biomolecules

MDPI

Review

Advanced sampling methods for multiscale simulation of disordered proteins and dynamic interactions

Xiping Gong [1], Yumeng Zhang [1] and Jianhan Chen [1, 2, *]

[1] Department of Chemistry, University of Massachusetts Amherst, Amherst, Massachusetts, USA
[2] Department of Biochemistry and Molecular Biology, University of Massachusetts Amherst, Amherst, Massachusetts, USA
* Correspondence: jianhanc@umass.edu

Received: date; Accepted: date; Published: date

Abstract: Intrinsically disordered proteins (IDPs) are highly prevalent and play important roles in biology and human diseases. It is now also recognized that many IDPs remain dynamic even in specific complexes and functional assemblies. Computer simulations are essential for deriving a molecular description of the disordered protein ensembles and dynamic interactions for mechanistic understanding of IDPs in biology, diseases, and therapeutics. Here, we provide an in-depth review of recent advances in multi-scale simulation of disordered protein states, with a particular emphasis on the development and application of advanced sampling techniques for studying IDPs. These techniques are critical for adequate sampling of the broad manifold of functionally relevant conformational space of IDPs. Together with dramatically improved protein force fields, these advanced simulation approaches have achieved substantial success and demonstrate significant promise towards quantitative and predictive modeling of IDPs and their dynamic interactions. We will also discuss important remaining challenges in atomistic simulation of larger systems and how various coarse-grained approaches may help to bridge the remaining gaps in the accessible time- and length-scales of IDP simulations.

Keywords: conformational ensemble; enhanced sampling; generalized Born; Gō-model; implicit solvent; liquid-liquid phase transition; replica exchange; protein force fields.

1. Introduction

Intrinsically disordered proteins (IDPs) or regions (IDRs), compared to well-structured proteins, do not have stable tertiary structures under physiological conditions. Nevertheless, IDPs or IDRs can be found in nearly a third of proteins encoded in the human proteome [1], and they play key roles in a variety of biological processes that underlie vital cellular functions ranging from signaling, regulation to transport [2,3]. The inherent thermodynamic instability of an IDP's conformation allows it to respond sensitively to numerous stimuli, including binding, changes in cellular environments (e.g., pH), and post-translational modifications [4-8]. Such conformational plasticity arguably enables IDPs to interact with multiple signaling pathways and serve as scaffolds to form multi-protein complexes [9]. Importantly, IDPs and IDRs house around 25% of disease-associated missense mutations [10]. They have been considered promising therapeutic targets for treating various diseases (such as chronic diseases) [11-13]. While many IDPs have been shown to undergo binding-induced folding transitions upon specific binding [3], many examples are also emerging to demonstrate that IDPs can remain unstructured even in specific complexes and functional assemblies [14-20]. Such a dynamic mode of specific protein interactions seems much more prevalent than previously thought [21-23].

Reliable description of the conformational ensembles of IDPs and IDRs have proven very challenging. A disordered state does not lend itself to traditional structural determination methods that are geared toward describing a coherent set of similar structures. Biophysical techniques such as NMR, SAXS and FRET can provide complementary information on various local and long-range structural organizations [7]. However, these ensemble-averaged measurements alone are not sufficient to unambiguously define the heterogeneous ensemble, due to the severely underdetermined nature of the structure calculation problem [8,24,25]. As a result, studies of IDPs have relied heavily in the traditional structure-function paradigm, by solving the folded structure of the bound state, analyzing coupled binding and folding mechanisms, or identifying putative pre-existing functional structures in the unbound state [3]. However, the disordered ensemble itself is arguably the central conduit of cellular signaling. The functional mechanism of an IDP is

encoded in how the disordered ensemble as a whole responds to various stimuli, may it be cooperative binding-induced folding or redistribution of conformational sub-states in dynamic interactions. Multiple cellular signals can be naturally integrated through cooperative responses of the whole dynamic ensemble [26-28]. Therefore, there is a critical need for reliable characterization of disordered protein conformation ensembles, in both bound and unbound states, in order to establish the molecular basis of IDPs and IDRs in various physiological and pathophysiological processes.

Given the fundamental challenges of characterizing disordered protein states based on ensemble-averaged measurements alone, molecular modeling and simulations have a crucial and unique role to play in mechanistic studies of IDPs and IDRs [29-33]. This is reflected in continuously increasing numbers of research articles that contain keywords "intrinsically disordered" and "molecular dynamics" published in the last 10 years (Figure 1). A particularly attractive approach is to first generate the disordered ensemble using transferable, physics-based force fields without any experimental restraints and then use the later for independent validation [7]. Such *de novo* simulations of disordered protein ensembles require both high force field accuracy and adequate sampling of relevant conformational space, pushing the limit of these two central ingredients of molecular dynamics (MD) and Monte Carlo (MC) simulations. The challenges of simulating disordered proteins have driven significant interest in developing better protein force fields and advanced sampling methods (Figure 1). In particular, important advances have been made in the state-of-the-art atomistic force fields for describing the conformational equilibria of ordered and disordered proteins [13]. Enhanced sampling techniques have played crucial roles in both the development and application of atomistic force fields, by allowing one to cross energy barriers faster and accelerate the conformational sampling of IDPs [34-41]. Nonetheless, atomistic simulations still have limited capability for describing large systems such as biological condensates[42]. For this, the multi-scale approaches are necessary to bridge the gaps in experimental and computational time- and length-scales, including implicit solvent models, which removes the solvent degrees of freedom [8], and various coarse-grained models, which significantly reduce both proteins and solvent degrees of freedom [43].

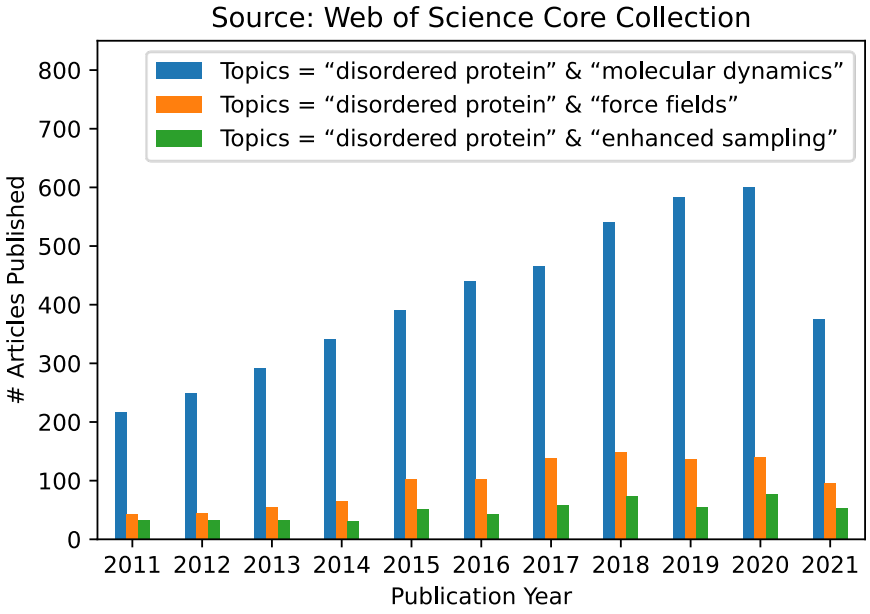


Figure 1. Number of articles identified with three different search keywords published from 2011 to 2021 based on a Web of Science core collection source (as of August 15, 2021).

In this review, we will start by highlighting the challenges of sampling IDP conformational ensembles and provide a summary on the state-of-the-art force fields available to describe the IDP conformations. It is noted that several excellent review papers have published recently that cover general theoretical and computational approaches for studying IDPs, in particular regarding protein-protein interactions and biological condensates [44-47]. This review will therefore focus on recent development of advanced sampling methods for simulating disordered conformational ensembles and dynamic interactions of IDPs. We will also discuss some of the key advances in multi-scale modeling of IDPs that greatly extend the accessible length- and time-scales of molecular simulations. Finally, we discuss future directions in developing a robust computational framework for simulating IDP conformational equilibria and interactions.

2. Challenges of simulating IDP conformational equilibria

Compared to the globular proteins that have one or a few well-defined global energy minima, the energy landscape of an IDP is more flat and generally includes many local energy minima separated by modest energy barriers [48]. IDPs and IDRs typically have fewer hydrophobic residues, but a larger number of polar or charged as well as disorder-promoting residues (such as glycine and proline) [44]. These sequence features hamper the formation of hydrophobic cores that drive protein folding and thus prevent the formation of stable tertiary structures. Instead, IDPs and IDRs favor forming an ensemble of unfolded or partially folded states. This presents a major challenge for simulation and depends critically on the ability of the force fields to accurately describe the energetics of relevant conformational states, especially for capturing both folded and unfolded states of an IDP. For example, one recent study tested atomistic simulations of IDPs for eight force fields and found marked differences in the describing the conformational ensembles of IDPs, in particular the secondary structure content [49]. Similar observations have also been made in other benchmark studies, consistently showing that protein force fields previously optimized for folded proteins are not suitable for simulating disordered protein states, largely due to over-stabilization of protein-protein interactions [50]. These benchmark studies also suggested that the key towards better protein force field was to rebalance protein-protein, protein-water, and water-water interactions.

Besides accurate force fields, reliable simulation of IDPs also hinges on sufficient sampling of many relevant conformation states within a reasonable simulation time. Standard MD simulations are generally insufficient to generate representative conformational ensembles, even using the most accurate protein force fields coupled with advance of GPU computing or specialized hardware such as ANTON supercomputer [51]. For example, a recent reanalysis of 30-μs ANTON trajectory of 40-residue Aβ40 peptide in explicit solvent revealed very limited convergence even at the secondary structure level [13]. This can be attributed to the diverse and large accessible conformational space of an IDP and the potentially high free energy barriers separating various sub-states that require exponentially longer time to cross. Note that typical simulation times on conventional hardware (such as GPUs) are at least one-order of magnitude shorter. There is thus great danger in relying on standard MD to calculate disordered protein conformational ensembles at the atomistic level. There is critical need to develop and leverage so-called enhanced sampling techniques, which aim to generate statistically meaningful conformational ensembles with dramatically less computation.

Computational studies of IDP interaction and assembly are even more demanding. The conformational equilibrium of an IDP can respond sensitively to specific and nonspecific binding, potentially shifting from a disordered to somewhat ordered state or fully folded state. In principle, simulations could provide the much-needed spatial and time resolutions to elucidate the kinetics and thermodynamics of coupled folding and binding processes and characterize the mechanistic features. However, the challenge is that this coupled process of folding and binding is a complex reaction involving the formation of many noncovalent interactions, which requires extremely long simulations generally beyond the current capabilities at the atomistic level. As such, coarse-grained models are generally required for computational studies of IDP interaction and assembly.

3. The state-of-the-art protein force fields for describing IDP conformations

Empirical protein force fields are potential energy functions that typically include physics-motivated bonded and non-bonded terms carefully parameterized based on a wide range of theoretical and experimental data [52]. These force fields can in principle be transferable between folded proteins and IDPs. To achieve this, it is also critical to develop suitable water models and better describe the water-protein interactions [53,54]. Two recent review articles have already provided comprehensive descriptions on the latest development of better protein force fields [52,55]. We therefore briefly summarize the status of the state-of-the-art nonpolarizable and polarizable force fields for IDP dynamics and interactions.

