Biophysical Journal
Article

Biophysical Society

Thermodynamics of Conformational Transitions in a Disordered Protein Backbone Model

Justin A. Drake[1] and B. Montgomery Pettitt[1,*]
[1]Sealy Center for Structural Biology and Molecular Biophysics, The University of Texas Medical Branch, Galveston, Texas

ABSTRACT Conformational entropy is expected to contribute significantly to the thermodynamics of structural transitions in intrinsically disordered proteins or regions in response to protein/ligand binding, posttranslational modifications, and environmental changes. We calculated the backbone (dihedral) conformational entropy of oligoglycine (Gly_N), a protein backbone mimic and model intrinsically disordered region, as a function of chain length ($N = 3$, 4, 5, 10, and 15) from simulations using three different approaches. The backbone conformational entropy scales linearly with chain length with a slope consistent with the entropy of folding of well-structured proteins. The entropic contributions of second-order dihedral correlations are predominantly through intraresidue ϕ-ψ pairs, suggesting that oligoglycine may be thermodynamically modeled as a system of independent glycine residues. We find the backbone conformational entropy to be largely independent of global structural parameters, like the end-to-end distance and radius of gyration. We introduce a framework referred to herein as "ensemble confinement" to estimate the loss (gain) of conformational free energy and its entropic component when individual residues are constrained to (released from) particular regions of the ϕ-ψ map. Quantitatively, we show that our protein backbone model resists ordering/folding with a significant, unfavorable ensemble confinement free energy because of the loss of a substantial portion of the absolute backbone entropy. Proteins can couple this free-energy reservoir to distal binding events as a regulatory mechanism to promote or suppress binding.

INTRODUCTION

Over the past 20 years, it has become clear that the classical structure-function paradigm is not a universal property of the eukaryotic proteome (1–3). Although structure dictates function for a large class of proteins like enzymes, a vast array of proteins employ highly dynamic, intrinsically disordered regions (IDRs) to carry out diverse functions, many of which are involved in regulation and control of signaling networks (1–5). Modular proteins, like nuclear transcription factors, rely on a combination of order and disorder to function, and a coupling between these regions provides additional regulatory mechanisms to fine tune binding affinity and downstream signal cascades (6–10). Mutations in IDRs can abrogate control of signaling networks, eventually leading to disease onset (3,11,12). A seemingly ubiquitous and functionally necessary property of IDRs is their ability to undergo structural transitions in response to a number of factors, including protein/ligand binding, posttranslational modifications, and environmental changes (3,5,10,13–17). These transitions may proceed to more ordered or disordered states but can also result in a redistribution of the disordered, structural ensemble (3,10,15,16). To successfully target drugs to IDRs or genetically engineer IDRs with certain therapeutic properties to treat various diseases, we need a more complete understanding of the thermodynamics associated with IDR conformational transitions.

IDRs directly responsible for facilitating protein interactions primarily do so with relatively short, contiguous amino acid sequences that become more ordered or structured upon binding. These functional elements or sequence motifs have been referred to as short linear motifs, eukaryotic linear motifs, and molecular recognition features (MoRFs) (3,12,18–20). Short linear motifs are typically around 3–10 amino acids (19), whereas MoRFs are longer at 10–70 amino acids and, by definition, form secondary structures upon binding (18). Transcription factors are enriched with these motifs (21). p53, for example, contains multiple MoRFs with individual MoRFs being able to bind many partners by forming different secondary structures (22), thus allowing it to serve as a hub protein or master regulator in signaling.

Protein, DNA, and small-molecule binding can also induce structural transitions or a redistribution of the

Submitted January 29, 2018, and accepted for publication April 16, 2018.
*Correspondence: mpettitt@utmb.edu
Editor: Alemayehu Gorfe.
https://doi.org/10.1016/j.bpj.2018.04.027

structural ensemble in regions distal to the binding interface. For example, Amemiya et al. (23) and Zea et al. (17) found that ligand binding increased or decreased disorder in regions that were typically less than 10 amino acids long from analyses of protein structure databases. There are also numerous examples in which proteins use conformational transitions in IDRs (or simply disorder in general) in an allosteric or cooperative mechanism to effect downstream signaling (7,10,15,24). These observations prompted the development of new allostery models based on the ensemble nature (i.e., structural diversity) of IDRs (10,15,25,26). Propagation of the allosteric signal does not necessarily require the folding or unfolding of an IDR; rather, it can also be achieved through the remodeling of the disordered ensemble (10).

Of particular interest is the potential thermodynamic coupling between effector and allosteric sites facilitated by IDR structural transitions. Conformational entropy is a critical property to describe or understand the thermodynamic origin of IDR structural transitions and how such transitions can modulate protein binding at distal sites (10,14,15). Recently, nuclear magnetic resonance methods have been developed to probe binding-induced conformational entropy changes in structured proteins. This so-called "entropy meter" (27,28), which relates changes in backbone and side-chain motions to changes in conformational entropy, has highlighted the importance of conformational entropy in suppressing or driving ligand binding (27,29–31). As a result, one emerging concept is that protein structural plasticity, or the capacity of a protein to alter its internal structural fluctuations, provides a reservoir of entropy and thus free energy that is available to the protein to carry out its function (32). So we consider IDRs as significant conformational entropy reservoirs from which free energy may be withdrawn or deposited via structural transitions (e.g., due to binding). With respect to protein or ligand binding, these transitions would mediate the thermodynamic connection between allosteric and effector sites. Qualitatively, our understanding of how proteins can tune conformational entropy to modulate IDR function continues to improve (10,14,15), yet a complete, quantitative description is lacking.

