Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Advances in the determination of disordered protein ensemble

Hamidreza Ghafouri[1], Silvio C. E. Tosatto[1,2] and Alexander Miguel Monzon[1]

Intrinsically disordered proteins (IDPs) play essential roles in regulation, signaling, and phase separation, yet their structural complexity cannot be captured by a single conformation. Instead, they populate dynamic ensembles that encode a context-dependent function. Recent advances in experimental techniques coupled with physics-based simulations, coarse-grained models, and machine learning, have transformed our ability to generate and interpret IDP ensembles. Integrative frameworks now combine complementary data with computational approaches to refine ensembles at both local and global levels. Nevertheless, challenges remain in benchmarking, error estimation, and modeling assemblies involving protein–protein and protein–nucleic acid interactions. We highlight recent progress and outline the emerging directions that will shape the next generation of ensemble determination methods.

Addresses
[1] Department of Biomedical Sciences, University of Padova, Via Ugo Bassi 58/B, 35131, Padova, Italy
[2] Institute of Biomembranes, Bioenergetics and Molecular Biotechnologies, National Research Council (CNR-IBIOM), Bari, 70126, Italy

Corresponding authors: Tosatto, Silvio C.E. (silvio.tosatto@unipd.it); Monzon, Alexander Miguel (alexander.monzon@unipd.it)

Current Opinion in Structural Biology 2026, **96**:103198

This review comes from a themed issue on **Folding, Binding and Protein Design (2026)**

Edited by **John Christodoulou** and **Michele Vendruscolo**

For complete overview of the section, please refer the article collection - Folding, Binding and Protein Design (2026)

Available online 11 December 2025

https://doi.org/10.1016/j.sbi.2025.103198

Introduction

Approximately 70 % of human proteins contain at least one segment of 30 or more amino acids lacking stable secondary or tertiary structure, and ∼5 % are entirely disordered [1]. These intrinsically disordered proteins and regions (IDPs/IDRs) play central roles in regulation, signaling, and phase separation, yet their behavior cannot be captured by a single static structure. Instead, they populate dynamic conformational ensembles that encode functional versatility and context-dependent interactions [1]. While state-of-the-art predictors such as AlphaFold [2] excel for folded proteins, they fail to represent IDP heterogeneity, underscoring the need for dedicated approaches. Ensemble determination requires an integrative workflow that combines experimental techniques, computational generation, and validation (Figure 1) [3]. Advances in biophysical methods, physics-based simulations, knowledge-driven strategies, and machine learning (ML) now allow ensembles constrained at multiple scales. By merging data integration frameworks with artificial intelligence (AI), the field is moving toward quantitative and predictive descriptions of disordered protein behavior. Here, we review recent progress in experimental and computational methodologies for IDP ensemble determination.

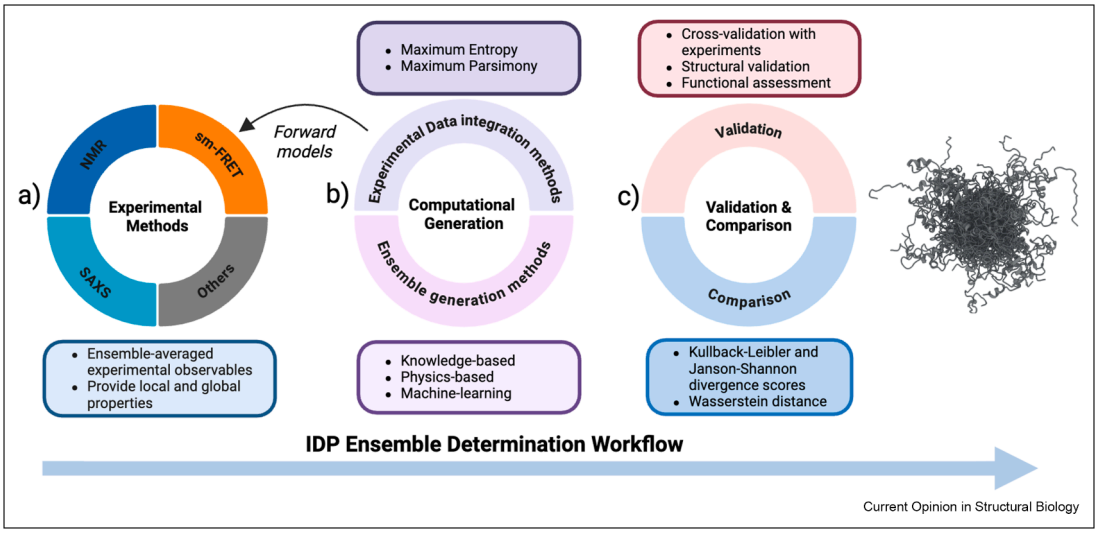
Recent advances in experimental techniques

The structural characterization of IDPs relies on experimental approaches that capture the broad conformational heterogeneity and dynamic nature of these systems. In recent years, methodological advances in nuclear magnetic resonance (NMR) spectroscopy, single-molecule Förster resonance energy transfer (smFRET), and small-angle scattering (SAS), complemented by other biophysical techniques, have significantly expanded our ability to determine conformational ensembles with higher resolution, and physiological relevance [3] (Figure 2).

NMR remains the gold standard for atomic-level insight into IDPs, capturing ensemble-averaged features such as secondary structure propensities, backbone torsion angles, and long-range contacts with minimal perturbation [3]. Recent developments in relaxation dispersion and paramagnetic techniques (e.g. PRE) have enhanced detection of sparsely populated or transiently folded states [4]. Moreover, in-cell NMR continues to mature, extending studies to physiologically relevant environments [5] (Figure 2b).

smFRET offers nanometer-scale distance measurements and uniquely resolves subpopulations and

Figure 1



Roadmap for the determination of IDP conformational ensembles. Experimental data (e.g. NMR, smFRET, SAXS) provide observables that guide computational ensemble generation. Integration methods refine ensembles using forward models, while comparison and validation ensure consistency with independent data. IDP, intrinsically disordered protein; NMR, nuclear magnetic resonance; SAXS, small-angle X-ray scattering; smFRET, single-molecule Förster resonance energy transfer.

dynamic transitions that are obscured in ensemble-averaged methods. Applied to IDPs, smFRET has characterized global chain dimensions, conformational substates, and structural rearrangements in conditions ranging from dilute solutions to crowded cellular milieus and biomolecular condensates [6]. Technical advances in site-specific fluorophore labeling, alternating/pulsed excitation schemes, and multiparameter detection have improved quantitative analysis, while integration with polymer physics models and molecular simulations has refined the extraction of full inter-dye distance distributions and dynamic reconfiguration times [7] (Figure 2a).