3.1. Nonpolarizable protein force fields

Many previous nonpolarizable force fields have significant shortcomings for describing the unfolded or disordered proteins. For example, they typically provide a poor description of the secondary structure content for IDPs and have a preference to give too compact conformations with respect to the experimentally measured dimension of IDPs [49,56]. These problems were likely attributed to the unbalanced parameterization of dihedral torsion space and description of protein-protein and protein-water interactions [57]. As a result, most of improved force fields managed to give more accurate secondary structure propensities by adjusting dihedral parameters or adding grid-based energy correction map (CMAP) parameters [55]. The over-compactness of disordered proteins can be alleviated by modifying protein-water van der Waals interactions or combining with refined water models [53]. Representative state-of-the-art

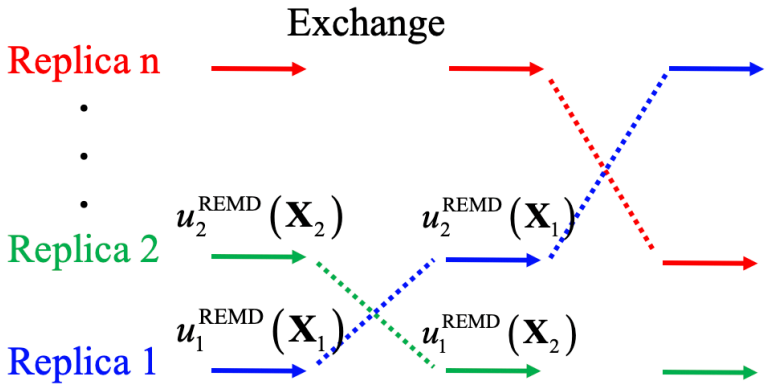
force fields includes the latest CHARMM36m/TIP3P* [58], ff19SB/OPC [59]and a99SB-disp/TIP4P-D [51]. Many benchmark studies have consistently demonstrated that these refined force fields do provide significant improvements in describing not only single folded and disordered proteins, but also the multiprotein systems that are either soluble or aggregate in the solution [56,60-63]. At the same time, these studies also identified significant remaining limitations in description of the noncovalent interactions in the multiprotein systems [61]. Recognizing limitations in the ability of a99SB-disp/TIP4P-D force field to accurately describe the protein-protein interactions, a new force field, DES-Amber, was recently developed to provide more accurate simulations of protein-protein complexes while maintaining reliable descriptions of both ordered and disordered single-chain proteins [62]. However, DES-Amber is still limited in reproducing the experimental protein-protein association free energies of some protein complexes, in particular for the systems with highly polar interfaces [62]. In the latter case, it was found that the charged sidechains were buried at the protein-protein interface instead of being solvent-exposed. It was further suggested that nonpolarizable force fields were fundamentally limited in achieving a balanced description of charged groups that were solvent-exposed or buried at a protein-protein interface.

3.2. Polarizable protein force fields

Polarizable force fields explicitly consider the electronic polarization using various empirical models to provide better description of charged and polar protein motifs in heterogeneous biomolecular environments [64]. Exciting progresses have been made in the last few years and several polarizable force fields are now available for stable simulation of proteins in both aqueous and membrane environments [65,66]. Simulations using the latest polarizable force fields have also showed a high level of consistency with experimental observations, particularly the ion solvation and binding thermodynamics, permeation free energy of ions or small charged molecules into the cell membrane, and protein-ligand binding [64]. For example, the Drude-2013 polarizable force field, compared to CHARMM36 force field, is more accurate to describe folding cooperativity of $(AAQAA)_3$ peptide, which can be attributed to enhanced backbone dipole moments in the helix state [67]. Additional studies are still needed to show the necessity of considering polarizable force fields in IDP simulations, where the significantly higher computational cost adds to the challenge of generating converged ensembles [64]. Existing comparisons suggest that polarizable force fields, including AMOEBA and Drude models, still frequently have problems in reproducing the nature structures and folding of proteins [68-70]. For example, stronger protein-water interactions in polarizable force fields can destabilize the native protein structure, opposite to the observations from nonpolarizable force fields where protein-water interactions have traditionally been underestimated [42]. Nonetheless, it can be anticipated that polarizable force fields will continue to improve and become increasingly important for simulating IDP structure and interactions.

4. Enhanced sampling methods for sampling IDP conformational ensembles

Enhanced sampling techniques generally accelerate the crossing of energy barriers to achieve better sampling efficiency, such as by introducing bias potentials, modifying the potential energy itself, and changing the effective temperature. These techniques have proven essential in atomistic simulations of IDPs [71,72], yielding levels of convergence that could not be achieved even with drastically longer standard constant-temperature MD simulations [13]. The central idea of biased MD simulations is similar to importance sampling in MC simulations, where a biased potential is introduced to construct a flat free energy landscape along single or multiple collective variables of interest, such that many states can be readily sampled due to the removal of free energy barriers. The replica-exchange (REX) class of sampling methods, in particular, replica exchange molecular dynamics (REMD), has been one of the most popular methods for simulating protein conformations. Figure 2 shows the general scheme of REMD simulations, where the key point is to first set up multiple replicas with different unitless unbiased or biased potentials, given as the energy over k_BT (T is the temperature), and then use the Metropolis rule to allow MC to exchange the replicas and maintain the detailed balance. A key advantage of using multiple replicas and maintaining detailed balance is avoiding the reweighting problem generally required for biased simulations. Note that virtually all biased sampling strategies can be readily incorporated within the REX framework to benefit from both classes of enhanced sampling, including metadynamics (MTD) [73,74], accelerated MD (aMD) [75], umbrella sampling (US) [76,77], integrated tempering sampling [78]. In practice, effective REMD protocols require proper choices of 1) the optimal number of replicas and proper distributions of conditions, to ensure a uniform exchange acceptance rate and efficient random walk in the condition space, and 2) the choice of those unitless (biased) potentials for effective conformational diffusion at each condition [79]. Here, we divide various enhanced sampling strategies into two general groups depending on the need for collective variables and summarize their recent applications to IDP conformational sampling.



$$u_m^{\mathrm{REMD}}(\mathbf{X}) = \beta_m E_m(\mathbf{X}), \quad m = 1, 2, \cdots, n$$

$$P_{\mathrm{acc}}(1 \to 2) = \min\left[1, e^{-\Delta_{12}}\right],$$

$$\Delta_{12} = \left[u_1^{\mathrm{REMD}}(\mathbf{X}_2) + u_2^{\mathrm{REMD}}(\mathbf{X}_1)\right] - \left[u_1^{\mathrm{REMD}}(\mathbf{X}_1) + u_2^{\mathrm{REMD}}(\mathbf{X}_2)\right].$$

Figure 2. The generalized replica exchange molecular dynamics protocol based on unitless potentials, where the initial condition of each replica could have a varied temperature or scaled potential. β_m is the inverse of temperature, $E_m(X)$ is the potential energy of m^{th} condition for given a configuration X.

4.1. Collective variables-based sampling methods and optimization

MTD and its variants have been considered one of the most important collective variables (CV)-based sampling methods for protein simulations [80]. MTD uses a history-dependent bias potential, which is generally a sum of Gaussians, to eventually construct a flat free energy landscape along the predetermined CV(s). Furthermore, the parallel tempering MTD (PT-WTD) and the combinations with other biased sampling methods have been also developed to increase the sampling efficiency and convergence of free energy calculations [81,82]. For example, the PT-WTD and bias-exchange MTD has been employed to obtain the conformational ensembles and coupled binding and folding of disordered pKID and KID proteins, using the α-score of helical structures as CVs [83]. It has also shown that the REMD-based MTD, compared to conventional MTD or T-REMD, can enhance the conformational sampling of N-Glycans using dihedral angles as CVs to characterize the global motions [84]. The binding mechanism of two disordered peptides, NRF2 and PTMA, were simulated by the well-tempered MTD (WT-MTD), and the results showed that the WT-MTD method could provide converged free energy profiles with 1.5 μs sampling time [85]. Together, these applications have shown that MTD-class of sampling methods can be effectively applied to IDP simulations. Beside MTD, another important class of CV-based sampling strategy is the US method [77]. US is not strictly an enhanced sampling method like MTD. It typically uses multiple harmonic potentials to focus sampling various states along the collective variables of interest. US is often combined with REMD in studies of IDPs, as illustrated in a recent 2D window-exchange US simulation of the coupled folding and binding mechanism of HdeA homodimer [86]. The simulation was able to capture rare unfolding transitions of the dimer at neutral pH and provided detailed description of the transition pathways.

A central limitation of CV-based sampling methods is the efficiency strongly depends on the quality of selected CV(s). For diffusion processes such as protein conformational fluctuation, it is often not clear that which CV(s) can best capture large-scale transitions or even if these transitions could be effectively described using one or a few CV(s) [87-89]. Another practical limitation is that the computational cost of MTD and US grows exponentially as a function of the number of CVs, generally limiting the maximum to 3. Parallel bias metadynamics (PBMetD) approaches have been proposed to overcome this limitation, by applying multiple low-dimension bias potentials in parallel [90,91]. Nonetheless, the efficacy of PBMetD for sampling complex (disordered) protein conformational space is yet to be demonstrated. Another recent work presented a temperature accelerated sliced sampling method to explore the high dimensional free energy landscape by combining TAMD/d-AFED, MTD and US methods to sample many CVs simultaneously [92]. However, the approach shares the limitation of PBMetD where the underlying bias potentials remain low dimensional in nature. To address the problem of determining the best CVs for a particular problem of interest, machine learning algorithms and deep learning network have been recently proposed analyze information from many candidate CVs, and construct the free energy landscape using low-dimensional representations [93,94]. On-the-fly discovery of optimal CV was also demonstrated using the artificial neural networks that has a strong capacity of learning and optimization

for given linear or nonlinear CVs [95]. In another recent study, an 8-dimensional optimal biased potential was constructed and applied to the free energy calculations of polypeptides using two machine learning algorithms, namely, nearest neighbor density estimator and artificial neural network [96]. Similar deep neural networks have also shown to be capable of constructing nontrivial biased potentials, for deep enhanced sampling of protein conformational space and overcoming so-called hidden barriers [97,98]. These are exciting developments that may greatly expand the applicability of MTD, US and other CV-based sampling techniques to problems of increasing complexity, including simulations of IDPs and their dynamic interactions, especially when combined with REX.

4.2. Collective variables-free sampling methods and optimziation

CV-free sampling avoids the need to identify a set of optimal CVs and can be highly desirable for simulating high-dimensional conformational fluctuation of IDPs. Many CV-free sampling methods have been also developed, including the tempering-based and energy-scaled biased methods. Tempering-based sampling methods rely on increasing the effective simulation temperature (i.e., tempering) to accelerate barrier crossing. Examples include the temperature cool walking [99], annealed importance sampling [100], simulated tempering [101], and temperature-based REMD [36]. T-REMD, in particular, has proven highly effective for protein folding and studies of IDP conformation ensembles, where multiple replicas are simulated at different temperatures in parallel to promote barrier crossing as the system undergo random walk in the temperature space (Figure 2). Nevertheless, one potential limitation is that the number of replicas required for T-REMD scales as the squared root of the number of degree of freedoms (DOFs) of whole system to maintain a reasonable exchange acceptance probability. This can dramatically increase the computational cost of the explicit solvent T-REMD simulations. Several methods have been proposed to overcome this limitation of T-REMD, such as adding energy-related terms (such as accelerated-MD or Gaussian accelerated MD, named GaMD) or scaling the potential energy function (including the scaled MD that scaled all energy terms and replica exchange solute tempering (REST) methods that scaled part of energy terms) [84,102-104].

aMD adds boost potentials to reduce the energy barriers and accelerate sampling [75]. However, it suffers from a serious energetic noise when reweighting [105]. The GaMD has been thus developed to reduce noise by introducing a new harmonic boost potential, to allow a new reweighting technique that could accurately recover the free energy landscape using a cumulant expansion to the second order [106]. GaMD has achieved some success in studying protein folding, protein-ligand binding, and protein-protein interactions [105]. In particular, specifically developed Ligand GaMD [107] and Peptide GaMD [108] can capture the binding and dissociation of molecular ligands and highly disordered peptides within microsecond simulations. Recently, this GaMD method has also been combined with the REMD protocol, which can avoid the energy reweighting problem [103]. A combination of replica-exchange umbrella sampling (REUS) and GaMD has also been designed for the conformational sampling and free energy calculations [104]. It is noted that the CVs-free enhanced sampling methods are more generally more suitable for simulating IDP conformations and dynamics, because the specific CVs are often unclear.

