

| n/a | Involved in the study |
|---|---|
| ☒ | ☐ ChIP-seq |
| ☒ | ☐ Flow cytometry |
| ☒ | ☐ MRI-based neuroimaging |

## Plants

| | |
|---|---|
| Seed stocks | N/A. No plant material used. |
| Novel plant genotypes | N/A. No plants generated or studied. |
| Authentication | N/A. No plants used. |