In this article, we are primarily interested in the thermodynamics underlying structural transitions of short IDRs, which we expect will provide insight into the numerous biological processes that rely on order-disorder transitions of IDRs of varying lengths and establish a framework to develop a more quantitative theory of IDRs as free-energy reservoirs. We use oligoglycine (Gly_N, where N is the chain length) as a protein backbone mimic and model IDR (33–36). IDRs are often enriched in glycine residues (13,37); a high glycine content has been associated with more compact IDRs (38), and stretches of oligoglycine can be found in large disordered domains (7,39). We begin first by calculating the conformational entropy of oligoglycine from backbone ϕ and ψ dihedral angles sampled by molecular dynamics (MD) as a function of chain length ($N = 3$ to 15 residues) comparing the quasi-harmonic analysis (QHA) (40,41), Boltzmann quasi-harmonic (BQH) (42–44), and mutual information expansion (MIE) methods (45). Because proteins containing IDRs likely impose structural constraints on the disordered region and order-disorder transitions may alter the global structural properties of an IDR, we calculate the backbone conformational entropy of Gly_{15} as a function of end-to-end distance and radius of gyration. Then, we consider how much conformational entropy is lost or gained as free energy when oligoglycine is constrained to or released from particular conformational states.

Theory

We briefly review the theory underlying the approaches we use to calculate the conformational entropy of various glycine polypeptides. We refer to conformational entropy as the contribution to the system entropy that depends only on the spatial coordinates of the solute molecule. Although this article primarily focuses on the dihedral angle contributions to conformational entropy (discussed in detail in Materials and Methods), the theories presented below are general and may be used with any coordinate system and any number of coordinates. Where appropriate, we refer the reader to more detailed discussions, and for consistency, we follow the notation of (44–46).

Solute entropy and coordinate system

The entropy of a solute can be expressed in terms of the probability density, $\rho(\mathbf{p}, \mathbf{r})$, of Cartesian coordinates ($\boldsymbol{r}$) and conjugate momenta ($\boldsymbol{p}$) as follows:

$$S^{\text{solute}} = -k_B \iint \rho(\mathbf{p},\mathbf{r}) \ln[h^s \rho(\mathbf{p},\mathbf{r})] \mathrm{d}\mathbf{p}\mathrm{d}\mathbf{r}, \qquad (1)$$

where k_B is Boltzmann's constant, h is Planck's constant, and s is the number of coordinate-momenta pairs that define solute phase space. In Cartesian coordinates, $s = 3N$, where N is the number of solute atoms. Because of independence, $\rho(\mathbf{p}, \mathbf{r})$ can be factored into marginal distributions of $\mathbf{p}$ and $\mathbf{r}$, which upon expanding Eq. 1 gives the following:

$$\begin{aligned} S^{\text{solute}} &= -k_B \int \rho(\mathbf{p}) \ln[h^s \rho(\mathbf{p})] \mathrm{d}\mathbf{p} - k_B \int \rho(\mathbf{r}) \ln[\rho(\mathbf{r})] \mathrm{d}\mathbf{r} \\ &= S_p + S_r, \end{aligned} \qquad (2)$$

with S_p and S_r being the momentum and conformational contributions to S^{solute}, respectively. By separating S^{solute}, neither S_p nor S_r is correctly dimensioned because of the factor h^s, and only upon their addition or considering changes in entropy will S^{solute} have the correct units (46).

It is often more convenient or appropriate to describe the conformation of a protein or polypeptide in an internal coordinate system, for example, using bonds, angles, and torsions (i.e., BAT coordinates) (47–49). Under a transformation of Cartesian coordinates to internal coordinates ($\boldsymbol{q}$), the conformational entropy becomes

$$S_r = S^{\text{int}} + S^{\text{ext}} + k_B \ln\langle J \rangle, \tag{3}$$

where S^{int} is the conformational entropy associated with the $3N-6$ internal coordinates that specify the relative atomic positions and is given by

$$S^{\text{int}} = -k_B \int \rho(\mathbf{q}) \ln \rho(\mathbf{q}) \mathrm{d}\mathbf{q}. \tag{4}$$

S^{ext} is the entropic contribution of the remaining six coordinates associated with the overall translation and rotation of the molecule that specify the absolute positions of the atoms. $\langle J \rangle$ is the ensemble average of the Jacobian of the coordinate transformation, $\boldsymbol{r} \rightarrow \boldsymbol{q}$. Expressions for S^{ext} and the Jacobian in the BAT coordinate system can be found in (46). In BAT coordinates, the "hard" bond and angle coordinates may be treated essentially independent of or separable from the "soft" torsions/dihedrals (47,48,50), and therefore, their contributions to S^{int} are additive. Depending on the system, it may be reasonable to assume that an isothermal process (e.g., protein binding) does not significantly perturb the bond and angle vibrations, which may be approximated at their equilibrium positions. Then, from Eqs. 2 to 4, the change in solute entropy can be approximated as $\Delta S^{\text{solute}} = \Delta S_r \approx \Delta S^{\text{int}} \approx \Delta S^{\text{dihed}}$.

A number of methods have been developed to estimate S^{int} or ΔS^{int} (44,51). Below, we briefly sketch the QHA (40,41), BQH (42–44), and MIE (45) methods used in this article.

QHA and BQH analysis

In the QHA method, the distribution of internal coordinates, $\rho(\boldsymbol{q})$, around an average conformation is assumed to be multivariate Gaussian. This allows the entropy integral in Eq. 4 to be evaluated analytically, giving

$$S^{\text{int}} \approx S^{\text{QHA}} = \frac{1}{2} k_B \ln[(2\pi e)^s |\boldsymbol{\sigma}|], \tag{5}$$

where $|\boldsymbol{\sigma}|$ is the determinant of the variance-covariance matrix of $\boldsymbol{q}$ about $\langle \boldsymbol{q} \rangle$, and $s = 3N-6$ is the number of internal coordinates or perhaps some subset of coordinates. Equation 5 can be expressed in terms of a correlation matrix, $\boldsymbol{C}$, by factoring out the diagonal terms from $\boldsymbol{\sigma}$ (42–44) as follows:

$$S^{\text{QHA}} = \frac{1}{2} k_B \sum_{i=1}^{s} \ln(2\pi e \sigma_i) + \frac{1}{2} k_B \ln|\mathbf{C}|. \tag{6}$$

The first term involves a sum over the variance ($\sigma_i \equiv \sigma_{ii}$) of each internal coordinate, whereas the second term accounts for correlations among coordinates through the determinant, $|\boldsymbol{C}|$. Each element of the $s \times s$ matrix $\boldsymbol{C}$ is $\sigma_{ij}/\sqrt{\sigma_i \sigma_j}$, where σ_{ij} is the covariance of coordinates i and j. Coordinate variances and the correlation matrix can be calculated from the trajectories of atomic positions from an MD simulation.

