ToposBio™

Topos-1

An all-atom foundation model for intrinsically disordered proteins

Abstract

Artificial intelligence is revolutionizing biomedicine due to groundbreaking advances in protein structure prediction, but many diseases remain out of reach because the proteins involved are too dynamic for current tools. While models like AlphaFold aim to predict a single, static 3D conformation, we introduce Topos-1, an all-atom generative model that predicts realistic conformational ensembles for the roughly one-third of the human proteome that lacks a single well-defined structure. Dynamic, flexible, and disordered proteins are at the root of many currently incurable neurodegenerative diseases and aggressive forms of cancer, including Alzheimer's, Parkinson's, and prostate cancer. Topos-1 sets a new benchmark for predicting conformational ensembles, outperforming AlphaFold, BioEmu, Chai, Boltz, and domain-specific models across extensive experimental and computational evaluations. Our exhaustive evaluations include both large-scale global properties as well as small-scale local properties that can be crucial for structure-based drug design. We also report progress using Topos-1 to prospectively screen and design novel small molecule drug candidates, showing that computational predictions are in excellent agreement with experimentally measured drug potencies for a fully disordered target. Topos-1 shows clear power-law scaling and will improve even further with additional data, compute, and model size. By integrating Topos-1 into our AI-powered drug design platform, we accelerate traditional drug design approaches that rely on slow and expensive computational and experimental tools. More importantly, Topos-1 also unlocks a capability to design new therapeutic strategies that address so-called "undruggable" diseases.

1. Introduction

Biomolecules are dynamic, but most experimental measurements only resolve the static elements of their structure [1, 2]. Proteins do not occupy a single state, but rather an ensemble of conformations that interconvert on timescales fast and slow [2, 3]. The advances in machine learning-based structure prediction models over the last five years [4–7] have exploited the extant experimental data [8] to predict the structures of largely static regions of biomolecules with an accuracy that approaches experimental resolution. However, many proteins consist of, or contain sequences that have, a high degree of *intrinsic disorder*, meaning that they do not adopt a single well-defined state and are not resolved in the most widely used experimental characterization methods [9, 10]. As a result, models like AlphaFold often do not produce physically reasonable predictions for these highly dynamic structural ensembles [11, 12]. Of course, we emphasize that the most widely used structure prediction models are not designed to sample conformational ensembles and that intrinsically disordered proteins (IDPs) constitute a negligible fraction of their training data [8, 13, 14]. Neglecting highly flexible regions of proteins is untenable for many important therapeutic targets. Indeed, estimates suggest nearly a third of the human proteome adopts no well-defined structure [15–17].

This class of highly flexible proteins is implicated in major neurodegenerative diseases, including Alzheimer's, Huntington's, ALS, and Parkinson's [18–21]. A variety of aggressive cancers have also resisted therapeutic interventions, stymied by a lack of insight into disordered domains [19, 22]. Even within structured proteins, dynamic regions contribute to functional elements and are widely implicated in molecular recognition [10]. For example, antibody complementarity-determining regions are highly flexible and pose a challenge to the current paradigm of structure prediction models [23]. Given the importance of flexible protein motifs in human diseases, new tools are needed to better capture these conformational ensembles. Efficient and accurate prediction of the conformational space of these flexible proteins would enable the application of drug design approaches that cannot be used without physically accurate structures [24, 25].

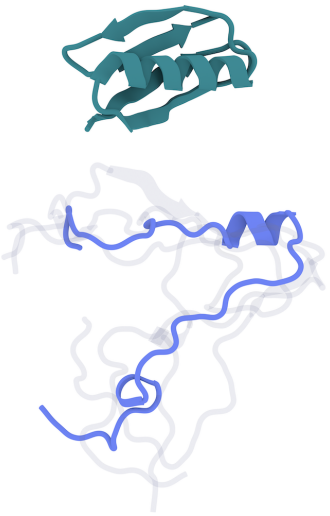

Figure 1 | Existing protein structure prediction models excel for structured proteins, which exhibit limited conformational diversity (top). Topos-1 predicts the complex conformational ensembles of *disordered* proteins, which represent roughly a third of the human proteome (bottom).

Here we introduce Topos-1, an all-atom, large-scale generative model that is purpose-built to efficiently produce physically realistic conformations of intrinsically disordered proteins. Topos-1 achieves state-of-the-art performance on experimental and computational benchmarks. Unlike models trained primarily on sequences of structured proteins and data from the Protein Data Bank [8], we designed both the architecture of our model and the training data generation pipeline to excel on highly dynamic motifs. By scaling training to include proprietary physics-based simulations and experimental data of IDPs, we show that Topos-1 generates ensembles that generalize to novel IDPs and obtains best-in-class accuracy on both global and local structural properties. Additionally, Topos-1 shows power-law scaling with model size and training data volume, demonstrating continued performance gains as compute and data increase. Topos-1 not only outperforms state-of-the-art structure prediction models (which again, we emphasize are not designed for this task) but also a variety of specialized IDP models.

In this report, we first demonstrate that Topos-1 produces conformations that set a new standard when evaluated on experimental observables and molecular dynamics (MD) simulation data. This is the case for both local and global properties, and Topos-1 is the top-performing all-atom model across all physical observables that we measure. Moreover, Topos-1 succeeds in contexts where conventional structure prediction tools struggle, including showing conformational sensitivity to perturbations in sequence. We show that ensembles generated by our model are a powerful tool with which to computationally screen and *design* small molecule drug candidates. We design and synthesize novel, undisclosed compounds that bind the intrinsically disordered androgen receptor N-terminal domain, a currently intractable prostate cancer target. We show that our computational predictions are in excellent agreement with the measured potencies that rival previous clinical candidates.

The improvement trajectory for Topos-1 is clear: the model demonstrates power-law scaling with data, compute, and model size. By continuing to collect both computational and experimental data that shed light on flexible protein motifs, we anticipate that Topos-1 will continue to set the standard for predicting the dynamical fluctuations of proteins. Model inference can be parallelized over arbitrary numbers of GPUs, massively accelerating sampling in comparison to physics-based methods such as molecular dynamics. Our best estimates suggest that Topos-1 generates conformational ensembles 1,000×

faster than atomistic molecular dynamics in explicit solvent. This acceleration unlocks a new paradigm for conformation-aware molecular design.

In this Technical Report,

1. We introduce **Topos-1**, an all-atom generative model that produces physically realistic conformational ensembles for intrinsically disordered proteins, *outperforming AlphaFold-2, Boltz-2, Chai-1, BioEmu, and prior IDP-specific methods* on all experimental benchmarks.
2. We demonstrate that Topos-1 generates tens of thousands of all-atom conformations in minutes—**orders of magnitude faster** than conventional molecular dynamics—and exhibits predictable improvements with increased training data and model capacity.
3. We develop a **physics-based data engine** and training pipeline to build the largest IDP-focused all-atom simulation corpus to date, which consists of both large-scale atomistic simulations and internal experimental measurements for model improvement and validation.
4. We show that Topos-1 captures **disease-relevant** conformational states that are not captured by standard structure prediction models, which we highlight with a study of α-synuclein, a key protein in Parkinson's disease whose misfolding and aggregation are central to pathology.
5. We demonstrate the **practical downstream utility** of Topos-1 for drug design in a case study on the androgen receptor, a clinically validated driver of aggressive prostate cancer, wherein potencies predicted using conformational ensembles generated by Topos-1 yield ligand rankings consistent with potencies measured experimentally in cell-based assays.

2. Results

2.1. Topos-1 outperforms AlphaFold-2, Boltz-2, Chai-1, and BioEmu on literature experimental IDP benchmarks

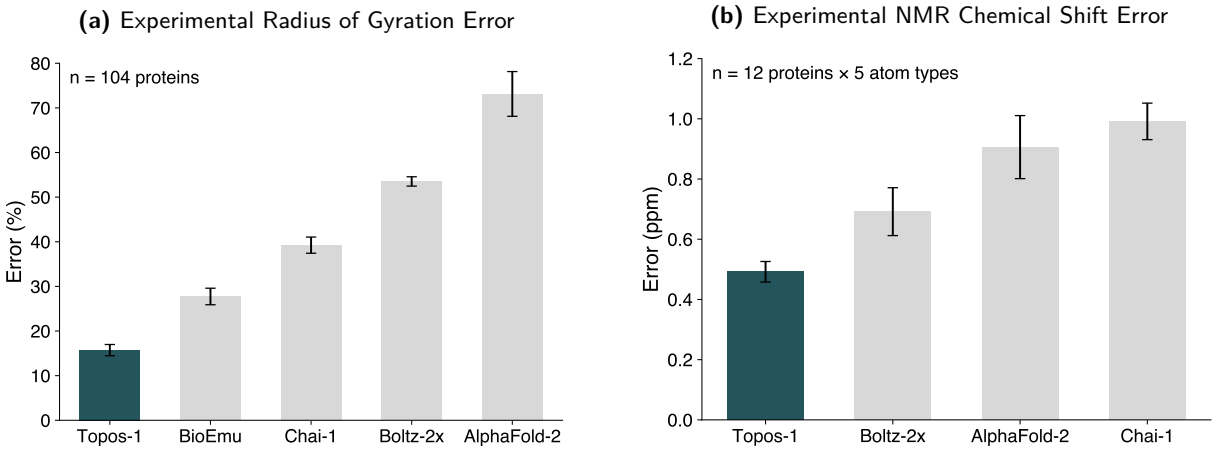

Figure 2 | Topos-1 outperforms leading protein structure prediction models on experimental benchmarks for IDPs in both global, large-scale properties and local geometric features. (a) Mean absolute percent error in radius of gyration (R_g). Topos-1 reduces the error in R_g by 43–79% relative to baseline models. Experimental R_g values determined via SAXS. **(b)** Mean absolute error in NMR chemical shift predictions across 12 IDPs and 5 atom types (carbonyl C, C_α, C_β, N, H). Topos-1 achieves 29–50% lower chemical shift prediction error than baseline models. Error bars denote the standard error of the mean across proteins, computed from $n = 200$ samples per protein for each model.

Intrinsically disordered proteins do not adopt a single dominant structure but rather an ensemble of conformations. Consequently, measuring the accuracy of a generated ensemble requires examining both global properties and local backbone and side chain conformational statistics. To evaluate global conformational properties, for each model we computed the ensemble-averaged radius of gyration R_g, a metric reflective of the spatial extent and compaction of an ensemble, and

compared it to experimental values determined with small-angle X-ray scattering (SAXS). To evaluate local conformational properties, for each model we compared conformational changes quantified by chemical shifts from NMR experiments.

To ensure a robust evaluation, we curated a dataset of SAXS measurements of IDPs, following previous work [26–28]. We restricted the analysis to IDPs with length $\leq$ 200 residues. All proteins in this evaluation set were excluded from Topos-1 training, along with any other sequences with 50% or more sequence similarity as measured with MMseqs2 linclust [29] (see Supplementary Section B.2).

For each model, we constructed ensembles by generating multiple conformational samples per protein, which were subsequently aggregated to compute ensemble-averaged observables; we generated 200 samples per protein. We evaluated both Boltz-2 and Boltz-2x, a variant that incorporates physics-based potentials. Boltz-2x outperformed Boltz-2 in both metrics, so in Fig. 2 we report only the results from Boltz-2x. Throughout this Technical Report, we refer to AlphaFold-Multimer v2.3 [30] as AlphaFold-2 (see Supplementary Section B.1.3 for further discussion).

For each protein, R_g was computed for each generated conformation and averaged. Errors were computed per protein and aggregated across proteins (see Fig. 2a and Supplementary Section B.1). Across 104 IDPs, Topos-1 reduces normalized R_g error by approximately 43% relative to BioEmu, 60% relative to Chai-1, 71% relative to Boltz-2x, and 79% relative to AlphaFold-2. These results indicate Topos-1 captures global ensemble properties of IDPs more accurately than these other models.

We assessed local backbone agreement with experiment by comparing predicted chemical shifts against experimentally derived NMR chemical shifts. We used a subset of IDPs for which high-quality NMR data was available. Chemical shifts were computed as the ensemble-averaged mean absolute error across backbone atom types (see Fig. 2b and Supplementary Section B.1). BioEmu does not predict side chain atoms and was therefore excluded from the chemical shift analysis. Across these IDPs, Topos-1 achieves approximately 29% lower chemical shift prediction error than Boltz-2x, 46% lower than AlphaFold-2, and 50% lower than Chai-1. We report further positive results comparing Topos-1 against additional models trained on coarse-grained MD data in Supplementary Section B.1.

2.2. Topos-1 generates physically valid ensembles that are consistent with molecular dynamics simulations

For many functionally important and disease-relevant proteins that lack a stable 3D structure, molecular dynamics (MD) simulations provide crucial insights into the conformational ensemble. Here, we show that Topos-1 generates conformations that closely match MD-derived ensembles while requiring orders of magnitude less computation. Whereas MD simulations typically require days of GPU time, Topos-1 produces comparable ensembles in minutes. To quantify agreement with MD, we evaluated ensemble-averaged global structural observables for 270 test IDPs (Fig. 3). The IDP ensembles were generated using internal MD simulations. IDP sequences were selected from a diverse set of intrinsically disordered regions (IDRs) in semi-ordered human proteins. Across all global metrics described in Supplementary Section B.3 (R_g, root-mean-squared fluctuations (RMSF), and secondary-structure propensities), Topos-1 outperforms all baseline models, underscoring the importance of architectures and training data explicitly designed to capture the statistics of intrinsically disordered conformational ensembles.

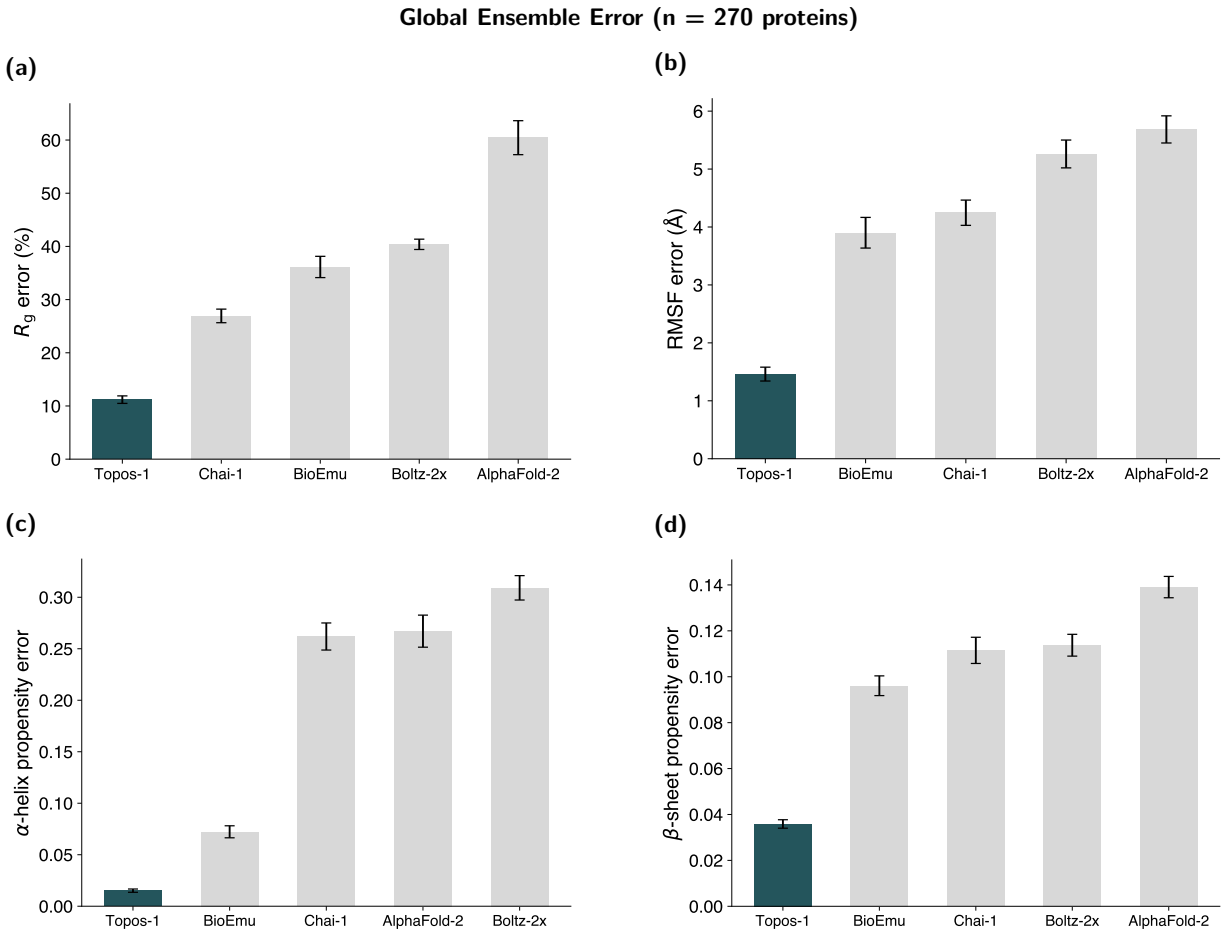

Figure 3 | Topos-1 accurately reproduces global ensemble properties obtained from physics-based molecular dynamics simulations. (a) Mean absolute percent error in radius of gyration (R_g). (b) Mean absolute error in root-mean-squared fluctuations (RMSF). (c) Mean absolute error in α-helix propensity, measured as the ensemble-averaged fraction of residues in helical conformations. (d) Mean absolute error in β-strand propensity, measured as the ensemble-averaged fraction of residues in β-strand conformations. All metrics are computed relative to molecular dynamics reference ensembles and averaged over 200 conformations per protein for 270 IDPs in the Topos-1 test set. Error bars indicate the standard error of the mean across IDPs.