REST is a special variant of T-REMD designed specifically to reduce the number of DOFs that contribute to the Metropolis criteria of replica exchange, such that smaller number of replicas are needed [37,109]. The basic idea of REST is to separate the system into two 'hot solute' and 'cold solvent' regions. The 'solvent' could be actual water molecules but could also be any region of the system where no tempering is to be applied. This offers great flexibility in tailoring REST for a specific system of interest. Even more generally, the 'solute' region can be defined to include only a subset of interaction terms within the 'solute' region, such as dihedral-angle energy or Lennard-Jones energy term in the generalized REST (gREST) method [110]. Temperature-dependent factors are used to scale the 'solute'-'solute' and 'solute'-'solvent' interactions, while keeping the 'solvent'-'solvent' interactions intact:

$$
\begin{aligned}
&u_m^{\mathrm{REST}}(\mathbf{X}) = \beta_0\lambda_m^{\mathrm{pp}}E_{\mathrm{pp}}(\mathbf{X}) + \beta_0\lambda_m^{\mathrm{pw}}E_{\mathrm{pw}}(\mathbf{X}) + \beta_0 E_{\mathrm{ww}}(\mathbf{X}),\\
&\text{REST1:}\quad \lambda_m^{\mathrm{pp}} = \frac{\beta_m}{\beta_0},\quad \lambda_m^{\mathrm{pw}} = \frac{\beta_0+\beta_m}{2\beta_0\beta_m},\\
&\text{REST2:}\quad \lambda_m^{\mathrm{pp}} = \frac{\beta_m}{\beta_0},\quad \lambda_m^{\mathrm{pw}} = \sqrt{\frac{\beta_m}{\beta_0}},
\end{aligned}
\tag{1}
$$

where X is the conformational coordinates and β_m is the inverse of $k_B T_m$. The scaling of 'solute'-'solute' interactions allows the 'solute' to be simulated with an effective temperature of T_m while maintaining the 'solvent' temperature at T_0. As a result, the exchange acceptance probability will be independent of 'solvent'-'solvent' interactions, which

reduces the effective system size and requires fewer replicas to cover the same temperature range. A key open choice in REST is how the 'solute'-'solvent' term is scaled (Equation 1). Different solute-solute and solute-solvent scaling factors can strongly affect the ability of driving conformational transitions of the selected 'solute' region. A strong solute-solute interaction favors to compact the protein conformations, whereas a strong solute-solvent interaction prefers the disordered, solvent-exposed conformations. Different scaling schemes lead to very different characteristics of REST1 (original) and REST2 (revised) protocols (Equation 1). High temperature conditions favor the unfolded conformations in REST1, while both folded and unfolded conformations were observed in REST2 model for the condition with the same effective 'solute' temperature. The reason is that REST2 was designed to have a weaker solute-solvent interactions to promote the sampling of folded conformations even at high temperatures [109]. While this could allow the sampling of reversible folding transitions at all temperatures in REST2, it could lead to conformational trapping and hampering the sampling of disordered conformations of IDPs. One important implication is that the performance of REST can be sensitive the balance of protein-protein and protein-water interactions of a given protein force field. For example, Liu et al. showed that, while REST2 was highly effective in generating converged ensembles of 61-residue p53 N-terminal transactivation domain (TAD) using a99sb-disp, it completely failed to converge even with ~1 μs/replica in CHARMM36m and CHARMM36mw force fields [111]. Separate standard MD simulations reveal that p53-TAD can readily escape the apparent trapped conformations observed during REST2, suggesting that these traps arise due to the imbalance of scaled protein-protein, protein-water and water-water interactions [111].

REST has proven to be one of the most reliable choices for enhanced sampling of protein folding and particularly disordered conformational ensembles [112,113]. Sugita and co-workers leveraged gREST to target the dihedral-angle energy term and successfully sampled folding transitions of beta-hairpins and Trp-cage in explicit water, using fewer replicas but cover wider conformational space compared to REST2 [110]. Walsh et al. applied REST to investigate n16N disordered peptide conformational ensembles [114]. The conformations obtained via REST methods showed a high consistency with NMR experimental data. Furthermore, REST are specifically appropriate in simulating IDRs as the disordered region can be targeted in REST without tempering the well-structured region (or water). Zhou and co-workers studied the disordered loop of Staphylococcus aureus sortase A (SrtA) to order transition upon binding to calcium [115]. Chen and Liu characterized Bcl-xL interfacial conformational dynamics in explicit solvent [116]. Both works directly showed that REST covered broader conformational spaces for intrinsically disordered regions and led to faster convergence compared to either standard MD or T-REMD simulations. REST simulations have also been successfully integrated with experiment to study how cancer-associated mutations and drug molecules may modulate the disordered ensembles of p53-TAD and Aβ peptides in recent years [117-120].

Despite the success of REST for CV-free enhanced sampling, it does not benefit from targeted acceleration along specific CVs that are known to be rate limiting. For this, REST (or REX in general) has been combined with CV-based enhanced sampling to maximize the efficiency of sampling complex, high dimensional conformational space of proteins. Some of the examples are discussed in the sections above. Here we note a couple additional recent examples. By integrating free energy perturbation (FEP) and REST methods, Abel et al. obtained more thorough samplings of different ligand conformations around the active site and realized relative binding affinity predictions [121]. Okamoto and co-workers have applied REUS/REST two-dimensional replica-exchange method to predict two protein-ligand complex systems with the help of REST to weaken the solute-solvent interactions but improve the binding events and REUS to enhance the sampling along with the reaction coordinates [122].

Multiscale enhanced sampling (MSES) is yet another fascinating example of CV-free enhanced sampling strategy. Protein folding and other cooperative transitions such as self-assembly are known to be dominated by entropy barriers, which renders tempering ineffective for driving faster transitions. Coupled with a lack of obvious CVs, sampling complex conformational transitions of IDPs and their interactions is challenging for both CV-based and REX-based CV-free methods. For this, an effective solution is to couple atomistic simulations with a coarse-grain (CG) model, such that one could benefit from both faster transitions of CG modeling and accuracy of atomistic force field [123]. A particularly attractive approach was first introduced by Kidera and coworkers, where restraint potentials were used to couple CG and atomistic conformational dynamics along "essential" DOFs shared by the two models [35]. The bias introduced by the coupling potential is removed using Hamiltonian REX (H-REX). Chen and coworkers further adapt the method to utilize topology-based CG models (see below), better coupling potential and advanced Hamiltonian/temperature REX (H/T REX) [124-126]. Coupling the CG and atomistic models using restraints is a key strength of these MSES protocols. It allows full control of the energetic impact of diverged structures at different resolutions, which improves exchange efficiency and provides superior scalability to large systems. MSES coupling also provides robust tolerance of CG defects by preventing the CG model from dictating the conformational dynamics. The efficacy of MSES has been illustrated using several systems. It was highly effective in simulate reversible transitions of small β-hairpins and helical IDPs [124-

126] and proved instrumental in further refinement of a GBMV2 implicit solvent protein force field for both ordered and disordered peptides [127]. Very recently, MSES was also observed to be effective in sampling the cis–trans transitions of lutein by coupling the atomistic model with the Martini CG model [128]. Nonetheless, the application of MSES to larger and more complex proteins has proven more challenging than originally expected, apparently due to difficulty in effective coupling of CG and atomistic conformational fluctuations of a larger protein.

Other tempering methods including integrated tempering and simulated tempering have also been combined with different biased potentials to enhance sampling [129,130]. For example, an integrated accelerated MD method has recently used to sample the conformations of pepX peptides, and it was shown that this method can improve the sampling efficiency and provide a good strategy for simulating IDPs [70,129]. The combination with the metadynamics has also been presented to sample the conformational space of silica, and the acceleration was increased by over one order of magnitude [130]. One significant benefit is that only a single replica is required and could be suitable for Anton specialized hardware [51]. However, one drawback is that we have to estimate the relative free energies of all conditions (or equivalently the density of states), which require recursive simulations and can be difficult to converge for complex systems such as large IDPs and complexes.

4.3. Reweighting techniques for generating unbiased ensembles

When bias potentials are used to enhance sampling, reweighting is often required to obtain the unbiased samples and construct statistically optimal unbiased free energy surfaces. Two reweighting methods are widely used for this, including the weighted histogram analysis method (WHAM) for the biased simulations with specific CVs and a more general multistate Bennett acceptance ratio (MBAR) approach [131,132]. Stability of both WHAM and MBAR can be susceptible to large energetic fluctuations due to exponential dependence of weights on the value of the unitless potentials. Large energy fluctuations among sampled conformations can lead to large uncertainties during reweighting and thus final unbiased distributions. Another population based reweighting method has been used for unbiasing the scaled MD simulations by making a multidimensional histogram of all sampled configurations [133]. However, the dimensionality of configurational space is usually very huge and thus be hardly completely described by some dimensionality reduction techniques (such as the principal component analysis). Recently, it was proposed that this energetic noise can be alleviated by truncating the cumulant expansion of the exponential average [106], which has been originally used in the accelerated molecular dynamics. It has shown that it can accurately recover the free energy profiles within an acceptable error (~k_BT), especially for the near-Gaussian biased unitless potentials [106]. This approximated reweighting methods have therefore been successfully used for reweighting several biased simulations [104]. It should be mentioned that those reweighting techniques can be used for reweighting any biased simulation, even for the REMD simulations. Nonetheless, all reweighting methods including MBAR relies on good overlap between the true conformational space and the region sampled by biased simulations. When the overlap is limited, the reweighted distributions will remain significantly different from the true result. Conformational space of even very short IDPs (*e.g.*, ~10 residues or longer) can be complex enough to present formidable challenges for recovering the true disordered ensemble from a biased trajectory, generated either at high temperatures or with modified Hamiltonian. Instead of analyzing self-convergence (as a function of simulation time), a more rigorous test of convergence is to analyze results obtained from simulations initiated from distinct and distal initial states (such as highly structured and fully disordered conformations [7]).

5. Multi-scale approaches for overcoming sampling problems of large systems

As discussed above, dramatic improvement in atomistic protein force fields coupled with enhanced samplings and GPU computing have now enabled us to generate the disordered conformational ensembles of increasingly complex IDPs in both bound and unbound states. Many important phenomena related to IDPs remain largely out of the reach of physics-based atomistic simulations, such as aggregation [134-136] and biological condensates [137-140]. Here, we review two of the key multi-scale approaches that allow one to simulate longer time-scale bioprocesses and more complex systems within current computational capability, namely, implicit solvent and coarse-grained (CG) models. Both approaches have been extensively studied and applied to globular proteins as well as IDPs.

5.1. Implicit solvent models for removing solvent DOFs

Implicit treatment of solvent is an effective approach to reduce the computational cost of atomistic IDP simulations. The basic idea is to directly estimate the solvation free energy to capture the mean effect of solvent on the thermodynamic properties of the solute [141]. Implicit solvent is essentially a multi-scale model, where the solvent is represented using certain physical model while keeping atomistic details of the solute. These models have emerged as

attractive alternatives for simulations of IDPs and their interactions compared to explicit solvent. In particular, many generalized Born (GB) based implicit solvent models have been developed, including the fast analytical continuum treatment of solvation (FACTS) [142], Amber GB models (such as GB-HCT[143], GB-OBC[144], and GB-Neck[145,146]), analytical generalized Born plus nonpolar (AGBNP) [147,148], and GB models implemented in CHARMM program (such as GBSW [149] and GBMV[150,151]). Several of these GB models can be optimized to provide a balance between computational efficiency and accuracy desired for IDP simulations [127,152,153], by systematic optimization of key physical parameters such as atomic radii to balance solvation and intramolecular interactions. Applied to various model IDPs with extensive experimental data, implicit solvent simulations have provided important insights on detailed conformational properties of the unbound state and how these properties may support function [32,33,154-156].

Despite many successes, implicit solvent models have not widely tested and applied to the studies of larger IDPs. Several factors likely contribute to this. Most implicit solvent models are built upon existing protein force fields, which until recent years have significant limitations in describing disordered protein conformations. Implicit treatment of solvent also relies on various approximations for computational efficiency, such as treating water as a continuous dielectric medium in GB models, limiting the ability of implicit solvent to accurately capture the conformational dependence of solvation free energy. A particular limitation is the common use of surface area (SA)-based model for describing nonpolar solvation energy, which has known limitations in describing the length-scale dependence as well as solvent screening of dispersion interactions [153]. These limitations can result in a systematic bias towards overly compact conformational ensemble, which is more pronounced for larger IDPs.

