Computer Simulations of Intrinsically Disordered Proteins

Song-Ho Chong, Prathit Chatterjee, and Sihyun Ham

Department of Chemistry, Sookmyung Women's University, Yongsan-Ku, Seoul 04310, Korea; email: sihyun@sookmyung.ac.kr

Annu. Rev. Phys. Chem. 2017. 68:117–34

First published online as a Review in Advance on February 6, 2017

The *Annual Review of Physical Chemistry* is online at physchem.annualreviews.org

https://doi.org/10.1146/annurev-physchem-052516-050843

Keywords

molecular dynamics, force field, statistical thermodynamics, configurational entropy, correlation entropy

Abstract

The investigation of intrinsically disordered proteins (IDPs) is a new frontier in structural and molecular biology that requires a new paradigm to connect structural disorder to function. Molecular dynamics simulations and statistical thermodynamics potentially offer ideal tools for atomic-level characterizations and thermodynamic descriptions of this fascinating class of proteins that will complement experimental studies. However, IDPs display sensitivity to inaccuracies in the underlying molecular mechanics force fields. Thus, achieving an accurate structural characterization of IDPs via simulations is a challenge. It is also daunting to perform a configuration-space integration over heterogeneous structural ensembles sampled by IDPs to extract, in particular, protein configurational entropy. In this review, we summarize recent efforts devoted to the development of force fields and the critical evaluations of their performance when applied to IDPs. We also survey recent advances in computational methods for protein configurational entropy that aim to provide a thermodynamic link between structural disorder and protein activity.

1. INTRODUCTION

There is increasing interest in intrinsically disordered proteins (IDPs). These proteins are fully functional yet lack well-defined three-dimensional structures, thereby breaking the conventional rigid rule of the structure–function paradigm (1). Fully or partially disordered proteins are abundant in eukaryotes; in particular, $\sim$50% of the sequences coded by the human genome are predicted to comprise disordered segments of $>$30 amino acids (2, 3). IDPs play a crucial role in gene regulation, signal transduction, and biomolecular recognition (4, 5). Conformational disorder is an essential structural ingredient of IDPs, which enables them to bind with multiple partners with high specificity but modest affinity (6, 7). IDPs also frequently serve as a hub in protein–protein interaction networks (8), and they are associated with a variety of human diseases such as cancer, diabetes, and neurodegenerative disorders (9–11). Their critical roles in cellular functions and networks as well as their association with various human diseases make IDPs attractive therapeutic targets. Thus, IDPs constitute a fascinating class of proteins whose investigation may not only offer new paradigms for how proteins function through disorder, but also facilitate the development of novel drug molecules to modulate protein–protein interactions.

Because of the absence of a single dominant structure, the structural features of IDPs must be characterized with an ensemble of interconverting conformations (12, 13). This poses a challenge for experimental methods that normally measure time- and space-averaged properties and, hence, have difficulty capturing inherently transient conformational order/disorder (14, 15). In contrast, computer simulations produce a time sequence of atomic-level configurations and offer a potentially powerful complement to experiments to elucidate the key conformational characteristics of IDPs. Indeed, atomistic simulations have been adopted to elucidate the inherent flexibility and heterogeneous ensemble of IDPs (16–20). However, achieving an accurate structural characterization of IDPs via simulations is challenging because simulation results crucially rely on the accuracy of the underlying potential energy functions or force fields. Indeed, protein force fields were developed mainly to target folded globular proteins, and their applicability to IDPs is not obvious. In Section 2, we survey recent efforts devoted to the development of force fields and critical evaluations of their performance when applied to IDPs.

Structural investigation alone is often insufficient to rationalize protein activity (21). Indeed, protein configurational entropy is receiving growing attention as a major factor that controls the activity of IDPs associated with a number of cellular functions (22–24). Thus, understanding the relationship between the configurational entropy and the degree of conformational disorder, as well as its variation upon conformational change and binding with partner(s), is of fundamental importance. Doing so entails fully characterizing the protein free-energy landscape because configurational entropy measures how much configuration space is accessible to a protein's internal degrees of freedom. However, this is a daunting task, in particular for IDPs, because configuration-space integration must be performed over the heterogeneous structural ensembles sampled by IDPs. Thus, certain approximations are inevitably introduced to evaluate this important thermodynamic parameter. Even though the quasi-harmonic approximation (25–29) has been the most popular approach to compute protein configurational entropy based on atomistic simulations, it does not capture the intricate features of the underlying landscape, such as the presence of multiple minima and the correlation effects between conformational coordinates. A significant effort has hence been put forth to go beyond the quasi-harmonic approximation, and Section 3 is devoted to a survey of recent developments in computational methods, particularly focusing on those that enable exploration of the thermodynamic descriptions of conformational disorder.

2. CHARACTERIZING CONFORMATIONAL DISORDER

In this section, we first provide an overview of the features of representative biomolecular force fields employed in atomistic molecular dynamics simulations (Section 2.1). We then outline recent efforts devoted to refining those force fields (Section 2.2) and to developing new water models (Section 2.3) that better capture the structural characteristics of IDPs. Finally, we survey recent critical evaluations of the performance of the force fields in simulations of IDPs (Section 2.4).

2.1. Biomolecular Force Fields

In principle, quantum mechanical calculations provide the complete potential energy surface of molecular systems as a function of the constituent atoms' coordinates. However, this is not practically feasible for complex macromolecules such as proteins, particularly when they are impacted by aqueous environments. Therefore, it is customary to employ molecular mechanics force fields, which refer to empirical functional forms and parameter sets, to calculate the potential energy of biomolecular systems (30, 31). Functional forms of potential energy can generally be written as $E_{\mathrm{total}} = E_{\mathrm{bonded}} + E_{\mathrm{nonbonded}}$. The bonded term ($E_{\mathrm{bonded}}$) typically consists of bond, angle, and dihedral-angle potentials that describe the interactions of the atoms linked by covalent bonds. The nonbonded term ($E_{\mathrm{nonbonded}}$) includes noncovalent van der Waals and electrostatic interactions. Parameter sets within empirical functional forms are determined by quantum mechanical calculations for small related systems (e.g., short peptides) and/or through fitting procedures to experimental observables.

Various force fields have been developed to describe biomolecules in aqueous environments; the representative force fields are Amber [ff99 (32), ff99SB (33), ff03 (34)], CHARMM (Chemistry at Harvard Molecular Mechanics) [CHARMM22 (35), CHARMM22/CMAP (grid-based energy correction maps) correction (36)], GROMOS96 (Groningen Molecular Simulation) [43a1 (37), 53a6 (38), 54a7 (39)], and OPLS (Optimized Potentials for Liquid Simulations) (40) (**Table 1**). However, because of their empirical and approximate nature, they exhibit certain problems, such as variations in structural propensities and compactness. For example, Amber ff99 (32) tends to overestimate the α-helical structures (33, 41), whereas Amber ff99SB (33) underestimates those structures (42); the CHARMM22/CMAP correction (36) favors helical structures (43, 44); GROMOS96 (43a1) (37) displays a tendency to form β-sheet structures (41, 45); and OPLS yields a better balance between helical and extended conformations (45).

The secondary structural propensities exhibited by different force fields impact IDP simulations. For example, in the simulation studies of amyloid-β protein dimers (an IDP implicated in Alzheimer's disease), use of the GROMOS96 (53a6) force field has led to a much higher average β-sheet content in the dimer structure (46) than that obtained with CHARMM and OPLS force fields (47–49). A similar preference for secondary structures, depending on the force fields, was observed in a simulation study on amylin, an IDP implicated in type 2 diabetes (50): GROMOS96 (53a6) tends to predict β-hairpins, CHARMM22/CMAP generates overly α-helical structures, and OPLS favors disordered structures. Accordingly, good balance among helical, strand, and coil structures is needed in force field developments.