Assuming Gaussian distributions of $\boldsymbol{q}$ can produce a poor approximation of S^{int} when, for example, BAT coordinates are used because torsion/dihedral angle probability distributions are typically multimodal (52). To improve estimates of S^{int} and account for non-Gaussian distributions, Di Nola et al. (42) proposed replacing the first term in Eq. 6 with the Boltzmann entropy, such that

$$S^{\text{int}} \approx S^{\text{BQH}} = -k_B \sum_{i=1}^{s} \int \rho(q_i) \ln \rho(q_i) \mathrm{d}q_i + \frac{1}{2} k_B \ln|\mathbf{C}|, \tag{7}$$

where i indexes one of the internal coordinates (q_i). The probability distribution, $\rho(q_i)$, and associated entropy integral (Eq. 4) can be numerically approximated from one-dimensional histograms collected over an MD trajectory. In QHA and BQH, the first terms in Eqs. 6 and 7 represent a first-order approximation of conformational entropy under the assumption of independent coordinates, whereas the second term contributes a negative correction because of second-order correlations among coordinates.

MIE

MIE (45) is a nonparametric approach that approximates the probability density of all coordinates in terms of a set of lower-order density distributions using the generalized Kirkwood superposition approximation from liquid theory. As a simple example, the second-order approximation of a three-dimensional density function of coordinates q_1, q_2, and q_3 is

$$\rho^{(2)}(q_1, q_2, q_3) = \frac{\rho_2(q_1, q_2)\rho_2(q_1, q_3)\rho_2(q_2, q_3)}{\rho_1(q_1)\rho_1(q_2)\rho_1(q_3)}, \tag{8}$$

where the numerator includes joint distributions (ρ_2) of all possible pairs of coordinates, and the denominator includes the marginal distributions (ρ_1) of each coordinate. Unlike liquid theory, the marginal and joint distributions are treated as unique and depend on the coordinates or coordinate pairs, respectively. Substituting Eq. 8 into Eq. 4 and expanding the logarithm gives the second-order MIE approximation of S^{int}:

$$\begin{aligned} S^{\text{int}} \approx S_2^{\text{MIE}} = {} & S_1(q_1) + S_1(q_2) + S_1(q_3) - I_2(q_1, q_2) \\ & - I_2(q_1, q_3) - I_2(q_2, q_3), \end{aligned} \tag{9}$$

where the mutual information (MI) term is defined as $I_2(q_i, q_j) = S_1(q_i) + S_1(q_j) - S_2(q_i, q_j)$. The general functional

forms of S_1 and S_2 are given by Eq. 4 but with their subscripts indicating that the entropy integral is over a marginal or joint distribution, respectively. Note that the sum over the S_1 terms in Eq. 9, which we will refer to as S_1^{MIE}, is equivalent to the first term in Eq. 7. Whereas the QHA and BQH methods account for correlations among coordinates via a harmonic approximation, MIE estimates the entropic contribution to S^{int} due to correlations directly from the joint distributions that define the MI terms. For a more complete discussion of MIE, its application to a number of molecular systems, and its generalization to higher-order approximations with any number of internal coordinates, we refer to (45). In this article, we will at most consider a third-order approximation of $\rho(\boldsymbol{q})$ with the corresponding estimate of S^{int} given by

$$S^{\mathrm{int}} \approx S_3^{\mathrm{MIE}} = S_2^{\mathrm{MIE}} + I_3(q_1, q_2, q_3). \tag{10}$$

For a system with more internal coordinates, Eq. 10 will include an I_3 term for each coordinate triplet. The various marginal and joint distributions of internal coordinates can be estimated by constructing the necessary histograms from MD simulation trajectories. Although in principle $\rho(\boldsymbol{q})$ may be approximated at higher orders, practically it may become computationally prohibitive to achieve the necessary conformational sampling required to converge estimates of S_3^{MIE} even for small systems because of the sparseness of the three-dimensional histograms. This problem may be mitigated to a certain extent by using coarser histogram bins or alternative binning strategies.

We next present a general framework to estimate the change in free energy and conformational entropy when an ensemble of disordered polypeptide structures is restricted or confined to particular conformational states. We use this process, hereafter referred to as ensemble confinement, to study the backbone contributions to the thermodynamics associated with order-disorder transitions of short IDRs.

Ensemble confinement free energy

We start by assuming that there are M states defined through a partition of conformation space. The probability of observing a conformation in state j is p_j. Each state has an internal energy (U_j) and intrastate entropy (S_j^{intra}), with the latter accounting for the degeneracy of state j. The conformational entropy, or entropy of the ensemble excluding rotation and translation, may be decomposed into two terms (53,54) as follows:

$$S^{\mathrm{ens}} = \sum_{j=1}^{M} p_j S_j^{\mathrm{intra}} - k_B \sum_{j=1}^{M} p_j \ln p_j = S^{\mathrm{intra}} + S^{\mathrm{state}}, \tag{11}$$

where S^{state} is the entropy arising from the partition of conformation space into the M predefined states. The free energy of restricting or confining the ensemble to a single state i is $\Delta A_i = \Delta U_i - T\Delta S_i$, where $\Delta U_i = U_i - \langle U \rangle$ and $\Delta S_i = S_i^{\mathrm{intra}} - S^{\mathrm{ens}}$. Substituting Eq. 11 into the expression for ΔA_i and rearranging the terms gives the following:

$$\Delta A_i = U_i - TS_i^{\mathrm{intra}} + \sum_{j=1}^{M} p_j \left(-U_j + TS_j^{\mathrm{intra}} - k_B T \ln p_j \right) \tag{12}$$

Using the fact that

$$p_j = \frac{e^{S_j^{\mathrm{intra}}/k_B - \beta U_j}}{Z}, \tag{13}$$

where Z is the partition function that normalizes p_j, Eq. 12 simplifies to

$$\Delta A_i = -k_B T \ln p_i. \tag{14}$$

Equation 14 indicates that the (ensemble) confinement free energy, ΔA_i, is almost always positive or unfavorable because $p_i \leq 1$. The average confinement free energy across the possible states is

$$\langle \Delta A \rangle = \sum_{i=1}^{M} p_i \Delta A_i = -k_B T \sum_{i=1}^{M} p_i \ln p_i = TS^{\mathrm{state}}, \tag{15}$$

which highlights the entropic nature of the confinement free energy. Although any collective variable may be used to define state i, we follow the approaches of (54,55) by defining the conformational state of oligoglycine in terms of backbone dihedral angles. As detailed in the Materials and Methods, we calculate the dihedral angle contribution to the conformational entropy (S^{int}) of successively longer oligoglycines using the QHA, BQH, and MIE approaches, and using the framework presented above, we quantify the extent of conformational entropy lost or free energy gained upon confining the ensemble of oligoglycine structures to particular conformational states.