Beyond global statistics, accurate prediction of local structure is essential for describing how IDPs interact with themselves, other biomolecules, and small-molecule ligands. To assess the ability of Topos-1 to reproduce MD-derived ensembles at atomic resolution, we evaluated ensemble-averaged agreement for backbone dihedral angle distributions and side chain rotamer populations (Fig. 4). These metrics are described in more detail in Supplementary Sections B.3.4 and B.3.5. Models trained primarily on static protein structures—including AlphaFold-2, Boltz-2x, and Chai-1—exhibit poor agreement with MD backbone ensembles, highlighting their limited capacity to capture local conformational heterogeneity. BioEmu, which incorporates MD data during training, performs substantially better than structure-only baselines. Consistent with the global-property analysis, Topos-1 achieves the lowest error across all local metrics, demonstrating its ability to capture both backbone and side chain statistics in IDPs. Because BioEmu does not generate side chain coordinates, it was excluded from the rotamer-based evaluation.

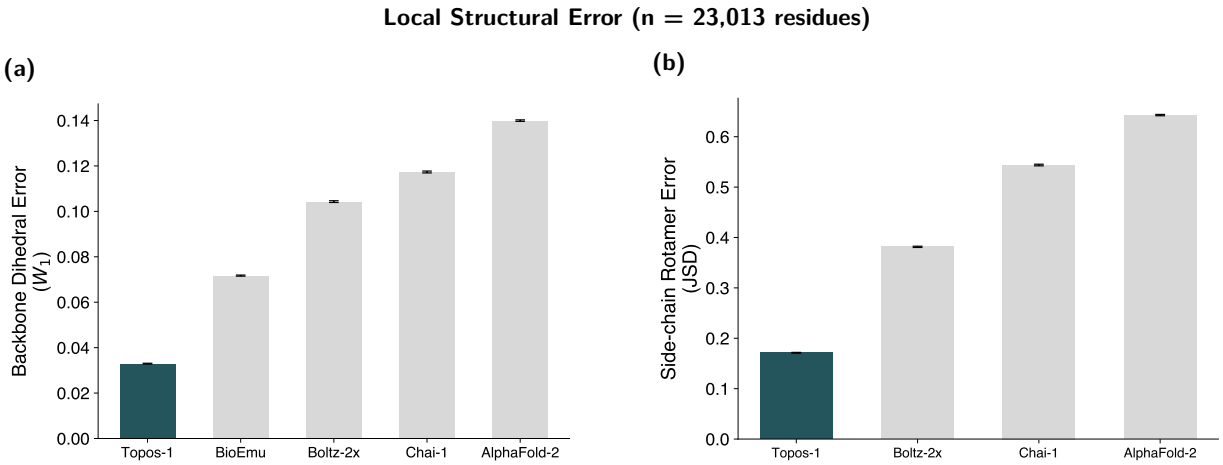

Figure 4 | Topos-1 accurately captures local backbone and side chain conformational ensembles. (a) Circular Wasserstein-1 distance (W_1) between backbone dihedral angle distributions of generated ensembles and molecular dynamics references, computed for ϕ and ψ angles and averaged across all residues from all proteins in the test set. (b) Mean Jensen-Shannon divergence (JSD) between side chain rotamer-state distributions of generated ensembles and MD references, averaged over all residues with defined χ angles from all proteins in the test set. Error bars indicate the standard error of the mean. Rotamer-based comparisons are restricted to all-atom models that explicitly generate side chain coordinates. Both calculations are described in greater detail in Supplementary Sections B.3.4 and B.3.5, respectively.

To further assess the agreement between ensembles generated with Topos-1 and MD simulation, we compared the collective variables (CVs) obtained using a nonlinear dimensionality reduction method developed for protein conformational analysis, ELViM [31] (Fig. 5). ELViM provides a useful framework for visualizing and comparing high-dimensional conformational ensembles and has previously been applied to visualize and compare IDP ensembles [32, 33]. The resulting CVs encode both local and global conformational features by embedding ensemble-wide pairwise similarities derived from backbone C_α coordinates or backbone dihedral angles. For each model, we estimated probability density functions in the ELViM CV space using a kernel density estimation and quantified the agreement with MD reference ensembles via the Jensen-Shannon divergence (JSD) (see Supplementary Section B.3.5). The JSD ranges from 0 to $\ln 2$, where 0 indicates indistinguishable distributions and $\ln 2$ indicates no shared information. As shown in Fig. 5, Topos-1 achieves the lowest mean JSD relative to MD across the test set, outperforming all other evaluated methods. In Fig. 5a, the CV distributions for six representative IDPs are shown for combined embeddings and for each model separately. The leftmost column shows joint embeddings. The second and third columns from the left show embeddings for the MD reference and Topos-1, where there is a strong agreement between CV distributions. IDP 6 (row 6) serves as an example in which all models struggle to match the MD reference. Figure 5 shows the JSD distributions for each model sorted by the mean JSD across the test set. Importantly, the trends in the CV space are fully consistent with those observed in the global and local ensemble metrics (see Figs. 3 and 4).

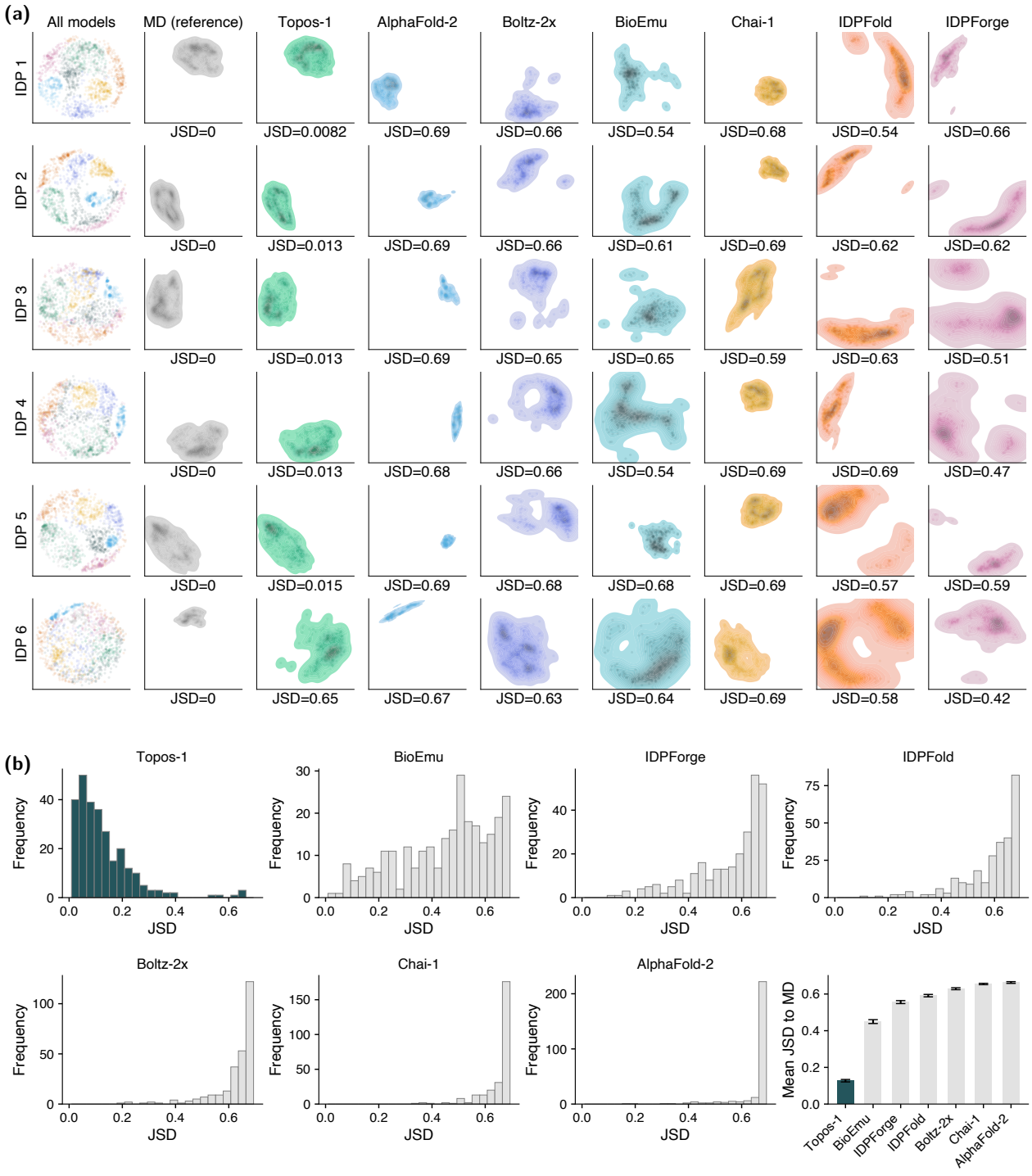

Figure 5 | Conformational landscapes generated by Topos-1 embed near molecular dynamics data in unsupervised landscape analysis. (a) ELViM-based collective variable (CV) embeddings comparing conformational ensembles generated by Topos-1 and baseline models against molecular dynamics (MD) reference ensembles for representative intrinsically disordered proteins (IDPs). Embeddings are computed using a shared reference space defined by the combined conformations from all models. Topos-1 exhibits the strongest overlap with MD-derived CV distributions, whereas structure-based prediction models (AlphaFold-2, Boltz-2x, Chai-1) produce overly compact ensembles and IDP-specific models show limited overlap despite greater dispersion. (b) Jensen-Shannon divergence (JSD) between model-generated and MD-derived CV distributions across 270 IDPs in the test set. Histograms show the distribution of JSD values, and the bar chart summarizes the mean JSD with the standard error of the mean for each model. Topos-1 achieves the lowest divergence from MD, consistent with global and local ensemble metrics.

2.3. Topos-1 performance improves with scale

Beyond strong domain performance, a crucial requirement of modern deep learning methods is scalability [34]. Empirical evidence suggests that general approaches that scale predictably and favorably with increasing compute tend to outperform more tailored methods in the long run [35]. Recent work on scaling protein models [7, 36, 37] has shown promising advances by training increasingly large models on distillation data from existing structure prediction models. Synthetic data have become necessary because high-quality experimental protein structure data from sources like the PDB [8] have been exhausted.

Training signal for Topos-1 comes from both synthetic and experimental data, and we hypothesized that scaling computationally derived ensembles would improve performance consistent with scaling laws in other domains. To quantify the scaling behavior of Topos-1, we compared 12 independently trained models spanning 3 model sizes and 4 distinct scales of data. We constructed four datasets by splitting our proprietary dataset by protein and subsampling progressively larger subsets, reflecting different synthetic data generation budgets. To ensure a consistent evaluation protocol, all models were evaluated on the minimum validation loss achieved after training on 4 A100 GPUs to convergence.

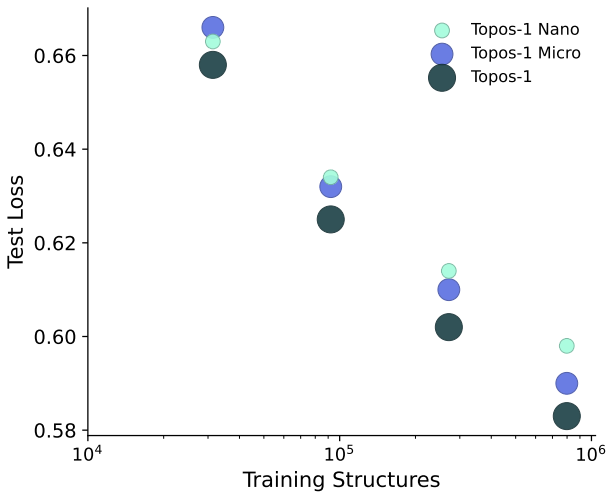

Figure 6 | Topos-1 shows power-law scaling, improving in performance with both model size and training data, demonstrating clear gains with increasing total compute. We performed 12 independent training runs with models spanning three sizes (136M, 182M, 319M parameters) and four training set sizes. Increasing either model size or training set size yields consistent reductions in test loss.

As shown in Fig. 6, increases in synthetic data generation compute and model size both lead to consistent and systematic reductions in validation loss. Importantly, there are no signs of convergence or diminishing returns for the scaling of data generation. Together, these results indicate that synthetic data generation compute is an effective lever for improving model performance, motivating further investment in large-scale data generation using our proprietary physics-based simulation engine.

3. Evaluation of Topos-1 With Internal Experimental Data

To complement computational ensemble predictions, we generated and analyzed results collected using an internal collection of IDPs. Our experiments probe structural features across multiple length scales to provide additional model evaluation data. In this section, we present data obtained with SAXS and circular dichroism (CD), two methods we use to provide additional ensemble-averaged observables to further inform Topos-1. SAXS measurements enable the evaluation of intramolecular distance distributions in IDPs beyond what is captured by R_g alone, and CD measurements quantify protein secondary structure content comprised of α-helix, β-strand, and coil features. In addition to evaluating model performance, our internal SAXS and CD experiments can be incorporated as supervision to train models whose ensemble statistics match experimental observables, or used to guide ensemble generation at inference time.

3.1. Topos-1 ensembles are consistent with internally-generated SAXS data

In addition to the curated dataset of literature SAXS results analyzed in Section 2.1, we performed SAXS measurements on an internal IDP library and compared the observed solution-phase structural properties with predictions from Topos-1 and other models. These comparisons quantify overall agreement while also revealing specific structural regimes where models deviate from experiment. Ensembles generated by Topos-1 for HOXB7, Katacalcin, PHF21A, and Exendin-4 are among the most accurate across the various models evaluated (Fig. 7). In particular, the tails of the pair-distance distribution function $P(r)$ for Topos-1 decay at rates nearly identical to experiment, indicating that Topos-1 achieves the proper balance between extended and collapsed conformers. In contrast, other models produce overly compacted or extended structures despite also

providing reasonable mid-range agreement. Departures from experiment at different distance regimes were readily identified by this analysis, even in cases where ensembles were indistinguishable based on the global R_g metric alone. Since SAXS is especially sensitive to these intramolecular distances, we leverage these measurements to develop more discriminative performance benchmarks and guide the acquisition of supplemental training data aimed at further enhancing the model.

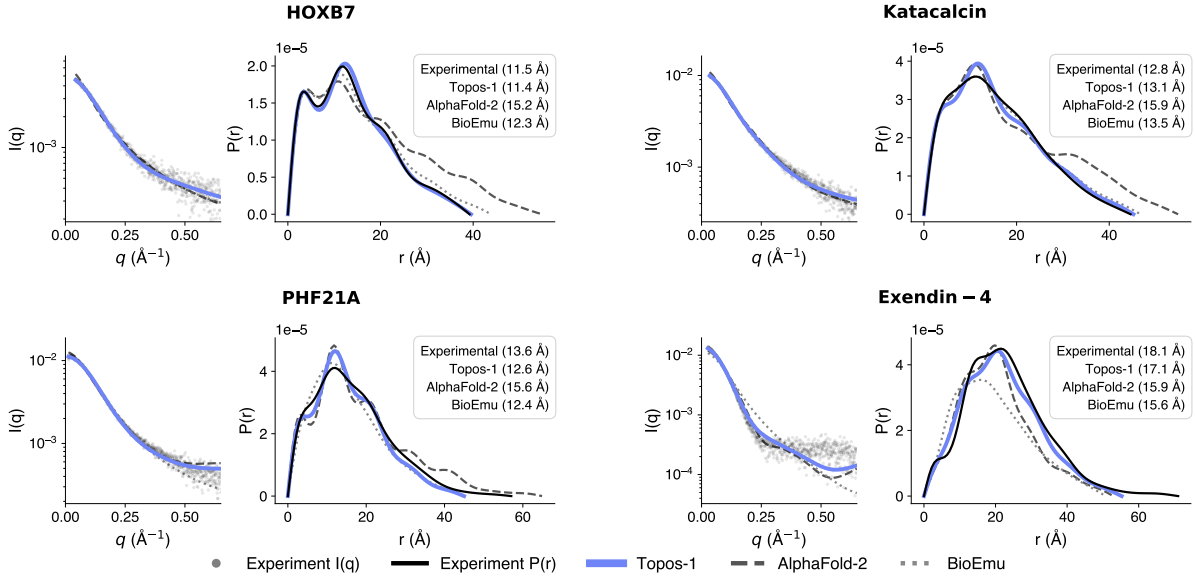

Figure 7 | SAXS analysis of selected IDRs. For each protein, the left panel shows the experimental scattering intensity I(q) versus the scattering vector q, overlaid with ensemble-averaged I(q) curves predicted by Topos-1, AlphaFold-2, and BioEmu. For each protein, the right panel shows the pair-distance distribution function P(r) for experimental data and each computational model. Values in parentheses in the legends denote the corresponding R_g. Topos-1 ensembles closely match experimental P(r) distributions across all distance regimes. The reported R_g values demonstrate that this single metric alone cannot distinguish between the accuracy of different models; the full P(r) distribution reveals important differences in conformational ensemble accuracy. We refer the reader to Fig. S1 for an expanded version of this figure.

3.2. Topos-1 ensembles are consistent with internally-generated circular dichroism data

Despite lacking a well-folded stable conformation, IDPs often exhibit varying degrees of secondary structure, defined as the percent of α-helix, β-strand, and coil-like structures. Secondary structure content can be quantified using CD spectroscopy, which we performed on our proprietary collection of IDPs (see Section A.4 for details). As illustrated in Fig. 8, ensembles generated by Topos-1 are in reasonable agreement with CD spectra. Still, minor disagreements with experiment are present, as is the case for Exendin-4, a peptide with higher helical content. These deviations from experiment provide informative signals for improving future models.