2.2. Recent Improvements

To achieve balance among the secondary structures, substantial effort has been invested in improving force fields, with primary focus on modifying the backbone and side-chain dihedral-angle potentials (see **Table 1**). Representative force fields include Amber ff99SB* (42), ff99SB-ILDN (51),

Table 1 Representative biomolecular force fields and their default water model

Force fields	Parameter sets	Developments (modifications)	Water model	Reference	Villin[a]	WW[a]
Amber	ff99	Amber base parameter set	TIP3P	32	NA	NA
	ff99SB	Improved backbone torsional parameters	TIP3P	33	NA	NA
	ff99SB*	Corrections to backbone energy terms	TIP3P	42	NA	NA
	ff99SB-ILDN	Improved side-chain torsion potentials	TIP3P	51	✓	✓
	ff99SB*-ILDN	ff99SB* + ILDN modifications	TIP3P	53	✓	✓
	ff99SB-ILDN-phi	Modifications to backbone ϕ angles	TIP4P-Ew	56	NA	NA
	ff99SB-ILDN-NMR	Modifications to backbone dihedrals based on NMR chemical shifts	TIP4P-Ew	57	NA	NA
	ff03	Another Amber base parameter set	TIP3P	34	✓	✗
	ff03*	Corrections to backbone energy terms	TIP3P	42	✓	✓
	ff03w	Corrections to backbone torsion potentials with improved water model	TIP4P/2005	52	NA	NA
	ff03ws	Modified short-range protein–water interaction potential ($\lambda = 1.10$)	TIP4P/2005	73	NA	NA
CHARMM	CHARMM22	CHARMM base parameter set	Modified TIP3P	35	✗	✗
	CHARMM22/ CMAP	CMAP backbone corrections	Modified TIP3P	36	✓	✗
	CHARMM22*	Corrections to backbone energy terms	Modified TIP3P	53	✓	✓
	CHARMM36	Modifications to backbone and side-chain torsion potentials	Modified TIP3P	54	NA	NA
GROMOS	GROMOS96 (43a1)	All-atom GROMOS parameter set	SPC	37	NA	NA
	GROMOS96 (53a6)	Accurate reproduction of hydration thermodynamics	SPC	38	NA	NA
	GROMOS96 (54a7)	Improvement in torsional potentials and hydration free energy	SPC	39	NA	NA
OPLS	OPLS-AA	All-atom OPLS parameter set	No default model	40	✗	✓

Abbreviations: CHARMM, Chemistry at Harvard Molecular Mechanics; CMAP, grid-based energy correction map; GROMOS, Groningen Molecular Simulation; NA, not available; NMR, nuclear magnetic resonance; OPLS, Optimized Potentials for Liquid Simulations; AA, all atom; SPC, simple point charge; TIP3P, three-site transferable intermolecular potential.

[a]A check mark (✓) indicates that simulations initiated from the unfolded state reached the folded state in 10 μs (villin headpiece subdomain) and 50 μs (WW domain), whereas a cross mark (✗) indicates that the folding did not occur within the respective simulation times.

their combination ff99SB*-ILDN, ff03* (42), ff03w (52), CHARMM22* (53), and CHARMM36 (54); other variants have also been developed (55–57). Systematic force field comparison studies have shown that, overall, force field modifications tend to show improvements that are relatively stable for different types of secondary structures (58–60). The ability of force fields to fold small α-helical (villin headpiece subdomain) and β-sheet (WW domain) proteins has also been tested (59), demonstrating a preference for ff99SB*-ILDN and CHARMM22* (**Table 1**).

Thus, several simulation studies provide remarkably accurate characterization of ordered folded protein states. Nevertheless, the unfolded states observed in simulations exhibit certain discrepancies. For example, the CHARMM22/CMAP force field, which provides an excellent description of folded protein states, generates unfolded states that are substantially more helical than those found experimentally (59). Furthermore, simulations of proteins larger than 20–30 amino acids tend to produce unfolded states that are more compact and structured than those suggested experimentally (44, 61). For example, Amber generally generates more compact unfolded states than does CHARMM (44). Several other studies have observed structures that were too compact, contained substantial secondary structures, and exaggerated the intramolecular hydrogen bonding networks of unfolded proteins (62–64).

2.3. Importance of Protein–Water Interactions

The choice of solvent model is often significant when quantitatively characterizing biomolecules in aqueous environments. In particular, compared with those of folded globular proteins, the structural properties of IDPs are more sensitive to protein–water interactions, as IDPs are more solvent-exposed. Therefore, validating the use of a particular water model with its corresponding force field(s) is necessary. To date, three-site models, such as TIP3P (three-site transferable intermolecular potential) (65) and SPC (simple point charge) (66), have been employed most widely (**Table 2**). Indeed, TIP3P and its slightly modified version are the default solvent models in the Amber and CHARMM force fields, and the SPC model is usually combined with the GROMOS force field. Four-site water models, such as TIP4P (67) and its additional modifications, TIP4P/2005 (68) and TIP4P-Ew (69), have also been developed in recent years to reproduce the structural, dynamical, and thermodynamic properties of water for better comparison with experiments (**Table 2**).

In fact, simulation studies of short disordered peptides demonstrate that adopting more refined water models yields more accurate conformational ensembles (52, 56, 70, 71). For example, combining the ff03w force field with TIP4P/2005 generates more realistic unfolded-state conformations than are produced using TIP3P water (52). In addition, a study of the amyloid-β_{21-30} peptide reports that combining Amber ff99SB with the TIP4P-Ew model, rather than the TIP3P model, provides better predictions for nuclear magnetic resonance (NMR) observables (70). This was also demonstrated for the full-length 42-residue amyloid-β protein in another study (72), which compared combinations of Amber ff99SB with TIP3P and with TIP4P-Ew. The ff99SB/TIP4P-Ew

Table 2 Representative water models

Parameters	TIP3P (65)	SPC (66)	TIP4P (67)	TIP4P-Ew (69)	TIP4P/2005 (68)	TIP4P-D (74)
r(OH), Å	0.9572	1.0	0.9572	0.9572	0.9572	0.9572
HOH, deg	104.52	109.47	104.52	104.52	104.52	104.52
r(OM), Å[a]	NA	NA	0.15	0.125	0.1546	0.1546
$A \times 10^{-3}$, (kcal · Å^{12})/mol[b]	582.0	629.4	600.0	656.1	731.3	904.7
B, (kcal · Å^{6})/mol[b]	595.0	625.5	610.0	653.5	736.0	900.0
q(O) or q(M)[c]	−0.834	−0.82	−1.04	−1.04844	−1.1128	−1.16
q(H)[c]	+0.417	+0.41	+0.52	+0.52422	+0.5564	+0.58

Corresponding references are provided in parentheses. Abbreviations: NA, not available; SPC, simple point charge; TIP, transferable intermolecular potential.

[a]M refers to the dummy atom in the four-site water models that is located near the oxygen along the bisector of the HOH angle.

[b]A and B are the parameters of the Lennard-Jones potential when it is represented by $A/r_{OO}^{12} - B/r_{OO}^{6}$, with r_{OO} denoting the oxygen–oxygen distance.

[c]Partial charges are in units of the electron charge.

combination showed stronger protein–water interactions with TIP4P-Ew than with TIP3P (72), thereby providing more extended protein conformations and yielding residue-resolved secondary structure contents in better agreement with NMR analysis.

As noted above, unfolded or disordered states are predicted to be too compact relative to experiments by current biomolecular force fields (44), implying that these force fields insufficiently expose proteins to water. Two approaches have been proposed to address this problem (73, 74). In one approach, the simplest possible change was introduced. The depth of the Lennard-Jones potential between the atoms in the protein and the oxygen atom of water is scaled by a factor of 1.1, thus leaving the water–water and protein–protein interactions untouched (73); the resulting force field is termed ff03ws (**Table 1**). Such minor strengthening of the protein–water interaction suffices to reproduce experimentally measured chain sizes of disordered and unfolded proteins. In another approach (74), a new TIP4P-D water model was introduced by modifying parameters in the TIP4P model (see **Table 2**) to correct for the deficiencies in water dispersion interactions. This new model yielded disordered-state protein structures that are more expanded and in better agreement with experiment than those obtained with traditional water models (74).

