Structural ensembles of disordered proteins from hierarchical chain growth and simulation
Lisa M. Pietrek¹,a, Lukas S. Stelzl²,3,4,a and Gerhard Hummer¹,5

Abstract
Disordered proteins and nucleic acids play key roles in cellular function and disease. Here, we review recent advances in the computational exploration of the conformational dynamics of flexible biomolecules. While atomistic molecular dynamics (MD) simulation has seen a lot of improvement in recent years, large-scale computing resources and careful validation are required to simulate full-length disordered biomolecules in solution. As a computationally efficient alternative, hierarchical chain growth (HCG) combines pre-sampled chain fragments in a statistically reproducible manner into ensembles of full-length atomically detailed biomolecular structures. Experimental data can be integrated during and after chain assembly. Applications to the neurodegeneration-linked proteins α-synuclein, tau, and TDP-43, including as condensate, illustrate the use of HCG. We conclude by highlighting the emerging connections to AI-based structural modeling including AlphaFold2.

Introduction
A significant fraction of the proteome in higher organisms consists of intrinsically disordered proteins (IDPs) that do not fold into well-defined structures and of proteins with intrinsically disordered regions (IDRs) [1]. Disordered segments are also present in nucleic acids. In particular, single-stranded RNAs (ssRNAs) such as messenger RNA (mRNA) feature regions that do not form double helices or other folded structures [2,3]. IDPs and IDRs are unfolded in solution and can transiently adopt secondary structure [4]. Binding to other biomolecules can induce IDRs to fold [5], though disorder can persist also in the bound state [6]. IDPs and IDRs have distinct functions, e.g., in the nuclear pore complex [7], are a major component of biomolecular condensates [8], and are closely linked to neurodegenerative diseases [9] with their interactions (dys)regulated by mutations and post-translational modifications [10,11].

The structural heterogeneity of IDPs is best represented by a broad structural ensemble [12]. Molecular dynamics (MD) simulation are well suited to investigate the underlying structural dynamics [13]. However, for flexible proteins one faces the challenge of sampling a vast energy landscape whose many shallow minima need to be represented accurately by the potential energy function. Exploring this landscape is thus both an entropic problem not easily accelerated by enhanced sampling and an enthalpic problem because of low-energy traps. Non-local interactions in IDPs are necessarily transient, unlike in folded proteins. As a consequence, the conformation space of IDPs is inherently hierarchical in the sense that, at any scale, the local conformational preference will be minimally impacted by regions distant in sequence. Building on this principle, we recently introduced hierarchical chain growth (HCG) [14] to explore the structural heterogeneity of IDPs.

Here, we briefly highlight some recent advances in MD simulations of IDPs and then focus our review on the concepts and applications of chain growth as an extension, alternative, and complement to atomistic MD simulations. By preserving the local structure across scales where
possible, chain growth is appealing not only because of high computational speed and flexibility, but also by the possibility to produce accurate representations of the structural ensembles even of large IDPs. Chain growth can be used to create a broad ensemble of structures that can, if needed, be refined by integrative modeling using experimental data and/or MD simulations.

**MD simulations of disordered proteins**

As reviewed by Wang [13], atomistic MD simulations of disordered proteins have advanced significantly over the challenging beginnings with inadequate sampling and often overly collapsed configurations as a result of inadequate force fields. A steady increase in the computing power now makes it possible to sample the vast conformational space of IDPs with unbiased atomistic MD simulations [15]. Thanks to concomitant improvements in the quality of the force fields describing the molecular interactions, MD simulations are becoming a powerful complement to experiments on disordered proteins [16]. Atomistic MD simulations have revealed important intermediates in protein aggregation in neurodegenerative disease [17,18]. The power of MD with atomically detailed representations becomes particularly apparent as IDRs move into the focus as direct drug targets [19]. Despite these advances, the high computational cost associated with sampling the myriad of states of long IDRs warrants the development of approaches to complement MD simulations [16].