SAS, including both X-ray (SAXS) and neutron (SANS) scattering, provides global information on molecular size, shape, and compactness with minimal perturbation to the system in solution. For IDPs, SAS delivers key global descriptors such as the radius of gyration (Rg), scaling exponent (ν), and pairwise distance distribution P(r), complementing the residue-level detail of NMR and smFRET [8]. Recent advances include enhanced beamline sensitivity, rapid data acquisition enabling kinetic studies from microseconds to hours, and improved analysis frameworks capable of extracting polymer scaling behavior directly from scattering profiles [9] (Figure 2c).

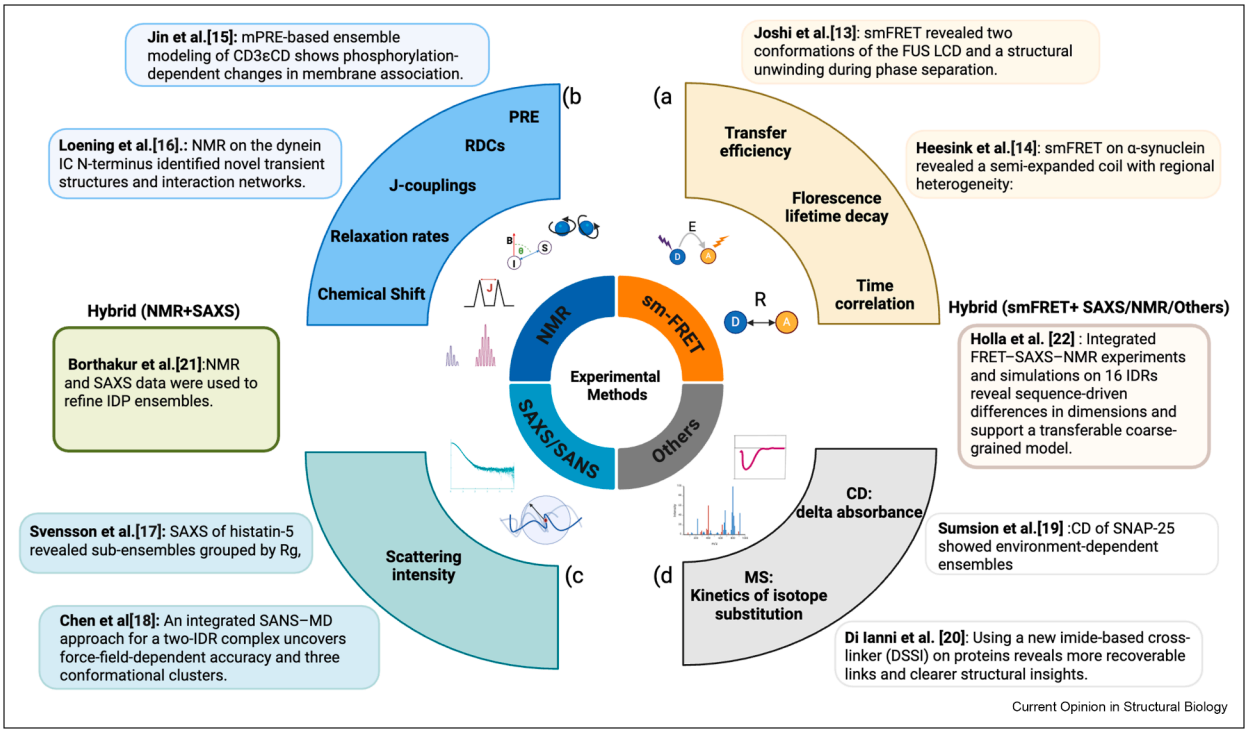
Complementary techniques have gained traction in probing additional facets of IDP structure and dynamics. Hydrogen–deuterium exchange, particularly coupled to mass spectrometry, maps transient secondary structure and conformational flexibility, while covalent labeling and hydroxyl radical footprinting capture solvent exposure at side chain resolution [10] (Figure 2d).

Collectively, these methodological advances have reshaped the experimental landscape of IDP research, fostering increasingly integrative approaches to decipher sequence-encoded interactions, conformational dynamics, and the resulting structural transitions and phase separation. Recent entries in the Protein Ensemble Database (PED) [11] often combine SAXS and NMR data and/or smFRET measurements [12], frequently complemented by other orthogonal techniques such as electron paramagnetic resonance and circular dichroism (CD), to generate ensembles constrained at both global and local levels.

Recent advances in computational methods

Computational modeling of IDPs as conformational ensembles is a widely used approach for interpreting experimental data and gaining insight into their conformational heterogeneity [23]. There are three main strategies for generating conformational ensembles of IDPs (Table 1): (a) physics-based methods, primarily using molecular dynamics (MD) simulations; (b) knowledge-based techniques, which rely on heuristics or statistical patterns derived from known protein structures; and (c) ML–based approaches, particularly

Figure 2



Experimental methods for the determination of conformational ensembles of IDPs. For each technique we highlight the most important observables and recent studies to showcase the potential of each method for the determination of conformational ensembles. **(a)** smFRET provides intramolecular distance distributions at the nanometer scale [13,14], **(b)** NMR yields residue-specific information on local structure and dynamics through chemical shifts, couplings, relaxation parameters [15,16], **(c)** SAXS/SANS reports on overall chain dimensions and global conformational properties [17,18], d) other techniques such as CD and MS-based techniques probe secondary structure content and long-range residue–residue distance restraints [19,20]. Moreover, **hybrid techniques** integrate complementary observables from different experimental methods to generate ensembles consistent with multiple structural restraints [21,22]. CD, circular dichroism; IDP, intrinsically disordered protein; MS, mass spectrometry; NMR, nuclear magnetic resonance; SANS, small-angle neutron scattering; SAXS, small-angle X-ray scattering; smFRET, single-molecule Förster resonance energy transfer.

generative models, which infer structural features from trained models without relying on explicit physical principles. Another category of computational techniques for determining IDP ensembles involves integrating experimental data into computational models to obtain ensembles that more accurately reflect experimental observations. These methods are broadly divided into two classes, maximum entropy and maximum parsimony [24], explained later in this section.

Ensemble generation techniques

Knowledge-based methods

Knowledge-based methods offer the fastest way to generate conformational ensembles. They construct conformers by sampling residue fragments from nonredundant, high-resolution experimental structures, using the observed torsion angle distributions in these datasets as empirical descriptors of allowable backbone geometries rather than as explicit statistical potentials or derived force fields. Early examples include TraDES [25], which used probabilistic sampling with secondary structure biases, and flexible-meccano [26], which added IDP-specific conformational propensities and back-calculation of experimental observables. Recent tools, such as IDPConformerGenerator [27], extend earlier approaches by incorporating fragment-based $\phi/\psi/\omega$ sampling, all-atom modeling, integration of experimental restraints, and pipelines for ensemble analysis and reweighting. The main limitation of such methods is their neglect of time-dependent physical dynamics, which prevents them from capturing kinetically relevant processes and may lead to ensembles that do not fully reflect the underlying energy landscape.

Physics-based methods

In physics-based approaches, MD simulations use force fields to model protein motions over time, offering detailed insights into the dynamic behavior of IDPs.