A natural concern here is whether and to what extent such alternations in the protein–water interaction or in the water model affect the folded-state characteristics, which already provide acceptable results without such modifications. Interestingly, strengthening the protein–water interactions in the ff03ws force field did not significantly influence protein folded states, despite marginal changes to the stability of the helical and sheet structures and larger amplitude dynamics exhibited by the loop regions (73). In contrast, the TIP4P-D water model somewhat destabilizes the folded states: The native states of the protein villin headpiece subdomain and WW domain near the melting temperature were destabilized by ∼2 kcal/mol in TIP4P-D versus TIP3P (74).

2.4. Case Studies and Further Necessary Improvements

Here, we provide a brief overview of IDP case studies, on the basis of which we indicate a need for further improvements to force fields to better capture the structural characteristics of IDPs. Earlier studies reported that atomistic simulations using state-of-the-art force fields yield conformational ensembles of IDPs that are in good agreement with various experimental observables (19, 75–80). However, the applicability of recent force fields has been called into question. For instance, a simulation study of Histatin 5, a 24-residue cationic salivary IDP with antimicrobial and antifungal properties, demonstrates that recent force fields (Amber ff99SB-ILDN, ff99SBNMR1-ILDN, GROMOS 53a6 and 54a7) are equally inappropriate for reproducing the experimental small-angle X-ray form factor (81). Indeed, overly compact conformational ensembles were generated from these force fields, and it was necessary to alter the protein–water interaction (thus adopting the parameters of ff03ws) (see **Table 1**) to obtain simulation results in agreement with experiments. Moreover, systematic simulation studies of IDPs for different force fields also indicated surprisingly large differences in the hydrogen bonding patterns, chain dimensions, and secondary structural contents (50, 77, 82).

Furthermore, there are contradictory results regarding the most/least accurate force fields for simulating IDPs (82, 83). In a simulation study of the disordered 24-residue arginine/serine peptide (82), IDP ensembles generated by several atomistic force fields (Amber ff99SB*-ILDN, ff03w, ff03ws, CHARMM22*, CHARMM36, OPLS) were compared against small-angle X-ray scattering and NMR data. The conformational ensemble obtained using CHARMM 22* agreed best with all available experimental data. In a separate study (83), unstructured peptides with sequence EGAAXAASS (X = G, W, I, D, and V) were investigated using Amber ff99SB*-ILDN, ff03w, and CHARMM22*, and the results were compared with those obtained via NMR spectroscopy. Here, simulations with CHARMM22* provided the poorest agreement with experimental measurements, whereas ff03w yielded the best agreement. Thus, two independent studies show CHARMM22* is in both agreement and disagreement with experimental data.

At present, owing to the somewhat inconsistent findings reported, there is no definite consensus on the most accurate force field for carrying out IDP simulations. Thus, force field parameters, including those of water models, need to be further improved. More recently, investigators proposed a new force field, termed ff99IDPs, that specifically targets IDPs. In ff99IDPs, CMAPs were added to the backbone dihedral-angle potentials of disorder-promoting residues (84). Compared with ff99SB-ILDN, this force field yielded results closer to experimental measurements for three representative IDP systems (α-synuclein, aspartic proteinase inhibitor IA_3, and arginine-rich HIV-1 Rev) (85). Furthermore, ff99IDPs maintains the secondary structure in ordered protein regions, indicating the importance of taking into account IDP structures during general-purpose force field development.

3. THERMODYNAMIC DESCRIPTION OF CONFORMATIONAL DISORDER

Quantitative measures of conformational disorder are of fundamental importance for elucidating the thermodynamic driving forces and molecular mechanisms by which IDPs perform their functions. To that end, protein configurational entropy—associated with a protein's internal degrees of freedom—is a potentially relevant thermodynamic parameter, and deriving its computation from atomistic simulations is among the central problems in physical chemistry. Here, we review recent developments in statistical thermodynamic methods for estimating this important quantity, particularly focusing on methods and their applications that aim to provide thermodynamic descriptions of conformational order/disorder.

3.1. Protein Configurational Entropy

Computing configurational entropy is key because this factor is central in determining protein stability. It is also an important constituent of protein–ligand and protein–protein binding affinities. However, configurational entropy is also the most difficult thermodynamic quantity to estimate. Therefore, significant effort has been devoted to developing appropriate computational methods (86–89). The relevance of configurational entropy in computational drug design is also receiving increased interest (90, 91).

The configurational entropy of a molecule is defined by

$$S_{\mathrm{config}} = -k_{\mathrm{B}} \int \mathrm{d}\mathbf{q}\, p(\mathbf{q}) \log p(\mathbf{q}), \qquad 1.$$

i.e., by the integration of the multidimensional ($3N-6$ dimensional when the number of constituent atoms is N) probability distribution function $p(\mathbf{q})$ over a molecule's internal degrees of freedom $\mathbf{q}$. Both Cartesian and bond-angle-torsion (BAT) internal coordinates can be adopted to represent the configuration vector $\mathbf{q}$. In the latter, the appropriate Jacobian, here omitted, needs to be included. Because the Jacobian depends only on the bond lengths and angles, which are rather rigid (92), it can reasonably be neglected in computing unimolecular entropy change. However, the Jacobian associated with the external coordinates, which leads to external entropy, must be considered for binding entropy (93, 94). Accurately estimating the full probability distribution function $p(\mathbf{q})$ from finite samples generated by simulations and performing the high-dimensional configuration integral over $\mathbf{q}$ for complex biomolecules such as proteins are formidable tasks. Therefore, researchers inevitably introduce certain simplifying approximations.

Most often, protein configurational entropy is determined using the quasi-harmonic method that assumes a multivariate Gaussian distribution for the probability distribution function $p(\mathbf{q})$. This can be done with both Cartesian (27–29) and BAT internal coordinates (25, 26). In this method, the variance of the $3N - 6$ distributions of principal coordinates is computed on the basis of the mass-weighted covariance matrix of the internal coordinates, which is then used to estimate configurational entropy.

The quasi-harmonic approach has been widely used because it requires only the covariance matrix from the simulations as input. This method is also being applied to proteins that include inherently flexible regions such as calmodulin (95), a calcium-binding messenger protein regulating diverse target proteins (22). It was shown that the computed configurational entropy changes that occur upon binding with various target peptides correlate reasonably well with experimental measurements (22).

3.2. Beyond the Quasi-Harmonic Approximation

A major drawback of the quasi-harmonic method is its lack of accuracy for systems possessing a multiple-occupied free-energy landscape (96, 97). This severely limits its applicability to proteins because multiple local wells are generally present in the protein free-energy landscape, in particular to IDPs whose conformational transitions among multiple minima are essential for their functions. Several theoretical tools (98–112) have been developed to improve the underlying basic assumptions, i.e., (*a*) assuming that the probability distributions of coordinates (including collective coordinates such as principal coordinates) are independent and (*b*) assuming the Gaussian functional form of the probability distribution along each independent coordinate. In the following, we survey some of these theoretical developments.

To avoid any assumption about the shape of the probability distribution function $p(\boldsymbol{q})$, nonparametric methods have been proposed to estimate $p(\mathbf{q})$ from finite samples generated by simulations. Building histograms in bins of some fixed size along each direction of the configuration vector $\mathbf{q}$ is the most commonly used method (100). However, care must be taken to avoid a possible bin-size dependence (89). More recently, a nonparametric method was introduced in which the probability distribution function $p(\mathbf{q})$ is estimated in terms of the nearest-neighbor distances between the sample points (101). This approach is a variant of the histogram method, in which a sample-point-centered histogram is constructed and the bin size is adjusted so the resulting configurational entropy is unbiased in the asymptotic limit of a large sample size. However, the computational complexity of this method increases markedly with the dimensionality of the configuration space; therefore, its applicability is limited to relatively small molecular systems. To overcome this difficulty, an adaptive kernel density estimation has been developed that extends the applicability of this method to a configuration space of higher dimensions (103).