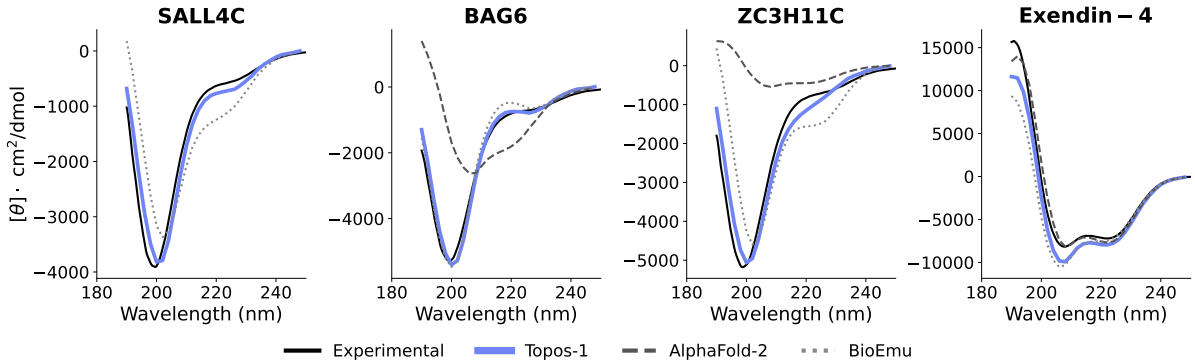

Figure 8 | Topos-1 ensembles are consistent with in-house circular dichroism data. Shown are three examples of in-house collected circular dichroism (CD) datasets for different proteins with different qualitative properties, illustrating strengths and possible areas of improvement. Topos-1 shows agreement with the spectra of the highly disordered proteins SALL4C and ZC3H11C, on which existing structure prediction models produce secondary structure inconsistent with measurement. On Exendin-4, a protein with moderate propensity for secondary structure, Topos-1 performs competitively but shows less strong agreement than AlphaFold-2, which is designed for structured proteins. AlphaFold-2 was omitted from SALL4C due to highly unphysical results. We refer the reader to Section A.4 for details on data collection and processing.

4. Designing Novel Drug-like Compounds with Topos-1

The successful application of Structure-Based Drug Discovery (SBDD) approaches to IDPs requires an efficient method to characterize their conformational space. Despite developments in force fields and enhanced sampling techniques, obtaining sufficiently distinct conformational minima of IDPs within a timeframe and computational cost that is compatible with SBDD remains a significant challenge. To test the performance of Topos-1 against highly challenging targets outside of our training data, we utilized Topos-1 to generate conformations of α-synuclein and the androgen receptor. For each target, we show that Topos-1 provides ensembles useful for SBDD with orders-of-magnitude speedup relative to MD.

4.1. Case study: Topos-1 captures the behavior of the Parkinson's disease target α-synuclein

α-synuclein is an intrinsically disordered protein that is a major therapeutic target for Parkinson's disease (PD) [38, 39]. In the 30 years since the link between α-synuclein and PD was established, no FDA-approved drugs have emerged for this target. We utilized Topos-1 to generate conformations of α-synuclein and compared our generated ensemble to those obtained with BioEmu and AlphaFold-2. To compare to a reference ensemble, we used literature values that include experimentally determined secondary structure propensities obtained via analysis of 2D NMR [40] as well as high-quality reference MD data [33]. Figure 10 shows the ensemble-averaged mean and per-residue mean absolute error of α-helical propensities in α-synuclein, relative to experimental data. Figures S6 and S7 show the calculated β-strand and random coil propensities, respectively, and again demonstrate excellent agreement between Topos-1 and MD simulation [33].

As can be seen in Fig. 10a, the mean absolute error of predicted α-helical propensity by Topos-1 is very low, comparable to that of high-quality MD simulations. BioEmu and AlphaFold-2 both have significantly larger prediction errors. We note that the large error for AlphaFold-2 can be explained by the presence of membrane-bound α-synuclein structures in the Protein Data Bank, a significant source of training data for AlphaFold-2, which are not representative of an isolated α-synuclein monomer in solution. Figure 10b breaks down this error by residue, showing that while there are a few localized

Figure 9 | α-synuclein is an intrinsically disordered protein that is implicated in Parkinson's disease. A subset of the numerous possible conformations of α-synuclein predicted by Topos-1 is shown in blue and gray.

hot spots of over-prediction, all but one of these are present in the ensemble predicted by MD. Taken together, Topos-1 exhibits significantly better agreement with both MD and experiment than BioEmu and AlphaFold-2 in predicting the total and per-residue secondary structure propensity of α-synuclein. This is critical for generating the correct ensemble of protein structures.

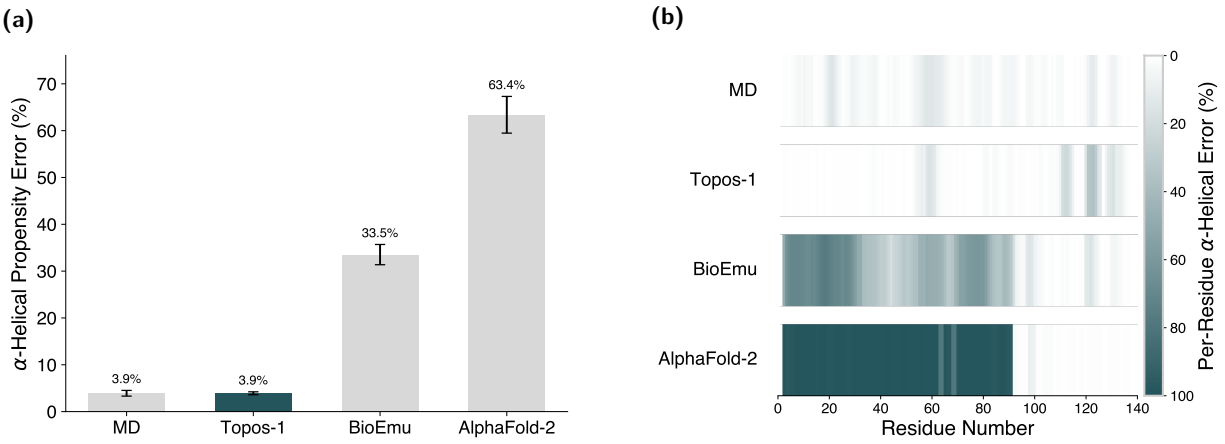

Figure 10 | Topos-1 generates conformational ensembles of Parkinson's disease target α-synuclein that outperform state-of-the-art structure prediction models in experimental fidelity. (a) Mean absolute error in helical propensity relative to experiment averaged across the α-synuclein sequence ($n = 140$ residues). (b) Residue-resolved mean absolute error in helical propensity reveals contiguous regions of excessive helical propensity. Experimental measurements indicate near-zero helical propensity across the sequence [40].

Another important metric in evaluating the physical validity of a generated conformational ensemble is the backbone dihedral angle distribution of the protein. Figure 11 shows the circular Wasserstein-1 distance (see Supplementary Section B.3.4) from the reference backbone dihedral angle distribution obtained from MD simulation of α-synuclein. As can be seen in Fig. 11, Topos-1 exhibits a significant improvement in prediction error compared to BioEmu and AlphaFold-2, with residue-averaged distances of 0.037, 0.090, and 0.162, respectively.

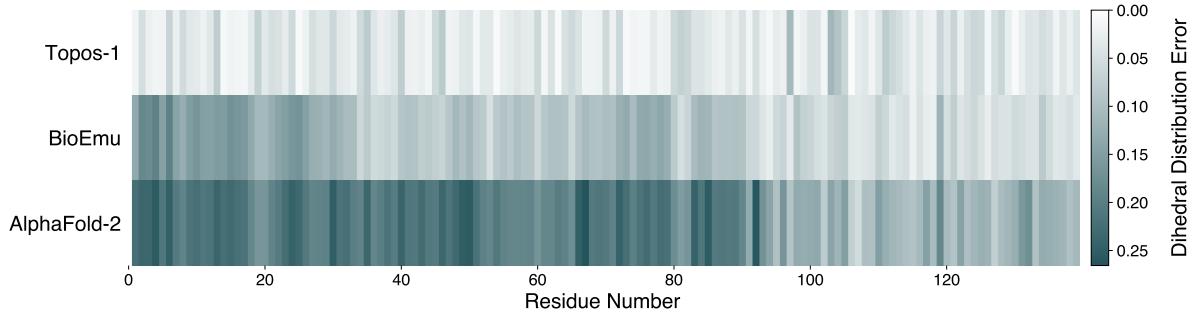

Figure 11 | Topos-1 produces conformational ensembles of α-synuclein in excellent agreement with 30 μs of molecular dynamics simulations conducted on the Anton supercomputer. The circular Wasserstein-1 distance between the backbone dihedral angle distributions of generated samples and conformations from molecular dynamics conducted in Ref. [33], shown per residue and averaged over ϕ and ψ angles.

To further evaluate the validity of the generated conformational ensembles of α-synuclein, in Fig. 12 we compare the CV distributions derived from ELViM embeddings for each model against reference MD simulations [33]. Both qualitative and quantitative (via calculation of JSDs) analysis of the ELViM embeddings reveals excellent agreement between Topos-1 and the reference MD data. An extended ELViM analysis with all models considered in this study is shown in Fig. S8. The results in Figs. 10, 11, 12, and S8 demonstrate that Topos-1 captures both the diversity and the accuracy of conformational states for a well-studied, disease-relevant IDP such as α-synuclein, a necessary step in the pursuit of SBDD. Moreover, it achieves this in a matter of minutes, orders of magnitude faster compared to conventional methods.

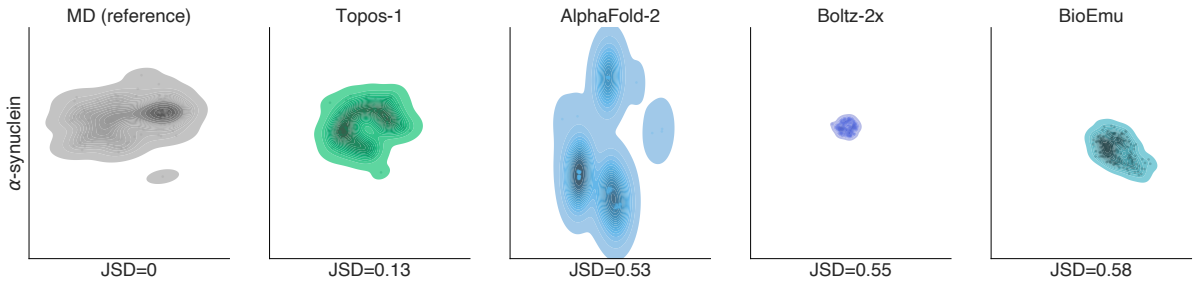

Figure 12 | Topos-1 conformational landscapes more accurately match those from molecular dynamics (MD) simulations for α-synuclein. Samples from Topos-1, AlphaFold-2, Boltz-2x, and BioEmu were combined with 200 conformations from a publicly available MD simulation of the disordered protein α-synuclein [33] and the same analysis was performed as in Fig. 5. Samples from Topos-1 more accurately reflect publicly available data for α-synuclein compared to state-of-the-art structure prediction models and IDP-specific models. All figures have equivalent axes.

4.2. Case study: Topos–1 demonstrates sensitivity to sequence perturbations

Sequence variants are prevalent in IDPs, which, in addition to their implication in cancer and neurodegenerative diseases discussed above, play a crucial role in viral proteins. However, most structure prediction pipelines only weakly differentiate mutants [41]. To test whether Topos-1 can capture engineered changes in IDR conformational ensembles, we benchmarked against a published protein engineering study that altered the charge pattern in two viral IDPs, measles virus NTAIL (127 aa) and Hendra virus PNT4 (107 aa), each in three variants (low-κ, WT, high-κ) where κ is a scalar measure of charge patterning along the sequence. In these variants, the amino-acid composition was held constant while the linear ordering of charged residues was permuted, isolating the effect of charge clustering from residue propensity [42].

For the sequence variants, the authors characterized the conformational response of the proteins to the permutations using a combination of experimental techniques, including size-exclusion chromatography-SAXS, and reported a consistent physical trend: increasing the charge clustering (higher κ) compacted the chain, with R_g and D_{max} decreasing as κ increased. Consistent with this observation, ensembles generated with Topos-1 reproduce the order in R_g of low-$\kappa \gtrsim$ WT > high-κ for both NTAIL and PNT4, with the value of R_g for high-κ being noticeably more compact, matching the rank ordering and relative magnitude observed experimentally (Fig. 13). This example highlights a practical use case for protein engineering: Topos-1 responds to sequence-level charge patterning beyond amino-acid propensity, even in short ($\approx$ 100 aa) IDPs.

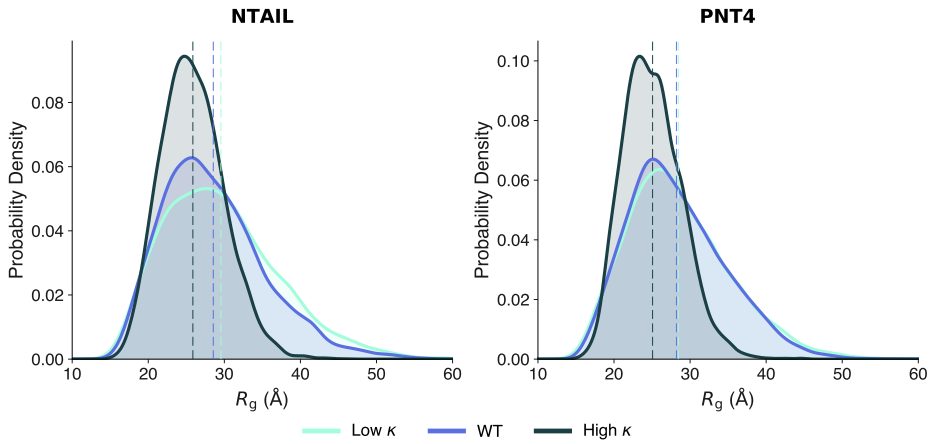

Figure 13 | Topos-1 correctly predicts the effects of charge-patterned mutations and reproduces ensemble-level trends in agreement with experiment. Predicted radius of gyration (R_g) distributions for low-κ, wild-type (WT), and high-κ variants of the measles virus NTAIL peptide (left) and the Hendra virus PNT4 peptide (right). Each distribution contains 6,000 samples from Topos-1. κ quantifies the sequence-wise arrangement of oppositely charged residues along the sequence, ranging from well-mixed (low-κ) to charge-segregated (high-κ). Topos-1 predicts the high-κ variants to have a lower mean R_g and the low-κ variants to be similar to WT, consistent with experimental observation [42]. Dashed vertical lines denote ensemble means.

4.3. Case study: Topos-1 facilitates structure-based discovery of small molecules for the prostate cancer target androgen receptor

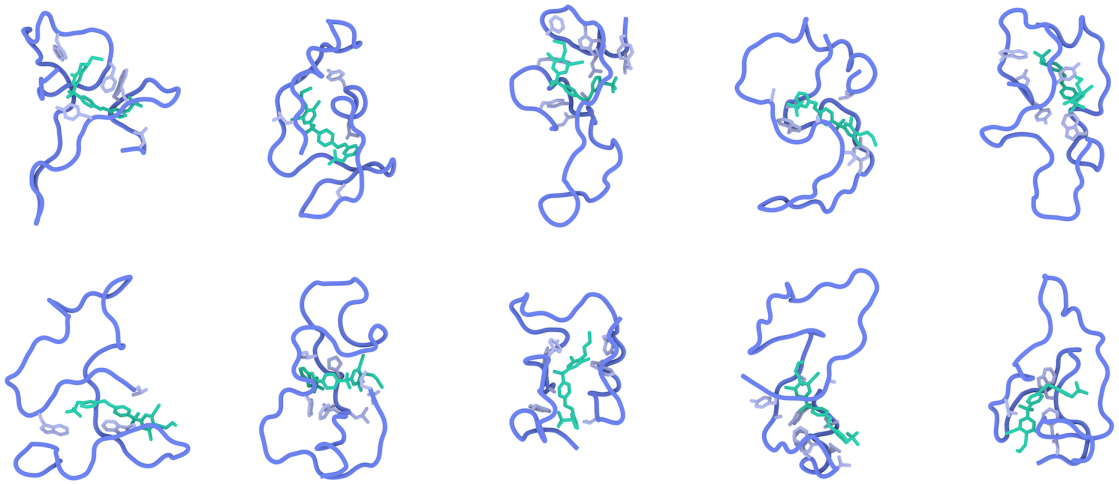

Figure 14 | Representative poses of ligand interaction with the conformational space of the Tau5 R2R3 region predicted by Topos-1-AP. Protein backbone, side chains, and ligand are shown in dark blue, light blue, and green, respectively.

The androgen receptor (AR) is a central driver of prostate cancer, and its intrinsically disordered N-terminal domain (NTD) participates in many therapeutically relevant interactions, yet remains far more challenging as a therapeutic target than the canonical ligand-binding domain or DNA-binding interface [43]. Previous work examined ligand binding to the Tau5 R2R3 region (residues 393 to 448) in human AR NTD via long timescale MD simulations [44]. As such, this system serves as a challenging and well-developed use case to assess whether Topos-1 can reduce the computational cost of ensemble generation, from tens of thousands of GPU hours to minutes on a single GPU with reliable downstream utility.

Here we discuss results using Topos-1-AP, a proprietary ensemble-based affinity prediction (AP) tool that works with ensembles generated with Topos-1. We utilized Topos-1-AP for two ligand sets and the Tau5 R2R3 region. Representative interaction poses from one of the ligands and the Tau5 R2R3 region are shown in Fig. 14. Details of the ligand sets and the cell-based assay are provided in Supplementary Section B.6. Ligand set L1 is comprised of 15 small molecules with literature-reported IC_{50} values [45–48]. Ligand set L2 is comprised of 27 small molecules, including proprietary molecules with cell-based IC_{50} values measured. Affinity scores predicted by Topos-1-AP were then used to assess linear correlation (by Pearson's r) and rank ordering (by Spearman's ρ) relative to experimental IC_{50} values.