MATERIALS AND METHODS

We use oligoglycine as a model to study the backbone contributions to the thermodynamics associated with order-disorder transitions of short IDRs. Oligoglycine (Gly_N, where N is the number of residues or chain length) is a protein backbone mimic and model disordered polypeptide (33,35,36,56–58). Tracts of oligoglycine can be found in IDRs (59), and high glycine content has been associated with compact IDRs (38). Using MD simulations, we sample the structural ensemble of successively longer oligoglycines and calculate the dihedral angle contributions to the absolute conformational entropy as a function of chain length. Units of entropy (eu) are cal/mol/K. Because a protein containing an IDR may impose some structural constraints on the disordered region, we consider these entropy estimates as a function of end-to-end distance and radius of gyration. Using the ensemble confinement framework presented in Theory, we determine the extent of backbone entropy lost or gained when oligoglycine is constrained to or released from particular conformational states. Below, we

provide details on the oligoglycine model, MD simulation parameters/protocol, and the application of the various methods discussed in Theory to calculate the backbone dihedral conformational entropy. Data, analysis scripts, and simulation input files are available upon request.

System and simulations

Gly_3, Gly_4, Gly_5, Gly_{10}, and Gly_{15} were built in an extended conformation with neutral acetyl and N-methylamide caps using XLeap in AmberTools (60). Each oligoglycine was solvated in a box of transferable intermolecular potential with three points water molecules with at least a 10 Å padding to the walls of the box. Simulations were performed with nanoscale molecular dynamics 2.9 and 2.10 (61) with Amber (Assisted Model Building with Energy Refinement) ff12SB (60) at constant temperature (300 K) and pressure (1 atm). MD trajectories of Gly_3 and Gly_{10} were taken from a previous study (62), and the same parameters were used here to simulate the remaining oligoglycines. Briefly, a steepest descent minimization was performed for each system followed by production simulations with a 2 fs time step using the velocity Verlet algorithm. A Langevin thermostat and barostat were used to maintain temperature and pressure, respectively. A cutoff of 12 Å was used for nonbonded interactions, and the van der Waals interactions were attenuated with a switching function beginning at 10 Å. Full electrostatic interactions were calculated every two steps using particle mesh Ewald method with a 1 Å grid spacing. To be consistent with Amber's nonbonded exclusion policy, the 1–4 scaling was set to 0.8333. Bonds involving hydrogen atoms were fixed with SHAKE. Gly_3, Gly_4, Gly_5, Gly_{10}, and Gly_{15} were simulated for 300, 550, 995, 950, and 1150 ns, respectively, after dropping at least 20 ns for equilibration. Coordinates were saved for analysis every 1 ps.

Dihedral angle conformational entropy

From simulations of each oligoglycine, we construct trajectories of the ϕ and ψ backbone dihedral angles, neglecting the nearly rigid ω angles, and calculate the backbone conformational entropy (Theory). We refer to the entropy estimates from QHA, BQH, and MIE as S^{QHA}, S^{BQH}, and S^{MIE}, respectively. By considering only dihedral angles, we are assuming that they, as a set, contribute to the conformational entropy, S^{int} (Eq. 4), independent of the bonds and angles that define the BAT coordinate system. That is, we assume any changes in the entropic contributions of dihedral-bond and dihedral-angle correlations are negligible. S^{QHA}, S^{BQH}, and S^{MIE} were calculated as a function of oligoglycine length.

To calculate S^{QHA} and S^{BQH} using Eqs. 6 and 7, dihedral angles were represented on the unit circle in the complex plane, which is convenient for constructing the variance-covariance ($\boldsymbol{\sigma}$) matrix and subsequently the correlation matrix ($\boldsymbol{C}$) for angular coordinates (63). Each element of $\boldsymbol{\sigma}$ is computed as $\sigma_{ij} = \langle Z_i Z_j^* \rangle - \langle Z_i \rangle \langle Z_j^* \rangle$, where Z is the complex representation of dihedral i or j, and Z^* is the complex conjugate. The diagonal terms were factored out of $\boldsymbol{\sigma}$ to give $\boldsymbol{C}$ (see Theory), which can then be directly used in Eqs. 6 and 7 to calculate S^{QHA} and S^{BQH}, respectively. For S^{BQH}, histograms of each dihedral angle were constructed on the range $[-\pi, \pi]$ using 180 bins to evaluate the exact Boltzmann entropy expression in Eq. 7. Note that the summations in Eqs. 6 and 7 are over the $2N_{\mathrm{res}}$ dihedral angles, where N_{res} is the number of glycine residues, and that the correlation matrix has dimensions of $2N_{\mathrm{res}} \times 2N_{\mathrm{res}}$.