Several recent efforts have been made to further improve implicit solvent models for IDP simulations. The GB-Neck2 model has been optimized to reproduce solvation energies for a variety of protein systems [146]. Recent benchmark studies have shown that the GB-Neck2 model can reasonably discriminate folded and disordered peptides and could be used for quantitative protein folding simulations up to millisecond time scales [157-159]. Recently, the GBMV2 model, which includes an analytical approximation of molecular volume and is arguably one of the best GB models, has been implemented on the CUDA platform using the CHARMM/OpenMM interface [160]. The ~2 order of magnitude GPU acceleration greatly enables GBMV2 to simulate the conformation and interaction of larger IDPs. The ABSINTH implicit solvent model focuses on recapitulating the polymer properties of peptides and has been successfully used for a variety of IDP simulations, including Aβ peptides and aggregation of phenylalanine [161,162] and sequence-conformation relationship of IDPs in general [8,163]. Recently, an ABSINTH-C model was developed to address the problem of overly shallow Ramachandran distributions of ABSINTH, by adding residue-specific correction terms [164]. The new model not only has a capacity to maintain stable native structures of α-/β-folded proteins, but also increase the reversible folding of β-hairpin peptides.

5.2. Coarse-grain models for reducing the DOFs of proteins

Notwithstanding the ever-improving atomistic modeling, coarse-graining has remained an attractive and often effective strategy for extending the accessible time and length-scales of MD simulations. By grouping multiple (protein) atoms into CG beads and using simplified potential energy functions, CG modeling does not only reduce the system size, often by ~10-fold, but also allows much larger MD integration time steps up to 10s of fs. Together, many CG models can be several orders of magnitude more efficient than atomistic ones. Numerous CG models have achieved varying levels of success in studies of protein folding, binding, and assembly [165,166]. Nonetheless, there are important distinctions between the conformational properties between globular proteins and IDPs, as well as the relative importance of electrostatic, hydrophobic, and hydrogen-bonding interactions in governing their conformational equilibria. Therefore, CG models optimized for the folded proteins are generally not suitable for the IDP simulations. It is often necessary to readjust the parameters of protein-protein and protein-solvent interactions or add new terms for more accurate description of IDP conformations (Figure 3). Here, we summarize several of these refined CG models for more efficient sampling of IDP conformation and interactions as well as their successes and limitations.

Gō/Gō-like models, also known as topology-based models, are based on the funneled energy landscape theory [167] and have been highly successful in describing the folding mechanism and pathway of structured proteins [168]. Somewhat surprisingly, Gō-like models have also proven effective for determining the mechanism and kinetics of IDP interactions, particularly the coupled binding and folding process [169-174]. The implication is that the binding and folding are governed by similar principles that require minimal frustration for efficiency. Note that Gō-like models generally require additional calibrations to provide a more quantitative description of the balance between intermolecular interactions and intrinsic conformational propensities [175]. A key limitation of topology-based modeling of IDPs is lack of the ability to capture the impacts of non-"native" structural features and nonspecific interactions, which could play important roles in IDP structure and function. This may be partially overcome by including new energy terms

(Figure 3) such as explicit charge-charge interactions, inert crowder molecules and confinement potentials. A particularly interesting discovery from these extended topology-based modeling of IDPs is the role long-range electrostatic interactions in promoting efficient coupled binding and folding, allowing IDPs to fold at timescales beyond the μs "folding speed limit" to avoid a potential kinetic bottleneck in specific recognition [171,172,176]. IDP-binding proteins have evolved to contain charges near the binding interface to complement those highly conserved ones on IDPs. Long-range electrostatic interactions between these charges do not only accelerate the encountering of IDPs but also promote the efficiency of IDP folding upon nonspecific encounter.

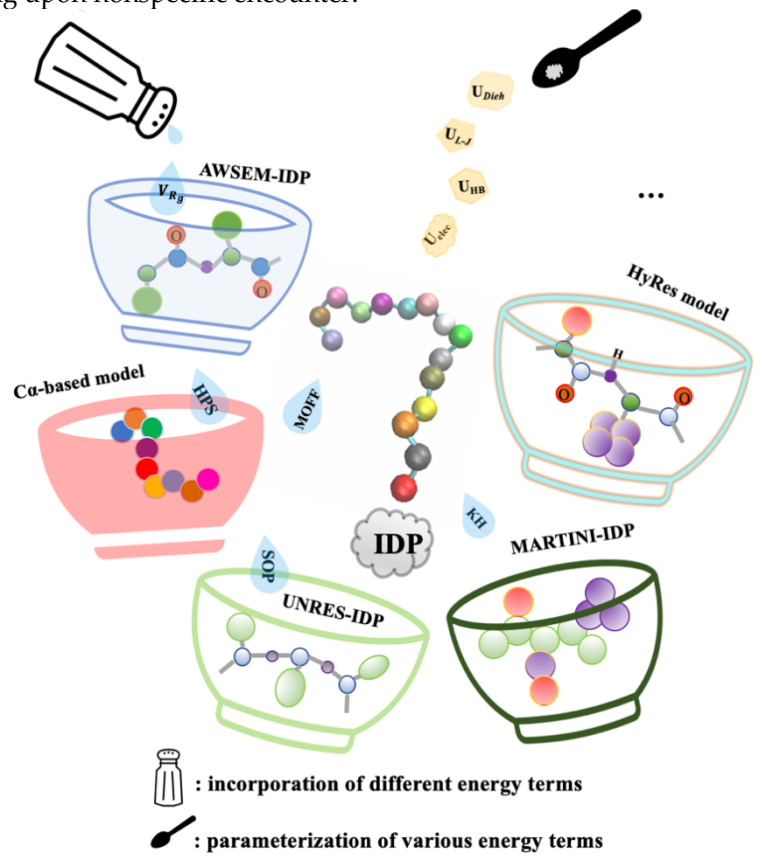


Figure 3. Coarse-grain modeling for addressing various IDPs-related challenges. These models can have a range of spatial resolutions and may be refined by introducing various effective potentials and/or re-calibrating the parameters of these energy terms.

Several higher resolution coarse-grained models have been also developed specifically for modeling IDPs. Thirumalai and co-workers reparametrize the two-bead self-organized polymer coarse-grained model (SOP-CG) to reproduce Rg values of a set of diverse IDPs with 20 to 441 residues [177]. The resulting SOP-IDP also accurately reproduce the small-angle X-ray scattering profiles for these IDPs. Nonetheless, SOP-IDP is designed for IDPs solely and lack the transferability and compatibility in describing even small globular proteins under the physiological conditions. Recognizing the limitation of Cα-only backbone representation in capturing the intrinsic conformational propensities of IDPs, Chen and Liu developed a hybrid resolution (HyRes) model that contains an atomistic description of the backbone, to provide a semi-qualitative description of the secondary structure propensities, and intermediate resolution side chains, to allow qualitative description of the overall peptide chain dimension and transient long-range interactions [178]. While HyRes was originally designed for driving faster atomistic sampling for MSES simulations, applications to a set of small and large IDPs including p53-TAD suggest that HyRes may be appropriate for simulating IDP structure and interactions by itself [178]. Papoian and co-workers have developed the AWSEM-IDP model that can be used to efficiently sample the large conformational space of IDPs and at the same time can distinguish the levels of peptide chain expansion of globular proteins and IDPs [179]. AWSEM-IDP includes only C_α, C_β and O atoms, and has been reparametrized for IDPs by adjusting the secondary structure-related potential energy terms as well as introducing a new parameter, V_{Rg} term, for controlling the collapse and size fluctuation of the protein.

An important application for CG models is to study liquid-liquid phase transitions (LLPS) that are frequently mediated by IDPs [44-47]. Dignon et al, proposed a residue-based C_α-only CG model to represent the disordered low

complexity domain of the RNA-binding protein FUS-LCD and the DEAD-box helicase protein LAF-1 in the formation of LLPS [179]. The model uses the Debye-Hückel approximation for long-range electrostatic interactions and the hydrophobicity scale model [180] or the Kim-Hummer model [181] to short-range residue-residue interactions. The results indicated that both two approaches could reproduce the experimentally observed phase behaviors and changes in phase diagrams caused by mutation. Although they mentioned that the temperature-dependent phase behaviors were not compatible with the experimental absolute temperature and the ionic strength dependence was not fully tested due to the breakdown of the Debye-Hückel electrostatic energy potentials. The model could be further refined. For example, more residue-type parameters were considered to account for phosphorylation and acetylation effects [182], which allows in-depth investigation of how post-translational modifications may control LLPS behaviors. Recently, Latham and Zhang re-tuned Dignon et al's model to better reproduce the Rg distributions of a set of folded and disordered proteins [183]. The resulting Maximum entropy Optimized force field (MOFF) includes a new residue-residue interaction matrix and is more transferable for modeling both globular proteins and IDPs. Hummer and co-workers modified the MARTINI model via re-scaling the solute-solute non-bonded Lennard-Jones potentials to reproduce the experimental transfer free energies of phase separation among dilute and dense liquid phases and proposed a more general approach in tuning CG models with MD for LLPS related studies by optimizing and balancing the solute-solute and solute-solvent interactions then matching the CG data to the atomistic simulation or experimental results. [180]. The resulting MARTINI-IDP model was shown to successfully simulate the droplet formation and capture reversible phase transformations. These are exciting progresses that highlights the strong potential for simple C_α-only CG models in molecular simulations of LLPS involving IDPs. Nonetheless, difficulty in describing local structure propensities (such as transient helices) with the Cα-only representation may be an important limitation for studying certain specific effects of IDPs in LLPS.

6. Concluding remarks

Effective and reliable molecular simulations are crucial for characterizing the details of disordered conformational ensembles of IDPs in isolation, dynamic complexes or biological condensates. Such computational capability, integrated with experimental studies, makes it possible to determine how the dynamic protein states may respond various cellular stimuli in signaling and regulation and more rigorously establish the (dynamic) structure-function relationship of IDPs and IDRs. In this review, we highlight recent advances in meeting two central requirements for reliable IDP simulations, namely, accurate force fields, for describing the energetics of protein conformations, and efficient MD simulation methods, for adequate sampling of relevant conformational space. The need to simulate disordered protein ensembles has played a key role in driving significant improvements in empirical protein force fields in recent years. Many of these force fields are now well balanced for both folded and disordered proteins. The force field development itself has directly benefited from many advanced sampling methods that allow accurate calculation of the conformational equilibria of model peptides and proteins during force field recalibration. These enhanced sampling techniques rely on carefully designed biasing potentials, modification to the original Hamiltonian and/or tempering to accelerate barrier crossing and generate statistically meaningful ensembles with far less computation. Many of the enhanced sampling strategies are complementary with each other, and can be readily integrated together to further improve the efficiency. Together, the improved protein force fields and powerful sampling techniques now allow realistic simulations of the conformation and interaction of at least modest-sized IDPs at atomistic level. Nonetheless, the high dimensionality and complex nature of disordered protein conformation continues to push the limits of the force field and sampling capability. Many proteins models with various levels of resolution are being developed and fine-tuned for IDP simulations, particularly for studying biological condensates. These models ranging from C_α-only single-bead protein models to implicit solvent ones with atomistic proteins. Many of the current models are geared towards modeling systems with minimal residual structures. A key challenge in these multi-scale modeling and simulation of IDPs is finding the optimal compromise between resolution, accuracy and efficiency for the particular problem of interest. Nonetheless, it can be expected that multi-scale simulations will continue to play a central role in studying IDPs and dynamic interactions.

Author Contributions: writing—original draft preparation, X.G. and Y.Z.; writing—review and editing, J.C.; visualization, X.G. and Y.Z.; All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Institutes of Health (GM114300) and National Science Foundation (MCB 1817332).