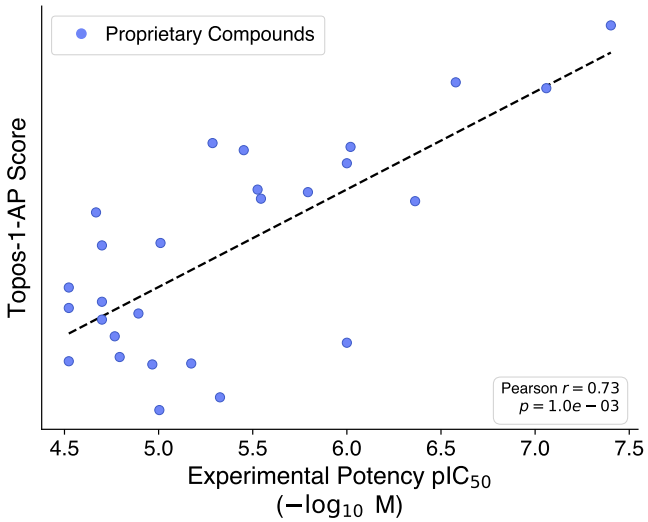

Figure 15 | Topos-1 conformational ensembles enable prediction of high-potency binding of novel small molecules. Correlation between the Topos-1-AP score and experimental potency from an in-house cell-based assay on 27 proprietary Topos Bio compounds. The dashed line shows the least-squares fit. The model yields a Pearson's r correlation coefficient of 0.73 with a two-sided p value of 1.3×10^{-5}. Higher computational affinity scores indicate stronger predicted binding, and higher pIC_{50} values indicate greater experimental potency. Topos-1 generates conformational ensembles that support downstream virtual screening, at a fraction of the computational cost of long-timescale molecular dynamics and without the need for significant ensemble post-processing.

Compared to classical MD-based affinity prediction workflows [49], Topos-1-AP achieves substantial computational speedups at three distinct stages. The first speedup comes from the rapid generation of representative conformational ensembles by Topos-1. The second speedup comes from the high statistical accuracy of the ensemble generated with Topos-1 compared to experimental measurements and MD data (see Section 2.2), avoiding the need for significant post-processing. This is in contrast to MD, which often requires clustering and corrective reweighting based on experimental references. The final speedup comes from the fact that samples from Topos-1 are globally and locally accurate (see Section 2.2). This global structural coherence over the entire IDP sequence is crucial to obtaining accurate affinity prediction of small molecule binding to IDPs, as information regarding binding site identity in IDPs is severely limited. Taken together, the preservation of intra-conformer structure quality and inter-conformer distribution by Topos-1 at ultra-low computational cost plays a central role in the performance of Topos-1-AP.

As shown in Fig. 15, without prior information about the ligand binding sites we observe medium-to-strong linear correlation and ranking power by Topos-1-AP in both ligand sets (see Section B.6.2). Furthermore, fivefold-replicated predictions with Topos-1-AP using different random seeds yielded consistent ligand rankings. Note that even a single run can reach satisfactory agreement with experiment (Pearson's $r = 0.58$, Spearman's $\rho = 0.64$). As such, these results establish Topos-1 as a computationally efficient and scalable alternative to MD-based ensemble generation for IDP-ligand affinity prediction. The resulting pipeline achieves orders-of-magnitude reductions in computational cost while maintaining reliable downstream performance in affinity prediction and virtual screening applications.

5. Conclusion

We introduced Topos-1, a structure prediction model that achieves state-of-the-art performance on intrinsically disordered proteins by designing a new generative model architecture with specific competencies for flexible proteins. While the model unambiguously performed the best on an exhaustive set of experimental and computational evaluations, we again note that AI structure prediction models like AlphaFold were trained primarily using experimental structures that fail to resolve the regions on which we focus. Nevertheless, we assert that focusing on these regions is important for both providing insight into the biology of and designing new drugs for intrinsically disordered proteins. In fact, we showed that novel compounds can be designed *de novo* using Topos-1 as a guide for high potency due to its highly physical and rapidly enumerated conformational ensembles.

Topos-1 currently focuses on accuracy for single-molecule properties of intrinsically disordered proteins, but many disordered sequences are co-expressed with structured regions. Building on the success of AlphaFold, as many others have, we plan to expand Topos-1 to accurately capture interactions between order and disorder. The biological function of disordered proteins is often driven by interactions that are modulated by post-translational modifications, and we continue to expand Topos-1 to better capture the often-dramatic effects of phenomena like phosphorylation and methylation [50]. In addition, cellular context, notably membrane interaction, plays a significant role in IDP structure, and we are expanding Topos-1 to differentiate these contexts.

Enabling the design of small molecule therapeutics is a major goal of Topos Bio, and while we show promising data for novel compounds, expanding the capabilities of our generative model to include co-folding would provide a significant acceleration to our platform. We anticipate that this can be achieved with additional scale of both experimental and synthetic data generation; indeed, our experiments have already shown dramatic model improvements with a growing dataset. This expanded biomolecular corpus will include small molecules, proteins, and nucleic acids, which are often highly flexible. Furthermore, we have found that test-time compute can enhance our model even further and expect scaling at inference time to become a major aspect of future versions of Topos-1.

Contributors

Tomas Salgado*, Andre Graubner*, Scott Leishman*, Malhar Kute, Amy He, Benjamin Rousseau, Connor Bybee, Anthony Giannetti, Cassie Lu, Anne Pipathsouk, Jake Drummond, Robert Galemmo, Fernando Martinez, Amir Khosrowshahi, Ryan Zarcone, Grant Rotskoff†

* *Lead authors*

† *Scientific Advisor, work performed as a consultant to Topos Bio*

Correspondence: topos-1@toposbio.ai

References

[1] Hans Frauenfelder, Stephen G Sligar, and Peter G Wolynes. The energy landscapes and motions of proteins. *Science*, 254(5038):1598–1603, 1991. http://dx.doi.org/10.1126/science.1749933. URL https://www.science.org/doi/abs/10.1126/science.1749933.

[2] Katherine Henzler-Wildman and Dorothee Kern. Dynamic personalities of proteins. *Nature*, 450(7172):964–972, 2007. http://dx.doi.org/10.1038/nature06522. URL https://doi.org/10.1038/nature06522.

[3] David D Boehr, Ruth Nussinov, and Peter E Wright. The role of dynamic conformational ensembles in biomolecular recognition. *Nature Chemical Biology*, 5(11):789–796, 2009. http://dx.doi.org/10.1038/nchembio.232. URL https://doi.org/10.1038/nchembio.232.

[4] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenstein, David Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Pushmeet Kohli, and Demis Hassabis. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873):583–589, Aug 2021. http://dx.doi.org/10.1038/s41586-021-03819-2. URL https://doi.org/10.1038/s41586-021-03819-2.

[5] Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, Lindsay Willmore, Andrew J Ballard, Joshua Bambrick, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 630(8016):493–500, 2024. http://dx.doi.org/10.1038/s41586-024-07487-w. URL https://doi.org/10.1038/s41586-024-07487-w.

[6] Chai Discovery, Jacques Boitreaud, Jack Dent, Matthew McPartlon, Joshua Meier, Vinicius Reis, Alex Rogozhnikov, and Kevin Wu. Chai-1: Decoding the molecular interactions of life, 2024. URL https://www.biorxiv.org/content/early/2024/10/15/2024.10.10.615955.

[7] Saro Passaro, Gabriele Corso, Jeremy Wohlwend, Mateo Reveiz, Stephan Thaler, Vignesh Ram Somnath, Noah Getz, Tally Portnoi, Julien Roy, Hannes Stark, David Kwabi-Addo, Dominique Beaini, Tommi Jaakkola, and Regina Barzilay. Boltz-2: Towards accurate and efficient binding affinity prediction, 2025. URL https://www.biorxiv.org/content/early/2025/06/18/2025.06.14.659707.

[8] Helen M Berman, John Westbrook, Zukang Feng, Gary Gilliland, Talapady N Bhat, Helge Weissig, Ilya N Shindyalov, and Philip E Bourne. The protein data bank. *Nucleic acids research*, 28(1):235–242, 2000. http://dx.doi.org/10.1093/nar/28.1.235. URL https://doi.org/10.1093/nar/28.1.235.

[9] H. Jane Dyson and Peter E. Wright. Intrinsically unstructured proteins and their functions. *Nature Reviews Molecular Cell Biology*, 6(3):197–208, 2005. http://dx.doi.org/10.1038/nrm1589. URL https://doi.org/10.1038/nrm1589.

[10] Alex S Holehouse and Birthe B Kragelund. The molecular basis for cellular function of intrinsically disordered protein regions. *Nature Reviews Molecular Cell Biology*, 25(3):187–211, 2024. http://dx.doi.org/10.1038/s41580-023-00673-0. URL https://doi.org/10.1038/s41580-023-00673-0.

[11] Kiersten M Ruff and Rohit V Pappu. Alphafold and implications for intrinsically disordered proteins. *Journal of molecular biology*, 433(20):167208, 2021. http://dx.doi.org/10.1016/j.jmb.2021.167208. URL https://doi.org/10.1016/j.jmb.2021.167208.

[12] Bi Zhao, Sina Ghadermarzi, and Lukasz Kurgan. Comparative evaluation of alphafold2 and disorder predictors for prediction of intrinsic disorder, disorder content and fully disordered proteins. *Computational and Structural Biotechnology Journal*, 21:3248–3258, 2023. http://dx.doi.org/10.1016/j.csbj.2023.06.001. URL https://doi.org/10.1016/j.csbj.2023.06.001.

[13] Alexandra Hatos et al. DisProt: intrinsic protein disorder annotation in 2020. *Nucleic Acids Research*, 48(D1): D269–D276, 2020. http://dx.doi.org/10.1093/nar/gkz975. URL https://doi.org/10.1093/nar/gkz975.

[14] Damiano Piovesan, Marco Necci, Natalia Escobedo, Alexander Miguel Monzon, Alexandra Hatos, Ivan Mičetić, Fabiola Quaglia, Lisanna Paladin, Priyadharshini Ramasamy, Zsuzsanna Dosztányi, and Silvio C. E. Tosatto. MobiDB: intrinsically disordered proteins in 2021. *Nucleic Acids Research*, 49(D1):D361–D367, 2021. http://dx.doi.org/10.1093/nar/gkaa1058. URL https://doi.org/10.1093/nar/gkaa1058.

[15] Robin Van Der Lee, Marija Buljan, Benjamin Lang, Robert J Weatheritt, Gary W Daughdrill, A Keith Dunker, Monika Fuxreiter, Julian Gough, Joerg Gsponer, David T Jones, et al. Classification of intrinsically disordered regions and proteins. *Chemical reviews*, 114(13):6589–6631, 2014. http://dx.doi.org/10.1021/cr400525m. URL https://doi.org/10.1021/cr400525m.

[16] Massimiliano Bonomi, Gabriella T Heller, Carlo Camilloni, and Michele Vendruscolo. Principles of protein structural ensemble determination. *Current opinion in structural biology*, 42:106–116, 2017. http://dx.doi.org/10.1016/j.sbi.2016.12.004. URL https://doi.org/10.1016/j.sbi.2016.12.004.

[17] Rakesh Trivedi and Hampapathalu Adimurthy Nagarajaram. Intrinsically disordered proteins: An overview. *International Journal of Molecular Sciences*, 23(22):14050, 2022. http://dx.doi.org/10.3390/ijms232214050. URL https://

www.mdpi.com/1422-0067/23/22/14050.

[18] Vladimir N Uversky, Christopher J Oldfield, and A Keith Dunker. Intrinsically disordered proteins in human diseases: introducing the D2 concept. *Annu. Rev. Biophys.*, 37(1):215–246, 2008. http://dx.doi.org/10.1146/annurev.biophys.37.032807.125924. URL https://doi.org/10.1146/annurev.biophys.37.032807.125924.

[19] Mateusz Biesaga, Marta Frigolé-Vivas, and Xavier Salvatella. Intrinsically disordered proteins and biomolecular condensates as drug targets. *Current Opinion in Chemical Biology*, 62:90–100, 2021. http://dx.doi.org/10.1016/j.cbpa.2021.02.009. URL https://doi.org/10.1016/j.cbpa.2021.02.009.

[20] Orkid Coskuner-Weber, Ozan Mirzanli, and Vladimir N Uversky. Intrinsically disordered proteins and proteins with intrinsically disordered regions in neurodegenerative diseases. *Biophysical Reviews*, 14(3):679–707, 2022. http://dx.doi.org/10.1007/s12551-022-00968-0. URL https://doi.org/10.1007/s12551-022-00968-0.

[21] Kagistia Hana Utami, Satoru Morimoto, Yasue Mitsukura, and Hideyuki Okano. The roles of intrinsically disordered proteins in neurodegeneration. *Biochimica et Biophysica Acta (BBA)-General Subjects*, 1869(4):130772, 2025. http://dx.doi.org/10.1016/j.bbagen.2025.130772. URL https://doi.org/10.1016/j.bbagen.2025.130772.

[22] Bálint Mészáros, Borbála Hajdu-Soltész, András Zeke, and Zsuzsanna Dosztányi. Mutations of intrinsically disordered protein regions can drive cancer but lack therapeutic strategies. *Biomolecules*, 11(3):381, 2021. http://dx.doi.org/10.3390/biom11030381. URL https://doi.org/10.3390/biom11030381.

[23] Monica L. Fernández-Quintero, Janik Kokot, Franz Waibl, Anna-Lena M. Fischer, Patrick K. Quoika, Charlotte M. Deane, and Klaus R. Liedl. Challenges in antibody structure prediction. *mAbs*, 15(1):2175319, 2023. http://dx.doi.org/10.1080/19420862.2023.2175319. URL https://doi.org/10.1080/19420862.2023.2175319.

[24] H Jane Dyson. Making sense of intrinsically disordered proteins. *Biophysical Journal*, 110(5):1013–1016, 2016. http://dx.doi.org/10.1016/j.bpj.2016.01.030. URL https://doi.org/10.1016/j.bpj.2016.01.030.

[25] Steven J Metallo. Intrinsically disordered proteins are potential drug targets. *Current opinion in chemical biology*, 14 (4):481–488, 2010. http://dx.doi.org/10.1016/j.cbpa.2010.06.169. URL https://doi.org/10.1016/j.cbpa.2010.06.169.

[26] Paul Robustelli, Stefano Piana, and David E. Shaw. Developing a molecular dynamics force field for both folded and disordered protein states. *Proceedings of the National Academy of Sciences*, 115(21):E4758–E4766, May 2018. http://dx.doi.org/10.1073/pnas.1800690115. URL https://doi.org/10.1073/pnas.1800690115.

[27] Oufan Zhang, Zi Hao Liu, Julie D. Forman-Kay, and Teresa Head-Gordon. Deep Learning of Proteins with Local and Global Regions of Disorder, March 2025. URL http://arxiv.org/abs/2502.11326.

[28] Borna Novak, Jeffrey M. Lotthammer, Ryan J. Emenecker, and Alex S. Holehouse. Accurate predictions of conformational ensembles of disordered proteins with STARLING, 2025. URL https://www.biorxiv.org/content/early/2025/02/25/2025.02.14.638373.

[29] Martin Steinegger and Johannes Söding. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature Biotechnology*, 35(11):1026–1028, 2017. http://dx.doi.org/10.1038/nbt.3988. URL https://doi.org/10.1038/nbt.3988.

[30] Richard Evans, Michael O'Neill, Alexander Pritzel, Natasha Antropova, Andrew Senior, Tim Green, Augustin Žídek, Russ Bates, Sam Blackwell, Jason Yim, Olaf Ronneberger, Sebastian Bodenstein, Michal Zielinski, Alex Bridgland, Anna Potapenko, Andrew Cowie, Kathryn Tunyasuvunakool, Rishub Jain, Ellen Clancy, Pushmeet Kohli, John Jumper, and Demis Hassabis. Protein complex prediction with AlphaFold-Multimer, 2022. URL https://www.biorxiv.org/content/early/2022/03/10/2021.10.04.463034.

[31] Rafael Giordano Viegas, Ingrid BS Martins, Murilo Nogueira Sanches, Antonio B Oliveira Junior, Juliana B de Camargo, Fernando V Paulovich, and Vitor BP Leite. ELViM: Exploring biomolecular energy landscapes through multidimensional visualization. *Journal of Chemical Information and Modeling*, 64(8):3443–3450, 2024. http://dx.doi.org/10.1021/acs.jcim.4c00034. URL https://doi.org/10.1021/acs.jcim.4c00034.

[32] Rafael G Viegas, Hao Wu, Murilo N Sanches, Garegin A Papoian, and Vitor BP Leite. Energy landscapes and structural plasticity of intrinsically disordered histones. *Journal of Chemical Information and Modeling*, 65(16):8679–8687, 2025. http://dx.doi.org/10.1021/acs.jcim.4c02269. URL https://doi.org/10.1021/acs.jcim.4c02269.

[33] Kaushik Borthakur, Thomas R Sisk, Francesco P Panei, Massimiliano Bonomi, and Paul Robustelli. Determining accurate conformational ensembles of intrinsically disordered proteins at atomic resolution. *Nature Communications*, 16(1):9036, 2025. http://dx.doi.org/10.1038/s41467-025-64098-3. URL https://doi.org/10.1038/s41467-025-64098-3.

[34] Jared Kaplan, Sam McCandlish, Tom Henighan, Tom B. Brown, Benjamin Chess, Rewon Child, Scott Gray, Alec Radford, Jeffrey Wu, and Dario Amodei. Scaling laws for neural language models, 2020. URL https://arxiv.org/abs/2001.08361.

[35] Rich Sutton. The bitter lesson. http://www.incompleteideas.net/IncIdeas/BitterLesson.html, 2019. Accessed: 2019-03-13.

[36] Tomas Geffner, Kieran Didi, Zuobai Zhang, Danny Reidenbach, Zhonglin Cao, Jason Yim, Mario Geiger, Christian Dallago, Emine Kucukbenli, Arash Vahdat, et al. Proteina: Scaling flow-based protein structure generative models, 2025. URL https://arxiv.org/abs/2503.00710.

[37] Yuyang Wang, Jiarui Lu, Navdeep Jaitly, Josh Susskind, and Miguel Angel Bautista. SimpleFold: Folding proteins is simpler than you think, 2025. URL https://arxiv.org/abs/2509.18480.

[38] Jacob T Bendor, Todd P Logan, and Robert H Edwards. The function of α-synuclein. *Neuron*, 79(6):1044–1066, 2013. http://dx.doi.org/10.1016/j.neuron.2013.09.004. URL https://doi.org/10.1016/j.neuron.2013.09.004.