With Eqs. 9 and 10, the second- (S_2^{MIE}) and third- (S_3^{MIE}) order MIE approximations of the full entropy were calculated from the marginal and joint distributions of the $2N_{\mathrm{res}}$ dihedral angles using the program Algorithm for Computing Configurational ENTropy (45). Marginal and two-dimensional probability distributions were approximated with histograms using 120 bins over the range $[-\pi, \pi]$. An analysis of S_2^{MIE} versus bin size suggested that 120 bins in each dimension were sufficient to converge S_2^{MIE} to within $\sim$ 1–2 eu of those calculated with 110 and 130 bins for all oligoglycines. However, S_3^{MIE} is much more sensitive with respect to the number of bins, and due to the sparseness of the 3D histograms, 20 bins were used in each dimension to calculate S_3^{MIE}. We also explored the use of an alternative, second-order MIE truncation strategy in which each residue in oligoglycine is assumed independent of one another. Here, $S_{\mathrm{res}}^{\mathrm{MIE}}$ is calculated using only the MI terms for the ϕ and ψ angles within each residue and ignoring those terms that include dihedrals from different residues. $S_{\mathrm{res}}^{\mathrm{MIE}}$ will always be greater than or equal to S_2^{MIE} because the latter includes more MI terms that may potentially reduce $S_{\mathrm{res}}^{\mathrm{MIE}}$. Equality is achieved only when each residue represents an independent subsystem. However, depending on the strength of correlated dihedral motions and the desired accuracy, $S_{\mathrm{res}}^{\mathrm{MIE}}$ may be a reasonable and less computationally demanding approximation of S_2^{MIE} because histograms of all pairs of dihedrals are not needed. In a similar fashion, if no third or higher-order correlations exist between dihedrals, then $S_2^{\mathrm{MIE}} = S_3^{\mathrm{MIE}}$ and so on.

To assess convergence, all entropy estimates were monitored as a function of simulation time. S_2^{MIE} and to a greater extent S_3^{MIE} showed poorer convergence than S^{QHA}, S^{BQH}, and $S_{\mathrm{res}}^{\mathrm{MIE}}$, especially for Gly_{10} and Gly_{15}. To improve accuracy, functions of S_2^{MIE} and S_3^{MIE} with respect to time (t) were individually fit to a hyperbolic function, $f(t) = a - b/t$, for all oligoglycines in a manner similar to the approaches of (45,64). The asymptotes (a) were used as an additional approximation of S_2^{MIE} and S_3^{MIE} and are referred to as $S_2^{\mathrm{MIE,fit}}$ and $S_3^{\mathrm{MIE,fit}}$, respectively. To estimate statistical uncertainty, simulation trajectories for each oligoglycine were split into five contiguous blocks, and the various entropy estimates were calculated for each block with the exception of the third-order MIE entropy estimates, which showed poor convergence. We take the SD of each entropy estimate across the blocks as a conservative error estimate. This approach was similarly taken for subsequent thermodynamic analyses.

Effects of structural constraints on entropy estimates

A preliminary analysis of the Gly_{10} and Gly_{15} MD trajectories suggested that constraints on the end-to-end distance (R), which is defined between carbons in the acetyl and N-methylamide caps, and on the radius of gyration (R_g) only weakly affected the dihedral angle populations and therefore likely the dihedral conformational entropy. Because of a lack of conformational sampling from the explicit solvent MD simulation, particularly at values of R and R_g far from average, eight implicit solvent simulations of Gly_{15} were performed with end-to-end distances restrained to 5, 10, 15, 20, 25, 30, 35, and 40 Å using a harmonic bias potential and the Amber ff12SB force field. Simulations were conducted at constant temperature (300 K) and volume with nanoscale molecular dynamics 2.11 (61) using the default Generalized Born implicit solvent model (65,66) and a descreening cutoff of 12 Å.

A force constant of 25 kcal/mol/Å^2 was used for the harmonic bias potential to restrain R. Simulations were run for 1 μs each with 10 ns dropped for equilibration. The trajectories were further partitioned by selecting conformations with an R_g centered at 6, 7, 8, 9, 10, 11, 12, and 13 $\pm$ 0.5 Å. $S_{\mathrm{res}}^{\mathrm{MIE}}$, $S_2^{\mathrm{MIE,fit}}$ and $S_3^{\mathrm{MIE,fit}}$ were calculated as a function of R and R_g. Additionally, we compute the one-dimensional entropy (S_1^{MIE}) from the marginal distributions of each dihedral angle (i.e., $\sum_{i=1}^{2N_{\mathrm{res}}} S_1(q_i)$ in Eq. 9).

Ensemble confinement free energy

We use the ensemble confinement framework presented in Theory to estimate the change in free energy (ΔA_i) upon limiting the ensemble of oligoglycine structures to some conformational state, i. To define the reference state, we follow an approach similar to that in (54,55,67). We begin by partitioning the two-dimensional ϕ-ψ map (Fig. 1) of a glycine residue into six regions, labeled 1–6, that correspond to a high-energy region (1), β-sheet (2), right (3) and left (4) ppII, and right (5) and left (6) α-helix regions. For a given oligoglycine conformation sampled from an MD simulation, each residue, $j \in [1 \ldots N_{\mathrm{res}}]$, is assigned to one of these six

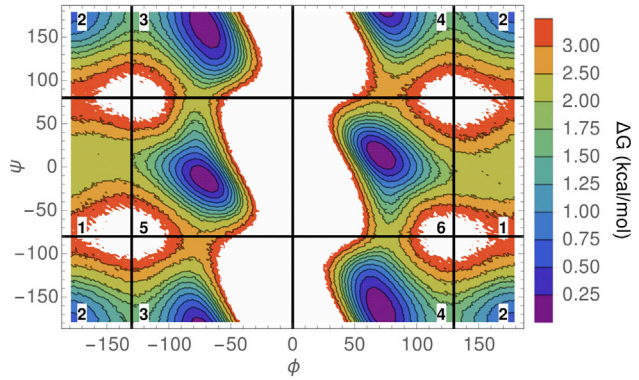

FIGURE 1 ϕ-ψ free-energy map. For illustrative purposes, a two-dimensional histogram of ϕ,ψ dihedral angles was constructed across all residues from an MD trajectory of Gly_{15}. The histogram was converted to a free-energy map relative to the most populated bin. Regions with values greater than 3.5 kcal/mol are colored white. The map is partitioned into six regions that are used to define a conformational state of oligoglycine. States 2, 3, 4, 5, and 6 correspond to the β, $ppII_R$, $ppII_L$, α_R, and α_L regions, respectively. State 1 corresponds to a high-energy region (Hi).

regions given its ϕ-ψ pair. The conformational state (i) is then defined as $i = \{d_1, d_2, \ldots, d_{N_{res}}\}$, where d_j takes on values of 1–6. The free-energy change of confining oligoglycine to state i is calculated using Eq. 14 as $\Delta A_i = -k_B T \ln p_i$. For Gly_{3-5}, p_i was calculated as $p_i = h_i/n_f$, where h_i is the number of times oligoglycine visited conformational state i, and n_f is the total number of observations or simulation frames.