**Chain growth**

Modeling of the global structure of polymers has long been approached by chain growth algorithms. For a biomolecule with internal structure, we imagine dividing its sequence into fragments (Figure 1). For each of these fragments, we generate a pool of structures, as illustrated schematically with the four urns in Figure 1. This pool may be filled with local structures taken from databases of experimental structures or from molecular dynamics simulations of chain fragments. The task is then to assemble these fragments by a chain-growth algorithm. Naively one might consider that one simply needs to grow polymer chains sequentially (Figure 1a). However, so not to introduce a bias, one would have to stop the growth of a chain as soon as a clash is encountered and start to grow an entirely new chain instead of simply redrawing a new fragment (Figure 1b). Otherwise, the outcome will depend on arbitrary choices such as the direction of chain growth, N-to-C versus C-to-N. Rosenbluth and Rosenbluth recognized this problem of detailed balance in chain growth early in the history of computer simulations, and addressed it by a careful reweighting of self-avoiding random walks (SAWs) on a lattice [20].

In combination with importance sampling, chain growth has become a powerful tool to create large ensembles for polymers, including biopolymers [21,22]. To grow a chain, one assembles short fragments that can be sampled very efficiently at high quality. For IDPs, the flexible-meccano model by Bernadó et al. [12] is widely used, also for proteins under physiological conditions [23]. It builds on the observation that the local structure in IDPs is captured well by coil models [24–29]. In flexible-meccano, chains are grown based on the backbone-dihedral statistics in the Protein Data Bank (PDB).

**Hierarchical chain growth**

In disordered proteins, local structure is determined primarily by the local amino acid sequence, lacking the cooperative interactions of folded proteins between regions distant in sequence. HCG [14–16] exploits this hierarchical nature. A protein chain is divided into overlapping sequence fragments. Fragment structures are sampled with replica-exchange molecular dynamics (REMD) simulations. From the resulting pools, the fragments are then chosen at random. Adjacent fragments are combined with a rigid body superimposition of the heavy atoms of their overlapping regions. If the corresponding root-mean-square distance (RMSD) is below a given cut-off and if there are no steric clashes, the fragment pair is entered into the respective pool at the next assembly level. This assembly process is continued hierarchically all the way up to the level of full-length chains (Figure 1c). At each level of the assembly process, the size of the chain fragments effectively doubles. The hierarchical assembly manifestly preserves detailed balance, which guarantees that arbitrary choices such as the order of the assembly do not affect the final ensemble. Hence, HCG grows ensembles of chains with a well-defined distribution. By construction, the members of the HCG ensemble are statistically independent. As a result, HCG produces broad ensembles of IDPs with highly diverse conformations in a computationally efficient manner, sampling a significantly broader conformational space than, say, one 2 µs-long MD simulation in case of α-synuclein (aS) [14–16]. A web application of HCG is available at https://bio-phys.pages.mpcdf.de/hcg-from-library/.

If needed, HCG can be complemented by MD simulations of solvated full-length chains. As shown for aS in Figure 2, the radius of gyration $R_g$ calculated for an HCG ensemble with 20,000 chains [14–16] is already in good agreement with the measured value from SEC-SAXS [30]. For three different combinations of protein force field and water models, we found that aS tended to collapse below the size seen in the SEC-SAXS measurements [30]. These findings highlight, first, that care must be taken to assess the collapse tendency. Second, as shown in Figure 2, even for the loosely packed aS with 140 amino acids, it takes many hundreds of nanoseconds of MD just to relax the chain size, consistent with measured chain reconfiguration times in the 100-ns regime [6]. Third, without any further simulations,
HCG appears to be at least on par with the three MD simulation models. HCG thus provides an excellent starting point for further inquiry.