[39] Paolo Calabresi, Alessandro Mechelli, Giuseppina Natale, Laura Volpicelli-Daley, Giulia Di Lazzaro, and Veronica Ghiglieri. Alpha-synuclein in Parkinson's disease and other synucleinopathies: from overt neurodegeneration back to early synaptic dysfunction. *Cell Death & Disease*, 14(3):176, 2023. http://dx.doi.org/10.1038/s41419-023-05672-9. URL https://doi.org/10.1038/s41419-023-05672-9.

[40] Carlo Camilloni, Alfonso De Simone, Wim F Vranken, and Michele Vendruscolo. Determination of secondary structure populations in disordered states of proteins using nuclear magnetic resonance chemical shifts. *Biochemistry*, 51(11): 2224–2231, 2012. http://dx.doi.org/10.1021/bi3001825. URL https://doi.org/10.1021/bi3001825.

[41] Claudio Mirabello, Björn Wallner, Björn Nystedt, Stavros Azinas, and Marta Carroni. Unmasking AlphaFold to integrate experiments and predictions in multimeric complexes. *Nature Communications*, 15(1):8724, 2024. http://dx.doi.org/10.1038/s41467-024-52951-w. URL https://doi.org/10.1038/s41467-024-52951-w.

[42] Giulia Tedeschi, Edoardo Salladini, Carlo Santambrogio, Rita Grandori, Sonia Longhi, and Stefania Brocca. Conformational response to charge clustering in synthetic intrinsically disordered proteins. *Biochimica et Biophysica Acta (BBA)-General Subjects*, 1862(10):2204–2214, 2018. http://dx.doi.org/10.1016/j.bbagen.2018.07.011. URL https://doi.org/10.1016/j.bbagen.2018.07.011.

[43] Ye Chen and Tian Lan. N-terminal domain of androgen receptor is a major therapeutic barrier and potential pharmacological target for treating castration resistant prostate cancer: a comprehensive review. *Frontiers in Pharmacology*, 15:1451957, 2024. http://dx.doi.org/10.3389/fphar.2024.1451957. URL https://doi.org/10.3389/fphar.2024.1451957.

[44] Jiaqi Zhu, Xavier Salvatella, and Paul Robustelli. Small molecules targeting the disordered transactivation domain of the androgen receptor induce the formation of collapsed helical states. *Nature Communications*, 13(1):6390, 2022. http://dx.doi.org/10.1038/s41467-022-34077-z. URL https://doi.org/10.1038/s41467-022-34077-z.

[45] Jingjing Xie, Hao He, Wenna Kong, Ziwen Li, Zhenting Gao, Daoqing Xie, Lin Sun, Xiaofei Fan, Xiangqing Jiang, Qiangang Zheng, et al. Targeting androgen receptor phase separation to overcome antiandrogen resistance. *Nature Chemical Biology*, 18(12):1341–1350, 2022. http://dx.doi.org/10.1038/s41589-022-01151-y. URL https://doi.org/10.1038/s41589-022-01151-y.

[46] R Le Moigne, NH Hong, P Pearson, V Lauriault, CA Banuelos, NR Mawji, T Tam, J Wang, P Virsik, RJ Andersen, et al. 545P preclinical profile of EPI-7386, a second-generation N-terminal domain androgen receptor inhibitor for the treatment of prostate cancer. *Annals of Oncology*, 31:S475, 2020. http://dx.doi.org/10.1016/j.annonc.2020.08.659. URL https://doi.org/10.1016/j.annonc.2020.08.659.

[47] Carmen A Banuelos, Yusuke Ito, Jon K Obst, Nasrin R Mawji, Jun Wang, Yukiyoshi Hirayama, Jacky K Leung, Teresa Tam, Amy H Tien, Raymond J Andersen, et al. Ralaniten sensitizes Enzalutamide-resistant prostate cancer to ionizing radiation in prostate cancer cells that express androgen receptor splice variants. *Cancers*, 12(7):1991, 2020. http://dx.doi.org/10.3390/cancers12071991. URL https://doi.org/10.3390/cancers12071991.

[48] Shaon Basu, Paula Martínez-Cristóbal, Marta Frigolé-Vivas, Mireia Pesarrodona, Michael Lewis, Elzbieta Szulc, C Adriana Bañuelos, Carolina Sánchez-Zarzalejo, Stasė Bielskutė, Jiaqi Zhu, et al. Rational optimization of a transcription factor activation domain inhibitor. *Nature structural & molecular biology*, 30(12):1958–1969, 2023. http://dx.doi.org/10.1038/s41594-023-01159-5. URL https://doi.org/10.1038/s41594-023-01159-5.

[49] Anjali Dhar, Thomas R. Sisk, and Paul Robustelli. Ensemble docking for intrinsically disordered proteins. *Journal of Chemical Information and Modeling*, 65(13):6847–6860, Jul 2025. http://dx.doi.org/10.1021/acs.jcim.5c00370. URL https://doi.org/10.1021/acs.jcim.5c00370.

[50] Alaji Bah and Julie D. Forman-Kay. Modulation of intrinsically disordered protein function by post-translational modifications. *Journal of Biological Chemistry*, 291(13):6696–6705, Mar 2016. http://dx.doi.org/10.1074/jbc.R115.695056. URL https://doi.org/10.1074/jbc.R115.695056.

[51] Jesse Bennett Hopkins, Richard E. Gillilan, and Søren Skou. BioXTAS RAW: improvements to a free open-source program for small-angle X-ray scattering data reduction and analysis. *Journal of Applied Crystallography*, 50(5):1545–1553, 2017. http://dx.doi.org/10.1107/S1600576717011438. URL https://doi.org/10.1107/S1600576717011438.

[52] D. Svergun, C. Barberato, and M. H. J. Koch. *CRYSOL* – a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *Journal of Applied Crystallography*, 28(6):768–773, Dec 1995. http://dx.doi.org/10.1107/S0021889895007047. URL https://doi.org/10.1107/S0021889895007047.

[53] Søren Hansen. Bayesian estimation of hyperparameters for indirect Fourier transformation in small-angle scattering. *Journal of Applied Crystallography*, 33(6):1415–1421, Dec 2000. http://dx.doi.org/10.1107/S0021889800012930. URL https://doi.org/10.1107/S0021889800012930.

[54] Sarah Lewis, Tim Hempel, José Jiménez-Luna, Michael Gastegger, Yu Xie, Andrew Y. K. Foong, Victor García Satorras, Osama Abdin, Bastiaan S. Veeling, Iryna Zaporozhets, Yaoyi Chen, Soojung Yang, Adam E. Foster, Arne Schneuing, Jigyasa Nigam, Federico Barbero, Vincent Stimper, Andrew Campbell, Jason Yim, Marten Lienen, Yu Shi,

Shuxin Zheng, Hannes Schulz, Usman Munir, Roberto Sordillo, Ryota Tomioka, Cecilia Clementi, and Frank Noé. Scalable emulation of protein equilibrium ensembles with generative deep learning. *Science*, 389(6761):eadv9817, 2025. http://dx.doi.org/10.1126/science.adv9817. URL https://www.science.org/doi/abs/10.1126/science.adv9817.

[55] Junjie Zhu, Zhengxin Li, Zhuoqi Zheng, Bo Zhang, Bozitao Zhong, Jie Bai, Taifeng Wang, Ting Wei, Jianyi Yang, and Hai-Feng Chen. Precise generation of conformational ensembles for intrinsically disordered proteins via fine-tuned diffusion models, 2024. URL https://www.biorxiv.org/content/early/2024/05/28/2024.05.05.592611.

[56] Gabor Nagy, Maxim Igaev, Nykola C Jones, Søren V Hoffmann, and Helmut Grubmüller. SESCA: predicting circular dichroism spectra from protein molecular structures. *Journal of Chemical Theory and Computation*, 15(9):5087–5102, 2019. http://dx.doi.org/10.1021/acs.jctc.9b00203. URL https://doi.org/10.1021/acs.jctc.9b00203.

[57] Gabor Nagy, Søren Vrønning Hoffmann, Nykola C. Jones, and Helmut Grubmüller. Reference data set for circular dichroism spectroscopy comprised of validated intrinsically disordered protein models. *Applied Spectroscopy*, 78(9):897–911, 2024. http://dx.doi.org/10.1177/00037028241239977. URL https://doi.org/10.1177/00037028241239977. PMID: 38646777.

[58] Elliott J. Stollar and David P. Smith. Correction: Uncovering protein structure. *Essays in Biochemistry*, 65 (2):407–407, 07 2021. http://dx.doi.org/10.1042/EBC-2019-0042C_COR. URL https://doi.org/10.1042/EBC-2019-0042C_COR.

[59] Fan Cao, Sören von Bülow, Giulio Tesei, and Kresten Lindorff-Larsen. A coarse-grained model for disordered and multi-domain proteins. *Protein Science*, 33(11):e5172, 2024. http://dx.doi.org/https://doi.org/10.1002/pro.5172. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/pro.5172.

[60] Sebastian Kmiecik, Dominik Gront, Michal Kolinski, Lukasz Wieteska, Aleksandra Elzbieta Dawid, and Andrzej Kolinski. Coarse-grained protein models and their applications. *Chemical Reviews*, 116(14):7898–7936, Jul 2016. ISSN 0009-2665. http://dx.doi.org/10.1021/acs.chemrev.6b00163. URL https://doi.org/10.1021/acs.chemrev.6b00163.

[61] The UniProt Consortium. UniProt: the universal protein knowledgebase in 2025. *Nucleic Acids Research*, 53(D1): D609–D617, 11 2024. http://dx.doi.org/10.1093/nar/gkae1010. URL https://doi.org/10.1093/nar/gkae1010.

[62] Hannah K. Wayment-Steele, Adedolapo Ojoawo, Renee Otten, Julia M. Apitz, Warintra Pitsawong, Marc Hömberger, Sergey Ovchinnikov, Lucy Colwell, and Dorothee Kern. Predicting multiple conformations via sequence clustering and alphafold2. *Nature*, 625(7996):832–839, Jan 2024. http://dx.doi.org/10.1038/s41586-023-06832-9. URL https://doi.org/10.1038/s41586-023-06832-9.

[63] Björn Wallner. Afsample: improving multimer prediction with alphafold using massive sampling. *Bioinformatics*, 39(9):btad573, 09 2023. http://dx.doi.org/10.1093/bioinformatics/btad573. URL https://doi.org/10.1093/bioinformatics/btad573.

[64] Yang Shen and Ad Bax. Sparta+: a modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network. *Journal of Biomolecular NMR*, 48(1):13–22, Sep 2010. http://dx.doi.org/10.1007/s10858-010-9433-9. URL https://doi.org/10.1007/s10858-010-9433-9.

[65] Aleksandra L. Ptaszek, Jie Li, Robert Konrat, Gerald Platzer, and Teresa Head-Gordon. UCBShift 2.0: Bridging the gap from backbone to side chain protein chemical shift prediction for protein structures. *Journal of the American Chemical Society*, 146(46):31733–31745, Nov 2024. http://dx.doi.org/10.1021/jacs.4c10474. URL https://doi.org/10.1021/jacs.4c10474.

[66] Robert T McGibbon, Kyle A Beauchamp, Matthew P Harrigan, Christoph Klein, Jason M Swails, Carlos X Hernández, Christian R Schwantes, Lee-Ping Wang, Thomas J Lane, and Vijay S Pande. MDTraj: a modern open library for the analysis of molecular dynamics trajectories. *Biophysical Journal*, 109(8):1528–1532, 2015. http://dx.doi.org/10.1016/j.bpj.2015.08.015. URL https://doi.org/10.1016/j.bpj.2015.08.015.

[67] Peter JA Cock, Tiago Antao, Jeffrey T Chang, Brad A Chapman, Cymon J Cox, Andrew Dalke, Iddo Friedberg, Thomas Hamelryck, Frank Kauff, Bartek Wilczynski, et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, 25(11):1422–1423, 2009. http://dx.doi.org/10.1093/bioinformatics/btp163. URL https://doi.org/10.1093/bioinformatics/btp163.

[68] Wolfgang Kabsch. A solution for the best rotation to relate two sets of vectors. *Acta Crystallographica Section A*, 32(5):922–923, 1976. http://dx.doi.org/10.1107/S0567739476001873. URL https://doi.org/10.1107/S0567739476001873.

[69] Patrick Kunzmann and Kay Hamacher. Biotite: a unifying open source computational biology framework in Python. *BMC Bioinformatics*, 19(1):346, 2018. http://dx.doi.org/10.1186/s12859-018-2367-z. URL https://doi.org/10.1186/s12859-018-2367-z.

[70] Rémi Flamary, Nicolas Courty, Alexandre Gramfort, Mokhtar Z. Alaya, Aurélie Boisbunon, Stanislas Chambon, Laetitia Chapel, Adrien Corenflos, Kilian Fatras, Nemo Fournier, Léo Gautheron, Nathalie T.H. Gayraud, Hicham Janati, Alain Rakotomamonjy, Ievgen Redko, Antoine Rolet, Antony Schutz, Vivien Seguy, Danica J. Sutherland, Romain Tavenard, Alexander Tong, and Titouan Vayer. POT: Python Optimal Transport, 2021. URL http://jmlr.org/papers/v22/20-451.html.

[71] Shayan Hundrieser, Marcel Klatt, and Axel Munk. The statistics of circular optimal transport, 2022. URL https://doi.org/10.1007/978-981-19-1044-9_4.

[72] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020. http://dx.doi.org/10.1038/s41592-019-0686-2. URL https://doi.org/10.1038/s41592-019-0686-2.

[73] David W Scott. *Multivariate density estimation: theory, practice, and visualization*. John Wiley & Sons, 2015.

A. Materials and Methods

To rigorously benchmark Topos-1 against both experimental observables and physics-based simulations, we developed an integrated biophysical pipeline combining SEC-purified peptide preparation with SAXS, CD, and dynamic light scattering (DLS) characterization under a unified buffer regime. This approach enables repeatable, high-confidence measurements of intrinsically disordered protein ensembles at scale and provides a unique training and validation resource for model development.

A.1. Peptides

A panel of peptide sequences ranging from 21–100 amino acids was selected based on low pLDDT scores from AlphaFold-2, together with additional sequence-derived metrics including hydrophobic residue content and distribution, net charge, charge patterning, predicted solubility, and isoelectric point. Peptides were obtained either as commercially available from Cayman Chemical or were custom synthesized by Lifetein. Custom peptides were delivered at >95% purity, contained an N-terminal biotin–PEG4 moiety to enhance solubility and support downstream assays, and were C-terminally blocked with an alcohol. All peptides were supplied lyophilized and reconstituted to initial stock concentrations of 20–40 mg mL^{-1}. All measurements were performed in a common buffer: 20 mM sodium phosphate, pH 7.4, 100 mM sodium fluoride. Sodium fluoride was used in place of chloride to avoid strong far-UV absorbance interference during CD spectroscopy. No DMSO or other organic solvents were used. Prior to further processing, samples were passed through 0.1 μm spin filters and purified by size-exclusion chromatography on an ÄKTA Avant FPLC equipped with a Cytiva Superdex 30 10/300 GL column. The flow rate was 1 mL min^{-1}. Absorbance was monitored at 205, 214, and 280 nm, enabling quantification of peptides lacking aromatic residues. 250 μL of sample was injected, and 0.3 mL fractions were collected in 96-well polypropylene plates. This workflow removed aggregates and provided complete buffer exchange into the working buffer. Conductivity measurements confirmed quantitative removal of residual trifluoroacetic acid from synthesis. Peptide loading concentrations were selected such that SEC peak fractions contained sufficient material for all downstream experiments without additional handling, ensuring consistent provenance across assays.

A.2. Dynamic light scattering (DLS)

Dynamic light scattering measurements were performed using a Wyatt DynaPro Plate Reader II in 384-well format. Samples were prepared directly from SEC peak fractions, and the SEC running buffer served as a no-scattering reference. Each well contained 50 μL of sample and was measured immediately to prevent evaporation artifacts. Measurements were collected at 20 °C, using 10 acquisitions per well with a 5-second acquisition time at the instrument's 830 nm laser and 173° backscatter detection geometry. Hydrodynamic radius (R_{h}), sample homogeneity, and polydispersity were determined using Wyatt Dynamics Software via cumulant analysis of the autocorrelation function. Where relevant, R_{h} values were compared to SAXS-derived R_{g} values to check for oligomeric states.

A.3. Small–Angle X–Ray Scattering (SAXS)

A.3.1. Data Collection

Synchrotron SAXS data were collected at Beamline 4-2 at the Stanford Synchrotron Radiation Lightsource (SSRL; Menlo Park, CA, USA) under proposals S-XV-RAS-12E0 and S-XV-RAS-12F0A. BL4-2 is part of the SSRL Structural Molecular Biology Program and is supported by the DOE Office of Biological and Environmental Research and the National Institutes of Health, National Institute of General Medical Sciences. Data were collected on a Pilatus3 X 1M detector positioned 1.2 m from the sample at an X-ray wavelength of $\lambda = 1.127$ Å. Scattering intensity was recorded as I(q) vs q, where $q = (4\pi/\lambda)\sin\theta$ where 2θ is the scattering angle. For each peptide, two buffer blanks were collected. SAXS samples were taken directly from SEC-purified peak fractions, and three two-fold dilutions were prepared to assess potential self-association. 30 μL samples were arrayed in 96-well PCR plates and loaded using the BL4-2 autosampler. For each condition, sixteen 1-second exposures were collected. Data were normalized to transmitted-beam intensity, radially averaged, and solvent-blank scattering was subtracted to generate final profiles. Frames were inspected for radiation damage, and exposures showing

systematic intensity decay or low-q deviation were excluded. Guinier analysis ensured that: $q * R_g < 1.3$ and confirmed the absence of aggregation or interparticle interference.

A.3.2. Data Processing and Analysis

Primary SAXS data processing was performed using BioXTAS RAW 2.3.0 [51] and the ATSAS 3.2.1 software suite [52]. Pair-distance distribution functions P(r) were computed using bIFT [53]. The theoretical SAXS I(q) curve for each member of predicted ensembles was computed using CRYSOL taking into account appropriate hydration models, contrast, excluded volume, and the presence (or absence) of model hydrogens. The ensemble I(q) was calculated by averaging the samples in reciprocal space. A 5% sigma-floor noise model was applied prior to P(r) calculations.