Although p_i may be accurately estimated for the short oligoglycines, it is much more challenging to obtain a statistically meaningful or physically representative estimate of p_i for long oligoglycines because, for example, there are 6^{15} possible conformational states of Gly_{15}, most of which are transiently populated. Therefore, we treated Gly_{10} and Gly_{15} as being composed of two and three independent yet connected systems of Gly_5, respectively. p_i can then be factored into joint probability distributions of the Gly_5 blocks, and approximated as follows:

$$p_i \approx \tilde{p}_i = \begin{cases} p(\{d_{1-5}\})p(\{d_{6-10}\}), & \text{for } Gly_{10} \\ p(\{d_{1-5}\})p(\{d_{6-10}\})p(\{d_{11-15}\}) & \text{for } Gly_{15}. \end{cases}$$

For Gly_{10} and Gly_{15}, ΔA_i is evaluated using the approximate distribution, $\tilde{p}_i$, with Eq. 14. The average confinement free energy over all observed conformational states is given by Eq. 15 as $\langle \Delta A \rangle = \sum_{i=1}^{M} p_i \Delta A_i$. Although $\tilde{p}_i$ was used to estimate ΔA_i for Gly_{10} and Gly_{15}, we use the observed probabilities, $p_i = h_i/n_f$, to calculate the weighted average, $\langle \Delta A \rangle$, in Eq. 15 to limit any systematic error introduced by assuming that they can be decomposed into independent blocks of Gly_5. Additionally, to study the effects of interresidue correlations, we computed $\langle \Delta A \rangle$ for all oligoglycines assuming each residue is independent of the others [i.e., $p_i \approx p(d_1)p(d_2)\cdots p(d_{N_{res}})$]. We refer to these estimates as $\langle \Delta A \rangle^{res}$ and those based on the joint-probability distributions (p_i or $\tilde{p}_i$) as $\langle \Delta A \rangle^{oligo}$.

RESULTS

We wish to better understand the thermodynamics associated with conformational transitions of short IDRs and in particular the free energy afforded to such transitions in the form of backbone conformational entropy. We begin by calculating the dihedral angle contribution to the absolute conformational entropy (hereafter simply referred to as conformational entropy) of successively longer oligoglycine polypeptides (Gly_{3-5}, Gly_{10}, and Gly_{15}). Comparing the QHA, BQH, and MIE methods, we calculated the conformational entropy (S) from the ϕ and ψ backbone dihedral angles sampled from MD simulations. Superscripts are used to denote the method used to calculate the conformational entropy, and for the MIE approach, subscripts indicate the order of approximation or truncation strategy (see Materials and Methods). Entropy units (eu) are cal/mol/K. Then, we investigate the conformational entropy as a function of end-to-end distance (R) and radius of gyration (R_g) of Gly_{15} using restrained, implicit solvent simulations because IDR-containing proteins likely impose global, structural constraints on the disordered region, and order-disorder transitions may alter R and R_g. Lastly, we consider the extent of conformational entropy lost or gained as free energy when oligoglycine is constrained to or released from particular conformational states.

Conformational entropy scaling with chain length

Fig. 2 shows the scaling of S^{QHA}, S^{BQH}, S^{MIE}_{res}, $S^{MIE,fit}_2$, and $S^{MIE,fit}_3$ as a function of oligoglycine chain length. $S^{MIE,fit}_2$ and $S^{MIE,fit}_3$ are taken as the asymptotes of hyperbolic fits of S^{MIE}_2 and S^{MIE}_3, respectively, as functions of simulation time for each oligoglycine individually. As discussed in Materials and Methods, S^{MIE}_{res} is an alternative, second-order MIE approximation in which each glycine residue is treated as independent of the others by ignoring MI terms that include dihedrals from two different residues. The data plotted in Fig. 2 are provided in Table S1. The various entropy estimates in Fig. 2 all scale linearly with the number of residues or chain length, with slopes ranging from 3.86–5.57 eu/res or 1.16–1.67 kcal/mol/res at 300 K. Slopes were estimated from least-square fits, all of which yielded values of $R^2 \geq 0.99$.

The difference or range in slopes can be attributed to the assumptions or approximations underlying each method

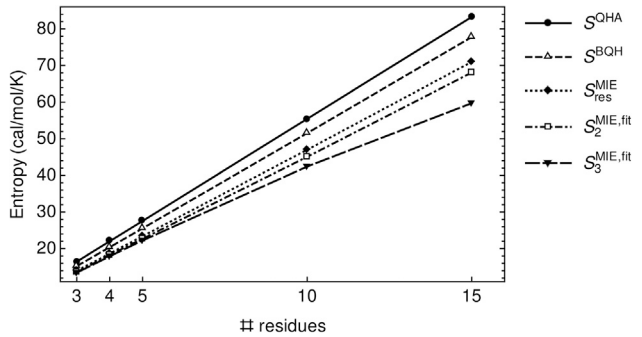

FIGURE 2 Backbone conformational entropy scaling with oligoglycine chain length or number of glycine residues. Superscripts indicate the method used. For the MIE entropy estimates, subscripts denote the order of approximation or truncation strategy. $S^{MIE,fit}_2$ and $S^{MIE,fit}_3$ are taken as the asymptotes of hyperbolic fits of S^{MIE}_2 and S^{MIE}_3, respectively, as functions of time (see Materials and Methods for more details).