The choice of method depends strongly on the biological question and the available computational resources. In particular, all-atom MD simulations with force fields

Table 1

Recent ensemble generation techniques, grouped into knowledge-based, physics-based, and machine learning approaches, with applications, strengths, and limitations.

Ensemble Generation Strategy	Technique	Underlying concept	Application context	Advantages	Drawbacks
Knowledge-based	MoMA [74]	Generates conformers through motion-planning algorithms that apply steric constraints and may integrate energy models.	When extensive sampling with realistic geometry is required	Incorporates all-atom representations with physically realistic structures.	Physical dynamics are not accounted for
	IDPConformer Generator [27]	Draws torsion angles from fragment libraries, potentially refined through restraints	When incorporating both sequence data and experimental constraints	Modular framework that allows reweighting and incorporation of SAXS/NMR constraints	Relies on the quality of fragments and the effectiveness of scoring functions.
	DIPEND [75]	Generates ensembles by sampling statistical dihedral angles conditioned on neighboring residues and applying structural filters	For rapid backbone ensemble generation directly from sequence with minimal requirements	Flexible and efficient, incorporating experimental priors with rapid clash resolution and refinement.	Relies on heuristics and rotamer libraries without modeling physical dynamics, offering only backbone-level sampling.
Physics-based	Amber ff99SB ws/ff03ws [76]	All-atom modeling framework refined to better capture protein−solvent interactions	For simulating IDPs where protein−water interactions are accurately balanced.	Enhanced accuracy of global dimensions and protein−protein interactions	Computationally intensive
	DES-amber [29]	All-atom force field optimized to capture protein−protein interactions	Suitable for simulating IDPs, folded proteins, and protein−protein complexes	Enhanced accuracy in global dimensions, transient helix formation, and protein−protein interactions	Limited to amber ecosystem
	MARTINI-3 [77]	Coarse-grained approach for studying complex biomolecular structures	Suitable for studying large biomolecular or membrane-associated systems	Scales efficiently to simulate large assemblies	Simplified interactions
	CALVADOS [40,41]	Implicit-solvent coarse-grained force field for IDPs and multi-domain proteins, parameterized using a top-down approach.	Efficient modeling approach that accounts for sequence effects in IDPs and multi-domain proteins	Efficient; accurate predictions of IDP compaction	Provides residue-level resolution without accounting for short-lived secondary structure or hydrogen bonds
Machine-Learning	idpGAN [42]	GAN trained on CG-MD ensembles to produce new conformers that reproduce the statistical properties of the training data	When rapid ensemble generation is needed	Reproduces realistic ensemble variability and is highly scalable	Constrained by training data patterns from CG simulations; generates Cα-only ensemble representations.
	idpSAM [78]	Latent diffusion approach (transformer autoencoder plus diffusion) trained on ABSINTH implicit-solvent ensembles to generate new conformers	Designed to generate ensembles of short IDRs (<60 residues) and generalize to unseen sequences.	Strong ability to generalize to new sequences with robust training stability.	Depends on large amounts of simulation-derived training data.
	BioEmu [79]	Diffusion-based model trained on extensive datasets combining simulations and experimental measurements	Fast production of equilibrium ensembles directly from the sequence.	Rapid and sequence-aware, generating diverse ensembles characteristic of disordered proteins.	Restricted to backbone sampling without side chains; tends to overestimate Rg for longer IDPs (>100 residues).
Hybrid	AlphaFold-Metainference [57]	Integrates AlphaFold-derived constraints with Bayesian ensemble	Aimed at integrating AI-based insights into the	Incorporates prior knowledge	Depends on AlphaFold-derived priors and confidence metrics.

Table 1 (*continued*)

Ensemble Generation Strategy	Technique	Underlying concept	Application context	Advantages	Drawbacks
		modeling, blending ML and physics-based methods.	refinement of ensembles.		
	DynamICE [50]	Employs a Generative recurrent neural network to create conformers, refined with experimental data using a Bayesian reward function (X-EISD).	For situations where fixed conformer pools are inadequate and diverse, data-driven IDP ensembles are required.	Allows *de novo* generation of conformers consistent with data, integrating physical and experimental constraints with iterative improvement.	Local torsion-based sampling limits accuracy in meeting distance restraints such as NOEs and PREs.
	bAles [56]	Incorporates AlphaFold2 distograms as distance restraints into a random coil prior using Bayesian integration.	Designed to efficiently generate atomistic IDP ensembles, particularly under limited experimental data.	Rapid, atomistic-level modeling that is resilient to AF2 errors and captures both local and long-range features.	Relies on AF2 accuracy and is less effective for environment-dependent or transient interactions.
	AFflecto [80]	Constructs IDP ensembles through fragment-based sampling guided by AlphaFold models, incorporating secondary strucutre elements (SSEs) and steric constraints.	For efficient conformer generation directed by AlphaFold predictions with optional control over SSEs.	Rapid and adjustable, integrates partial structural information without relying on MD or force fields.	Lacks dynamic sampling; side chains are represented by pseudo-atoms.

AF2, AlphaFold2; CD, circular dichroism; IDP, intrinsically disordered protein; IDR, intrinsically disordered proteins and region; MD, molecular dynamics; NMR, nuclear magnetic resonance; Rg, radius of gyration; SAXS, small-angle X-ray scattering; smFRET, single-molecule Förster resonance energy transfer. GAN, generative adverserial network; X-EISD, extended experimental inferential structure determination; NOEs, nuclear overhauser effects; PRE, paramagnetic relaxation enhancement.

adapted for IDPs [28–30] have proven valuable for unraveling binding mechanisms with protein partners [31], clarifying sequence–ensemble–function relationships [32], assessing the impact of small-molecule drugs [33], and elucidating the molecular processes underlying liquid–liquid phase separation [34]. Although computationally demanding, recent advances in GPU-accelerated MD simulations have greatly increased the accessibility of this approach [35]. In addition, enhanced sampling techniques now allow a more efficient exploration of the complex free energy landscapes of IDPs, yielding improved convergence and more reliable ensemble descriptions [36]. Altogether, ongoing research is addressing force field inaccuracies and expanding conformational sampling through AI and ML techniques [37,38]. For further details on atomistic MD simulations applied to IDP systems, we refer readers to this review paper [39].

On the other hand, coarse-grained (CG) MD simulations provide an efficient approach to study global parameters of IDPs across larger timescales and system sizes. Recent CG models, particularly those employing data-driven force fields optimized with SAXS and NMR data [40,41], have proven highly effective in capturing realistic global ensemble features such as radius of gyration (Rg), end-to-end distance (Ree), and the apparent Flory scaling exponent ($\boldsymbol{\nu}$). However, their reduced resolution limits the accuracy of local structural features and specific residue–residue interactions. Taken together, physics-based methods remain constrained by limited accessible timescales and the inherent approximations of current force fields which make it challenging to achieve convergence in IDP simulations.