A.3.3. Experimental Investigations

Comparison of the experimental data with predictions was performed using the same ensemble sets generated for the comparisons in Section 3.1. Comparison statistics of model performance versus the data were carried out in both reciprocal and real space. This analysis was carried out with ensembles generated by Topos-1, AlphaFold-2 [4], BioEmu [54], Boltz-2x [7], Chai-1 [6], IDPFold [55], and IDPForge [27]

The data shown were obtained from four intrinsically disordered peptides (Table S1). The first is the N-terminal transactivation domain of the homeodomain transcription factor HOXB7, a prognostic marker indicating poor clinical outcomes due to its role in driving aggressive cancers by stimulating oncogenic pathways such as MAPK, PI3K-Akt, and EGFR. The other proteins include Katacalcin, a 21-amino acid peptide hormone derived from the C-terminal cleavage of procalcitonin in thyroid C-cells; the C-terminal IDR region of PHF21A, an epigenetic repressor of gene expression during brain development; Exendin-4, a 39-amino acid glucagon-like peptide-1 (GLP-1) receptor agonist originally isolated from Gila monster (*Heloderma suspectum*) venom; BAG6, a co-chaperone involved in protein synthesis and folding, ZC3H11C, a protein predicted to be involved in the nuclear export of mRNA; SALL4C, a transcription factor with a role in embryonic development.

Table S1 | Subset of our internal dataset of intrinsically disordered proteins used in the SAXS and CD measurements performed and reported in Figs. 7 and 8, respectively.

Peptide	Residues	pLDDT	Sequence
HOXB7 TAD	18	57	NTLFSKYPASSSVFATGA
Katacalcin	21	43	DMSSDLERDHRPHVSMPQNAN
PHF21A C-terminal IDR	58	42	GIDLSKPVDSEATVGAISNGPDCTPPANAATSTPAPSPSSQSCTANCNQGEETK
Exendin-4	39	74	GEGTFTSDLSKQMEEEAVRLFIEWLKNGGPSSGAPPPS
BAG6	63	46	MGQAGSTISNSHAQPFDFPDDNQNSKKLAAGHELQPLAIVDQRPSSRASSRASSRPRPDDLEI
ZC3H11C	73	52	SEEQGEGSSGVSSLLLHPEPVPGPEKENVRTVVRTVTLSTKQGEEPLVRLGLTETLGKRKFSTGGDSDPPLKR
SALL4C	45	70	ANNNSARRGRKLAIENTMALLGTDGKRVSEIFPKEILAPSVNVDP

A.4. Circular dichroism (CD) spectroscopy

Far-UV CD spectra were collected using a JASCO J-1500 CD spectrometer equipped with temperature control and operated at 20 °C. Measurements were performed in 0.1 cm path-length quartz cuvettes. Peptides in 20 mM sodium phosphate, 100 mM NaF, pH 7.4 were used directly from peak SEC fractions as in the SAXS and DLS experiments. Sample concentrations were determined directly on the CD instrument from A_{205} and A_{214} absorbance measurements and were used to convert CD intensities from millidegrees to mean residue molar ellipticity. Spectra were collected from 180–300 nm at 0.1 nm step size. Four sequential scans were recorded and averaged per sample. Matched-buffer spectra were collected and subtracted using JASCO Spectra Manager / SpectraMax software. High-tension voltage and absorbance channels were monitored to confirm spectral quality.

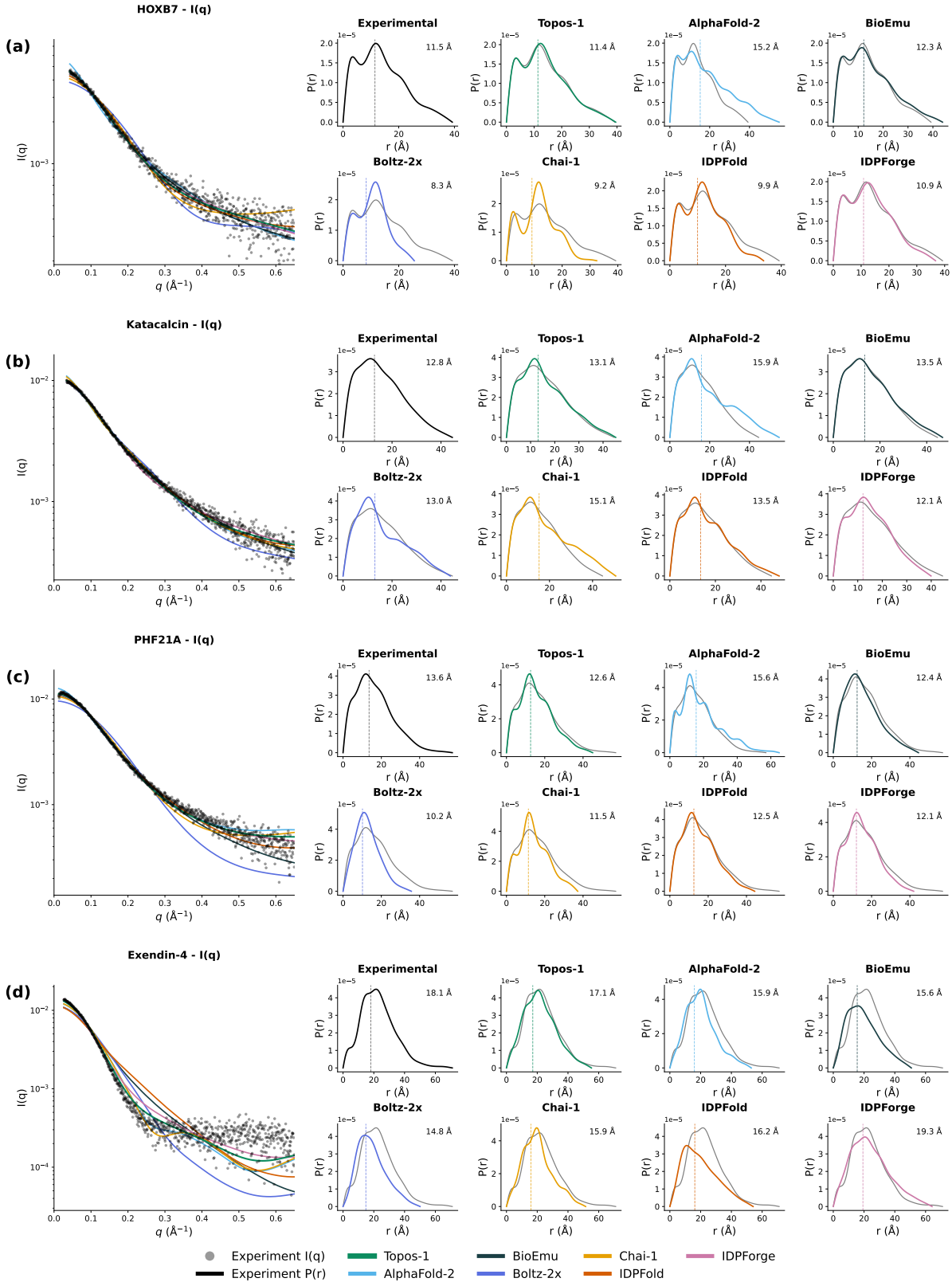

Figure S1 | Extended SAXS analysis of selected IDRs. Expanded version of Fig. 7 including additional model comparisons. Experimental I(q) data (left panels; black points) are overlaid with the ensemble-averaged I(q) predicted by calculating the scattering curve for each ensemble member and then averaging the curves. The P(r) curves derived from experiment and all 7 models tested in this report are presented on the right. For ease of comparison the experimental data are superimposed on each plot. The experimental or calculated R_g value is indicated by the vertical dashed line, with the corresponding R_g value reported in the top-right corner of each panel. The proteins used are HOXB7 (**a**), Katacalcin (**b**), PHF21A (**c**), and Exendin-4 (**d**).

A.4.1. Computation of CD spectra from predicted ensembles

In order to compare the predicted ensembles from different structure prediction models directly to the observed CD spectra, we use SESCA [56] with the DS-B4R1 basis set to forward-compute the spectra from the predicted structural ensembles.

SESCA computes the CD signal of a structural ensemble by linearly combining a set of *basis-functions* $B_i(\lambda)$ corresponding to different secondary structure motifs based on their relative occupancy f_i:

$$\theta_{\mathrm{pred}}(\lambda) = \sum_{i=1}^{M} f_i \, B_i(\lambda)$$

We perform intensity rescaling [56, 57] by applying a scalar correction to every computed spectrum to minimize the RMSD to the experimental spectrum before plotting:

$$s^* = \frac{\sum_\lambda \theta_{\mathrm{exp}}(\lambda)\,\theta_{\mathrm{pred}}(\lambda)}{\sum_\lambda \theta_{\mathrm{pred}}(\lambda)^2}, \qquad \theta_{\mathrm{fit}}(\lambda) = s^* \theta_{\mathrm{pred}}(\lambda).$$

A.4.2. Conversion of CD signal to mean residue molar ellipticity

Raw CD signals were recorded in millidegrees (mdeg) and converted to ellipticity (θ) in degrees. Mean residue molar ellipticity $[\theta]_{\mathrm{MRE}}$ was calculated as:

$$[\theta]_{\mathrm{MRE}} = \frac{\theta \times 100}{c \times \ell \times N}.$$

where θ is the measured ellipticity in degrees, c is the molar peptide concentration in M, ℓ is the optical path length in cm, and N is the number of residues [58]. The result is in deg M^{-1} m^{-1}, which is equivalent to the commonly used units of deg cm^2 $dmol^{-1}$. This normalization enables direct comparison of spectra across peptides of different sizes or under different conditions such as path length or concentration.

B. Supplementary Information

B.1. Additional benchmarking analyses

B.1.1. Comparison to coarse-grained IDP models

In Section B.1 we focus on comparisons to well-known, atomistic structure prediction models. Recently, several ensemble prediction methods trained on coarse-grained MD simulations such as CALVADOS [59] have emerged. Here, we compare Topos-1 to IDPFold and IDPForge, two recent models that utilize CALVADOS trajectories as part of their training data. We note that the results presented in Section B.1 are based on unrelaxed and unminimized model outputs, which was done to quantify the agreement of raw model outputs with experiment. In contrast, IDPForge by default post-processes its outputs with MD relaxation. To provide a fair comparison, we also minimized Topos-1 outputs as described in Section B.1.3. IDPFold does not predict side chains, so it was excluded from the NMR shift benchmark. CALVADOS is trained to show excellent agreement with global experimental measurements such as R_g while only approximately capturing local properties by using an effective coarse-grained potential [60]. Topos-1 outperforms both IDPFold and IDPForge when evaluated against both experimentally measured local and global ensemble properties and our simulation-based evaluation suite.

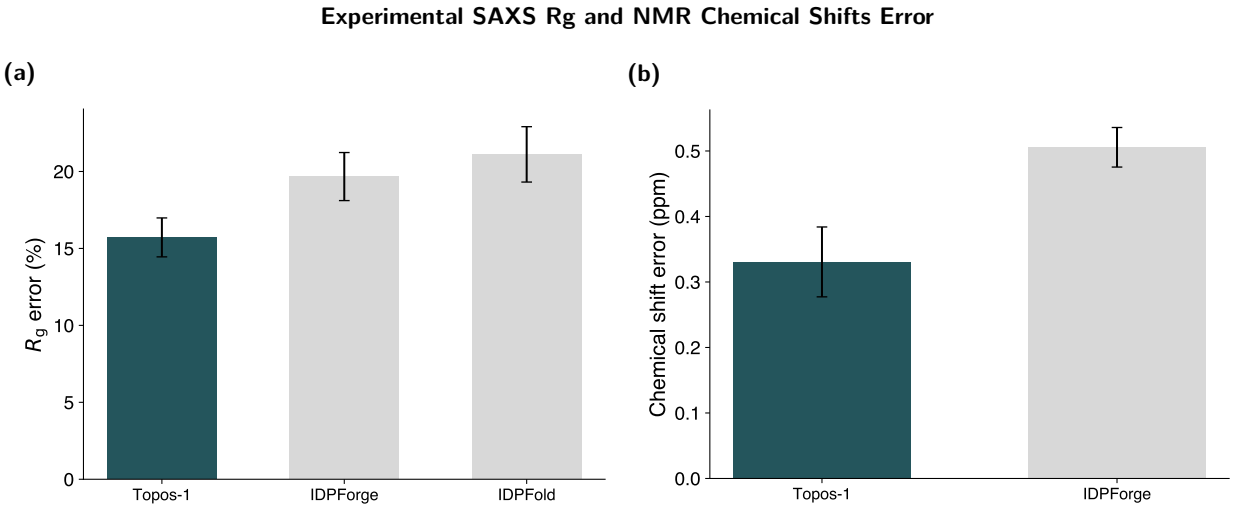

Figure S2 | Topos-1 outperforms models trained on coarse-grained MD simulations. (a) Mean absolute percent error in radius of gyration (R_g). Experimental R_g values determined via SAXS. **(b)** Mean absolute error in NMR chemical shift predictions across 12 IDPs and 5 atom types (carbonyl C, C_α, C_β, N, H). Error bars denote the standard error of the mean across proteins, computed from $n = 200$ samples per protein for each model.

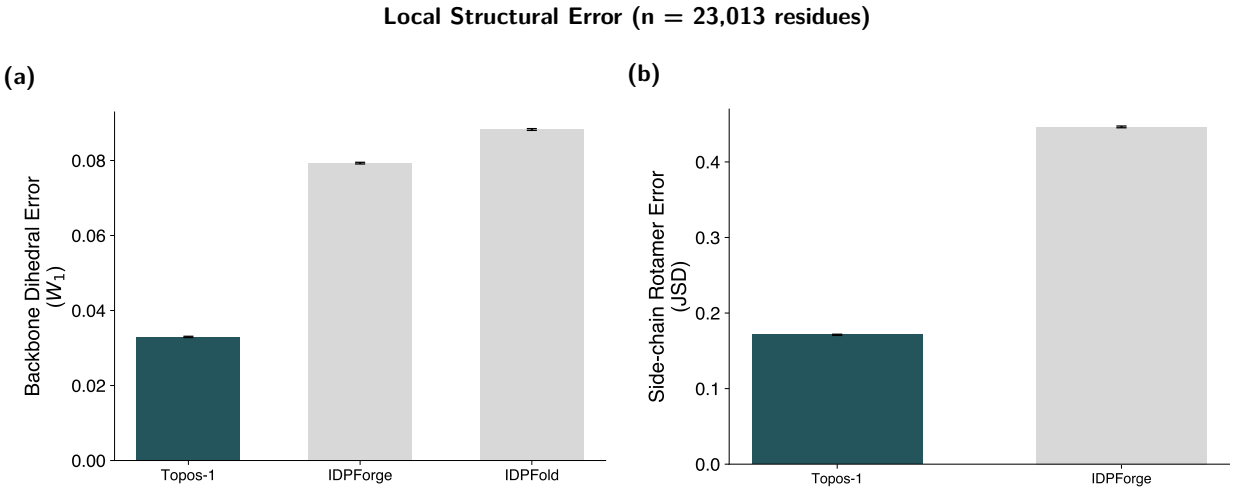

Figure S4 | Comparison between Topos-1 and models trained on coarse-grained MD data on local backbone and side chain conformational ensembles metrics. For details, we refer the reader to Sections 2.2 and Supplementary Sections B.3.4 and B.3.5. Because IDPFold does not generate side chain coordinates, it was excluded from the rotamer-based evaluation.

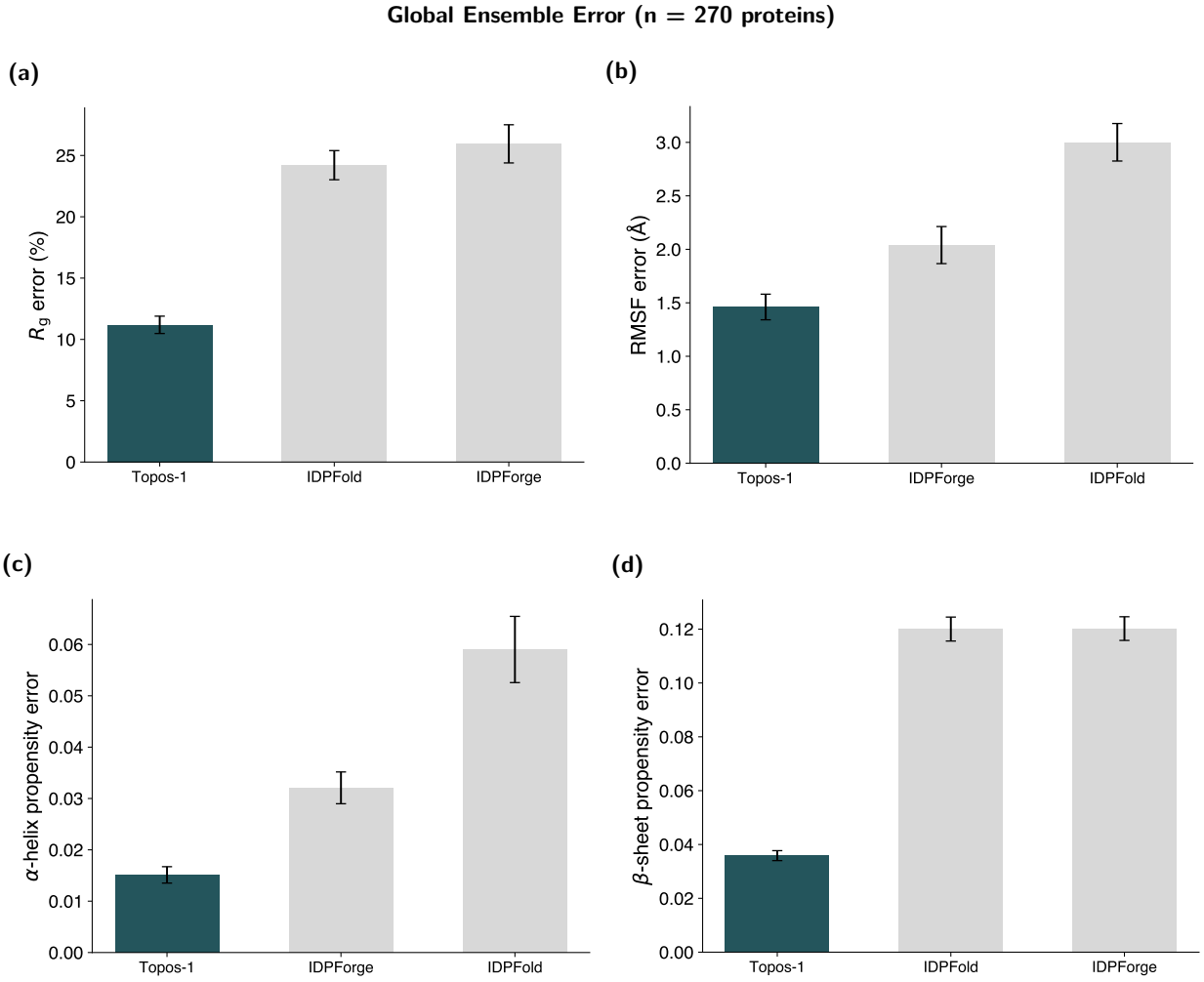

Figure S3 | Comparison between Topos-1 and models trained on coarse-grained MD data on global ensemble properties obtained from physics-based molecular dynamics simulations. For details, we refer the reader to Section 2.2.