Another focus has been to incorporate the correlation effects between the coordinates that the quasi-harmonic method did not take into account (**Figure 1*a*,*b***). Mutual information represents the correlation effects on configurational entropy—termed correlation entropy (113). For example, the pairwise mutual information $I_2(q_1, q_2)$ between two coordinates q_1 and q_2 is defined as $I_2(q_1, q_2) = S_1(q_1) + S_2(q_2) - S_2(q_1, q_2)$ in terms of the marginal entropies $S_1(q_1)$ and $S_2(q_2)$ and the joint entropy $S_2(q_1, q_2)$; $I_2(q_1, q_2)$ is nonnegative and becomes zero only if q_1 and q_2 are independent. Higher-order mutual information involving more than two coordinates can also be introduced (113). To incorporate those correlations systematically, a mutual information expansion (MIE) has been proposed (104, 105). In this expansion, configurational entropy is expressed as a series of mutual information terms representing successively higher-order correlations among the conformational coordinates. Although MIE can be formally derived to the full order and

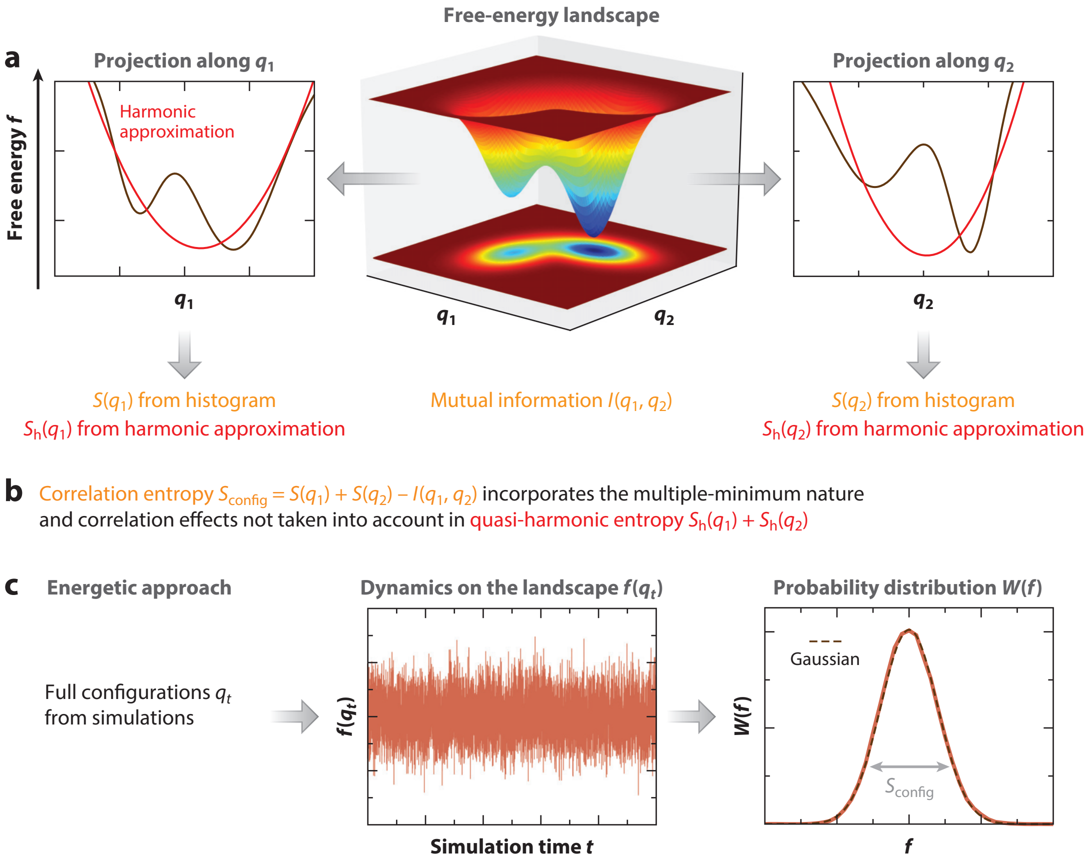

Figure 1

(*a*) Minimal model of the free-energy landscape consisting of just two local minima and their projections along coordinates q_1 (*left panel*) and q_2 (*right panel*); red curves denote the harmonic approximation. Marginal entropies $S(q_1)$ and $S(q_2)$, their harmonic approximations $S_h(q_1)$ and $S_h(q_2)$, and mutual information $I(q_1, q_2)$ can be obtained from these plots. (*b*) Correlation entropy incorporates the multiple-minimum nature of this landscape through $S(q_1)$ and $S(q_2)$ and the correlation effects via $I(q_1, q_2)$, which are not taken into account in $S_h(q_1) + S_h(q_2)$ entropy under the quasi-harmonic approximation. (*c*) In the energetic approach, time-dependent full-dimensional configuration vectors ($\mathbf{q}_t$) are first taken from simulations. For each configuration, effective energy $f(\mathbf{q}_t)$ is computed, from which the distribution function $W(f)$ is constructed. When $W(f)$ is well approximated by the Gaussian distribution (*dashed curve*), the configurational entropy can be estimated from the width of $W(f)$.

more accurate entropy estimation is achieved by including increasingly higher-order terms, terms higher than second order are usually neglected because of their high computational cost. Even the computation of the pairwise contributions can be challenging for large systems such as proteins.

More efficient computational methods for correlation entropy have therefore been developed, for example, the maximum information spanning tree (MIST) method (106, 107). MIST is another systematic mutual-information expansion, but unlike MIE (in which all the correlations between coordinates are incorporated), the expansion in MIST identifies dominant couplings between coordinates and, thus, considers only a subset of correlations. For this reason, MIST is computationally more efficient than MIE, which is an important advantage in handling large systems.

The minimally coupled subspace approach has also been proposed to efficiently incorporate correlation entropy (103). This method avoids directly applying MIE to large systems. Instead, highly coupled degrees of freedom are first clustered into minimally coupled low-dimensional ($\sim$15) clusters; then, MIE is applied. As a result, this method is applicable to large macromolecules including proteins.

3.3. Correlation Entropy and Conformational Order/Disorder

Correlation entropy is a crucial quantity for relating configurational entropy to conformational order/disorder. This is understood, e.g., by considering the change in configurational entropy upon secondary structure formation from a disordered conformation. Because each secondary structure, such as an α-helix or β-sheet, can be characterized as a special region in the Ramachandran (ϕ, ψ) plot, its formation can be viewed as a restriction imposed between the backbone dihedral angles ϕ and ψ. Hence, it appears as a negative correlation entropy contribution to configurational entropy. Thus, correlation entropy serves as an invaluable thermodynamic parameter that quantifies the conformational characteristics.

Correlation effects on the loss of configurational entropy upon folding originating from backbone dihedral angles, side-chain dihedral angles, and couplings between backbone and side-chain angles have been addressed via molecular dynamics simulations (114). Configurational entropy and contributions from correlations were computed for native- and denatured-state ensembles of ubiquitin. Standard molecular dynamics simulations were used to generate the native-state ensemble. In contrast, restrained simulations were used to yield the denatured-state ensemble so that the ensemble's radius of gyration and NMR parameters agree with experimental data. This was necessary because unrestrained simulations tend to generate excessive collapsed denatured-state structures, as mentioned in Section 2.2, which is indeed the case for ubiquitin (115). Backbone entropy largely accounts for the change in configurational entropy upon folding, and α-helix formation provides a more negative contribution to backbone entropy than does β-sheet formation.

The connection between correlation entropy and NMR order parameters characterizing the site-resolved motional disorder is also of particular interest (116). Indeed, the microscopic origins of the empirical relationships between configurational entropies and NMR order parameters (21, 22, 117) have been explored via molecular dynamics simulations (118). In this study (118), seven proteins ranging from quite rigid to internally flexible were investigated. The configurational entropy of each side-chain methyl group was computed by taking into account the correlated motions of the side chains. The computed entropy was then compared with the simulated order parameter of the corresponding methyl side chain. A significant correlation was observed between these two quantities, and this held even when all the data from different proteins were simultaneously considered (118), suggesting a universal relation between site-resolved conformational dynamics (disorder) and configurational entropy.