used to calculate conformational entropy (Theory). Here, S^{QHA} provides an upper bound on the true conformational entropy, as it effectively merges conformational states (52). As expected, $S^{QHA} > S^{BQH} > S^{MIE}_{res} > S^{MIE,fit}_2 > S^{MIE,fit}_3$ across chain lengths because each method in the given order more accurately captures the multimodal distributions of and/or correlations among dihedral angles than those preceding. With respect to the MIE approach, MI terms may effectively reduce the entropy if there are substantial correlations among dihedrals along the oligoglycine backbone. We find that differences among S^{MIE}_{res}, $S^{MIE,fit}_2$, and $S^{MIE,fit}_3$ are at most $\sim$ 1.1 eu for Gly_3, Gly_4, and Gly_5, suggesting that the most significant correlations exist between ϕ and ψ dihedrals within each residue and that S^{MIE}_{res} largely accounts for the entropic contributions of these correlations. For Gly_{10} and Gly_{15}, differences between S^{MIE}_{res} and $S^{MIE,fit}_2$ increase slightly to $\sim$ 2 eu, whereas a much greater difference is observed between $S^{MIE,fit}_2$ and $S^{MIE,fit}_3$. This large difference between the latter two is likely due to insufficient conformational sampling and the large bin widths required to estimate S^{MIE}_3. All second-order MI terms or entropies were less than 0.02 eu for pairs of dihedrals separated by two or more residues in Gly_{15}, suggesting that long-range correlations are minimal in Gly_N systems even at this longer chain length (data not shown). That significant correlations do not span multiple glycine residues is consistent with various estimates of the persistence length being on the order of 1–2 residues for denatured or unfolded proteins (34,68–70). When excluding $S^{MIE,fit}_3$ for Gly_{15}, linear fits of S^{MIE}_{res}, $S^{MIE,fit}_2$, and $S^{MIE,fit}_3$ yielded slopes of 4.73, 4.52, and 4.11 eu/res, respectively.

For the oligoglycine systems considered here, we find that S^{QHA}, S^{BQH}, and S^{MIE}_{res} converge considerably faster with respect to time than S^{MIE}_2 and, to a greater extent, S^{MIE}_3. As an example, Fig. 3 shows the trajectories of these entropy estimates for Gly_{15}. We also provide trajectories of the MIE entropies for all oligoglycines in Fig. S1. Hyperbolic fits of S^{MIE}_2 and S^{MIE}_3 with respect to simulation time yielded

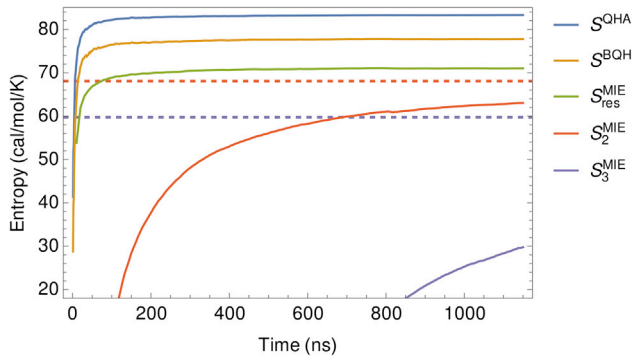

FIGURE 3 Trajectories of Gly_{15} conformational entropy estimates with respect to simulation time. The dashed lines indicate the asymptotes, $S^{MIE,fit}_2$ and $S^{MIE,fit}_3$, of the hyperbolic fits of S^{MIE}_2 and S^{MIE}_3, respectively, and are colored accordingly.

$R^2 \geq 0.99$ for all oligoglycines, allowing us to probe the convergence properties of these entropy estimates. For Gly_{3-5}, sampling was sufficient to converge S^{MIE}_2 and S^{MIE}_3 to within less than 1 eu of their asymptotic values ($S^{MIE,fit}_2$ and $S^{MIE,fit}_3$, respectively). However, for Gly_{15}, we estimate that it would take $\sim$ 5.9 and $\sim$ 34.6 μs of simulation time to converge S^{MIE}_2 and S^{MIE}_3 to within 1 eu of their asymptotes (Table S1). The hyperbolic nature of the S^{MIE}_2 and S^{MIE}_3 trajectories may only be relevant for the oligoglycine systems considered here, in which the residues are highly uncoupled. However, future work is warranted to determine if S^{MIE}_2 and S^{MIE}_3 follow hyperbolic trajectories for other short, disordered polypeptides, as relatively short simulations could be run to establish the initial scaling of the hyperbola after which a least-square fit could be performed to conveniently extract the asymptotes of S^{MIE}_2 and S^{MIE}_3. Error estimates of S^{QHA}, S^{BQH}, and the various first- and second-order MIE entropies were all less than 1 eu with the exception of $S^{MIE,fit}_2$ for Gly_{15} which was $\sim$ 2 eu (Table S1).

Effects of structural constraints on dihedral conformational entropy

To determine if structural transitions altering the end-to-end distance (R) or radius of gyration (R_g) of an IDR elicit a significant entropic change, we performed microsecond implicit solvent simulations of Gly_{15} constrained to eight different values of R. We further partitioned conformations by R_g as described in Materials and Methods. Then, we calculated the conformational entropy using the MIE approach as a function of R and R_g. We find that S^{MIE}_1 (one-dimensional entropy), S^{MIE}_{res}, $S^{MIE,fit}_2$, and $S^{MIE,fit}_3$ are largely independent of R and R_g (Fig. 4; Tables S2 and S3) with the greatest effects observed at their extreme values. Furthermore, these entropy estimates are also very consistent with those obtained from the unconstrained, explicit solvent simulations (Table S1). It appears, then, that global, structural changes in R or R_g produce only a minimal change in backbone entropy when compared to the absolute conformational entropy that could potentially be lost or gained up such structural transitions. In the following section, we explore the extent of conformational entropy lost as free energy when individual residues are constrained to particular regions of the ϕ-ψ map (i.e., local versus global constraints).