Applications of HCG extend beyond the sampling of IDP ensembles. For instance, HCG has shed light on the role of the disordered Atg9 termini in controlling membrane access for Atg8 lipidation during the early stages of autophagy [34]. Interestingly, some of the principles used in chain growth also find their application in other approaches to model important biological systems such as glycoproteins. For instance, Glyco-SHIELD [35] attaches glycan conformers onto proteins of interest. In another variant, Turoňová et al. [36] resampled the hip and knee joints of SARS-CoV-2 spike stalk to probe the full extent of its mobility.

Interactions between distant parts of the chain other than steric exclusion can be taken into account [22] [37*], including electrostatics, at least at the level of implicit solvent descriptions. Including electrostatic forces in HCG may be important for growing structures of highly charged biomolecules [6]. A pragmatic way forward can be to use larger chain fragments for HCG sampled in MD simulations using explicit ionic solutions.

Integration of experimental data
An ensemble representation establishes a sound foundation for the interpretation of experimental data in case of structural disorder in a molecular system [3,38–41]. As a first line of attack to improve the consistency between measured and calculated observables, one can reweight the members of the unbiased ensemble obtained from MD simulation, chain growth, or other sampling methods, rather than adjust their structure [42–45]. In a Bayesian view, the initial ensemble can be considered a sample of the prior distribution. By imposing restraints derived from experiment already in the creation of the ensemble [44,46,47], this sample can be enriched. Combinations with enhanced sampling techniques such as metadynamics [48] or replica exchange [44] further improve the sampling efficiency. Uncertainties in measurements and their modeling are readily taken care of in a Bayesian framework [44]. However, the integration of data is no panacea: for comparably poor force fields, the overlap with the “true” ensemble may not be sufficient for reweighting to establish meaningful ensembles [49]. In other words, the quality of the Bayesian prior matters, which may not surprise considering the vast conformational space to be sampled.

In chain growth, experimental data can be integrated already during the ensemble generation in a form of integrative modeling [28]. The flexible-meccano approach and its extension ASTEROIDS have been successfully used to account for different types of NMR data and single-molecule FRET and SAXS data [50*]. Biased fragment choice, with fragment weights derived from a Bayesian formulation, has been shown to be
Reweighted hierarchical chain growth (RHCG) is an extension of HCG to integrate experimental data by assigning weights to the fragment conformations [37•]. RHCG is designed to counteract the problem of systematic biases in the fragment pool. Consider, for instance, a systematic force-field error in the energetic balance between locally extended and helical peptide conformers. As the size of the molecules increases, it becomes less likely that all parts of a chain are drawn from the relevant subspace. Consequently, after global reweighting only a few chains may end up dominating the final ensemble. RHCG counteracts this tendency by using suitable fragment weights, which can be assigned, for instance, by Bayesian inference [39,44,45]. In a global reweighting of the ensemble after chain assembly, the fragment weights are fully accounted for [37•]. In this way, RHCG generates a well-defined and diverse output ensemble that has high overlap with the true ensemble.

Applied to tau K18 in solution, RHCG accounted for local and global structure as probed by NMR, single-molecule FRET, and small-angle X-ray scattering experiments [37•]. The disease-associated mutations P301L, P301T, and P301S were found to shift the balance away from the turn-like conformations associated with functional microtubule binding to the extended conformations seen in tau fibrils associated with neurodegeneration (Figure 3).

Teixeira et al. [53•] recently published a software suite that samples IDP ensembles following the principles of data-driven coil models and contains tools for further
analysis and ensemble refinement. Interestingly, their approach also captured shifts in local structure propensities in response to the neurodegeneration-linked P301L mutations in accordance with the RHCG ensemble [37].