3.4. Approaches Based on Nonstructural Variables

Researchers have also developed approaches to determine configurational entropy that focus on physical quantities other than structural variables such as Cartesian and BAT coordinates. These methods have conceptual similarity to the quasi-harmonic method because they are based on Gaussian statistics of the variables of interest, but they effectively take into account nonharmonic coordinate distributions. For example, a method has been proposed to estimate configurational entropy from atomic forces computed by molecular dynamics simulations (119). Similar to the quasi-harmonic method, harmonic approximation is employed in this method, but using the mass-weighted force covariance matrix. Using forces instead of coordinates is advantageous because force distributions are highly harmonic (120) and forces capture atomic correlations more directly, thereby overcoming the inherent limitation of the quasi-harmonic method. The force-based method is also more efficient and accurate than the quasi-harmonic method, making it an attractive method for computing protein configurational entropy.

A new computational approach that focuses on energy has also been developed to determine configurational entropy (121–123) (**Figure 1*c***). In this approach, configurational entropy is expressed in terms of the canonical configuration integral Z as

$$TS_{\text{config}} = \langle f \rangle + k_{\text{B}} T \log Z, \quad Z = \int \mathrm{d}\mathbf{q}\; e^{-\beta f(\mathbf{q})}, \tag{2.}$$

which follows from Equation 1 by recognizing that $\rho(\mathbf{q}) = e^{-\beta f(\mathbf{q})}/Z$ (121). Here, $f = E_{\text{u}} + G_{\text{solv}}$ comprises the solute energy (E_{u}) and the solvation free energy (G_{solv}). The function f, also called the effective energy, is the genuine identity defining the free-energy landscape (124). The key point in this approach is to introduce the distribution function $W(f)$ of the effective energy f (hence, the energetic approach), with which Equation 2 can be rewritten in a useful form (for a detailed discussion, see 122, 123). When $W(f)$ is close to a Gaussian distribution, this approach yields $TS_{\text{config}} = (\beta/2)\ \overline{\delta f^2}$ in terms of the variance $\overline{\delta f^2}$ of the effective energy. $W(f)$ obeys Gaussian statics even when the underlying free-energy landscape exhibits multiple minima (97), and this holds for numerous systems, including IDPs (121–123, 125), because of the central limit theorem.

Based on the energetic approach, a possible connection between the configurational entropies and residual structures of IDPs has been addressed (125). For this purpose, the wild-type 42-residue amyloid-β protein and its five familial mutants and two synthetic mutants were studied. To elucidate a link between the structural features and entropy, the configurational entropies of these proteins were modeled using the amounts of helical structures, sheet structures, and salt bridges, which were obtained from respective molecular dynamics simulations. A significant correlation was observed between computed and modeled configurational entropy, indicating an intimate link between conformational order/disorder and configurational entropy.

3.5. Conformational Versus Vibrational Entropies

A more ambitious challenge is to separate protein configurational entropy (S_{config}) into conformational (S_{conf}) and vibrational (S_{vib}) components, which are respectively associated with the number of accessible free-energy wells and the average width of the individual wells of the protein free-energy landscape (126). (Although the terms configurational entropy and conformational entropy are often used interchangeably in the literature, conformational entropy is considered as a subcategory of configurational entropy in this review.) This partitioning of configurational entropy is formally exact because the entropy contributions from the high-free-energy regions that separate the individual wells can be neglected (127). Such separation enables characterization of the modulations of the free-energy landscape caused, e.g., by ligand binding and posttranslational modification in simple terms. It will also be of great practical value in elucidating the molecular mechanisms that underlie protein activities.

This partitioning scheme also serves as a guide to develop computational methods to determine configurational entropy. For example, in the mining minima method, low-free-energy conformations are first identified to incorporate the multiple-minimum nature of the free-energy landscape, and vibrational properties in individual wells are incorporated via the harmonic approximation with anharmonic corrections (127, 128). However, enumerating the entire set of minima is not computationally feasible for complex systems such as proteins, and the applicability of this method is limited to relatively simple molecular systems.

More recently, investigators proposed a protocol in which the vibrational entropy of a rigid-rotor harmonic oscillator is combined with direct sampling of dihedral-angle distributions (129).

In this protocol, the conformational substates are identified by discretizing each of the one-dimensional dihedral-angle distributions, and the conformational entropy that takes into account correlation effects is then computed by employing the MIE. The applicability of this combined protocol to multiple systems has been demonstrated (129, 130).

Researchers have also suggested a method to dissect protein configurational entropy on the basis of the energetic approach introduced in Section 3.4 (123). In this method, the time variation of the effective energy $f(\mathbf{q}_t)$ is first computed according to the protein configurations $\mathbf{q}_t$ generated by molecular dynamics simulations. Because $f(\mathbf{q})$ is the defining quantity of the protein free-energy landscape, $f(\mathbf{q}_t)$ characterizes the landscape protein dynamics. Slowly varying and quickly oscillating components were observed from the time-dependent $f(\mathbf{q}_t)$ curves computed for the folded and unfolded states of the protein villin headpiece subdomain. This reflects the dynamics on the rugged free-energy landscape, which consist of rapid vibrations in individual free-energy wells and slow conformational transitions between them. These two components, of disparate timescales, can be dissected using the detrending technique known as Hodrick–Prescott filtering (131); this dissection leads to a natural separation of configurational entropy into conformational and vibrational terms. In accordance with previous empirical estimations (126), the change in S_{config} upon folding is dominated by S_{conf}, even though the magnitude of S_{vib} is significantly larger in each of the folded and unfolded states. Because of the general applicability of the energetic approach, it is straightforward to apply this dissection method to IDPs.

Interestingly, the greater relevancy of vibrational entropy (S_{vib}) versus conformational entropy (S_{conf}) for the change in configurational entropy has been suggested for events including the association of small molecules (127, 128) and protein–ligand and protein–protein binding (132–134). In contrast, conformational entropy is considered the dominant term in protein folding. Thus, if researchers are to elucidate in detail the role of entropy in IDPs, which often exhibit coupled folding and binding (135), a method that accesses both the conformational and vibrational components is indispensible. The computational methods reviewed here will therefore find fascinating applications in studies of IDPs.

4. CONCLUSIONS

Computer simulations are an invaluable tool for studying a variety of complex biomolecular systems in atomistic detail. Recent simulations have demonstrated substantial success in obtaining conformational ensembles of IDPs. However, researchers have also reported several shortcomings that can be largely attributed to inaccuracies in the currently available biomolecular force fields and solvent models. This indicates a need to improve force fields further. It also reflects the importance of targeting IDPs and of properly including solvent effects while developing force fields. An accurate characterization of the conformational ensembles of IDPs is another indispensable component of their thermodynamic descriptions. Recent statistical thermodynamic methods of describing configurational entropy take into account the multiple-minimum nature and the correlation effects inherent to the free-energy landscape of IDPs. A combination of these computational tools will yield new insights into the relationship between the structural disorder and function of IDPs.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

This work was supported by the Samsung Science and Technology Foundation under project SSTF-BA1401-13.