B.1.2. Benchmark dataset construction

To assess performance, several high-quality experimentally measured data sources largely focused on disordered proteins were curated.

This included a 104 protein SAXS-derived R_g dataset reported by Novak et al. [28] and made available via GitHub repository[1]. The original 133 protein set was filtered to include only those sequences containing ≤ 200 residues, and for each protein the average measured R_g was used as a reference.

A separate 12 protein dataset containing both NMR-derived chemical shift information and SAXS-derived R_g values was constructed from Zhang et al. [27] and Robustelli et al. [26]. Though Zhang reports experimental R_g results on 28 proteins, only 14 of these have experimentally measured chemical shifts, and we discarded 2 of those because they contained only a subset of atom types (His5 was hydrogen only, and rs1 was carbon and C_α only) to arrive at our final 12 protein set.

Lastly, a 270 protein sequence test dataset that was omitted from Topos-1 model development was constructed, which contained a significant corpus of proprietary MD simulation data. To prevent potential data leakage when splitting this dataset, we utilized MMSeqs2 linclust-based sequence similarity clustering with a 50% identity threshold [29]. See Section B.2 for further details.

B.1.3. Model inference and ensemble generation

For comparative purposes, the peptide structure prediction models AlphaFold-2 [4], BioEmu [54], Boltz-2x [7], Chai-1 [6], IDPFold [55], and IDPForge [27] were all considered in addition to our Topos-1 model. We note that STARLING [28] was omitted from final consideration, as we are interested in drug discovery applications, a task for which coarse-grained structure prediction models are ill-equipped. Each model was given an input peptide sequence (either in FASTA format or as a pre-computed multi-sequence alignment generated via MMSeqs2 [29] from Uniref100 clusters [61]) and configured to generate 200 conformational samples from that sequence. Models were executed with out-of-the-box default parameter settings, except as further specified below.

AlphaFold-2 was selected over AlphaFold-3 [5] because license restrictions prevent us (as a commercial entity) from using it in evaluations. Since AlphaFold-2 was designed to generate a single confident structure per model output and cannot be configured to generate diverse samples without introducing architectural changes such as MSA subsampling [62] or dropout layers [63], we use AlphaFold-Multimer [30], an extension of AlphaFold-2, as it supports generating multiple samples per model by running the model multiple times with different random seeds.

BioEmu samples were generated using the `bioemu-v1.1` model.

To produce the reported Boltz-2x outputs, the configuration flag `--use-potentials` was added to introduce physics-based steering into the generated outputs. Results using Boltz-2 without steering were also generated, but not reported as Boltz-2x performed better in our evaluations.

In addition to the raw, unbiased outputs produced by each model, further post-processing stages were optionally performed depending on the model and specific test to ensure a more equitable basis for comparison. This included adding missing C-terminal oxygen atoms, adding missing hydrogen atoms, running energy minimization with particular atom restraints (none, C_α only, all heavy atoms, all atoms), and running MD relaxation without atom restraints. Langevin dynamics and 1000 steps of 2 fs simulation operating at 300 K were used.

B.1.4. Radius of gyration (R_g) computation

Radius of gyration (R_g) is a global metric that quantifies the spatial extent of the conformational ensemble. R_g is computed for each conformation in an ensemble using the $C\alpha$ coordinates. For a conformation with N $C\alpha$ atoms at positions $\{\mathbf{r}_i\}_{i=1}^{N}$ and center of mass $\mathbf{r}_{\mathrm{cm}}$, the radius of gyration is defined as

$$R_g = \sqrt{\frac{1}{N}\sum_{i=1}^{N} \|\mathbf{r}_i - \mathbf{r}_{\mathrm{cm}}\|^2}.$$

[1] github.com/holehouse-lab/supportingdata/tree/master/2026/starling_2026/analysis/experimental_comparison/saxs_rg

B.1.5. NMR chemical shift evaluation

Nuclear magnetic resonance (NMR) chemical shifts capture the variation in the electron distribution for a given type of nucleus and can be used as a measure of local molecular geometric differences. To quantify the error in chemical shifts relative to experimentally measured references, we ran the conformational samples of each model through the SPARTA+ chemical shift prediction software `Version 2.9 Rev 2017.143.12.12` [64]. SPARTA+ predicts, for each residue, the value of the chemical shift for each backbone atom type (C_α, C, N, H) as well as side chain atom type C_β.

Chemical shift errors were calculated per-residue and per-atom type by taking the predicted outputs and comparing them with experimental reference values. Residue level alignment was performed in instances where the sequences differed and the raw error deltas per residue and atom type were first averaged across samples to get ensemble average errors. These ensemble-averaged errors were then aggregated to provide overall errors per atom type by taking the mean absolute error across residues and then further aggregating these across proteins by again taking the mean of the absolute errors. This is reported in Fig. S5. Finally these errors per atom type were aggregated to get a single protein level mean absolute error for each model, which is reported in Fig. 2.

Since the side chain atom type C_β is included in the references and predicted outputs, models that did not include side chain atoms were excluded from comparison. This eliminated the backbone-only BioEmu and IDPFold models from consideration.

SPARTA+ was selected over other chemical shift prediction packages like UCBShift2 [65] primarily based on its inference speed and integration into popular MD analysis software such as MDTraj [66].

B.1.6. Additional supplementary figures

A breakdown per atom type of the mean absolute errors in chemical shifts from Fig. 2 is reported in Fig. S5.

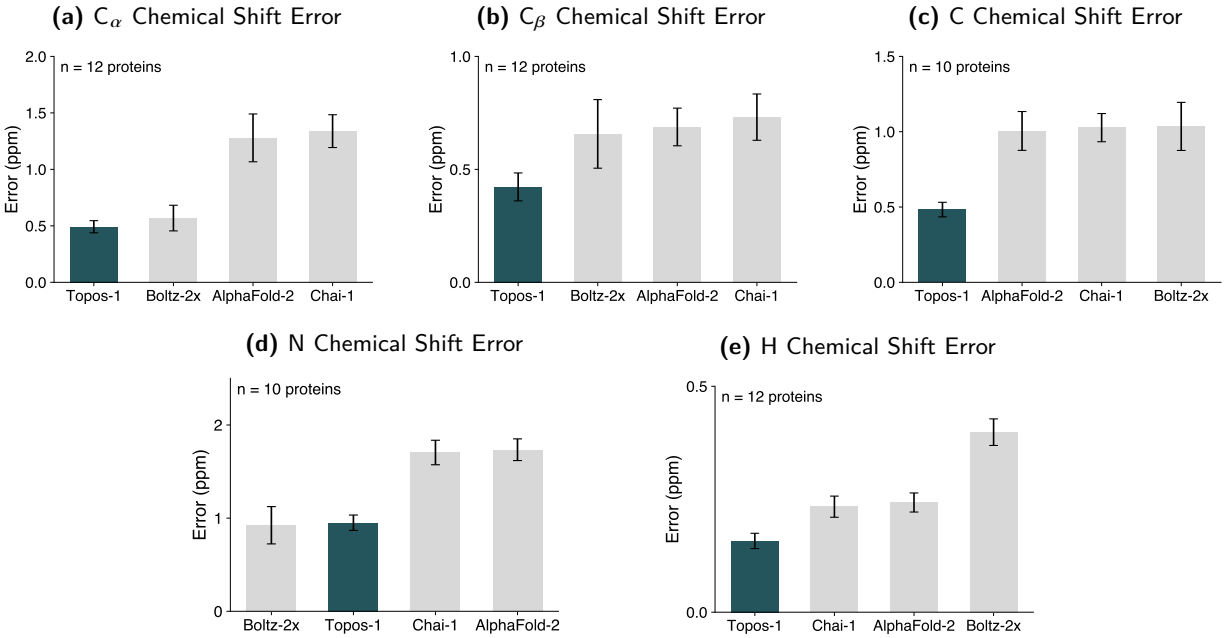

Figure S5 | Extended results for Fig. 2. Topos-1 outperforms leading protein structure prediction models on NMR experimental chemical shifts across atom types. Sample sizes vary per atom type due to limited experimental NMR data. Error bars denote standard error of the mean across proteins. **(a)** Mean absolute error in PPM for C_α atom predictions across 12 IDPs **(b)** Mean absolute error in PPM for C_β atom predictions across 12 IDPs **(c)** Mean absolute error in PPM for C (carbonyl) atom predictions across 10 IDPs **(d)** Mean absolute error in PPM for N atom predictions across 10 IDPs **(e)** Mean absolute error in PPM for H atom predictions across 12 IDPs.

B.2. Data leakage prevention

To prevent data leakage, we first clustered all sequences using MMSeqs2 linclust at a 50% sequence identity threshold [29]. We then performed embedding-based clustering using FAISS with a cosine similarity threshold of 0.995 on mean residue-level embeddings from our protein language model.

All sequences assigned to the same sequence- or embedding-based cluster were placed in the same dataset split. None of the proteins evaluated in this study, nor any proteins with high sequence or embedding similarity to them, appear in the Topos-1 training set. This includes all proteins that appear anywhere in the paper's results, including benchmark metrics, model comparisons, and case studies.

B.3. Structural metrics and distance definitions and calculations

B.3.1. Normalized radius of gyration error

For each protein, we calculate the ensemble-averaged R_g as the arithmetic mean across all conformations. The normalized R_g error is defined by

$$\text{error}_{R_g} = \frac{\left|\bar{R}_g^{\text{model}} - \bar{R}_g^{\text{ref}}\right|}{\bar{R}_g^{\text{ref}}},$$

where $\bar{R}_g^{\text{ref}}$ denotes the ensemble-averaged radius of gyration from the reference (MD) ensemble. For each baseline model, we report the mean normalized R_g error across all proteins, expressed as a percentage. R_g is calculated as described in Supplementary Section B.1.4.

B.3.2. RMSF error

Root-mean-square fluctuation (RMSF) quantifies the positional variability of each residue across the structures in an ensemble and is a standard measure of local flexibility.

RMSF is computed using only the $C\alpha$ coordinates. Prior to calculating RMSF, we align all structures in an ensemble to a common reference frame using Biopython's SVDSuperImposer [67], which performs an optimal rigid-body superposition via singular value decomposition (Kabsch algorithm [68]). The reference frame is defined as the ensemble-mean structure, which is the element-wise average of the coordinates across all N conformations. After alignment, the RMSF is computed as

$$\text{RMSF}(i) = \sqrt{\frac{1}{N}\sum_{j=1}^{N}\left\|\mathbf{r}_i^{(j)} - \bar{\mathbf{r}}_i\right\|^2},$$

where $\mathbf{r}_i^{(j)}$ denotes the aligned 3D position of residue i in conformation j, and $\bar{\mathbf{r}}_i$ represents the mean position of the residue i across the ensemble, calculated as

$$\bar{\mathbf{r}}_i = \frac{1}{N}\sum_{j=1}^{N}\mathbf{r}_i^{(j)},$$

and $\|\cdot\|$ denotes the Euclidean norm.

For each protein ensemble, the mean RMSF is computed as the arithmetic mean across all L residues:

$$\overline{\text{RMSF}} = \frac{1}{L}\sum_i \text{RMSF}(i).$$

For each protein p, we compute the ensemble-level mean RMSF for both the model-generated ensemble $\overline{\mathrm{RMSF}}^{(p)}_{\mathrm{model}}$ and the corresponding reference ensemble $\overline{\mathrm{RMSF}}^{(p)}_{\mathrm{ref}}$. The RMSF error for protein p is defined as

$$\mathrm{error_p} = \overline{\mathrm{RMSF}}^{(p)}_{\mathrm{model}} - \overline{\mathrm{RMSF}}^{(p)}_{\mathrm{ref}}.$$

Across a set of P proteins, we determine the overall RMSF error of the model using the mean absolute error (MAE):

$$\mathrm{MAE} = \frac{1}{\mathrm{P}} \sum_{\mathrm{P}} |\mathrm{error_p}|.$$

B.3.3. α–helix and β–strand propensity errors

Secondary structure propensity is calculated using the $\mathrm{C}\alpha$ coordinates. We use Biotite's annotate_sse function [69] to assign secondary structure elements (SSEs) as α-helix or β-strand on a per-residue, per-conformation basis. Per-residue propensity is defined as the fraction of conformations across an ensemble in which a residue adopts the corresponding SSE type. The mean propensity for a protein is calculated as the average of per-residue propensities across the sequence.

We calculate a model's secondary structure propensity error as the mean absolute error (MAE) across per-protein differences in mean secondary structure propensity between model-generated ensembles and corresponding reference ensembles derived from molecular dynamics (MD). We treat α-helix and β-strand propensities separately.

B.3.4. Backbone dihedral distance

The backbone dihedral distance was calculated by comparing the ϕ and ψ angle distributions between the generated ensembles and the MD ensembles. For each residue of each protein in the test set, we determined both ϕ and ψ using built-in MDTraj methods [66].

We mapped the dihedral angles to the circle $S^1 \cong [0, 1)$, and compared the resulting empirical measures using the circular 1-Wasserstein (W_1) distance as implemented by `ot.wasserstein_circle` in POT [70]. For two probability measures u and v on $[0, 1)$ with cumulative distribution functions

$$F_u(t) = u([0, t]), \qquad F_v(t) = v([0, t]), \qquad t \in [0, 1),$$

the circular W_1 distance can be written as [71]

$$W_1(u, v) = \int_0^1 |F_u(t) - F_v(t) - \mathrm{LevMed}(F_u - F_v)|\, dt,$$

where the level median LevMed is defined by

$$\mathrm{LevMed}(f) := \min\left\{\arg\min_{\alpha \in \mathbb{R}} \int_0^1 |f(t) - \alpha|\, dt\right\}.$$

We used W_1 (rather than W_2) because it has a more direct interpretation and is less sensitive to small differences in distribution tails.

The scores reported in Fig. 4 were first averaged over ϕ and ψ, and then over all residues in the test set, noting that the N- and C-terminal residues lack defined values of ϕ and ψ, respectively.

B.3.5. Rotamer dihedral distance

The rotamer distance metric was calculated by comparing the distribution of side chain dihedral states between the generated ensembles and the MD ensembles. Side chain conformations are quantified by discretizing each χ dihedral angle, as calculated by MDTraj methods [66], into three canonical rotamer states: trans ($\approx 180°$), gauche+ ($\approx 60°$), and gauche– ($\approx -60°$). Each residue is represented by the joint state of its rotatable side-chain dihedrals (e.g., $\chi_1 = t$, $\chi_2 = g+$, $\chi_3 = g-$ for a residue with three χ angles). Residue-wise rotamer distributions for each model were then compared to the

corresponding MD distributions using the Jensen–Shannon divergence (JSD) and averaged across residues in the test set. Residues with no χ angles (GLY, ALA) are ignored in this calculation. Reported scores are shown in Fig. 4.

To quantify the similarity between two probability distributions, we used the Jensen–Shannon divergence, a symmetric and bounded measure derived from the Kullback–Leibler (KL) divergence. Given two discrete or continuous probability distributions $P(x)$ and $Q(x)$ defined over the same support, we first define their mixture distribution

$$M(x) = \frac{1}{2}\big(P(x) + Q(x)\big).$$

The Jensen–Shannon divergence is then given by

$$\mathrm{JSD}(P \parallel Q) = \frac{1}{2}\mathrm{KL}(P \parallel M) + \frac{1}{2}\mathrm{KL}(Q \parallel M)\,,$$

where the Kullback–Leibler divergence is defined as

$$\mathrm{KL}(P \parallel Q) = \int P(x)\,\log\frac{P(x)}{Q(x)}\,dx,$$

with the integral replaced by a sum in the discrete case.

The JSD is non-negative, symmetric, and bounded such that $0 \le \mathrm{JSD}(P \parallel Q) \le \ln 2$, where a value of zero indicates identical distributions and $\ln 2$ corresponds to maximally distinct distributions with no shared information.

B.4. Case study: Topos–1 captures the behavior of the Parkinson's disease target α–synuclein (extended analyses)

B.4.1. Additional secondary structure analyses

In addition to the α-helical propensities reported in Fig. 10, Figs. S6 and S7 show the calculated β-strand and random coil propensities, respectively, for the ensembles generated with Topos-1, AlphaFold-2, and BioEmu. In both cases, Topos-1 again demonstrates excellent agreement with the ensemble obtained with MD simulation [33].