Machine learning methods

Machine learning–based methods have almost resolved the challenge of predicting stable folded protein structures, whose conformational ensembles exhibit only limited heterogeneity. Thanks to AlphaFold [2] and other state-of-the-art models, it is now possible to predict the 3D structures of folded proteins with near-experimental accuracy using only the primary sequence. The next major challenge for AI-based approaches, however, lies in predicting protein conformational dynamics, where each sequence maps to a highly heterogeneous distribution of structural ensembles. A

growing number of ML models have recently been developed to overcome the limitations of conventional experimental and computational approaches in determining IDP ensembles. Broadly, these models can be classified into three main categories:

a. **Sampling techniques** – it includes models that generate structural ensembles either de *novo* or by enhancing molecular simulations. These range from generative models that directly sample conformational space [42–44] to ML approaches that accelerate or guide MD trajectories by learning effective energy landscapes or proposing long timescale transitions [45,46].
b. **Feature predictors** – it comprises models that learn to link sequence features to ensemble-averaged properties. Rather than generating full ensembles, these models predict global or local descriptors of conformational behavior, such as chain compaction, phase separation propensity, or transient secondary structure content [47–49].
c. **Experimentally-aware models** – it includes models that integrate experimental restraints into the ML framework, either during training (e.g. incorporating NMR, SAXS, or spectroscopy data) or at the inference time by biasing the sampling process. Such approaches allow generated ensembles to be directly tuned toward agreement with experimental observables [50,51].

Detailed descriptions about these methods are available in these two reviews [52,53].

Despite these advances, ML-based approaches still struggle to fully capture IDP conformational heterogeneity as they depend heavily on training data quality and may miss physically realistic dynamics or rare yet functionally important states.

Experimental data integration methods

After generating conformational ensembles with the aforementioned methods, a common practice is to assess their agreement with experimental data (Figure 3). This process involves computing experimental observables from each conformer using *forward models* (Figure 3a) and comparing the ensemble-averaged predictions with experimental measurements (Figure 3b). In practice, however, the calculated ensembles often show poor agreement with experimental data. This mismatch typically arises from three main factors: insufficient sampling of the protein energy landscape, inaccuracies and approximations in force fields, and uncertainties or sparsity in both forward models and experimental measurements [54]. To overcome these limitations, experimental data integration techniques are employed to reweight ensembles or guide their generation, thereby ensuring that the resulting conformers more accurately reflect experimental observations (Figure 3c). Notably, two main approaches exist for integrating experimental data with ensemble generation: maximum entropy (ME) and maximum parsimony (MP). ME methods reweight ensembles to remain as close as possible to the original prior distribution while enforcing agreement with experimental observations. This strategy preserves the inherent heterogeneity of IDPs and is particularly powerful in maintaining ensemble diversity in high-entropy systems. In contrast, MP methods aim to identify the smallest subset of conformers that can explain the data, thereby simplifying the representation of conformational space. While ME emphasizes unbiased refinement, MP favors minimal models and is best suited for systems dominated by a few major conformations. Further details on these approaches can be found in a recent review [24]. Suppl. Table 1 summarizes the recent integration methods and their respective use cases.

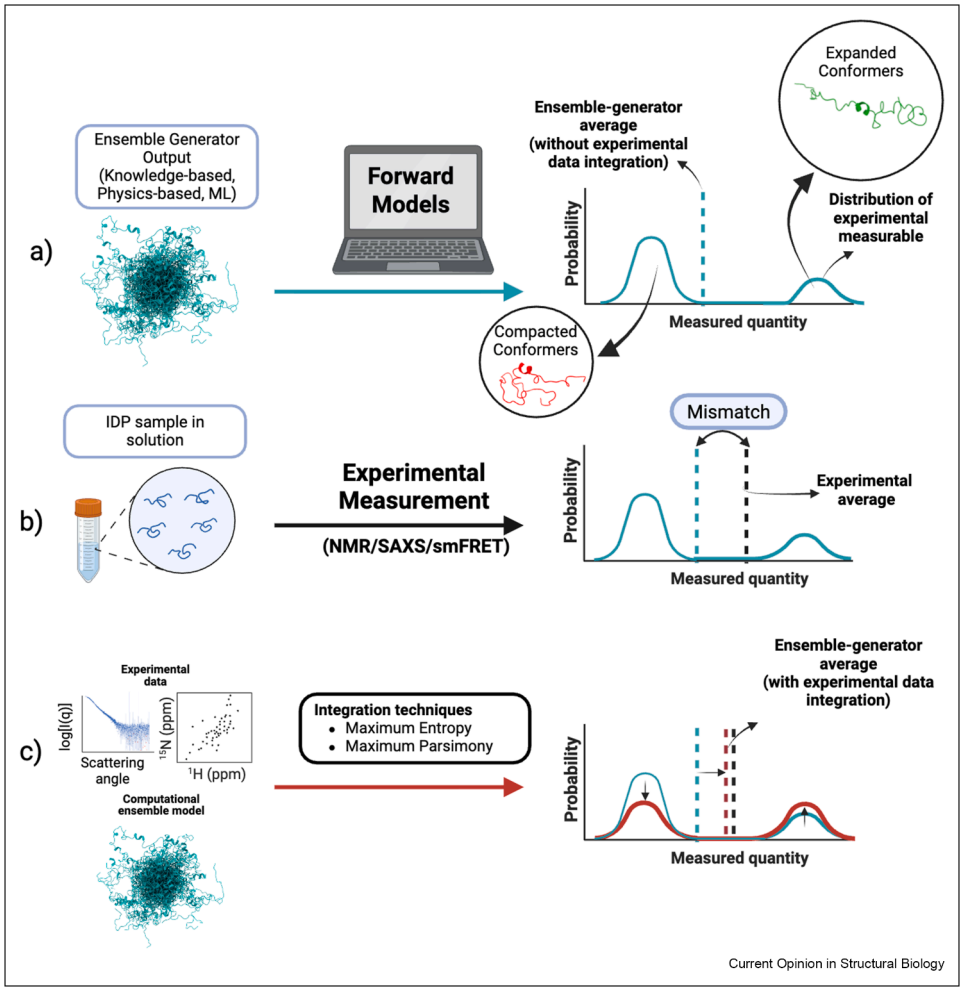
Hybrid techniques

Some computational strategies for modeling IDPs adopt hybrid approaches that combine complementary methods for ensemble generation and data integration. One direction integrates physics-based and knowledge-based sampling to balance computational efficiency with conformational diversity. A common example is fragment-based protocols refined with energy functions, which leverage both empirical knowledge and physical accuracy [55]. Another strategy merges ML with physics-based modeling, where deep learning predictions are refined within Bayesian or simulation frameworks [56,57]. A further class of approaches directly couples ensemble generation with experimental restraints, iteratively adjusting conformers to fit NMR or other measurements [50]. An additional hybrid framework combines hierarchical chain growth with MD simulations, enabling statistically rigorous and computationally efficient sampling of IDP ensembles that can be refined through integrative modeling or experimental reweighting [58]. Such flexible pipelines exploit the complementary strengths of different paradigms, enabling more realistic ensembles in cases where a single approach proves insufficient.