LITERATURE CITED

1. Chouard T. 2011. Breaking the protein rules. *Nature* 471:151–53
2. Oates ME, Romero P, Ishida T, Ghalwash M, Mizianty MJ, et al. 2013. D^2P^2: database of disordered protein predictions. *Nucleic Acids Res.* 41:D508–16
3. van der Lee R, Buljan M, Lang B, Weatheritt RJ, Daughdrill GW, et al. 2014. Classification of intrinsically disordered regions and proteins. *Chem. Rev.* 114:6589–631
4. Dyson HJ, Wright PE. 2005. Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* 6:197–208
5. Wright PE, Dyson HJ. 2015. Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol.* 16:18–29
6. Dunker AK, Brown CJ, Lawson JD, Iakoucheva LM, Obradović Z. 2002. Intrinsic disorder and protein function. *Biochemistry* 41:6573–82
7. Uversky VN. 2002. Natively unfolded proteins: a point where biology waits for physics. *Protein Sci.* 11:739–56
8. Dunker AK, Cortese MS, Romero P, Iakoucheva LM, Uversky VN. 2005. Flexible nets. The roles of intrinsic disorder in protein interaction networks. *FEBS J.* 272:5129–48
9. Chiti F, Dobson CM. 2006. Protein misfolding, functional amyloid, and human disease. *Annu. Rev. Biochem.* 75:333–66
10. Uversky VN, Oldfield CJ, Dunker AK. 2008. Intrinsically disordered proteins in human diseases: introducing the D^2 concept. *Annu. Rev. Biophys.* 37:215–46
11. Babu MM, van der Lee R, de Groot NS, Gsponer J. 2011. Intrinsically disordered proteins: regulation and disease. *Curr. Opin. Struct. Biol.* 21:432–40
12. Dunker AK, Gough J. 2011. Sequences and topology: intrinsic disorder in the evolving universe of protein structure. *Curr. Opin. Struct. Biol.* 21:379–81
13. Habchi J, Tompa P, Longhi S, Uversky VN. 2014. Introducing protein intrinsic disorder. *Chem. Rev.* 114:6561–88
14. Eliezer D. 2009. Biophysical characterization of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* 19:23–30
15. Bowler BE. 2012. Residual structure in unfolded proteins. *Curr. Opin. Struct. Biol.* 22:4–13
16. Carballo-Pacheco M, Strodel B. 2016. Advances in the simulation of protein aggregation at the atomistic scale. *J. Phys. Chem. B* 120:2991–99
17. Morriss-Andrews A, Shea JE. 2014. Simulations of protein aggregation: insights from atomistic and coarse-grained models. *J. Phys. Chem. Lett.* 5:1899–908
18. Nasica-Labouze J, Nguyen PH, Sterpone F, Berthoumieu O, Buchete N-V, et al. 2015. Amyloid β protein and Alzheimer's disease: when computer simulations complement experimental studies. *Chem. Rev.* 115:3518–63
19. Rosenman DJ, Wang C, García AE. 2016. Characterization of Aβ monomers through the convergence of ensemble properties among simulations with multiple force fields. *J. Phys. Chem. B* 120:259–77
20. Stanley N, Esteban-Martín S, De Fabritiis G. 2015. Progress in studying intrinsically disordered proteins with atomistic simulations. *Prog. Biophys. Mol. Biol.* 119:47–52
21. Tzeng SR, Kalodimos CG. 2012. Protein activity regulation by conformational entropy. *Nature* 488:236–40
22. Frederick KK, Marlow MS, Valentine KG, Wand AJ. 2007. Conformational entropy in molecular recognition by proteins. *Nature* 448:325–29
23. Flock T, Weatheritt RJ, Latysheva NS, Babu MM. 2014. Controlling entropy to tune the functions of intrinsically disordered regions. *Curr. Opin. Struct. Biol.* 26:62–72

24. Wand AJ. 2013. The dark energy of proteins comes to light: conformational entropy and its role in protein function revealed by NMR relaxation. *Curr. Opin. Struct. Biol.* 23:75–81
25. Karplus M, Kushick JN. 1981. Method for estimating the configurational entropy of macromolecules. *Macromolecules* 14:325–32
26. Levy RM, Karplus M, Kushick J, Perahia D. 1984. Evaluation of the configurational entropy for proteins: application to molecular dynamics simulations of an α-helix. *Macromolecules* 17:1370–74
27. Schlitter J. 1993. Estimation of absolute and relative entropies of macromolecules using the covariance matrix. *Chem. Phys. Lett.* 215:617–21
28. Schäfer H, Mark AE, van Gunsteren WF. 2000. Absolute entropies from molecular dynamics simulation trajectories. *J. Chem. Phys.* 113:7809–17
29. Andricioaei I, Karplus M. 2001. On the calculation of entropy from covariance matrices of the atomic fluctuations. *J. Chem. Phys.* 115:6289–92
30. Ponder JW, Case DA. 2003. Force fields for protein simulations. *Adv. Protein Chem.* 66:27–85
31. Mackerell AD. 2004. Empirical force fields for biological macromolecules: overview and issues. *J. Comput. Chem.* 25:1584–604
32. Wang J, Cieplak P, Kollman PA. 2000. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* 21:1049–74
33. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. 2006. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins Struct. Funct. Bioinform.* 65:712–25
34. Duan Y, Wu C, Chowdhury S, Lee MC, Xiong G, et al. 2003. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* 24:1999–2012
35. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, et al. 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* 102:3586–616
36. Mackerell AD, Feig M, Brooks CL. 2004. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* 25:1400–15
37. van Gunsteren WF, Billeter SR, Eising AA, Hünenberger PH, Krüger P, et al. 1996. *Biomolecular Simulation: The GROMOS96 Manual and User Guide*. Zürich: vdf Hochsch.
38. Oostenbrink C, Villa A, Mark AE, van Gunsteren WF. 2004. A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *J. Comput. Chem.* 25:1656–76
39. Schmid N, Eichenberger AP, Choutko A, Riniker S, Winger M, et al. 2011. Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *Eur. Biophys. J.* 40:843–56
40. Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL. 2001. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B* 105:6474–87
41. Nguyen PH, Li MS, Derreumaux P. 2011. Effects of all-atom force fields on amyloid oligomerization: replica exchange molecular dynamics simulations of the $A\beta_{16-22}$ dimer and trimer. *Phys. Chem. Chem. Phys.* 13:9778–88
42. Best RB, Hummer G. 2009. Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides. *J. Phys. Chem. B* 113:9004–15
43. Freddolino PL, Park S, Roux B, Schulten K. 2009. Force field bias in protein folding simulations. *Biophys. J.* 96:3772–80
44. Piana S, Klepeis JL, Shaw DE. 2014. Assessing the accuracy of physical models used in protein-folding simulations: quantitative evidence from long molecular dynamics simulations. *Curr. Opin. Struct. Biol.* 24:98–105
45. Yoda T, Sugita Y, Okamoto Y. 2004. Comparisons of force fields for proteins by generalized-ensemble simulations. *Chem. Phys. Lett.* 386:460–67
46. Sun Y, Qian Z, Wei G. 2016. The inhibitory mechanism of a fullerene derivative against amyloid-β peptide aggregation: an atomistic simulation study. *Phys. Chem. Chem. Phys.* 18:12582–91