Modeling disordered proteins in dense molecular systems
IDPs are often associated with the formation of protein condensates. An exciting perspective is to build molecularly detailed models of such crowded solutions of largely disordered biomolecules. The combination of high-resolution experiments, theory, and atomistic as well as coarse-grained modeling has already started to yield vital insights into the drivers of liquid–liquid phase separation [54,55], as recently reviewed by Fawzi et al. [56]. Coarse-grained simulation models parameterized using large sets of high-resolution experimental data can capture trends in the global arrangements of disordered proteins as well as their propensities to phase separate [57]. Another interesting direction is the simulation of dense solutions of disordered proteins or their fragments at sub-critical concentrations [58]. Such simulations can provide critical insights into molecular driving forces for condensation.

Fragment assembly can also be used to model dense systems such as condensates. Individual conformations are drawn from an ensemble of single chains grown with HCG and assembled in a simulation box as starting point for MD simulations. For the low-complexity domain (LCD) of the neurodegeneration-linked RNA-binding protein TDP-43, we generated models of condensates with atomic detail (Figure 4) using a variant of HCG and then ran MD simulations from this initial system [61]. In the simulations, phosphomimicking mutations led to a loss of protein–protein interactions and an increase in protein solvent interactions in the C-terminus of the TDP-43 LCD that destabilized the condensates, complementing coarse-grained simulations of the phase behavior of phosphomimicking mutants and phosphorylated TDP-43. Thus the simulations provided a molecular basis for the anti-aggregation effect of phosphorylation observed in experiments by Dormann and colleagues [61]. Disease-linked phosphorylation of TDP-43 may not be a driver of the progression of neurodegenerative disease and could rather be a bystander or even a cell-protective mechanism.

Outlook
On the methods side, the emerging connections of chain growth to machine learning and artificial intelligence (AI) deserve special attention. Historically, coil models have been an attempt to collect and represent statistical information about protein structure. As such, coil models and HCG have a natural connection to machine learning and AI.

AlphaFold2 [62] showcases the power of AI to predict three dimensional protein structure. The resulting acceleration in structural studies of complex assemblies [63] raises the intriguing question as to what can be learned about disordered regions from AlphaFold predictions. Currently AlphaFold2 does not capture disordered regions as a properly weighted ensemble. Hence, an exciting prospect is the combination of AlphaFold2 models of the folded protein and conformations from IDP/IDR ensembles using, e.g., HCG, molecular
dynamics or knowledge-based approaches. Interestingly, segments in IDRs often appear structured in AlphaFold2 predictions, possibly in reflection of their binding to distinct partner proteins [64], which had been used effectively to map and model the interactions of short linear motifs (SLiMs) with structured nucleoporins in the scaffold of the nuclear pore complex [63]. One potential problem is that AlphaFold2 may capture, in the same model, structures an IDR may adopt in different complexes, as has been shown for conditionally folded proteins by comparison to experimental structures [64]. Thus, a critical assessment of the thousands of local structures predicted for IDPs/IDRs is advisable even for proteins where AlphaFold2 produces high-quality models of the folded domains.

AI methods have also been used to characterize structural ensembles of IDPs [17]. Gupta and colleagues recently developed an AI based approach that learns IDP conformational space from short MD simulations to then generate broad IDP ensembles [65]. It is interesting to speculate to what extent this approach can be combined with ensembles sampled with HCG. Zhang et al. [66] are developing a neural network that learns structural ensembles of disordered proteins from experimental information. In fact, the neural network generates and learns torsion-angle probability distributions for independent neighboring residues, while also biasing the probability distribution towards experimental data, using a Bayesian formalism. Even more ambitiously, a recent preprint shows how a coarse-grained representation of an
atomistic ensemble can be learned by a neural network, which reproduces the equilibrium densities of the input ensemble [67]. HCG ensembles usually extend beyond the conformations sampled by direct MD simulations, at least for long chains, and should thus provide a valuable reference in these endeavors.