Comparison and validation techniques

As discussed earlier, conformational ensembles of IDPs can be derived from a variety of experimental and computational approaches. However, resolving these ensembles is inherently an ill-posed problem [3] as sparse experimental data often permit multiple solutions. This highlights the need for robust methods to compare ensembles and quantify differences between those obtained from distinct data sources. In the context of IDPs, an effective comparison metric should capture both global ensemble properties and local structural variations by reflecting differences in the underlying

Figure 3



Schematic representation of experimental data integration strategies. **(a)** Forward models calculate observables from the coordinates of each conformer in the ensemble and provide the distribution of experimental measurable in the ensemble. In this figure, we have shown Rg as an example **(b)** Ensemble-averaged value is compared to experimental measurement. In this step, we normally observe mismatches between them. The reasons are mainly insufficient sampling, uncertainties in experimental measurements and forward models, and approximations in force fields **(c)** Integration methods reweight or refine ensembles to fit the computational ensemble model to experimental data. Two main approaches for integration of experimental data are maximum entropy and maximum parsimony.

distributions rather than solely ensemble-averaged information. Such approaches are essential not only for validating ensembles derived from diverse experimental inputs but also for extracting mechanistic insights, for example, identifying binding regions or dynamic linkers, comparing homologous IDPs, and evaluating the effects of solution conditions. Next, we outline recent advances in IDP ensemble comparison techniques and introduce complementary approaches for validating conformational ensembles.

Ensemble comparison methods

A variety of approaches have been developed to compare conformational ensembles of IDPs, each capturing different aspects of their structural variability. Scalar descriptors provide a simple measure of conformational heterogeneity, allowing ensembles to be compared via single values, though at the cost of detail [59]. Distribution-based methods extend this by analyzing differences in probability distributions of structural features, such as inter-residue distances, using similarity measures like Kullback–Leibler or Jensen–Shannon divergence, Hellinger distance, or Wasserstein distance, which also account for geometric relationships but at higher computational cost [60]. Descriptor-based approaches rely on specific structural parameters: atomic coordinates with root-mean-square deviation (RMSD), torsion angles, or other local properties, sometimes combining local and global descriptors for a more integrated view [61]. Finally, alternative

frameworks employ concepts such as topology or the detection of local energy traps to highlight unique structural features [62]. Together, these strategies enable ensemble comparison across multiple levels of resolution and complexity.

Conformational ensemble validation

Ensemble validation techniques focus on testing the reliability of IDP ensembles by combining structural checks, cross-validation against experimental data, and assessments of biological relevance. Structural validation normally begins with stereochemical checks such as bond lengths, torsion angles, Ramachandran distributions, and clash avoidance, criteria not always satisfied in PED ensembles [11]. It then extends to global parameters like radius of gyration, end-to-end distance, asphericity, and scaling exponents, derived from SAXS, smFRET, and NMR measurements [63]. Local descriptors, including inter-residue contact probabilities [64], secondary structure content (CD and NMR), and residue-specific torsion preferences (NMR: CSs, RDCs, and NOEs), provide finer evaluation. Cross-validation with experimental data can validate IDP ensembles by testing robustness to data removal [65] and by verifying the ability to predict independent observables [66]. Applications to α-synuclein, tau, and viral IDPs using NMR, SAXS, and smFRET highlight this strategy [67]. Finally, functional validation examines biological relevance, such as binding interfaces [68], conserved linker dimensions [69], or disease mechanisms revealed through small-molecule binding and structure-activity relationship (SAR) analysis [70]. Emerging in-cell approaches [71] and sequence conservation analyses [72] provide ultimate physiological benchmarks.

Future perspectives

Despite recent progress, several challenges remain before conformational ensembles of IDPs can be routinely determined and applied across biology. A first priority is the establishment of robust community-wide critical assessment (e.g. IDP-bench [3]), which will allow systematic evaluation of generation methods, forward models, and integration strategies. Notably, a recent study proposed a benchmarking protocol based on a curated set of experimental data, primarily SAXS and NMR, to assess the performance of several ensemble generation approaches including machine learning and physics-based techniques [73]. Equally important is the development of data standards for reporting and exchanging experimental information, including rigorous annotation of uncertainties and error estimates. Without explicit error modeling, ensemble refinement risks overfitting and limited reproducibility. Finally, the field must expand beyond monomeric systems to include the structural heterogeneity of protein−protein, protein−RNA, and protein−DNA assemblies, which are central to cellular signaling and phase separation. Addressing these challenges will require coordinated efforts across experimentalists, biocurators and computational scientists, ultimately enabling more predictive, transferable, and biologically relevant ensemble descriptions.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work has been supported by: ELIXIR, the research infrastructure for life-science data; COST Action ML4NGP [CA21160], supported by the COST (European Cooperation in Science and Technology); European Union through NextGenerationEU; PNRR project ELIXIRxNextGenIT [IR0000010]; the National Center for Gene Therapy and Drugs based on RNA Technology [CN00000041]; co-funded by the European Union under grant agreement no. 101182949 (HORIZON-MSCA-SE project IDPfun2); views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them. Funding for open access charge: University of Padua.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.sbi.2025.103198.

References

Papers of particular interest, published within the period of review, have been highlighted as:

* of special interest
** of outstanding interest

1. Holehouse AS, Kragelund BB: **The molecular basis for cellular function of intrinsically disordered protein regions**. *Nat Rev Mol Cell Biol* 2024, **25**:187–211.

2. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, *et al.*: **Highly accurate protein structure prediction with AlphaFold**. *Nature* 2021, **596**:583–589.

3. ** Ghafouri H, Kadeřávek P, Melo AM, Aspromonte MC, Bernadó P, Cortes J, Dosztányi Z, Erdos G, Feig M, Janson G, *et al.*: *Towards a unified framework for determining conformational ensembles of disordered proteins*. 2025, https://doi.org/10.48550/arXiv.2504.03590.
This community-driven perspective outlines a unified framework for determining conformational ensembles of disordered proteins by integrating experimental data, computational modeling, and validation, highlighting key challenges such as force field accuracy, sampling efficiency, and benchmarking standards.

4. Mateos B, Konrat R, Pierattelli R, Felli IC: **NMR characterization of long-range contacts in intrinsically disordered proteins from paramagnetic relaxation enhancement in 13 C direct-**

detection experiments. *Chembiochem Eur J Chem Biol* 2019, **20**:335–339.

5. Adamski W, Salvi N, Maurin D, Magnat J, Milles S, Jensen MR, Abyzov A, Moreau CJ, Blackledge M: **A unified description of intrinsically disordered protein dynamics under physiological conditions using NMR spectroscopy**. *J Am Chem Soc* 2019, **141**:17817–17829.