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Csizmok, V.; Follis, A.V.; Kriwacki, R.W.; Forman-Kay, J.D. Dynamic Protein Interaction Networks and New Structural Paradigms in Signaling. *Chemical Reviews* **2016**, *116*, 6424-6462, doi:10.1021/acs.chemrev.5b00548.
2. Oldfield, C.J.; Dunker, A.K. Intrinsically Disordered Proteins and Intrinsically Disordered Protein Regions. *Annual Review of Biochemistry* **2014**, *83*, 553-584, doi:10.1146/annurev-biochem-072711-164947.
3. Wright, P.E.; Dyson, H.J. Intrinsically disordered proteins in cellular signalling and regulation. *Nat Rev Mol Cell Bio* **2015**, *16*, 18-29, doi:10.1038/nrm3920.
4. Uversky, V.N. Intrinsically disordered proteins and their (disordered) proteomes in neurodegenerative disorders. *Frontiers in Aging Neuroscience* **2015**, *7*, 18.
5. Dyson, H.J.; Wright, P.E. Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Bio* **2005**, *6*, 197-208, doi:10.1038/nrm1589.
6. Owen, I.; Shewmaker, F. The Role of Post-Translational Modifications in the Phase Transitions of Intrinsically Disordered Proteins. *International Journal of Molecular Sciences* **2019**, *20*, doi:10.3390/ijms20215501.
7. Chen, J. Towards the physical basis of how intrinsic disorder mediates protein function. *Arch Biochem Biophys* **2012**, *524*, 123-131, doi:10.1016/j.abb.2012.04.024.
8. Das, R.K.; Ruff, K.M.; Pappu, R.V. Relating sequence encoded information to form and function of intrinsically disordered proteins. *Curr Opin Struct Biol* **2015**, *32*, 102-112, doi:10.1016/j.sbi.2015.03.008.
9. Hatos, A.; Hajdu-Soltesz, B.; Monzon, A.M.; Palopoli, N.; Alvarez, L.; Aykac-Fas, B.; Bassot, C.; Benitez, G.I.; Bevilacqua, M.; Chasapi, A.; et al. DisProt: intrinsic protein disorder annotation in 2020. *Nucleic Acids Res* **2020**, *48*, D269-D276, doi:10.1093/nar/gkz975.
10. Vacic, V.; Iakoucheva, L.M. Disease mutations in disordered regions--exception to the rule? *Mol Biosyst* **2012**, *8*, 27-32, doi:10.1039/c1mb05251a.
11. Kulkarni, P.; Uversky, V.N. Intrinsically Disordered Proteins in Chronic Diseases. *Biomolecules* **2019**, *9*, doi:10.3390/biom9040147.
12. Oldfield, C.J.; Cheng, Y.; Cortese, M.S.; Brown, C.J.; Uversky, V.N.; Dunker, A.K. Comparing and Combining Predictors of Mostly Disordered Proteins. *Biochemistry-Us* **2005**, *44*, 1989-2000, doi:10.1021/bi047993o.
13. Chen, J.; Liu, X.; Chen, J. Targeting Intrinsically Disordered Proteins through Dynamic Interactions. *Biomolecules* **2020**, *10*, doi:10.3390/biom10050743.
14. Mittag, T.; Marsh, J.; Grishaev, A.; Orlicky, S.; Lin, H.; Sicheri, F.; Tyers, M.; Forman-Kay, J.D. Structure/Function Implications in a Dynamic Complex of the Intrinsically Disordered Sic1 with the Cdc4 Subunit of an SCF Ubiquitin Ligase. *Structure* **2010**, *18*, 494-506, doi:10.1016/j.str.2010.01.020.
15. McDowell, C.; Chen, J.; Chen, J. Potential Conformational Heterogeneity of p53 Bound to S100B(betabeta). *Journal of Molecular Biology* **2013**, doi:10.1016/j.jmb.2013.01.001.
16. Wu, H.; Fuxreiter, M. The Structure and Dynamics of Higher-Order Assemblies: Amyloids, Signalosomes, and Granules. *Cell* **2016**, *165*, 1055-1066, doi:10.1016/j.cell.2016.05.004.
17. Krois, A.S.; Ferreon, J.C.; Martinez-Yamout, M.A.; Dyson, H.J.; Wright, P.E. Recognition of the disordered p53 transactivation domain by the transcriptional adapter zinc finger domains of CREB-binding protein. *Proc Natl Acad Sci U S A* **2016**, doi:10.1073/pnas.1602487113.
18. Csizmok, V.; Orlicky, S.; Cheng, J.; Song, J.; Bah, A.; Delgoshaie, N.; Lin, H.; Mittag, T.; Sicheri, F.; Chan, H.S.; et al. An allosteric conduit facilitates dynamic multisite substrate recognition by the SCF(Cdc4) ubiquitin ligase. *Nat Commun* **2017**, *8*, 13943, doi:10.1038/ncomms13943.

19. Borgia, A.; Borgia, M.B.; Bugge, K.; Kissling, V.M.; Heidarsson, P.O.; Fernandes, C.B.; Sottini, A.; Soranno, A.; Buholzer, K.J.; Nettels, D.; et al. Extreme disorder in an ultrahigh-affinity protein complex. *Nature* **2018**, *555*, 61-+, doi:10.1038/nature25762.
20. Clark, S.; Myers, J.B.; King, A.; Fiala, R.; Novacek, J.; Pearce, G.; Heierhorst, J.; Reichow, S.L.; Barbar, E.J. Multivalency regulates activity in an intrinsically disordered transcription factor. *Elife* **2018**, *7*, doi:10.7554/eLife.36258.
21. Fuxreiter, M. Fuzziness in Protein Interactions-A Historical Perspective. *Journal of Molecular Biology* **2018**, *430*, 2278-2287, doi:10.1016/j.jmb.2018.02.015.
22. Weng, J.; Wang, W. Dynamic multivalent interactions of intrinsically disordered proteins. *Curr Opin Struct Biol* **2019**, *62*, 9-13, doi:10.1016/j.sbi.2019.11.001.
23. Miskei, M.; Antal, C.; Fuxreiter, M. FuzDB: database of fuzzy complexes, a tool to develop stochastic structure-function relationships for protein complexes and higher-order assemblies. *Nucleic Acids Research* **2017**, *45*, D228-D235, doi:10.1093/nar/gkw1019.
24. Ganguly, D.; Chen, J. Structural interpretation of paramagnetic relaxation enhancement-derived distances for disordered protein states. *J Mol Biol* **2009**, *390*, 467-477, doi:10.1016/j.jmb.2009.05.019.
25. Fisher, C.K.; Stultz, C.M. Constructing ensembles for intrinsically disordered proteins. *Curr Opin Struc Biol* **2011**, *21*, 426-431, doi:10.1016/j.sbi.2011.04.001.
26. Ferreon, A.C.; Ferreon, J.C.; Wright, P.E.; Deniz, A.A. Modulation of allostery by protein intrinsic disorder. *Nature* **2013**, *498*, 390-394, doi:10.1038/nature12294.
27. Garcia-Pino, A.; Balasubramanian, S.; Wyns, L.; Gazit, E.; De Greve, H.; Magnuson, R.D.; Charlier, D.; van Nuland, N.A.; Loris, R. Allostery and intrinsic disorder mediate transcription regulation by conditional cooperativity. *Cell* **2010**, *142*, 101-111, doi:10.1016/j.cell.2010.05.039.
28. Berlow, R.B.; Dyson, H.J.; Wright, P.E. Expanding the Paradigm: Intrinsically Disordered Proteins and Allosteric Regulation. *Journal of Molecular Biology* **2018**, *430*, 2309-2320, doi:10.1016/j.jmb.2018.04.003.
29. Levine, Z.A.; Shea, J.E. Simulations of disordered proteins and systems with conformational heterogeneity. *Curr Opin Struct Biol* **2017**, *43*, 95-103, doi:10.1016/j.sbi.2016.11.006.
30. Knott, M.; Best, R.B. A preformed binding interface in the unbound ensemble of an intrinsically disordered protein: evidence from molecular simulations. *Plos Computational Biology* **2012**, *8*, e1002605, doi:10.1371/journal.pcbi.1002605.
31. Mao, A.H.; Crick, S.L.; Vitalis, A.; Chicoine, C.L.; Pappu, R.V. Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. *Proceedings of the National Academy of Sciences of the United States of America* **2010**, *107*, 8183-8188, doi:10.1073/pnas.0911107107.
32. Ganguly, D.; Chen, J. Atomistic details of the disordered states of KID and pKID. implications in coupled binding and folding. *Journal of the American Chemical Society* **2009**, *131*, 5214-5223.
33. Zhang, W.; Ganguly, D.; Chen, J. Residual structures, conformational fluctuations, and electrostatic interactions in the synergistic folding of two intrinsically disordered proteins. *Plos Computational Biology* **2012**, *8*, e1002353, doi:10.1371/journal.pcbi.1002353.
34. Zhang, W.H.; Chen, J.H. Accelerate Sampling in Atomistic Energy Landscapes Using Topology-Based Coarse-Grained Models. *Journal of Chemical Theory and Computation* **2014**, *10*, 918-923, doi:10.1021/ct500031v.
35. Moritsugu, K.; Terada, T.; Kidera, A. Scalable free energy calculation of proteins via multiscale essential sampling. *J Chem Phys* **2010**, *133*, 224105, doi:10.1063/1.3510519.

36. Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters* **1999**, *314*, 141-151, doi:https://doi.org/10.1016/S0009-2614(99)01123-9.
37. Liu, P.; Kim, B.; Friesner, R.A.; Berne, B.J. Replica exchange with solute tempering: A method for sampling biological systems in explicit water. *Proceedings of the National Academy of Sciences of the United States of America* **2005**, *102*, 13749, doi:10.1073/pnas.0506346102.
38. Mittal, A.; Lyle, N.; Harmon, T.S.; Pappu, R.V. Hamiltonian Switch Metropolis Monte Carlo Simulations for Improved Conformational Sampling of Intrinsically Disordered Regions Tethered to Ordered Domains of Proteins. *Journal of Chemical Theory and Computation* **2014**, *10*, 3550-3562, doi:Doi 10.1021/Ct5002297.
39. Peter, E.K.; Shea, J.E. A hybrid MD-kMC algorithm for folding proteins in explicit solvent. *Physical Chemistry Chemical Physics* **2014**, *16*, 6430-6440, doi:Doi 10.1039/C3cp55251a.
40. Zhang, C.; Ma, J. Enhanced sampling and applications in protein folding in explicit solvent. *J Chem Phys* **2010**, *132*, 244101, doi:10.1063/1.3435332.
41. Zheng, L.Q.; Yang, W. Practically Efficient and Robust Free Energy Calculations: Double-Integration Orthogonal Space Tempering. *Journal of Chemical Theory and Computation* **2012**, *8*, 810-823, doi:Doi 10.1021/Ct200726v.
42. Best, R.B. Computational and theoretical advances in studies of intrinsically disordered proteins. *Curr Opin Struc Biol* **2017**, *42*, 147-154, doi:https://doi.org/10.1016/j.sbi.2017.01.006.
43. Kmiecik, S.; Gront, D.; Kolinski, M.; Wieteska, L.; Dawid, A.E.; Kolinski, A. Coarse-Grained Protein Models and Their Applications. *Chem Rev* **2016**, *116*, 7898-7936, doi:10.1021/acs.chemrev.6b00163.
44. Bhattacharya, S.; Lin, X. Recent Advances in Computational Protocols Addressing Intrinsically Disordered Proteins. *Biomolecules* **2019**, *9*, doi:10.3390/biom9040146.
45. Levine, Z.A.; Shea, J.-E. Simulations of disordered proteins and systems with conformational heterogeneity. *Curr Opin Struc Biol* **2017**, *43*, 95-103, doi:https://doi.org/10.1016/j.sbi.2016.11.006.
46. Wang, W. Recent advances in atomic molecular dynamics simulation of intrinsically disordered proteins. *Physical Chemistry Chemical Physics* **2021**, *23*, 777-784, doi:10.1039/D0CP05818A.
47. Shea, J.-E.; Best, R.B.; Mittal, J. Physics-based computational and theoretical approaches to intrinsically disordered proteins. *Curr Opin Struc Biol* **2021**, *67*, 219-225, doi:https://doi.org/10.1016/j.sbi.2020.12.012.
48. Arai, M. Unified understanding of folding and binding mechanisms of globular and intrinsically disordered proteins. *Biophysical Reviews* **2018**, *10*, 163-181, doi:10.1007/s12551-017-0346-7.
49. Rauscher, S.; Gapsys, V.; Gajda, M.J.; Zweckstetter, M.; de Groot, B.L.; Grubmüller, H. Structural Ensembles of Intrinsically Disordered Proteins Depend Strongly on Force Field: A Comparison to Experiment. *Journal of Chemical Theory and Computation* **2015**, *11*, 5513-5524, doi:10.1021/acs.jctc.5b00736.
50. Best, R.B.; Zhu, X.; Shim, J.; Lopes, P.E.M.; Mittal, J.; Feig, M.; MacKerell, A.D. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ, ψ and Side-Chain χ1 and χ2 Dihedral Angles. *Journal of Chemical Theory and Computation* **2012**, *8*, 3257-3273, doi:10.1021/ct300400x.
51. Robustelli, P.; Piana, S.; Shaw, D.E. Developing a molecular dynamics force field for both folded and disordered protein states. *Proceedings of the National Academy of Sciences* **2018**, *115*, E4758, doi:10.1073/pnas.1800690115.
52. Huang, J.; MacKerell, A.D. Force field development and simulations of intrinsically disordered proteins. *Curr Opin Struc Biol* **2018**, *48*, 40-48, doi:https://doi.org/10.1016/j.sbi.2017.10.008.