In recent years, we have witnessed a lot of progress in sampling structural ensembles of flexible (bio)polymers. However, efficient sampling of the vast conformational diversity still remains challenging. Approaches that model conformational ensembles based on local structure statistics, i.e., coil models, have been shown to be promising. The hierarchical chain growth (HCG) builds on the basic ideas of coil models. Using HCG one can sample ensembles with highly diverse conformations in a computationally efficient manner. In the cases studied, the ensemble properties agreed well with available experimental data. The quality of the generated ensemble could be further improved by integrating experimental information, producing richly detailed structural ensembles consistent with experiments across scales.

Conflict of interest statement
Nothing declared.

Data availability
No data was used for the research described in the article.

Acknowledgments
We acknowledge financial support from the German Research Foundation (CRC392: Molecular Principles of RNA Based Regulation) and the Max Planck Society. L.S.S thanks for support by ReALity (Resilience, Adaptation and Longevity), M'ODEL (Mainz Institute of Multiscale Modeling), the Forschungsgemeinschaft des Landes Rheinland-Pfalz, and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - Project number 233630050 - TRR 146.

We thank Jürgen Köfinger, Iva Pritisić, Matthieu Chavent, Dorothee Dormann, Ben Schuler and Markus Zweckstetter for many insightful discussions.

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

- of special interest
- of outstanding interest

1. Wright PE, Dyson HJ: Intrinsically unstructured proteins: Assessing the protein structure-function paradigm. J Mol Biol 1999, 293:321–331.

2. Grotz KK, Nuevech MF, Holmstrom ED, Heinz M, Stelzl LS, Schuler B, Hummer G: Dispersion correction alleviates dye stacking of single-stranded DNA and RNA in simulations of single-molecule fluorescence experiments. J Phys Chem B 2018, 122:11626–11639.

3. Bottaro S, Bussi G, Kennedy SD, Turner DH, Lindorff-Larsen K: Conformational ensembles of RNA oligonucleotides from integrating NMR and molecular simulations. Sci Adv 2018, 4: eaar8521.

4. Mukrasch MD, Bibow S, Korukottu J, Jeganathan S, Biernat J, Griesinger C, Mandelkow E, Zweckstetter M: Structural polymorphism of 441-residue tau at single residue resolution. PLoS Biol 2009, 7, e1000334.

5. Ulmer TS, Bax A, Cole NB, Nussbaum RL: Structure and dynamic of micelle-bound human a-synuclein. J Biol Chem 2005, 280:9595–9603.

6. Borgia A, Borgia MB, Bugge K, Kissling VM, Heidarsson PO, Fernandes CB, Sottini A, Soranno A, Buholzer KJ, Neeltels D, et al.: Deciphering the conformations and dynamics of FG-nucleoporins in situ. bioRxiv 2022, https://doi.org/10.1101/2022.07.07.499201.

7. Yu M, Heidari M, Mikhailova S, Tan P, Mingu S, Ruan H, Reinkemeier C, Obarska-Kosinska A, Siggel M, Beck M, et al.: Phase transitions in the assembly of multivalent signalling proteins. Nature 2012, 483:336–340.

8. Chen D, Drombosky KW, Hou Z, Sari L, Kashmir OM, Ryder BD, Perez VA, Woodward DR, Lin MM, Diamond MI, et al.: Tau local structure shields an amyloid-forming motif and controls aggregates. Nat Commun 2019, 10:2493.

9. Ambadijoudi S, Biernat J, Riedel D, Mandelkow E, Zweckstetter M: Liquid-liquid phase separation of the microtubule-binding repeats of the Alzheimer-related protein tau. Nat Commun 2017, 8:275.

10. Wegmann S, Erfekhazadeh B, Pecker T, Zolotowska KM, Bennett RE, Dujardins S, Laskowski PR, MacKenzie D, Kamath T, Commens C, et al.: Tau protein liquid-liquid phase separation can initiate tau aggregation. EMBO J 2018, 37, e98049.