6. * Chowdhury A, Nettels D, Schuler B: **Interaction dynamics of intrinsically disordered proteins from single-molecule spectroscopy**. *Annu Rev Biophys* 2023, **52**:433–462.
Chowdhury et al. (2023) reviewed how single-molecule fluorescence spectroscopy provides unique insights into the conformational dynamics and interaction mechanisms of IDPs, revealing disorder-facilitated binding modes, kinetic pathways, and links to processes such as aggregation and phase separation.

7. Nettels D, Galvanetto N, Ivanović MT, Nüesch M, Yang T, Schuler B: **Single-molecule FRET for probing nanoscale biomolecular dynamics**. *Nat Rev Phys* 2024, **6**:587–605.

8. Bernadó P, Svergun DI: **Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering**. *Mol Biosyst* 2012, **8**:151–167.

9. Trewhella J: **Recent advances in small-angle scattering and its expanding impact in structural biology**. *Struct Lond Engl* 1993 2022, **30**:15–23.

10. Mitra G: **Emerging role of mass spectrometry-based structural proteomics in elucidating intrinsic disorder in proteins**. *Proteomics* 2021, **21**, e2000011.

11. Ghafouri H, Lazar T, Del Conte A, Tenorio Ku LG, Consortium PED, Tompa P, Tosatto SCE, Monzon AM: **PED in 2024: improving the community deposition of structural ensembles for intrinsically disordered proteins**. *Nucleic Acids Res* 2024, **52**:D536–D544.

12. Lincoff J, Haghighatlari M, Krzeminski M, Teixeira JMC, Gomes G-NW, Gradinaru CC, Forman-Kay JD, Head-Gordon T: **Extended experimental inferential structure determination method in determining the structural ensembles of disordered protein states**. *Commun Chem* 2020, **3**:1–12.

13. Joshi A, Walimbe A, Avni A, Rai SK, Arora L, Sarkar S, Mukhopadhyay S: **Single-molecule FRET unmasks structural subpopulations and crucial molecular events during FUS low-complexity domain phase separation**. *Nat Commun* 2023, **14**:7331.

14. Heesink G, Marseille MJ, Fakhree MAA, Driver MD, van Leijenhorst-Groener KA, Onck PR, Blum C, Claessens MMAE: **Exploring Intra- and inter-regional interactions in the IDP α-Synuclein using smFRET and MD simulations**. *Biomacromolecules* 2023, **24**:3680–3688.

15. Jin H, Liu D, Ni Y, Wang H, Long D: **Quantitative ensemble interpretation of membrane paramagnetic relaxation enhancement (mPRE) for studying membrane-associated intrinsically disordered proteins**. *J Am Chem Soc* 2024, **146**:791–800.

16. Loening NM, Jara KA, Barbar EJ: **NMR approaches to identify transient structure and interactions of intrinsically disordered dynein intermediate chain**. *J Mol Biol* 2025, https://doi.org/10.1016/j.jmb.2025.169380.

17. * Svensson O, Bakker MJ, Skepö M: **Deeper insight of the conformational ensemble of intrinsically disordered proteins**. *J Chem Inf Model* 2024, **64**:6105–6114.
Svensson et al. (2024) combined SAXS with atomistic MD simulations to dissect the conformational ensemble of the IDP histatin-5 by grouping conformers by radius of gyration, providing detailed insight into structural heterogeneity and secondary structure propensities within the ensemble.

18. Chen SH, Weiss KL, Stanley C, Bhowmik D: **Structural characterization of an intrinsically disordered protein complex using integrated small-angle neutron scattering and computing**. *Protein Sci Publ Protein Soc* 2023, **32**, e4772.

19. Sumsion JS, Shumway SW, Blocker TM, Weed TD, Chambers TM, Poland RJ, Woodbury DJ: **Multiple structural states in an intrinsically disordered protein, SNAP-25, using circular dichroism**. *Biophys J* 2025, **124**:1828–1842.

20. Ianni AD, Ihling CH, Vranka T, Matoušek V, Sinz A, Iacobucci C: **Evaluating imide-based mass spectrometry-cleavable cross-linkers for structural proteomics studies**. *JACS Au* 2024, **4**:2936–2943.

21. ** Borthakur K, Sisk TR, Panei FP, Bonomi M, Robustelli P: **Determining accurate conformational ensembles of intrinsically disordered proteins at atomic resolution**. *Nat Commun* 2025, **16**:903.
Borthakur et al. (2024) introduced a maximum entropy reweighting procedure that integrates extensive experimental data with MD simulations to refine atomic-resolution ensembles of IDPs, achieving force field–independent conformational distributions consistent with NMR and SAXS data.

22. * Holla A, Martin EW, Dannenhoffer-Lafage T, Ruff KM, König SLB, Nüesch MF, Chowdhury A, Louis JM, Soranno A, Nettels D, *et al.*: **Identifying sequence effects on chain dimensions of disordered proteins by integrating experiments and simulations**. *JACS Au* 2024, **4**:4729–4743.
Holla et al. (2024) systematically investigated how sequence composition shapes the conformational ensembles of IDRs by combining single-molecule FRET, SAXS, CD, NMR, and simulations, revealing sequence-dependent differences in chain dimensions and advancing coarse-grained modeling frameworks.

23. Bottaro S, Lindorff-Larsen K: **Biophysical experiments and biomolecular simulations: a perfect match?** *Science* 2018, **361**:355–360.

24. * Stöckelmaier J, Oostenbrink C: **Combining simulations and experiments - a perspective on maximum entropy methods**. *Phys Chem Chem Phys PCCP* 2025, https://doi.org/10.1039/d5cp01263e.
Stöckelmaier and Oostenbrink (2025) reviewed maximum entropy methods for integrating simulations with experimental data to optimize conformational ensembles of proteins, providing a unified mathematical framework to enhance transparency and understanding of these approaches.

25. Feldman HJ, Hogue CWV: **Probabilistic sampling of protein conformations: new hope for brute force?** *Proteins: Struct, Funct, Bioinf* 2002, **46**:8–23.

26. Ozenne V, Bauer F, Salmon L, Huang J-R, Jensen MR, Segard S, Bernadó P, Charavay C, Blackledge M: **Flexible-meccano: a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables**. *Bioinforma Oxf Engl* 2012, **28**:1463–1470.

27. Teixeira JMC, Liu ZH, Namini A, Li J, Vernon RM, Krzeminski M, Shamandy AA, Zhang O, Haghighatlari M, Yu L, *et al.*: **IDPConformerGenerator: a flexible software suite for sampling the conformational space of disordered protein states**. *J Phys Chem A* 2022, **126**:5985–6003.

28. Huang J, Rauscher S, Nawrocki G, Ran T, Feig M, de Groot BL, Grubmüller H, MacKerell AD: **CHARMM36m: an improved force field for folded and intrinsically disordered proteins**. *Nat Methods* 2017, **14**:71–73.