53. Piana, S.; Donchev, A.G.; Robustelli, P.; Shaw, D.E. Water Dispersion Interactions Strongly Influence Simulated Structural Properties of Disordered Protein States. *The Journal of Physical Chemistry B* **2015**, *119*, 5113-5123, doi:10.1021/jp508971m.
54. Wu, H.-N.; Jiang, F.; Wu, Y.-D. Significantly Improved Protein Folding Thermodynamics Using a Dispersion-Corrected Water Model and a New Residue-Specific Force Field. *The Journal of Physical Chemistry Letters* **2017**, *8*, 3199-3205, doi:10.1021/acs.jpclett.7b01213.
55. Mu, J.; Liu, H.; Zhang, J.; Luo, R.; Chen, H.-F. Recent Force Field Strategies for Intrinsically Disordered Proteins. *Journal of Chemical Information and Modeling* **2021**, *61*, 1037-1047, doi:10.1021/acs.jcim.0c01175.
56. Song, D.; Liu, H.; Luo, R.; Chen, H.-F. Environment-Specific Force Field for Intrinsically Disordered and Ordered Proteins. *Journal of Chemical Information and Modeling* **2020**, *60*, 2257-2267, doi:10.1021/acs.jcim.0c00059.
57. Yang, S.; Liu, H.; Zhang, Y.; Lu, H.; Chen, H. Residue-Specific Force Field Improving the Sample of Intrinsically Disordered Proteins and Folded Proteins. *Journal of Chemical Information and Modeling* **2019**, *59*, 4793-4805, doi:10.1021/acs.jcim.9b00647.
58. Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B.L.; Grubmüller, H.; MacKerell, A.D. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature Methods* **2017**, *14*, 71-73, doi:10.1038/nmeth.4067.
59. Tian, C.; Kasavajhala, K.; Belfon, K.A.A.; Raguette, L.; Huang, H.; Migues, A.N.; Bickel, J.; Wang, Y.; Pincay, J.; Wu, Q.; et al. ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *Journal of Chemical Theory and Computation* **2020**, *16*, 528-552, doi:10.1021/acs.jctc.9b00591.
60. Rahman, M.U.; Rehman, A.U.; Liu, H.; Chen, H.-F. Comparison and Evaluation of Force Fields for Intrinsically Disordered Proteins. *Journal of Chemical Information and Modeling* **2020**, *60*, 4912-4923, doi:10.1021/acs.jcim.0c00762.
61. Abriata, L.A.; Dal Peraro, M. Assessment of transferable forcefields for protein simulations attests improved description of disordered states and secondary structure propensities, and hints at multi-protein systems as the next challenge for optimization. *Computational and Structural Biotechnology Journal* **2021**, *19*, 2626-2636, doi:https://doi.org/10.1016/j.csbj.2021.04.050.
62. Piana, S.; Robustelli, P.; Tan, D.; Chen, S.; Shaw, D.E. Development of a Force Field for the Simulation of Single-Chain Proteins and Protein–Protein Complexes. *Journal of Chemical Theory and Computation* **2020**, *16*, 2494-2507, doi:10.1021/acs.jctc.9b00251.
63. Song, D.; Luo, R.; Chen, H.-F. The IDP-Specific Force Field ff14IDPSFF Improves the Conformer Sampling of Intrinsically Disordered Proteins. *Journal of Chemical Information and Modeling* **2017**, *57*, 1166-1178, doi:10.1021/acs.jcim.7b00135.
64. Jing, Z.; Liu, C.; Cheng, S.Y.; Qi, R.; Walker, B.D.; Piquemal, J.-P.; Ren, P. Polarizable Force Fields for Biomolecular Simulations: Recent Advances and Applications. *Annual Review of Biophysics* **2019**, *48*, 371-394, doi:10.1146/annurev-biophys-070317-033349.
65. Bedrov, D.; Piquemal, J.-P.; Borodin, O.; MacKerell, A.D.; Roux, B.; Schröder, C. Molecular Dynamics Simulations of Ionic Liquids and Electrolytes Using Polarizable Force Fields. *Chemical Reviews* **2019**, *119*, 7940-7995, doi:10.1021/acs.chemrev.8b00763.
66. Inakollu, V.S.S.; Geerke, D.P.; Rowley, C.N.; Yu, H. Polarisable force fields: what do they add in biomolecular simulations? *Curr Opin Struc Biol* **2020**, *61*, 182-190, doi:https://doi.org/10.1016/j.sbi.2019.12.012.

67. Huang, J.; MacKerell, Alexander D. Induction of Peptide Bond Dipoles Drives Cooperative Helix Formation in the (AAQAA)3 Peptide. *Biophysical Journal* **2014**, *107*, 991-997, doi:https://doi.org/10.1016/j.bpj.2014.06.038.
68. Kamenik, A.S.; Handle, P.H.; Hofer, F.; Kahler, U.; Kraml, J.; Liedl, K.R. Polarizable and non-polarizable force fields: Protein folding, unfolding, and misfolding. *The Journal of Chemical Physics* **2020**, *153*, 185102, doi:10.1063/5.0022135.
69. Wang, A.; Zhang, Z.; Li, G. Higher Accuracy Achieved in the Simulations of Protein Structure Refinement, Protein Folding, and Intrinsically Disordered Proteins Using Polarizable Force Fields. *The Journal of Physical Chemistry Letters* **2018**, *9*, 7110-7116, doi:10.1021/acs.jpclett.8b03471.
70. Wang, A.; Peng, X.; Li, Y.; Zhang, D.; Zhang, Z.; Li, G. Quality of force fields and sampling methods in simulating pepX peptides: a case study for intrinsically disordered proteins. *Physical Chemistry Chemical Physics* **2021**, *23*, 2430-2437, doi:10.1039/D0CP05484D.
71. Yang, Y.I.; Shao, Q.; Zhang, J.; Yang, L.; Gao, Y.Q. Enhanced sampling in molecular dynamics. *The Journal of Chemical Physics* **2019**, *151*, 070902, doi:10.1063/1.5109531.
72. Wang, A.-h.; Zhang, Z.-c.; Li, G.-h. Advances in enhanced sampling molecular dynamics simulations for biomolecules. *Chinese Journal of Chemical Physics* **2019**, *32*, 277-286, doi:10.1063/1674-0068/cjcp1905091.
73. Barducci, A.; Bonomi, M.; Parrinello, M. Metadynamics. *WIREs Computational Molecular Science* **2011**, *1*, 826-843, doi:https://doi.org/10.1002/wcms.31.
74. Barducci, A.; Bussi, G.; Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Physical Review Letters* **2008**, *100*, 020603, doi:10.1103/PhysRevLett.100.020603.
75. Hamelberg, D.; Mongan, J.; McCammon, J.A. Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. *The Journal of Chemical Physics* **2004**, *120*, 11919-11929, doi:10.1063/1.1755656.
76. Torrie, G.M.; Valleau, J.P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics* **1977**, *23*, 187-199, doi:https://doi.org/10.1016/0021-9991(77)90121-8.
77. Kästner, J. Umbrella sampling. *WIREs Computational Molecular Science* **2011**, *1*, 932-942, doi:https://doi.org/10.1002/wcms.66.
78. Gao, Y.Q. An integrate-over-temperature approach for enhanced sampling. *The Journal of Chemical Physics* **2008**, *128*, 064105, doi:10.1063/1.2825614.
79. MacCallum, J.L.; Muniyat, M.I.; Gaalswyk, K. Online Optimization of Total Acceptance in Hamiltonian Replica Exchange Simulations. *The Journal of Physical Chemistry B* **2018**, *122*, 5448-5457, doi:10.1021/acs.jpcb.7b11778.
80. Bussi, G.; Laio, A. Using metadynamics to explore complex free-energy landscapes. *Nature Reviews Physics* **2020**, *2*, 200-212, doi:10.1038/s42254-020-0153-0.
81. Galvelis, R.; Sugita, Y. Replica state exchange metadynamics for improving the convergence of free energy estimates. *Journal of Computational Chemistry* **2015**, *36*, 1446-1455, doi:https://doi.org/10.1002/jcc.23945.
82. Piana, S.; Laio, A. A Bias-Exchange Approach to Protein Folding. *The Journal of Physical Chemistry B* **2007**, *111*, 4553-4559, doi:10.1021/jp067873l.
83. Liu, N.; Guo, Y.; Ning, S.; Duan, M. Phosphorylation regulates the binding of intrinsically disordered proteins via a flexible conformation selection mechanism. *Communications Chemistry* **2020**, *3*, 123, doi:10.1038/s42004-020-00370-5.

84. Galvelis, R.; Re, S.; Sugita, Y. Enhanced Conformational Sampling of N-Glycans in Solution with Replica State Exchange Metadynamics. *Journal of Chemical Theory and Computation* **2017**, *13*, 1934-1942, doi:10.1021/acs.jctc.7b00079.
85. Do, T.N.; Choy, W.-Y.; Karttunen, M. Binding of Disordered Peptides to Kelch: Insights from Enhanced Sampling Simulations. *Journal of Chemical Theory and Computation* **2016**, *12*, 395-404, doi:10.1021/acs.jctc.5b00868.
86. Dickson, A.; Ahlstrom, L.S.; Brooks III, C.L. Coupled folding and binding with 2D Window-Exchange Umbrella Sampling. *Journal of Computational Chemistry* **2016**, *37*, 587-594, doi:https://doi.org/10.1002/jcc.24004.
87. Guo, J.; Zhou, H.X. Protein Allostery and Conformational Dynamics. *Chem Rev* **2016**, *116*, 6503-6515, doi:10.1021/acs.chemrev.5b00590.
88. Gianni, S.; Freiberger, M.I.; Jemth, P.; Ferreiro, D.U.; Wolynes, P.G.; Fuxreiter, M. Fuzziness and Frustration in the Energy Landscape of Protein Folding, Function, and Assembly. *Acc Chem Res* **2021**, *54*, 1251-1259, doi:10.1021/acs.accounts.0c00813.
89. Neupane, K.; Manuel, A.P.; Woodside, M.T. Protein folding trajectories can be described quantitatively by one-dimensional diffusion over measured energy landscapes. *Nature Physics* **2016**, *12*, 700-703, doi:10.1038/nphys3677.
90. Pfaendtner, J.; Bonomi, M. Efficient Sampling of High-Dimensional Free-Energy Landscapes with Parallel Bias Metadynamics. *J Chem Theory Comput* **2015**, *11*, 5062-5067, doi:10.1021/acs.jctc.5b00846.
91. Prakash, A.; Fu, C.D.; Bonomi, M.; Pfaendtner, J. Biasing Smarter, Not Harder, by Partitioning Collective Variables into Families in Parallel Bias Metadynamics. *J Chem Theory Comput* **2018**, *14*, 4985-4990, doi:10.1021/acs.jctc.8b00448.
92. Awasthi, S.; Nair, N.N. Exploring high dimensional free energy landscapes: Temperature accelerated sliced sampling. *The Journal of Chemical Physics* **2017**, *146*, 094108, doi:10.1063/1.4977704.
93. Sidky, H.; Chen, W.; Ferguson, A.L. Machine learning for collective variable discovery and enhanced sampling in biomolecular simulation. *Molecular Physics* **2020**, *118*, e1737742, doi:10.1080/00268976.2020.1737742.
94. Chen, W.; Tan, A.R.; Ferguson, A.L. Collective variable discovery and enhanced sampling using autoencoders: Innovations in network architecture and error function design. *The Journal of Chemical Physics* **2018**, *149*, 072312, doi:10.1063/1.5023804.
95. Chen, W.; Ferguson, A.L. Molecular enhanced sampling with autoencoders: On-the-fly collective variable discovery and accelerated free energy landscape exploration. *Journal of Computational Chemistry* **2018**, *39*, 2079-2102, doi:https://doi.org/10.1002/jcc.25520.
96. Galvelis, R.; Sugita, Y. Neural Network and Nearest Neighbor Algorithms for Enhancing Sampling of Molecular Dynamics. *Journal of Chemical Theory and Computation* **2017**, *13*, 2489-2500, doi:10.1021/acs.jctc.7b00188.
97. Salawu, E.O. DESP: Deep Enhanced Sampling of Proteins' Conformation Spaces Using AI-Inspired Biasing Forces. *Frontiers in Molecular Biosciences* **2021**, *8*, doi:10.3389/fmolb.2021.587151.
98. Zhang, J.; Chen, M. Unfolding Hidden Barriers by Active Enhanced Sampling. *Physical Review Letters* **2018**, *121*, 010601, doi:10.1103/PhysRevLett.121.010601.
99. Brown, S.; Head-Gordon, T. Cool walking: A new Markov chain Monte Carlo sampling method. *Journal of Computational Chemistry* **2003**, *24*, 68-76, doi:https://doi.org/10.1002/jcc.10181.