11. Bernadó P, Bertocnini CW, Griesinger C, Zweckstetter M, Blockledge M: Defining long-range order and local disorder in native a-synuclein using residual dipolar couplings. J Am Chem Soc 2005, 127:17968–17969.

12. Shrestha UR, Smith JC, Petridis L: Full structural ensembles of intrinsically disordered proteins from unbiased molecular dynamics simulations. Commun. Biol. 2021.

13. Wang W: Recent advances in atomic molecular dynamics simulation of intrinsically disordered proteins. Phys Chem Chem Phys 2021, 23:777–784.

14. Pietrek LM, Stelzl LS, Hummer G: Hierarchical ensembles of intrinsically disordered proteins at atomic resolution in molecular dynamics simulations. J Chem Theor Comput 2020, 16:725–737.

15. Shrestha UR, Smith JC, Petridis L: Full structural ensembles of intrinsically disordered proteins from unbiased molecular dynamics simulations. Commun. Biol. 2021.

16. Shea JE, Best RB, Mittal J: Physics-based computational and theoretical approaches to intrinsically disordered proteins. Curr Opin Struct Biol 2021, 67:219–225.

17. Löhr T, Kohlhoff K, Heller GT, Camilloni C, Vendruscolo M: The authors show how to generate diverse, representative, and atomically detailed ensembles of disordered proteins with chain growth. The key advance is the development of hierarchical chain growth as a Monte Carlo algorithm to grow full-length chains of disordered proteins from libraries of fragment structures. The algorithm generates ensembles with a well-defined distribution.

18. Fatafta H, Khaled M, Owen MC, Sayyed-Ahmad A, Strodel B: Liquid-liquid phase separation of the microtubule-binding repeats of the Alzheimer-related protein tau. Nat Commun 2017, 8:275.

19. Pietrek LM, Stelzl LS, Hummer G: Hierarchical ensembles of intrinsically disordered proteins at atomic resolution in molecular dynamics simulations. J Chem Theor Comput 2020, 16:725–737.

20. Rosenbluth MN, Rosenbluth AW: The authors show how to generate diverse, representative, and atomically detailed ensembles of disordered proteins with chain growth. The key advance is the development of hierarchical chain growth as a Monte Carlo algorithm to grow full-length chains of disordered proteins from libraries of fragment structures. The algorithm generates ensembles with a well-defined distribution.

21. Ilja Siepmann J, Frenkel D: Configurational bias Monte Carlo: a new sampling scheme for flexible chains. Mol Phys 1992, 75:59–70.
22. Lettieri S, Mamonov AB, Zuckerman DM: Extending fragment-based free energy calculations with library Monte Carlo simulation: annealing in interaction space. J Comput Chem 2011, 32:1135–1143.

23. 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.

24. Schwalbe H, Flebig KM, Buck M, Jones JA, Grishmaw SB, Spencer A, Glaser SJ, Smith LJ, Dobson CM: Structural and dynamical properties of a denatured protein. Heteronuclear 3D NMR experiments and theoretical simulations of lysozyme in 8 M urea. Biochemistry 1997, 36:8977–8991.

25. Feldman HJ, Hogue CW: A fast method to sample real protein conformational space. Proteins Struct Funct Genet 2000, 39:112–131.

26. Dobson CM: The structural basis of protein folding and its links with human disease. Philos Trans R Soc Lond B Biol Sci 2001, 356:133–145.

27. Fitzkee NC, Rose GD: Reassessing random coil statistics in unfolded proteins. Proc Natl Acad Sci U S A 2004, 101:12497–12502.

28. Krzeminski M, Marsh JA, Neale C, Choy WY, Forman-Kay JD: Theory and Simulation/Computational Methods/C15. Filling the gap between experimental data and theory in protein folding and stability. Int Rev Biochem 2011, 3:5113–5124.

29. Lettieri S, Mamonov AB, Zuckerman DM: Characterization of disordered proteins with ENSEMBLE. Bioinformatics 2013, 29:398–399.