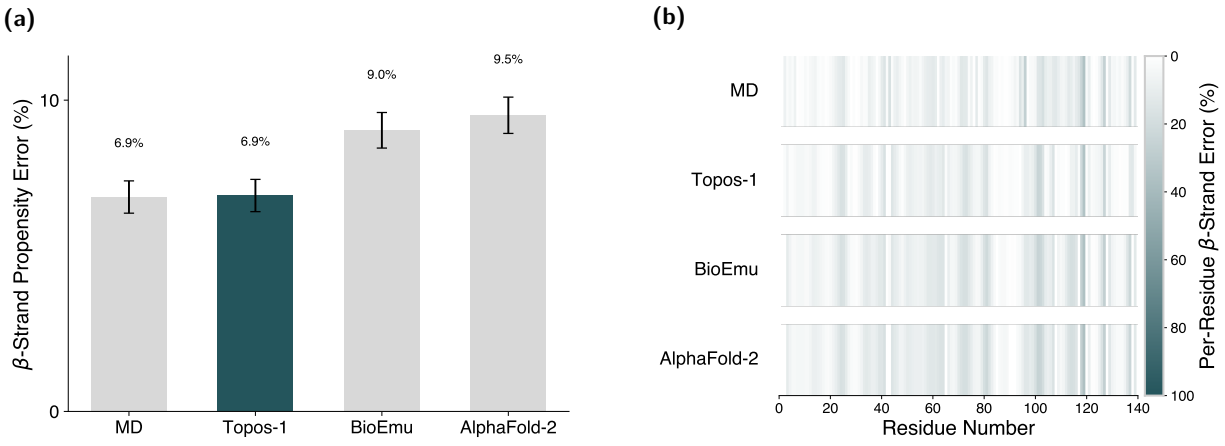

Figure S6 | Topos-1 generates conformational ensembles of Parkinson's disease target α-synuclein that outperform state-of-the-art structure prediction models in experimental fidelity. (a) Mean absolute error in β-strand propensity relative to experiment, averaged across the α-synuclein sequence ($n = 140$ residues). (b) Residue-resolved mean absolute error in β-strand propensity indicates largely similar regions of error across the models, with a lower overall error obtained with MD and Topos-1. Experimental measurements indicate moderate (~10%) β-strand propensity across the sequence [40].

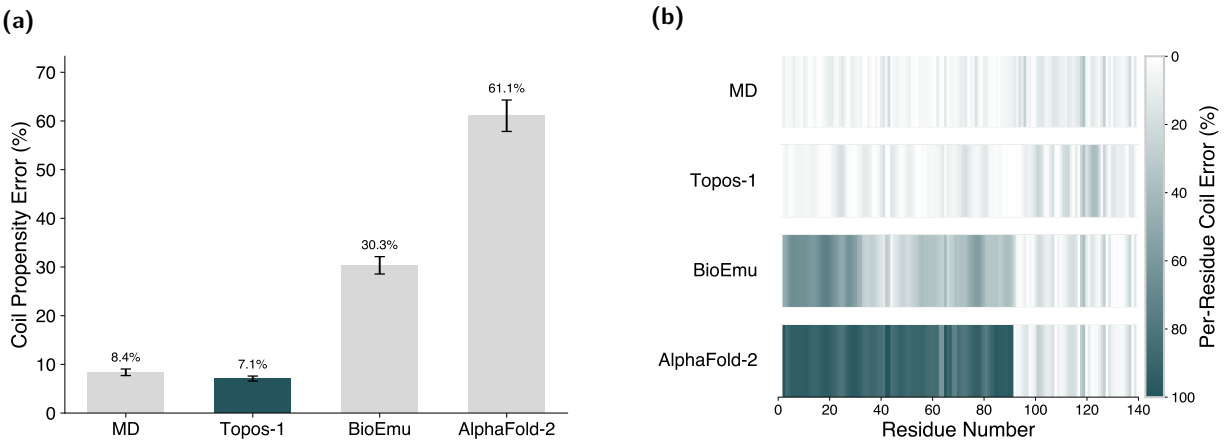

Figure S7 | Topos-1 generates conformational ensembles of Parkinson's disease target α-synuclein that outperform state-of-the-art structure prediction models in experimental fidelity. (a) Mean absolute error in random coil propensity relative to experiment, averaged across the α-synuclein sequence ($n = 140$ residues). (b) Residue-resolved mean absolute error in random coil propensity indicates largely similar regions of error across the models, with a lower overall error obtained with MD and Topos-1. Experimental measurements indicate significant random coil propensity across the sequence [40]. Experimental random coil propensities obtained by subtracting α-helix and β-strand propensities from unity, which is valid for α-synuclein.

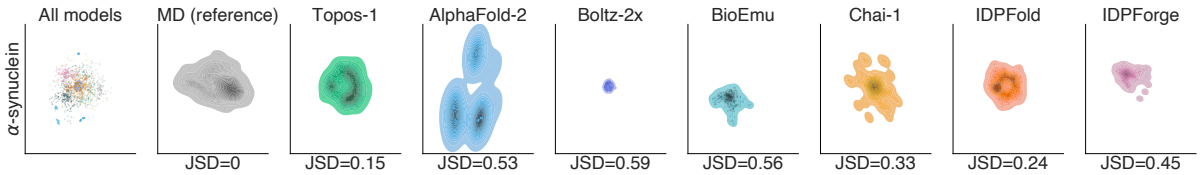

Figure S8 | Extended results for Fig. 12. The kernel density estimates and JSD to the reference MD [33] are shown for each model based on CVs obtained from ELViM embeddings. The embeddings are calculated by combining samples from each model. Consistent with the findings in Figs. 10, 11, and 12, Topos-1 has the strongest agreement with data obtained from publicly available MD simulations for α-synuclein.

B.4.2. Extended ELViM analyses

The results in Fig. 12 were obtained based on samples from AlphaFold-2, Boltz-2x, and BioEmu. The results in Figs. 5 and S8 were obtained based on samples from all models. In all cases, embeddings were calculated using default parameters and 200 samples from each model.

In this work, probability distributions were estimated using a kernel density estimation in the relevant collective variable space, and the JSD (see Section B.3.5) was computed between model-generated and MD reference distributions to quantify ensemble-level agreement. Gaussian kernel density estimations were performed using the SciPy [72] software package. Scott's rule [73] was used to estimate the kernel bandwidth.

B.5. Case study: Topos–1 demonstrates sensitivity to sequence perturbations (extended analyses)

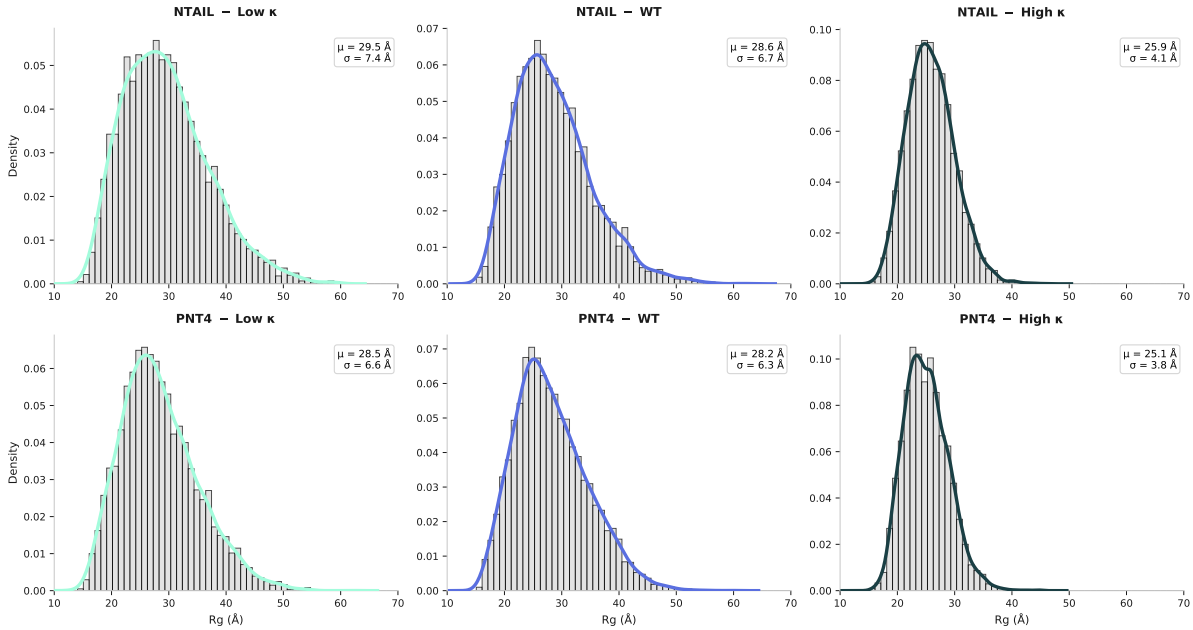

Figure S9 | Extended results for Fig. 13. We present the full R_g distributions for low κ, wild type, and high κ for each of NTAIL and PNT4 as modeled by Topos-1.

B.5.1. Extended R_g and D_{max} distributions

In addition to the radius of gyration (R_g), the maximum particle dimension (D_{max}) is commonly reported from SAXS experiments. However, D_{max} is inferred from the long-distance tail of the pair-distance distribution function P(r) and therefore comes with substantial uncertainty, particularly for flexible or disordered systems. Computationally, the proxies for D_{max}, usually the maximum or high-percentile interatomic distance within an ensemble, are also dominated by rare, extended conformations and thus exhibit high variance and strong sensitivity to sampling noise.

As such, R_g is reported as the primary measurement to compare to experimental SAXS results in the main text. However, accurate prediction of R_g does not guarantee accurate prediction of D_{max}. Nevertheless, when examining the high-distance tail of the ensembles, we observe consistency with experimental trends. Specifically, we report the distribution of interatomic distances as a function of percentile and use the extreme high-percentile regime (99% and above) as a computational proxy for D_{max}. For both NTAIL and PNT4, we recover the experimental rank low-$\kappa \gtrsim$ WT > high-κ, with WT closely resembling the low-κ variant and the high-κ variant exhibiting a markedly more compact ensemble.

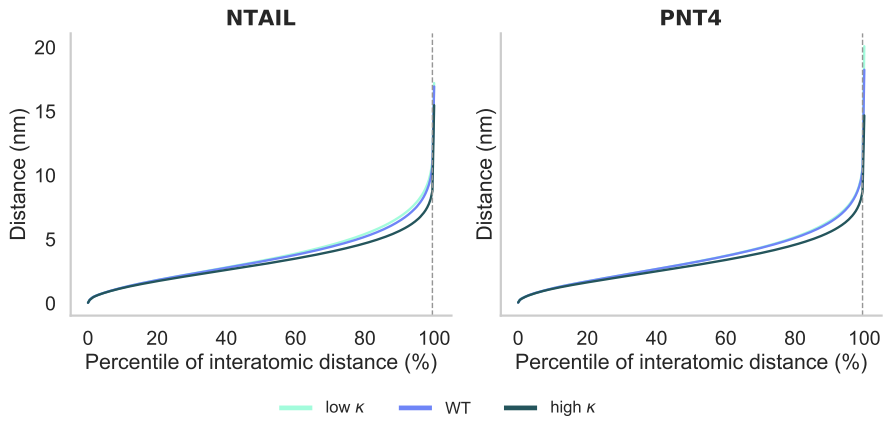

Figure S10 | Interatomic distance percentiles for NTAIL and PNT4 variants. Curves plot pooled heavy-atom distance versus percentile for low-κ (cyan), WT (blue), and high-κ (dark green), with 99–100% insets and a dashed line marking 99.5%, a possible percentile to make $D_{\max}$ proxies, highlighting the tail distribution that corresponds to the maximum particle dimension.

B.6. Case study: Topos–1 facilitates structure–based discovery of small molecules for the prostate cancer target androgen receptor (extended results and assays)

B.6.1. Androgen receptor trans–activation assay (L2 set)

A trans-activation assay was used to measure androgen receptor (AR) activity and compute IC_{50} values for compounds of interest. This assay used a Firefly Luciferase gene under the control of an Androgen Response Element, stably expressed in a prostate cancer cell line (22Rv1). The cell line AR-Luc-22Rv1 was acquired from a commercial source (BPS Bioscience) and maintained according to the manufacturer's instructions. The cell culture media consisted of RPMI 1640, ATCC modification (ThermoFisher Scientific) supplemented with 10% Fetal Bovine Serum (FBS) (Corning) and 1% Penicillin-Streptomycin (Pen-Strep) (Corning). Cells were passaged by enzymatic dissociation (Accutase) (Corning) at approximately 90% confluence on standard tissue culture plastics.

In preparation for the trans-activation assay, cells were enzymatically dissociated as above and re-plated into assay media (RPMI 1640, No Phenol Red) (ThermoFisher Scientific) supplemented with 10% Charcoal Stripped FBS (Corning) and 1% Pen-Strep. Re-plating was done onto white, clear-bottom 96-well plates (Greiner) at a density of 20,000 cells per well. After 24 hours, media was exchanged with assay media treated with Dimethyl Sulfoxide (DMSO) (Sigma-Aldrich) or 2 nM Dihydrotestosterone (DHT) (Sigma-Aldrich) and a dilution series of compounds of interest in DMSO. The final concentration of DMSO in all wells was 0.2%. Each plate contained Positive Control (DMSO only) and Negative Control (DMSO plus 2 nM DHT) wells. All compound-treated wells contained 2 nM DHT. Positive and negative control wells were replicated six times per plate. Each compound dilution was replicated three times per plate.

Cells were incubated with compounds for 16 hours and then treated with luciferase detection reagent (ONE-Glo EX Luciferase Assay System) (Promega) according to the manufacturer's instructions. Plates were read on a GloMax (Promega) plate reader in luminescence mode with a 1 second integration time. Data were expressed as percent inhibition, and IC_{50} values were calculated from a 4-parameter Hill equation as follows:

$$\text{Percent Inhibition} = \text{Positive Control} + \frac{\text{Negative Control} - \text{Positive Control}}{1 + (\text{IC}_{50}/[\text{Compound}])^{\text{Hill Slope}}},$$

where the maximum of the curve was constrained to 100% inhibition.

B.6.2. Extended affinity prediction results (L1 and L2 sets)

Ligand set L1 was assembled primarily from the compound series reported by Basu *et al.* [48], which to our knowledge constitutes the largest publicly available collection of small-molecule structure-activity relationship (SAR) data targeting the AR Tau5 R2R3 segment with reported IC_{50} values. The compounds in this series are structurally closely related, posing a challenge for conventional docking-based ranking due to their limited chemical diversity and compressed activity range. To provide additional reference points, a small number of representative compounds (ET516, EPI-7386, EPI-7170 and EPI-002) reported in other studies were included [45–47], although their IC_{50} values were measured under different assay conditions. As a result, some deviations from the overall trend are expected and likely reflect inter-assay variability.

Despite these sources of experimental heterogeneity, Topos-1-AP scores exhibit statistically significant agreement with experimental IC_{50} values, yielding a Pearson correlation coefficient of 0.611 and a Spearman rank correlation of 0.757. These results indicate that the ranking predicted by Topos-1-AP is robust even in the presence of structurally similar ligands and partially inconsistent experimental references.

Ligand set L2 consists of 27 proprietary compounds designed and evaluated in-house, with activities measured using a consistent cell-based assay. Topos-1-AP scores remain in good agreement with experimental measurements, yielding a Pearson correlation coefficient of 0.734 and a Spearman rank correlation of 0.593. These results suggest that the affinity prediction model retains meaningful ranking and predictive power, supporting its potential utility for forward prediction and prioritization in future compound design.

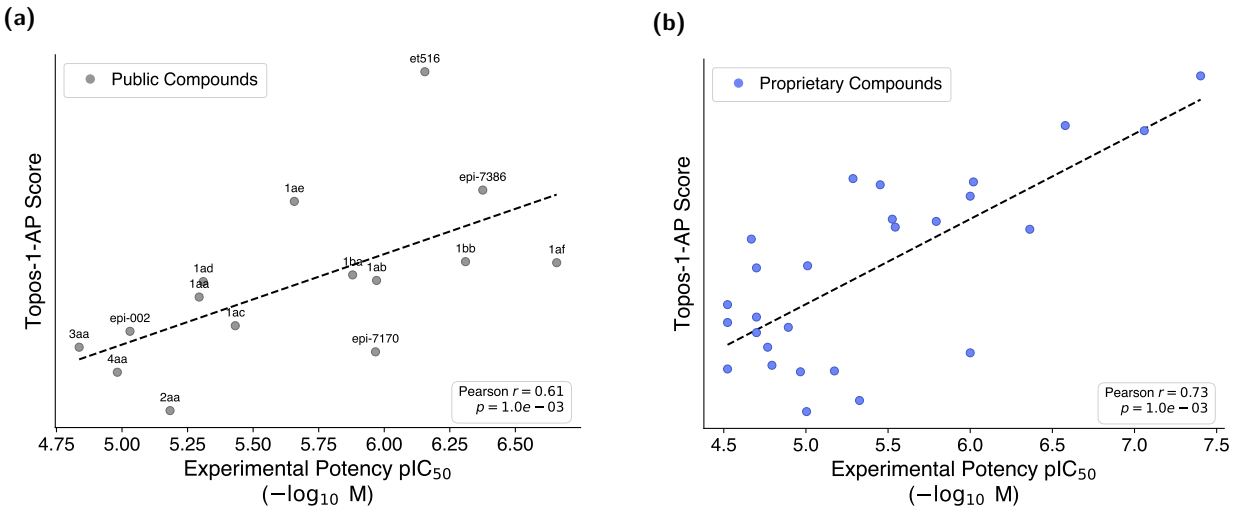

Figure S11 | (a) Correlation between the Topos-1-AP score and experimental potency of literature-reported compounds (set L1, 15 compounds). The experimental IC_{50} values of ET516, EPI-7386, EPI-7170 and EPI-002 were taken from several different reports [45–47]. All other experimental IC_{50} values were taken from Basu et al. [48]. **(b) Correlation between the Topos-1-AP score and experimental potency from an in-house cell-based assay, including proprietary compounds (set L2, 27 compounds).** Dots denote measured IC_{50}. The dashed line is the least-squares fit. Pearson's r and Spearman's ρ describe the linear relationship and ranking with two-sided p values. A larger Topos-1-AP score reflects more favorable binding.