100. Neal, R.M. Annealed importance sampling. *Statistics and Computing* **2001**, *11*, 125-139, doi:10.1023/A:1008923215028.
101. Marinari, E.; Parisi, G. Simulated Tempering: A New Monte Carlo Scheme. *Europhysics Letters (EPL)* **1992**, *19*, 451-458, doi:10.1209/0295-5075/19/6/002.
102. Fukunishi, H.; Watanabe, O.; Takada, S. On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction. *The Journal of Chemical Physics* **2002**, *116*, 9058-9067, doi:10.1063/1.1472510.
103. Huang, Y.-m.M.; McCammon, J.A.; Miao, Y. Replica Exchange Gaussian Accelerated Molecular Dynamics: Improved Enhanced Sampling and Free Energy Calculation. *Journal of Chemical Theory and Computation* **2018**, *14*, 1853-1864, doi:10.1021/acs.jctc.7b01226.
104. Oshima, H.; Re, S.; Sugita, Y. Replica-Exchange Umbrella Sampling Combined with Gaussian Accelerated Molecular Dynamics for Free-Energy Calculation of Biomolecules. *Journal of Chemical Theory and Computation* **2019**, *15*, 5199-5208, doi:10.1021/acs.jctc.9b00761.
105. Wang, J.; Arantes, P.R.; Bhattarai, A.; Hsu, R.V.; Pawnikar, S.; Huang, Y.-m.M.; Palermo, G.; Miao, Y. Gaussian accelerated molecular dynamics: Principles and applications. *WIREs Computational Molecular Science* **2021**, *11*, e1521, doi:https://doi.org/10.1002/wcms.1521.
106. Miao, Y.; Sinko, W.; Pierce, L.; Bucher, D.; Walker, R.C.; McCammon, J.A. Improved Reweighting of Accelerated Molecular Dynamics Simulations for Free Energy Calculation. *Journal of Chemical Theory and Computation* **2014**, *10*, 2677-2689, doi:10.1021/ct500090q.
107. Miao, Y.; Bhattarai, A.; Wang, J. Ligand Gaussian Accelerated Molecular Dynamics (LiGaMD): Characterization of Ligand Binding Thermodynamics and Kinetics. *Journal of Chemical Theory and Computation* **2020**, *16*, 5526-5547, doi:10.1021/acs.jctc.0c00395.
108. Wang, J.; Miao, Y. Peptide Gaussian accelerated molecular dynamics (Pep-GaMD): Enhanced sampling and free energy and kinetics calculations of peptide binding. *The Journal of Chemical Physics* **2020**, *153*, 154109, doi:10.1063/5.0021399.
109. Wang, L.; Friesner, R.A.; Berne, B.J. Replica Exchange with Solute Scaling: A More Efficient Version of Replica Exchange with Solute Tempering (REST2). *The Journal of Physical Chemistry B* **2011**, *115*, 9431-9438, doi:10.1021/jp204407d.
110. Kamiya, M.; Sugita, Y. Flexible selection of the solute region in replica exchange with solute tempering: Application to protein-folding simulations. *The Journal of Chemical Physics* **2018**, *149*, 072304, doi:10.1063/1.5016222.
111. Liu, X.; Chen, J. Residual Structures and Transient Long-Range Interactions of p53 Transactivation Domain: Assessment of Explicit Solvent Protein Force Fields. *Journal of Chemical Theory and Computation* **2019**, *15*, 4708-4720, doi:10.1021/acs.jctc.9b00397.
112. Shrestha, U.R.; Smith, J.C.; Petridis, L. Full structural ensembles of intrinsically disordered proteins from unbiased molecular dynamics simulations. *Communications Biology* **2021**, *4*, 243, doi:10.1038/s42003-021-01759-1.
113. Hicks, A.; Zhou, H.-X. Temperature-induced collapse of a disordered peptide observed by three sampling methods in molecular dynamics simulations. *The Journal of Chemical Physics* **2018**, *149*, 072313, doi:10.1063/1.5027409.

114. Brown, A.H.; Rodger, P.M.; Evans, J.S.; Walsh, T.R. Equilibrium Conformational Ensemble of the Intrinsically Disordered Peptide n16N: Linking Subdomain Structures and Function in Nacre. *Biomacromolecules* **2014**, *15*, 4467-4479, doi:10.1021/bm501263s.
115. Pang, X.; Zhou, H.-X. Disorder-to-Order Transition of an Active-Site Loop Mediates the Allosteric Activation of Sortase A. *Biophysical Journal* **2015**, *109*, 1706-1715, doi:https://doi.org/10.1016/j.bpj.2015.08.039.
116. Liu, X.; Jia, Z.; Chen, J. Enhanced Sampling of Intrinsic Structural Heterogeneity of the BH3-Only Protein Binding Interface of Bcl-xL. *The Journal of Physical Chemistry B* **2017**, *121*, 9160-9168, doi:10.1021/acs.jpcb.7b06768.
117. Liang, C.; Savinov, S.N.; Fejzo, J.; Eyles, S.J.; Chen, J. Modulation of Amyloid-beta42 Conformation by Small Molecules Through Nonspecific Binding. *J Chem Theory Comput* **2019**, *15*, 5169-5174, doi:10.1021/acs.jctc.9b00599.
118. Liu, X.; Chen, J. Modulation of p53 Transactivation Domain Conformations by Ligand Binding and Cancer-Associated Mutations. *Pac. Symp. Biocomput.* **2020**, *25*, 195-206.
119. Schrag, L.G.; Liu, X.; Thevarajan, I.; Prakash, O.; Zolkiewski, M.; Chen, J. Cancer-Associated Mutations Perturb the Disordered Ensemble and Interactions of the Intrinsically Disordered p53 Transactivation Domain. *J Mol Biol* **2021**, *433*, 167048, doi:10.1016/j.jmb.2021.167048.
120. Zhao, J.; Blayney, A.; Liu, X.; Gandy, L.; Jin, W.; Yan, L.; Ha, J.H.; Canning, A.J.; 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, doi:10.1038/s41467-021-21258-5.
121. Wang, L.; Wu, Y.; Deng, Y.; Kim, B.; Pierce, L.; Krilov, G.; Lupyan, D.; Robinson, S.; Dahlgren, M.K.; Greenwood, J.; et al. Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by Way of a Modern Free-Energy Calculation Protocol and Force Field. *Journal of the American Chemical Society* **2015**, *137*, 2695-2703, doi:10.1021/ja512751q.
122. Kokubo, H.; Tanaka, T.; Okamoto, Y. Two-dimensional replica-exchange method for predicting protein–ligand binding structures. *Journal of Computational Chemistry* **2013**, *34*, 2601-2614, doi:https://doi.org/10.1002/jcc.23427.
123. Zhou, H.X. Theoretical frameworks for multiscale modeling and simulation. *Curr Opin Struc Biol* **2014**, *25C*, 67-76, doi:10.1016/j.sbi.2014.01.004.
124. Zhang, W.; Chen, J. Accelerate Sampling in Atomistic Energy Landscapes Using Topology-Based Coarse-Grained Models. *Journal of Chemical Theory and Computation* **2014**, *10*, 918-923, doi:10.1021/ct500031v.
125. Lee, K.H.; Chen, J.H. Multiscale Enhanced Sampling of Intrinsically Disordered Protein Conformations. *Journal of Computational Chemistry* **2016**, *37*, 550-557, doi:10.1002/jcc.23957.
126. Liu, X.R.; Gong, X.P.; Chen, J.H. Accelerating atomistic simulations of proteins using multiscale enhanced sampling with independent tempering. *Journal of Computational Chemistry* **2021**, *42*, 358-364, doi:10.1002/jcc.26461.
127. Lee, K.H.; Chen, J.H. Optimization of the GBMV2 implicit solvent force field for accurate simulation of protein conformational equilibria. *Journal of Computational Chemistry* **2017**, *38*, 1332-1341, doi:10.1002/jcc.24734.
128. Liu, Y.; Pezeshkian, W.; Barnoud, J.; de Vries, A.H.; Marrink, S.J. Coupling Coarse-Grained to Fine-Grained Models via Hamiltonian Replica Exchange. *Journal of Chemical Theory and Computation* **2020**, *16*, 5313-5322, doi:10.1021/acs.jctc.0c00429.

129. Peng, X.; Zhang, Y.; Li, Y.; Liu, Q.; Chu, H.; Zhang, D.; Li, G. Integrating Multiple Accelerated Molecular Dynamics To Improve Accuracy of Free Energy Calculations. *Journal of Chemical Theory and Computation* **2018**, *14*, 1216-1227, doi:10.1021/acs.jctc.7b01211.
130. Yang, Y.I.; Niu, H.; Parrinello, M. Combining Metadynamics and Integrated Tempering Sampling. *The Journal of Physical Chemistry Letters* **2018**, *9*, 6426-6430, doi:10.1021/acs.jpclett.8b03005.
131. Shirts, M.R.; Chodera, J.D. Statistically optimal analysis of samples from multiple equilibrium states. *The Journal of Chemical Physics* **2008**, *129*, 124105, doi:10.1063/1.2978177.
132. Kumar, S.; Rosenberg, J.M.; Bouzida, D.; Swendsen, R.H.; Kollman, P.A. THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *Journal of Computational Chemistry* **1992**, *13*, 1011-1021, doi:https://doi.org/10.1002/jcc.540130812.
133. Sinko, W.; Miao, Y.; de Oliveira, C.A.F.; McCammon, J.A. Population Based Reweighting of Scaled Molecular Dynamics. *The Journal of Physical Chemistry B* **2013**, *117*, 12759-12768, doi:10.1021/jp401587e.
134. Ilie, I.M.; Caflisch, A. Simulation Studies of Amyloidogenic Polypeptides and Their Aggregates. *Chemical Reviews* **2019**, *119*, 6956-6993, doi:10.1021/acs.chemrev.8b00731.
135. Zhou, H.X.; Pang, X. Electrostatic Interactions in Protein Structure, Folding, Binding, and Condensation. *Chem Rev* **2018**, *118*, 1691-1741, doi:10.1021/acs.chemrev.7b00305.
136. Fassler, J.S.; Skuodas, S.; Weeks, D.L.; Phillips, B.T. Protein Aggregation and Disaggregation in Cells and Development. *J Mol Biol* **2021**, 167215, doi:10.1016/j.jmb.2021.167215.
137. Alberti, S.; Gladfelter, A.; Mittag, T. Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. *Cell* **2019**, *176*, 419-434, doi:10.1016/j.cell.2018.12.035.
138. Holehouse, A.S.; Pappu, R.V. Functional Implications of Intracellular Phase Transitions. *Biochemistry-Us* **2018**, *57*, 2415-2423, doi:10.1021/acs.biochem.7b01136.
139. Brangwynne, C.P.; Tompa, P.; Pappu, R.V. Polymer physics of intracellular phase transitions. *Nature Physics* **2015**, *11*, 899-904, doi:10.1038/Nphys3532.
140. Mathieu, C.; Pappu, R.V.; Taylor, J.P. Beyond aggregation: Pathological phase transitions in neurodegenerative disease. *Science* **2020**, *370*, 56-60, doi:10.1126/science.abb8032.
141. Chen, J.; Brooks, C.L.; Khandogin, J. Recent advances in implicit solvent based methods for biomolecular simulations. *Curr Opin Struc Biol* **2008**, *18*, 140-148.
142. Haberthür, U.; Caflisch, A. FACTS: Fast analytical continuum treatment of solvation. *Journal of Computational Chemistry* **2008**, *29*, 701-715, doi:https://doi.org/10.1002/jcc.20832.
143. Hawkins, G.D.; Cramer, C.J.; Truhlar, D.G. Parametrized Models of Aqueous Free Energies of Solvation Based on Pairwise Descreening of Solute Atomic Charges from a Dielectric Medium. *The Journal of Physical Chemistry* **1996**, *100*, 19824-19839, doi:10.1021/jp961710n.
144. Onufriev, A.; Bashford, D.; Case, D.A. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins: Structure, Function, and Bioinformatics* **2004**, *55*, 383-394, doi:https://doi.org/10.1002/prot.20033.
145. Mongan, J.; Simmerling, C.; McCammon, J.A.; Case, D.A.; Onufriev, A. Generalized Born Model with a Simple, Robust Molecular Volume Correction. *Journal of Chemical Theory and Computation* **2007**, *3*, 156-169, doi:10.1021/ct600085e.
146. Nguyen, H.; Roe, D.R.; Simmerling, C. Improved Generalized Born Solvent Model Parameters for Protein Simulations. *Journal of Chemical Theory and Computation* **2013**, *9*, 2020-2034, doi:10.1021/ct3010485.