29. Robustelli P, Piana S, Shaw DE: **Developing a molecular dynamics force field for both folded and disordered protein states**. *Proc Natl Acad Sci U S A* 2018, **115**:E4758–E4766.

30. Piana S, Robustelli P, Tan D, Chen S, Shaw DE: **Development of a force field for the simulation of single-chain proteins and protein–protein complexes**. *J Chem Theor Comput* 2020, **16**:2494–2507.

31. * Ghosh C, Nagpal S, Muñoz V: **Molecular simulations integrated with experiments for probing the interaction dynamics and binding mechanisms of intrinsically disordered proteins**. *Curr Opin Struct Biol* 2024, **84**, 102756.
Ghosh, Nagpal, and Muñoz (2024) reviewed how integrative approaches combining molecular simulations with experiments reveal the transient structures, interaction dynamics, and binding mechanisms of IDPs, highlighting implications for drug design and biomolecular engineering.

32. Hicks A, Escobar CA, Cross TA: **Zhou H-X: Seqsequence-dndent correlated segments in the intrinsically disordered region of ChiZ**. *Biomolecules* 2020, **10**:946.

33. Zhao J, Blayney A, Liu X, Gandy L, Jin W, Yan L, Ha J-H, Canning AJ, Connelly M, Yang C, *et al.*: **EGCG binds intrinsically disordered N-terminal domain of p53 and disrupts p53-MDM2 interaction**. *Nat Commun* 2021, **12**:986.

34. Zhang Y, Prasad R, Su S, Lee D, Zhou H-X: **Amino acid-dependent phase equilibrium and material properties of tetrapeptide condensates**. *Cell Rep Phys Sci* 2024, **5**, 102218.

35. Salomon-Ferrer R, Götz AW, Poole D, Le Grand S, Walker RC: **Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh ewald**. *J Chem Theor Comput* 2013, **9**:3878–3888.

36. Gong X, Zhang Y, Chen J: **Advanced sampling methods for multiscale simulation of disordered proteins and dynamic interactions**. *Biomolecules* 2021, **11**:1416.

37. Gupta A, Dey S, Hicks A, Zhou H-X: **Artificial intelligence guided conformational mining of intrinsically disordered proteins**. *Commun Biol* 2022, **5**:610.

38. Zhu J-J, Zhang N-J, Wei T, Chen H-F: **Enhancing conformational sampling for intrinsically disordered and ordered proteins by variational autoencoder**. *Int J Mol Sci* 2023, **24**: 6896.

39. Muhammedkutty FNK, MacAinsh M, Zhou H-X: **Atomistic molecular dynamics simulations of intrinsically disordered proteins**. *Curr Opin Struct Biol* 2025, **92**, 103029.

40. Tesei G, Schulze TK, Crehuet R, Lindorff-Larsen K: **Accurate model of liquid-liquid phase behavior of intrinsically disordered proteins from optimization of single-chain properties**. *Proc Natl Acad Sci U S A* 2021, **118**, e2111696118.

41. Cao F, von Bülow S, Tesei G, Lindorff-Larsen K: **A coarse-grained model for disordered and multi-domain proteins**. *Protein Sci Publ Protein Soc* 2024, **33**, e5172.

42. Janson G, Valdes-Garcia G, Heo L, Feig M: **Direct generation of protein conformational ensembles via machine learning**. *Nat Commun* 2023, **14**:774.

43. Zhu J, Li Z, Zheng Z, Zhang B, Zhong B, Bai J, Hong X, Wang T, Wei T, Yang J, *et al.*: *Precise generation of conformational ensembles for intrinsically disordered proteins via fine-tuned diffusion models*. 2024, https://doi.org/10.1101/2024.05.05.592611.

44. Zhang O, Liu ZH, Forman-Kay JD, Head-Gordon T: *Deep learning of proteins with local and global regions of disorder*. 2025, https://doi.org/10.48550/arXiv.2502.11326.

45. * Lewis S, Hempel T, Jiménez-Luna J, Gastegger M, Xie Y, Foong AYK, *et al.*: **Scalable emulation of protein equilibrium ensembles with generative deep learning**. *Science* 2025, **389**: eadv9817.

Lewis et al. (2025) presented BioEmu, a generative deep learning system that emulates protein equilibrium ensembles by integrating MD simulations, static structures, and stability data, enabling scalable prediction of conformational landscapes and thermodynamic properties with near-experimental accuracy.

46. Janson G, Jussupow A, Feig M: **Deep generative modeling of temperature-dependent structural ensembles of proteins**. *BioRxiv Prepr Serv Biol* 2025, https://doi.org/10.1101/2025.03.09.642148.

47. ** Tesei G, Trolle AI, Jonsson N, Betz J, Knudsen FE, Pesce F, Johansson KE, Lindorff-Larsen K: **Conformational ensembles of the human intrinsically disordered proteome**. *Nature* 2024, https://doi.org/10.1038/s41586-023-07004-5.

Tesei et al. (2024) developed a coarse-grained molecular modeling framework to predict conformational ensembles for nearly all human IDRs, uncovering links between sequence features, chain compaction, cellular function, and evolutionary conservation, and releasing a proteome-wide database as a community resource.

48. * Lotthammer JM, Ginell GM, Griffith D, Emenecker RJ, Holehouse AS: **Direct prediction of intrinsically disordered protein conformational properties from sequences**. *Nat Methods* 2024, https://doi.org/10.1038/s41592-023-02159-5.

Lotthammer et al. (2024) introduced ALBATROSS, a deep learning model that predicts conformational properties of IDRs—such as radius of gyration, end-to-end distance, and scaling exponent—directly from sequence, enabling proteome-wide analysis of sequence–ensemble relationships.

49. von Bülow S, Tesei G, Zaidi FK, Mittag T, Lindorff-Larsen K: **Prediction of phase-separation propensities of disordered proteins from sequence**. *Proc Natl Acad Sci U S A* 2025, **122**, e2417920122.

50. Zhang O, Haghighatlari M, Li J, Liu ZH, Namini A, Teixeira JMC, Forman-Kay JD, Head-Gordon T: **Learning to evolve structural ensembles of unfolded and disordered proteins using experimental solution data**. *J Chem Phys* 2023, **158**, 174113.

51. Fadini A, Li M, McCoy AJ, Terwilliger TC, Read RJ, Hekstra D, AlQuraishi M: **AlphaFold as a prior: experimental structure determination conditioned on a pretrained neural network**. *BioRxiv Prepr Serv Biol* 2025, https://doi.org/10.1101/2025.02.18.638828.

52. Janson G, Feig M: **Generation of protein dynamics by machine learning**. *Curr Opin Struct Biol* 2025, **93**, 103115.

53. von Bülow S, Tesei G, Lindorff-Larsen K: **Machine learning methods to study sequence-ensemble-function relationships in disordered proteins**. *Curr Opin Struct Biol* 2025, **92**, 103028.

54. Bonomi M, Vendruscolo M: **Determination of protein structural ensembles using cryo-electron microscopy**. *Curr Opin Struct Biol* 2019, **56**:37–45.