47. Barz B, Urbanc B. 2012. Dimer formation enhances structural differences between amyloid β-protein (1–40) and (1–42): an explicit-solvent molecular dynamics study. *PLOS ONE* 7:e34345
48. Zhang T, Zhang J, Derreumaux P, Mu Y. 2013. Molecular mechanism of the inhibition of EGCG on the Alzheimer Aβ_{1-42} dimer. *J. Phys. Chem. B* 117:3993–4002
49. Tarus B, Tran TT, Nasica-Labouze J, Sterpone F, Nguyen PH, Derreumaux P. 2015. Structures of the Alzheimer's wild-type Aβ_{1-40} dimer from atomistic simulations. *J. Phys. Chem. B* 119:10478–87
50. Hoffmann KQ, McGovern M, Chiu CC, de Pablo JJ. 2015. Secondary structure of rat and human amylin across force fields. *PLOS ONE* 10:e0134091
51. Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, et al. 2010. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins Struct. Funct. Bioinform.* 78:1950–58
52. Best RB, Mittal J. 2010. Protein simulations with an optimized water model: cooperative helix formation and temperature-induced unfolded state collapse. *J. Phys. Chem. B* 114:14916–23
53. Piana S, Lindorff-Larsen K, Shaw David E. 2011. How robust are protein folding simulations with respect to force field parameterization? *Biophys. J.* 100:L47–49
54. Best RB, Zhu X, Shim J, Lopes PEM, Mittal J, et al. 2012. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain χ_1 and χ_2 dihedral angles. *J. Chem. Theory Comput.* 8:3257–73
55. Doshi U, Hamelberg D. 2009. Reoptimization of the AMBER force field parameters for peptide bond (omega) torsions using accelerated molecular dynamics. *J. Phys. Chem. B* 113:16590–95
56. Nerenberg PS, Head-Gordon T. 2011. Optimizing protein-solvent force fields to reproduce intrinsic conformational preferences of model peptides. *J. Chem. Theory Comput.* 7:1220–30
57. Li D-W, Brüschweiler R. 2010. NMR-based protein potentials. *Angew. Chem. Int. Ed.* 49:6778–80
58. Beauchamp KA, Lin YS, Das R, Pande VS. 2012. Are protein force fields getting better? A systematic benchmark on 524 diverse NMR measurements. *J. Chem. Theory Comput.* 8:1409–14
59. Lindorff-Larsen K, Maragakis P, Piana S, Eastwood MP, Dror RO, Shaw DE. 2012. Systematic validation of protein force fields against experimental data. *PLOS ONE* 7:e32131
60. Lange OF, van der Spoel D, de Groot BL. 2010. Scrutinizing molecular mechanics force fields on the submicrosecond timescale with NMR data. *Biophys. J.* 99:647–55
61. Adhikari AN, Freed KF, Sosnick TR. 2013. Simplified protein models: predicting folding pathways and structure using amino acid sequences. *Phys. Rev. Lett.* 111:028103
62. Best RB, Mittal J. 2011. Free-energy landscape of the GB1 hairpin in all-atom explicit solvent simulations with different force fields: similarities and differences. *Proteins Struct. Funct. Bioinform.* 79:1318–28
63. Lindorff-Larsen K, Trbovic N, Maragakis P, Piana S, Shaw DE. 2012. Structure and dynamics of an unfolded protein examined by molecular dynamics simulation. *J. Am. Chem. Soc.* 134:3787–91
64. Skinner JJ, Yu W, Gichana EK, Baxa MC, Hinshaw JR, et al. 2014. Benchmarking all-atom simulations using hydrogen exchange. *PNAS* 111:15975–80
65. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926–35
66. Berendsen HJC, Postma JPM, van Gunsteren WF, Hermans J. 1981. Interaction models for water in relation to protein hydration. In *Intermolecular Forces*, ed. B Pullman, pp. 331–42. Dordrecht, Neth.: Springer
67. Jorgensen WL, Madura JD. 1985. Temperature and size dependence for Monte Carlo simulations of TIP4P water. *Mol. Phys.* 56:1381–92
68. Abascal JLF, Vega C. 2005. A general purpose model for the condensed phases of water: TIP4P/2005. *J. Chem. Phys.* 123:234505
69. Horn HW, Swope WC, Pitera JW, Madura JD, Dick TJ, et al. 2004. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J. Chem. Phys.* 120:9665–78
70. Fawzi NL, Phillips AH, Ruscio JZ, Doucleff M, Wemmer DE, Head-Gordon T. 2008. Structure and dynamics of the Aβ_{21-30} peptide from the interplay of NMR experiments and molecular simulations. *J. Am. Chem. Soc.* 130:6145–58
71. Wickstrom L, Okur A, Simmerling C. 2009. Evaluating the performance of the ff99SB force field based on NMR scalar coupling data. *Biophys. J.* 97:853–56

72. Chong SH, Ham S. 2013. Assessing the influence of solvation models on structural characteristics of intrinsically disordered protein. *Comput. Theor. Chem.* 1017:194–99
73. Best RB, Zheng W, Mittal J. 2014. Balanced protein–water interactions improve properties of disordered proteins and non-specific protein association. *J. Chem. Theory Comput.* 10:5113–24
74. Piana S, Donchev AG, Robustelli P, Shaw DE. 2015. Water dispersion interactions strongly influence simulated structural properties of disordered protein states. *J. Phys. Chem. B* 119:5113–23
75. Sgourakis NG, Merced-Serrano M, Boutsidis C, Drineas P, Du Z, et al. 2011. Atomic-level characterization of the ensemble of the Aβ_{1-42} monomer in water using unbiased molecular dynamics simulations and spectral algorithms. *J. Mol. Biol.* 405:570–83
76. Ball KA, Phillips AH, Nerenberg PS, Fawzi NL, Wemmer DE, Head-Gordon T. 2011. Homogeneous and heterogeneous tertiary structure ensembles of amyloid-β peptides. *Biochemistry* 50:7612–28
77. Rosenman DJ, Connors CR, Chen W, Wang C, García AE. 2013. Aβ monomers transiently sample oligomer and fibril-like configurations: ensemble characterization using a combined MD/NMR approach. *J. Mol. Biol.* 425:3338–59
78. Ball KA, Phillips AH, Wemmer DE, Head-Gordon T. 2013. Differences in β-strand populations of monomeric Aβ_{40} and Aβ_{42}. *Biophys. J.* 104:2714–24
79. Yedvabny E, Nerenberg PS, So C, Head-Gordon T. 2015. Disordered structural ensembles of vasopressin and oxytocin and their mutants. *J. Phys. Chem. B* 119:896–905
80. Jose JC, Chatterjee P, Sengupta N. 2014. Cross dimerization of amyloid-β and α-synuclein proteins in aqueous environment: a molecular dynamics simulations study. *PLOS ONE* 9:e106883
81. Henriques J, Cragnell C, Skepö M. 2015. Molecular dynamics simulations of intrinsically disordered proteins: force field evaluation and comparison with experiment. *J. Chem. Theory Comput.* 11:3420–31
82. Rauscher S, Gapsys V, Gajda MJ, Zweckstetter M, de Groot BL, Grubmüller H. 2015. Structural ensembles of intrinsically disordered proteins depend strongly on force field: a comparison to experiment. *J. Chem. Theory Comput.* 11:5513–24
83. Palazzesi F, Prakash MK, Bonomi M, Barducci A. 2015. Accuracy of current all-atom force-fields in modeling protein disordered states. *J. Chem. Theory Comput.* 11:2–7
84. Wang W, Ye W, Jiang C, Luo R, Chen HF. 2014. New force field on modeling intrinsically disordered proteins. *Chem. Biol. Drug Des.* 84:253–69
85. Ye W, Ji D, Wang W, Luo R, Chen HF. 2015. Test and evaluation of ff99IDPs force field for intrinsically disordered proteins. *J. Chem. Inf. Model.* 55:1021–29
86. Brady GP, Sharp KA. 1997. Entropy in protein folding and in protein–protein interactions. *Curr. Opin. Struct. Biol.* 7:215–21
87. Meirovitch H. 2007. Recent developments in methodologies for calculating the entropy and free energy of biological systems by computer simulation. *Curr. Opin. Struct. Biol.* 17:181–86
88. Zhou HX, Gilson MK. 2009. Theory of free energy and entropy in noncovalent binding. *Chem. Rev.* 109:4092–107
89. Polyansky AA, Zubac R, Zagrovic B. 2012. Estimation of conformational entropy in protein-ligand interactions: a computational perspective. In *Computational Drug Discovery and Design*, ed. R Baron, pp. 327–53. New York: Springer
90. Diehl C, Engström O, Delaine T, Håkansson M, Genheden S, et al. 2010. Protein flexibility and conformational entropy in ligand design targeting the carbohydrate recognition domain of galectin-3. *J. Am. Chem. Soc.* 132:14577–89
91. Silver NW, King BM, Nalam MNL, Cao H, Ali A, et al. 2013. Efficient computation of small-molecule configurational binding entropy and free energy changes by ensemble enumeration. *J. Chem. Theory Comput.* 9:5098–115
92. Gō N, Scheraga HA. 1976. On the use of classical statistical mechanics in the treatment of polymer chain conformation. *Macromolecules* 9:535–42
93. Gilson MK, Given JA, Bush BL, McCammon JA. 1997. The statistical-thermodynamic basis for computation of binding affinities: a critical review. *Biophys. J.* 72:1047–69
94. Chong SH, Ham S. 2016. New computational approach for external entropy in protein-protein binding. *J. Chem. Theory Comput.* 12:2509–16