30. Estafania A, Sible N, Delatorre E, Vaisset M, Cortes J, Berard P: Realistic ensemble models of intrinsically disordered proteins using a structure-encoding coil database. Structure 2019, 27:381–391.

31. Araki K, Yagi N, Nakatani R, Sekiguchi H, So M, Yagi H, Ohta N, Nagai Y, Goto Y, Mochizuki H: A small-angle X-ray scattering study of alpha-synuclein from human red blood cells. Sci Rep 2016, 6.

32. Piana S, Donchev AG, Robustelli P, Shaw DE: Water dispersion interactions strongly influence simulated structural properties of disordered protein states. J Phys Chem B 2015, 119:5113–5123.

33. 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.

34. Sawa-Makarska J, Baumann V, Coudevylle N, von Bülow S, Nagellova V, Abyzov A, Moreau CJ, Blackledge M, Martens S: Reconstitution of autophagosome nucleation de novo and its structural basis. J Phys Chem B 2020, 124:6003–6012.

35. Gecht M, von Bülow S, Penet C, Hummer G, Hanus C, Sikora M: GlycoSHIELD: a versatile pipeline to assess glycan impact on protein structures. bioRxiv 2022, https://doi.org/10.1101/2021.08.04.455134.

36. Tuurová B, Sikora M, Schümann C, Hagen WJH, Welsch S, Blanc FEC, von Bülow S, Gecht M, Bagola K, Hörner C, et al.: In situ structural analysis of SARS-CoV-2 spike reveals flexibility mediated by three hinges. Science 2020, 370:203–208.

37. Stelzl LS, Pietrek LM, Holla A, Oroz J, Sikora M, Köfinger J, Schuler B, Zweckstetter M, Hummer G: Global structure of the intrinsically disordered protein tau emerges from its local structure. JACS Au 2022, 2:673–686.

38. Reichen K, Stelzl LS, Köfinger J, Hummer G: Precision DEER distances from spin-label ensemble refinement. J Phys Chem Lett 2018, 9:5748–5752.

39. Köfinger J, Stelzl LS, Reuter K, Allande C, Reichel K, Hummer G: Efficient ensemble refinement by reweighting. J Chem Theor Comput 2019, 15:3390–3401.

40. Larsen AH, Wang Y, Bottaro S, Grudinin S, Arleth L, Lindorff-Larsen K: Combining molecular dynamics simulations with small-angle X-ray and neutron scattering data to study multidomain proteins in solution. PLoS Comput Biol 2020, 16:e1007870.

41. Tesei G, Martins JM, Kunze MBA, Wang Y, Crehuet R, Lindorff-Larsen K: DEER-REpredict: software for efficient calculation of spin-labeling EPR and NMR data from conformational ensembles. PLoS Comput Biol 2021, 17:1–18.

42. Röżycki B, Kim YC, Hummer G: SAXS ensemble refinement of ESCRT-III CHMP3 conformational transitions. Structure 2011, 19:109–116.

43. Boomsma W, Ferkinghoff-Borg J, Lindorff-Larsen K: Combining molecular simulation and experimental data: a Bayesian/molecular entropy reweighting approach. In Structural bioinformatics: methods and protocols. Edited by Gaspár Z, New York, NY: Springer US, 2020:219–240. Edited by.

44. Hummer G, Köfinger J: Bayesian ensemble refinement by replica simulations and reweighting. J Chem Phys 2015, 143, 243150.

45. Bottaro S, Bengtson T, Lindorff-Larsen K: Integrating molecular simulation and experimental data: a Bayesian/molecular entropy reweighting approach. In Structural bioinformatics: methods and protocols. Edited by Gaspár Z, New York, NY: Springer US, 2020:219–240. Edited by.

46. Lindorff-Larsen K, Best RB, DePristo MA, Dobson CM, Vendruscolo M: Simultaneous determination of protein structure and dynamics. Nature 2005, 433:128–132.