55. Ferrie JJ, Petersson EJ: **A Unified De Novo Approach for Predicting the Structures of Ordered and Disordered Proteins**. *J Phys Chem B* 2020, **124**:5538–5548.

56. * Schnapka V, Morozova T, Sen S, Bonomi M: *Atomic resolution ensembles of intrinsically disordered and multi-domain proteins with alphafold*. 2025, https://doi.org/10.1101/2025.06.18.660298.

Schnapka et al. (2025) introduced a Bayesian framework that integrates AlphaFold2 predictions with molecular mechanics force fields to generate atomic-resolution ensembles of disordered and multi-domain proteins, achieving experimental accuracy at substantially reduced computational cost.

57. ** Brotzakis ZF, Zhang S, Murtada MH, Vendruscolo M: **AlphaFold prediction of structural ensembles of disordered proteins**. *Nat Commun* 2025, **16**:1632.

Brotzakis et al. (2025) developed AlphaFold-Metainference, which integrates AlphaFold-derived restraints into MD simulations to generate structural ensembles of ordered and disordered proteins, extending deep learning approaches to capture IDP conformational heterogeneity.

58. Pietrek LM, Stelzl LS, Hummer G: **Structural ensembles of disordered proteins from hierarchical chain growth and simulation**. *Curr Opin Struct Biol* 2023, **78**, 102501.

59. Lyle N, Das RK, Pappu RV: **A quantitative measure for protein conformational heterogeneity**. *J Chem Phys* 2013, **139**, 121907.

60. Aina A, Hsueh SCC, Plotkin SS: **PROTHON: a local order parameter-based method for efficient comparison of protein ensembles**. *J Chem Inf Model* 2023, **63**:3453–3461.

61. González-Delgado J, Sagar A, Zanon C, Lindorff-Larsen K, Bernadó P, Neuvial P, Cortés J: **WASCO: a Wasserstein-based statistical tool to compare conformational ensembles of intrinsically disordered proteins**. *J Mol Biol* 2023, **435**, 168053.

62. Lotthammer JM, Holehouse AS: **Disentangling folding from energetic traps in simulations of disordered proteins**. *J Chem Inf Model* 2025, **65**:2897–2910.

63. Schuler B, Soranno A, Hofmann H, Nettels D: **Single-molecule FRET spectroscopy and the polymer physics of unfolded and intrinsically disordered proteins**. *Annu Rev Biophys* 2016, **45**:207–231.

64. Del Conte A, Camagni GF, Clementel D, Minervini G, Monzon AM, Ferrari C, Piovesan D, Tosatto SCE: **RING 4.0: faster residue interaction networks with novel interaction types across over 35,000 different chemical structures**. *Nucleic Acids Res* 2024, **52**:W306–W312.

65. Thomasen FE, Lindorff-Larsen K: **Conformational ensembles of intrinsically disordered proteins and flexible multidomain proteins**. *Biochem Soc Trans* 2022, **50**:541–554.

66. Tompa P, Varadi M: **Predicting the predictive power of IDP ensembles**. *Struct Lond Engl 1993* 2014, **22**:177–178.

67. Schwalbe M, Ozenne V, Bibow S, Jaremko M, Jaremko L, Gajda M, Jensen MR, Biernat J, Becker S, Mandelkow E, *et al.*: **Predictive atomic resolution descriptions of intrinsically disordered hTau40 and α-synuclein in solution from NMR and small angle scattering**. *Struct Lond Engl* 2014, **22**: 238–249. *1993*.

68. Piovesan D, Monzon AM, Tosatto SCE: **Intrinsic protein disorder and conditional folding in AlphaFoldDB**. *Protein Sci Publ Protein Soc* 2022, **31**, e4466.

69. González-Foutel NS, Glavina J, Borcherds WM, Safranchik M, Barrera-Vilarmau S, Sagar A, Estaña A, Barozet A, Garrone NA, Fernandez-Ballester G, *et al.*: **Conformational buffering underlies functional selection in intrinsically disordered protein regions**. *Nat Struct Mol Biol* 2022, **29**:781–790.

70. Basu S, Martínez-Cristóbal P, Frigolé-Vivas M, Pesarrodona M, Lewis M, Szulc E, Bañuelos CA, Sánchez-Zarzalejo C, Bielskutė S, Zhu J, *et al.*: **Rational optimization of a transcription factor activation domain inhibitor**. *Nat Struct Mol Biol* 2023, **30**:1958–1969.

71. Sciolino N, Burz DS, Shekhtman A: **In-Cell NMR spectroscopy of intrinsically disordered proteins**. *Proteomics* 2019, **19**, e1800055.

72. Bolognesi B, Faure AJ, Seuma M, Schmiedel JM, Tartaglia GG, Lehner B: **The mutational landscape of a prion-like domain**. *Nat Commun* 2019, **10**:4162.

73. Invernizzi M, Bottaro S, Streit JO, Trentini B, Venanzi NAE, Reidenbach D, Lee Y, Dallago C, Sirelkhatim H, Jing B, *et al.*: **Advancing protein ensemble predictions across the order–disorder**. *Continuum* 2025, https://doi.org/10.1101/2025.10.18.680935.

74. Barozet A, Molloy K, Vaisset M, Zanon C, Fauret P, Siméon T, Cortés J: **MoMA-LoopSampler: a web server to exhaustively sample protein loop conformations**. *Bioinformatics* 2022, **38**: 552–553.

75. Harmat Z, Dudola D, Gáspári Z: **DIPEND: an open-source pipeline to generate ensembles of disordered segments using neighbor-dependent backbone preferences**. *Biomolecules* 2021, **11**:1505.

76. Best RB, Zheng W, Mittal J: **Balanced protein-water interactions improve properties of disordered proteins and non-specific protein association**. *J Chem Theor Comput* 2014, **10**:5113–5124.

77. Souza PCT, Alessandri R, Barnoud J, Thallmair S, Faustino I, Grünewald F, Patmanidis I, Abdizadeh H, Bruininks BMH, Wassenaar TA, *et al.*: **Martini 3: a general purpose force field for coarse-grained molecular dynamics**. *Nat Methods* 2021, **18**:382–388.

78. Janson G, Feig M: **Transferable deep generative modeling of intrinsically disordered protein conformations**. *PLoS Comput Biol* 2024, **20**, e1012144.

79. Lewis S, Hempel T, Jiménez-Luna J, Gastegger M, Xie Y, Foong AYK, Satorras VG, Abdin O, Veeling BS, Zaporozhets I, *et al.*: **Scalable emulation of protein equilibrium ensembles with generative deep learning**. *Science* 2025, **389**. eadv9817.

80. Pajkos M, Clerc I, Zanon C, Bernadó P, Cortés J: **AFflecto: a web server to generate conformational ensembles of flexible proteins from AlphaFold models**. *J Mol Biol* 2025, https://doi.org/10.1016/j.jmb.2025.169003.