95. Smith DMA, Straatsma TP, Squier TC. 2012. Retention of conformational entropy upon calmodulin binding to target peptides is driven by transient salt bridges. *Biophys. J.* 103:1576–84
96. Chang CE, Chen W, Gilson MK. 2005. Evaluating the accuracy of the quasiharmonic approximation. *J. Chem. Theory Comput.* 1:1017–28
97. Chong SH, Ham S. 2015. Structural versus energetic approaches for protein conformational entropy. *Chem. Phys. Lett.* 627:90–95
98. Baron R, van Gunsteren WF, Hünenberger PH. 2006. Estimating the configurational entropy from molecular dynamics simulations: anharmonicity and correlation corrections to the quasi-harmonic approximation. *Trends Phys. Chem.* 11:87–122
99. Baron R, Hünenberger PH, McCammon JA. 2009. Absolute single-molecule entropies from quasi-harmonic analysis of microsecond molecular dynamics: correction terms and convergence properties. *J. Chem. Theory Comput.* 5:3150–60
100. Edholm O, Berendsen HJC. 1984. Entropy estimation from simulations of non-diffusive systems. *Mol. Phys.* 51:1011–28
101. Hnizdo V, Darian E, Fedorowicz A, Demchuk E, Li S, Singh H. 2007. Nearest-neighbor nonparametric method for estimating the configurational entropy of complex molecules. *J. Comput. Chem.* 28:655–68
102. Hnizdo V, Tan J, Killian BJ, Gilson MK. 2008. Efficient calculation of configurational entropy from molecular simulations by combining the mutual-information expansion and nearest-neighbor methods. *J. Comput. Chem.* 29:1605–14
103. Hensen U, Lange OF, Grubmüller H. 2010. Estimating absolute configurational entropies of macromolecules: the minimally coupled subspace approach. *PLOS ONE* 5:e9179
104. Matsuda H. 2000. Physical nature of higher-order mutual information: intrinsic correlations and frustration. *Phys. Rev. E* 62:3096–102
105. Killian BJ, Yundenfreund Kravitz J, Gilson MK. 2007. Extraction of configurational entropy from molecular simulations via an expansion approximation. *J. Chem. Phys.* 127:024107
106. King BM, Tidor B. 2009. MIST: maximum information spanning trees for dimension reduction of biological data sets. *Bioinformatics* 25:1165–72
107. King BM, Silver NW, Tidor B. 2012. Efficient calculation of molecular configurational entropies using an information theoretic approximation. *J. Phys. Chem. B* 116:2891–904
108. Li DW, Khanlarzadeh M, Wang J, Huo S, Brüschweiler R. 2007. Evaluation of configurational entropy methods from peptide folding-unfolding simulation. *J. Phys. Chem. B* 111:13807–13
109. Li DW, Brüschweiler R. 2009. *In silico* relationship between configurational entropy and soft degrees of freedom in proteins and peptides. *Phys. Rev. Lett.* 102:118108
110. Li DW, Showalter SA, Brüschweiler R. 2010. Entropy localization in proteins. *J. Phys. Chem. B* 114:16036–44
111. Harpole KW, Sharp KA. 2011. Calculation of configurational entropy with a Boltzmann-quasiharmonic model: the origin of high-affinity protein-ligand binding. *J. Phys. Chem. B* 115:9461–72
112. Cukier RI. 2015. Dihedral angle entropy measures for intrinsically disordered proteins. *J. Phys. Chem. B* 119:3621–34
113. Cover TM, Thomas JA. 2006. *Elements of Information Theory*. Hoboken, NJ: Wiley
114. Baxa MC, Haddadian EJ, Jumper JM, Freed KF, Sosnick TR. 2014. Loss of conformational entropy in protein folding calculated using realistic ensembles and its implications for NMR-based calculations. *PNAS* 111:15396–401
115. Piana S, Lindorff-Larsen K, Shaw DE. 2013. Atomic-level description of ubiquitin folding. *PNAS* 110:5915–20
116. Fenley AT, Killian BJ, Hnizdo V, Fedorowicz A, Sharp DS, Gilson MK. 2014. Correlation as a determinant of configurational entropy in supramolecular and protein systems. *J. Phys. Chem. B* 118:6447–55
117. Marlow MS, Dogan J, Frederick KK, Valentine KG, Wand AJ. 2010. The role of conformational entropy in molecular recognition by calmodulin. *Nat. Chem. Biol.* 6:352–58
118. Kasinath V, Sharp KA, Wand AJ. 2013. Microscopic insights into the NMR relaxation-based protein conformational entropy meter. *J. Am. Chem. Soc.* 135:15092–100
119. Hensen U, Gräter F, Henchman RH. 2014. Macromolecular entropy can be accurately computed from force. *J. Chem. Theory Comput.* 10:4777–81

120. Klefas-Stennett M, Henchman RH. 2008. Classical and quantum Gibbs free energies and phase behavior of water using simulation and cell theory. *J. Phys. Chem. B* 112:9769–76
121. Chong SH, Ham S. 2011. Configurational entropy of protein: a combined approach based on molecular simulation and integral-equation theory of liquids. *Chem. Phys. Lett.* 504:225–29
122. Chong SH, Ham S. 2014. Protein folding thermodynamics: a new computational approach. *J. Phys. Chem. B* 118:5017–25
123. Chong SH, Ham S. 2015. Dissecting protein configurational entropy into conformational and vibrational contributions. *J. Phys. Chem. B* 119:12623–31
124. Lazaridis T, Karplus M. 2002. Thermodynamics of protein folding: a microscopic view. *Biophys. Chem.* 100:367–95
125. Chong SH, Ham S. 2013. Conformational entropy of intrinsically disordered protein. *J. Phys. Chem. B* 117:5503–9
126. Karplus M, Ichiye T, Pettitt BM. 1987. Configurational entropy of native proteins. *Biophys. J.* 52:1083–85
127. Chang CEA, Chen W, Gilson MK. 2007. Ligand configurational entropy and protein binding. *PNAS* 104:1534–39
128. Chang CE, Gilson MK. 2004. Free energy, entropy, and induced fit in host-guest recognition: calculations with the second-generation mining minima algorithm. *J. Am. Chem. Soc.* 126:13156–64
129. Suárez E, Díaz N, Suárez D. 2011. Entropy calculations of single molecules by combining the rigid-rotor and harmonic-oscillator approximations with conformational entropy estimations from molecular dynamics simulations. *J. Chem. Theory Comput.* 7:2638–53
130. Suárez D, Díaz N. 2014. Sampling assessment for molecular simulations using conformational entropy calculations. *J. Chem. Theory Comput.* 10:4718–29
131. Hodrick RJ, Prescott EC. 1997. Postwar U.S. business cycles: an empirical investigation. *J. Money Credit Bank.* 29:1–16
132. Killian BJ, Kravitz JY, Somani S, Dasgupta P, Pang Y-P, Gilson MK. 2009. Configurational entropy in protein–peptide binding: computational study of Tsg101 ubiquitin E2 variant domain with an HIV-derived PTAP nonapeptide. *J. Mol. Biol.* 389:315–35
133. Thorpe IF, Brooks CL. 2007. Molecular evolution of affinity and flexibility in the immune system. *PNAS* 104:8821–26
134. Chang CEA, McLaughlin WA, Baron R, Wang W, McCammon JA. 2008. Entropic contributions and the influence of the hydrophobic environment in promiscuous protein–protein association. *PNAS* 105:7456–61
135. Sugase K, Dyson HJ, Wright PE. 2007. Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature* 447:1021–25

Annual Review of Physical Chemistry

Volume 68, 2017

Contents

Indexes

Errata

An online log of corrections to *Annual Review of Physical Chemistry* articles may be found at http://www.annualreviews.org/errata/physchem