147. Gallicchio, E.; Levy, R.M. AGBNP: An analytic implicit solvent model suitable for molecular dynamics simulations and high-resolution modeling. *Journal of Computational Chemistry* **2004**, *25*, 479-499, doi:https://doi.org/10.1002/jcc.10400.
148. Gallicchio, E.; Paris, K.; Levy, R.M. The AGBNP2 Implicit Solvation Model. *Journal of Chemical Theory and Computation* **2009**, *5*, 2544-2564, doi:10.1021/ct900234u.
149. Im, W.; Lee, M.S.; Brooks Iii, C.L. Generalized born model with a simple smoothing function. *Journal of Computational Chemistry* **2003**, *24*, 1691-1702, doi:https://doi.org/10.1002/jcc.10321.
150. Lee, M.S.; Salsbury, F.R.; Brooks, C.L. Novel generalized Born methods. *The Journal of Chemical Physics* **2002**, *116*, 10606-10614, doi:10.1063/1.1480013.
151. Lee, M.S.; Feig, M.; Salsbury Jr, F.R.; Brooks Iii, C.L. New analytic approximation to the standard molecular volume definition and its application to generalized Born calculations. *Journal of Computational Chemistry* **2003**, *24*, 1348-1356, doi:https://doi.org/10.1002/jcc.10272.
152. Chen, J.; Im, W.; Brooks, C.L. Balancing solvation and intramolecular interactions: Toward a consistent generalized born force field. *Journal of the American Chemical Society* **2006**, *128*, 3728-3736.
153. Chen, J.; Brooks, C.L. Implicit modeling of nonpolar solvation for simulating protein folding and conformational transitions. *Physical Chemistry Chemical Physics* **2008**, *10*, 471-481.
154. Chen, J. Intrinsically disordered p53 extreme C-terminus binds to S100B(betabeta) through "fly-casting". *Journal of the American Chemical Society* **2009**, *131*, 2088-2089.
155. Wang, Y.; Fisher, J.C.; Mathew, R.; Ou, L.; Otieno, S.; Sublet, J.; Xiao, L.; Chen, J.; Roussel, M.F.; Kriwacki, R.W. Intrinsic disorder mediates the diverse regulatory functions of the Cdk inhibitor p21. *Nat Chem Biol* **2011**, *7*, 214-221, doi:10.1038/nchembio.536.
156. Ganguly, D.; Chen, J. Modulation of the disordered conformational ensembles of the p53 transactivation domain by cancer-associated mutations. *PLoS Comput Biol* **2015**, *11*, e1004247, doi:10.1371/journal.pcbi.1004247.
157. Nguyen, H.; Maier, J.; Huang, H.; Perrone, V.; Simmerling, C. Folding Simulations for Proteins with Diverse Topologies Are Accessible in Days with a Physics-Based Force Field and Implicit Solvent. *Journal of the American Chemical Society* **2014**, *136*, 13959-13962, doi:10.1021/ja5032776.
158. Maffucci, I.; Contini, A. An Updated Test of AMBER Force Fields and Implicit Solvent Models in Predicting the Secondary Structure of Helical, β-Hairpin, and Intrinsically Disordered Peptides. *Journal of Chemical Theory and Computation* **2016**, *12*, 714-727, doi:10.1021/acs.jctc.5b01211.
159. Tao, P.; Xiao, Y. Using the generalized Born surface area model to fold proteins yields more effective sampling while qualitatively preserving the folding landscape. *Physical Review E* **2020**, *101*, 062417, doi:10.1103/PhysRevE.101.062417.
160. Gong, X.; Chiricotto, M.; Liu, X.; Nordquist, E.; Feig, M.; Brooks Iii, C.L.; Chen, J. Accelerating the Generalized Born with Molecular Volume and Solvent Accessible Surface Area Implicit Solvent Model Using Graphics Processing Units. *Journal of Computational Chemistry* **2020**, *41*, 830-838, doi:https://doi.org/10.1002/jcc.26133.
161. Vitalis, A.; Pappu, R.V. ABSINTH: A new continuum solvation model for simulations of polypeptides in aqueous solutions. *Journal of Computational Chemistry* **2009**, *30*, 673-699, doi:https://doi.org/10.1002/jcc.21005.
162. Vitalis, A.; Caflisch, A. Micelle-Like Architecture of the Monomer Ensemble of Alzheimer's Amyloid-β Peptide in Aqueous Solution and Its Implications for Aβ Aggregation. *Journal of Molecular Biology* **2010**, *403*, 148-165, doi:https://doi.org/10.1016/j.jmb.2010.08.003.

163. Mittal, A.; Holehouse, A.S.; Cohan, M.C.; Pappu, R.V. Sequence-to-Conformation Relationships of Disordered Regions Tethered to Folded Domains of Proteins. *J Mol Biol* **2018**, *430*, 2403-2421, doi:10.1016/j.jmb.2018.05.012.
164. Choi, J.-M.; Pappu, R.V. Improvements to the ABSINTH Force Field for Proteins Based on Experimentally Derived Amino Acid Specific Backbone Conformational Statistics. *Journal of Chemical Theory and Computation* **2019**, *15*, 1367-1382, doi:10.1021/acs.jctc.8b00573.
165. Pak, A.J.; Voth, G.A. Advances in coarse-grained modeling of macromolecular complexes. *Curr. Opin. Struct. Biol.* **2018**, *52*, 119-126, doi:10.1016/j.sbi.2018.11.005.
166. Kmiecik, S.; Gront, D.; Kolinski, M.; Wieteska, L.; Dawid, A.E.; Kolinski, A. Coarse-Grained Protein Models and Their Applications. *Chem. Rev.* **2016**, *116*, 7898-7936, doi:10.1021/acs.chemrev.6b00163.
167. Wolynes, P.G. Recent successes of the energy landscape theory of protein folding and function. *Q. Rev. Biophys.* **2005**, *38*, 405-410, doi:10.1017/s0033583505004075.
168. Hills, R.D.; Brooks, C.L. Insights from Coarse-Grained Gō Models for Protein Folding and Dynamics. *International Journal of Molecular Sciences* **2009**, *10*, doi:10.3390/ijms10030889.
169. Law, S.M.; Gagnon, J.K.; Mapp, A.K.; Brooks, C.L., 3rd. Prepaying the entropic cost for allosteric regulation in KIX. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 12067-12072, doi:10.1073/pnas.1405831111.
170. Chu, X.; Wang, J. Position-, disorder-, and salt-dependent diffusion in binding-coupled-folding of intrinsically disordered proteins. *Phys. Chem. Chem. Phys.* **2019**, *21*, 5634-5645, doi:10.1039/c8cp06803h.
171. Ganguly, D.; Otieno, S.; Waddell, B.; Iconaru, L.; Kriwacki, R.W.; Chen, J. Electrostatically Accelerated Coupled Binding and Folding of Intrinsically Disordered Proteins. *J. Mol. Biol.* **2012**, *422*, 674-684, doi:10.1016/j.jmb.2012.06.019.
172. Ganguly, D.; Zhang, W.; Chen, J. Electrostatically Accelerated Encounter and Folding for Facile Recognition of Intrinsically Disordered Proteins. *PLoS Comput. Biol.* **2013**, *9*, e1003363.
173. Liu, X.; Chen, J.; Chen, J. Residual Structure Accelerates Binding of Intrinsically Disordered ACTR by Promoting Efficient Folding upon Encounter. *J. Mol. Biol.* **2019**, *431*, 422-432, doi:10.1016/j.jmb.2018.12.001.
174. Liu, Z.R.; Huang, Y.Q. Advantages of proteins being disordered. *Protein Sci.* **2014**, *23*, 539-550, doi:Doi 10.1002/Pro.2443.
175. Ganguly, D.; Chen, J. Topology-based modeling of intrinsically disordered proteins: balancing intrinsic folding and intermolecular interactions. *Proteins* **2011**, *79*, 1251-1266, doi:10.1002/prot.22960.
176. Chu, W.T.; Shammas, S.L.; Wang, J. Charge Interactions Modulate the Encounter Complex Ensemble of Two Differently Charged Disordered Protein Partners of KIX. *J. Chem. Theory Comput.* **2020**, *16*, 3856-3868, doi:10.1021/acs.jctc.9b01264.
177. Baul, U.; Chakraborty, D.; Mugnai, M.L.; Straub, J.E.; Thirumalai, D. Sequence Effects on Size, Shape, and Structural Heterogeneity in Intrinsically Disordered Proteins. *The Journal of Physical Chemistry B* **2019**, *123*, 3462-3474, doi:10.1021/acs.jpcb.9b02575.
178. Liu, X.; Chen, J. HyRes: a coarse-grained model for multi-scale enhanced sampling of disordered protein conformations. *Physical Chemistry Chemical Physics* **2017**, *19*, 32421-32432, doi:10.1039/C7CP06736D.
179. Wu, H.; Wolynes, P.G.; Papoian, G.A. AWSEM-IDP: A Coarse-Grained Force Field for Intrinsically Disordered Proteins. *The Journal of Physical Chemistry B* **2018**, *122*, 11115-11125, doi:10.1021/acs.jpcb.8b05791.
180. Ashbaugh, H.S.; Hatch, H.W. Natively Unfolded Protein Stability as a Coil-to-Globule Transition in Charge/Hydropathy Space. *Journal of the American Chemical Society* **2008**, *130*, 9536-9542, doi:10.1021/ja802124e.

181. Kim, Y.C.; Hummer, G. Coarse-grained Models for Simulations of Multiprotein Complexes: Application to Ubiquitin Binding. *Journal of Molecular Biology* **2008**, *375*, 1416-1433, doi:https://doi.org/10.1016/j.jmb.2007.11.063.
182. Dignon, G.L.; Zheng, W.; Best, R.B.; Kim, Y.C.; Mittal, J. Relation between single-molecule properties and phase behavior of intrinsically disordered proteins. *Proceedings of the National Academy of Sciences* **2018**, *115*, 9929, doi:10.1073/pnas.1804177115.
183. Latham, A.P.; Zhang, B. Maximum Entropy Optimized Force Field for Intrinsically Disordered Proteins. *Journal of Chemical Theory and Computation* **2020**, *16*, 773-781, doi:10.1021/acs.jctc.9b00932.