47. Papaleo E, Carmillo C, Telium K, Vendruscolo M, Lindorff-Larsen K: Molecular dynamics ensemble refinement of the heterogeneous native state of NCBD using chemical shifts and NOEs. PeerJ 2018, 6:e61525.

48. Bonomi M, Carmillo C, Vendruscolo M: Metadynamics meta-inference: enhanced sampling of the meta-inference ensemble using metadynamics. Sci Rep 2016, 6, 31232.

49. Ahmed MC, Skanning LK, Jussupow A, Newcombe EA, Kragelund BB, Carmillo C, Langkilde AE, Lindorff-Larsen K: Refinement of α-synuclein ensembles against SAXS data: comparison of force fields and methods. Front Mol Biosci 2021, 8, 654333.

50. Nadi-Fabra S, Tengo M, Jensen MR, Blackledge M, Milles S: Quantitative description of intrinsically disordered proteins using single-molecule FRET, NMR, and SAXS. J Am Chem Soc 2021, 143:128–132.

51. Fisher CK, Huang A, Stultz CM: Modeling intrinsically disordered proteins with Bayesian statistics. J Am Chem Soc 2010, 132:14919–14927.

52. Boomsma W, Tian PF, Frlesen J, Ferkinghoff-Borg J, Hamelryck T, Lindorff-Larsen K, Vendruscolo M: Equilibrium simulations of proteins using molecular fragment replacement and NMR chemical shifts. Proc Natl Acad Sci U S A 2014, 111:13852–13857.

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

54. Zheng W, Dignon GL, Jovic N, Xu X, Regy RM, Fawzi NL, Kim YC, Best RB, Mittal J: Molecular details of protein condensates probed by micosecond-long atomistic simulations. J Phys Chem B 2020, 124:11671–11679.
Hierarchical chain growth provided atomically-resolved models of phase-separated TDP-43 condensates. Atomic molecular dynamics simulations of WT and 12D mutant condensates highlighted how phosphorylmimicking mutations enhance the solvation of the disordered domain of TDP-43. Hierarchical chain growth complemented coarse-grained simulations of phase behavior and led to a microscopic understanding of the experimental observations.

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

63. Mosalaganti S, Obanska-Kosinska A, Siggel M, Taniguchi R, Turoňová B, Zimeri CE, Buccak K, Schmidt FH, Margiotta E, Mackmull MT, et al.: AI-based structure prediction empowers integrative structural analysis of human nuclear pores. Science 2022, 376:1176.

64. Alderson TR, Pritisanac I, Moses AM, Forman-Kay JD: Systematic identification of conditionally folded intrinsically disordered regions by AlphaFold2. bioRxiv 2022.

The authors evaluate the presence of secondary structure in the disordered regions of AlphaFold models. They find evidence that AlphaFold2 captures conditional folding upon binding even when the binding partner is not included in the modeling, which can provide important clues for researchers but also needs to be interpreted carefully. Cautionary examples highlight that for disordered proteins AlphaFold2 may predict a mixture of the two different structures a protein adopts with two different binding partners.

65. Gupta A, Dey S, Hicks A, Zhou HX: Artificial intelligence guided conformational mining of intrinsically disordered proteins. Commun Biol 2022, 5:610.

66. Zhang O, Haghighatari M, Li J, Teixeira JMC, Namini A, Liu ZH, Forman-Kay JD, Head-Gordon T: Learning to evolve structural ensembles of unfolded and disordered proteins using experimental solution data. arXiv 2022.

A neural network is used to generate coils models. Experimental data are integrated in a form of reinforcement learning.

67. Köhler J, Chen Y, Krämer A, Clementi C, Noé F: Flow-matching efficient coarse-graining molecular dynamics without forces. arXiv 2022.

The authors develop a neural network representation to learn a coarse-grained force field from a fine-grained ensemble.