Ensemble confinement free energy

The ensemble confinement free energy (ΔA_i) presented in Theory and Materials and Methods estimates the gain (loss) of conformational free energy when oligoglycine is constrained to (released from) a conformational state (i), which is defined by assigning each glycine residue to one

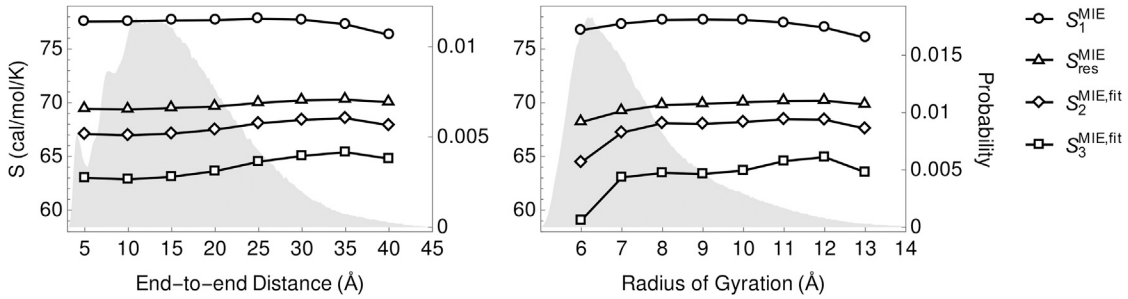

FIGURE 4 MIE conformational entropy estimates of Gly_{15} as a function of end-to-end distance (*left*) and radius of gyration (*right*). Implicit solvent simulations of Gly_{15} were performed with the distance between terminal carbons constrained to eight different values with a harmonic bias potential. Trajectories were then partitioned by radius of gyration. The first-order MIE approximation is given as $\sum_{i=1}^{2N_{res}} S_1(q_i)$ (i.e., the sum over one-dimensional terms in Eq. 9). Probability distributions of end-to-end distance and radius of gyration from explicit solvent simulations of Gly_{15} are shaded in gray.

of six regions of the ϕ-ψ map (Fig. 1). Equation 14 gives $\Delta A_i = -k_B T \ln p_i$, where p_i is the probability of observing oligoglycine in state i. For Gly_{3-5}, p_i is calculated directly from the observed frequencies of state i from simulation, whereas for Gly_{10} and Gly_{15}, p_i is approximated by assuming that they are composed of two and three independent, contiguous blocks of Gly_5, respectively. ΔA_i calculated in this manner is denoted as ΔA_i^{oligo}. Additionally, to study the effects of correlations among residues, we calculate ΔA_i by factorizing p_i as the product of the probabilities of each residue being in one of the six predefined regions of the ϕ-ψ map. These estimates carry the superscript "res." An ensemble average of ΔA_i^{oligo} and ΔA_i^{res} across observed conformational states (Eq. 15) gives $\langle \Delta A \rangle^{oligo} = TS_{oligo}^{state}$ and $\langle \Delta A \rangle^{res} = TS_{res}^{state}$, respectively, where S^{state} is the entropy related to the number of possible oligoglycine conformers.

We find that both $\langle \Delta A \rangle^{oligo}$ and $\langle \Delta A \rangle^{res}$ scale linearly with oligoglycine chain length (Table 1). Least-square fits of $\langle \Delta A \rangle^{oligo}$ and $\langle \Delta A \rangle^{res}$ yielded similar slopes of 0.95 and 0.99 kcal/mol/res, respectively. The corresponding slopes of S_{oligo}^{state} and S_{res}^{state} are 3.15 and 3.28 eu/res, which represent a significant fraction of the absolute conformational entropy per residue (Table S1). In other words, a large portion of the conformational entropy, on average, is lost as free energy when the structural ensemble is confined to particular states. For example, the slope of S_{res}^{state} is roughly 69% of that estimated for S_{res}^{MIE}, and the remaining entropy can be attributed to the internal entropy within the conformational states, which is often referred to as vibrational entropy (53,54). The slope of S_{oligo}^{state} is 85% of that measured for $S_3^{MIE,fit}$ and 80% when excluding Gly_{15}. Whereas for well-structured proteins the loss of backbone vibrational entropy upon binding may dominate that associated with the loss of the number of rotamers, the opposite may be true for IDRs. That $\langle \Delta A \rangle^{oligo}$ and $\langle \Delta A \rangle^{res}$ exhibit similar scaling profiles with respect to oligoglycine length again suggests that oligoglycine can be reasonably modeled as a system of uncoupled glycine residues over the length scale considered here. These results were also independently confirmed using the program CENCALC (71).

TABLE 1 Average Confinement Free Energy and Corresponding Conformational Entropy as a Function of Oligoglycine Chain Length

Gly	$\langle \Delta A \rangle^{oligo}$	$\langle \Delta A \rangle^{res}$	S_{oligo}^{state}	S_{res}^{state}
3	2.87 (0.01)	2.89 (0.02)	9.57	9.63
4	3.80 (0.02)	3.85 (0.03)	12.67	12.83
5	4.72 (0.01)	4.82 (0.02)	15.73	16.07
10	9.44 (0.03)	9.74 (0.03)	31.47	32.47
15	14.20 (0.11)	14.70 (0.04)	47.33	49.00
Slope	0.95	0.99	3.15	3.28
R^2	0.99	0.99		

Free energies are reported in kcal/mol and entropy is measured in cal/mol/K. Slopes were estimated from linear least-squares fit. For Gly_{3-5}, $\langle \Delta A \rangle^{oligo}$ was calculated from the joint distributions of ϕ-ψ dihedral state assignments across residues, whereas those for Gly_{10} and Gly_{15} were approximated using the joint distributions of two and three consecutive segments of Gly_5, respectively. $\langle \Delta A \rangle^{res}$ is estimated from the product of the marginal distributions of the per residue state assignments (see Materials and Methods). Errors in the ensemble confinement free energy were approximated as the SD of $\langle \Delta A \rangle^{oligo}$ or $\langle \Delta A \rangle^{res}$ calculated from five equally sized blocks of the MD trajectory. These are provided in parentheses. S_{oligo}^{state} and S_{res}^{state} are calculated from $\langle \Delta A \rangle^{oligo}$ and $\langle \Delta A \rangle^{res}$, respectively, with Eq. 15.

Next, we provide more detail on the results for Gly_{15}. Over the ~ 1.1 μs simulation, Gly_{15} visited roughly 400,000 unique conformational states. Fig. 5 (*left*) shows a plot of ΔA_i^{oligo} across the visited states, i, which are arbitrarily numbered and ordered by increasing values of ΔA_i^{oligo}. ΔA_i^{oligo} is large and unfavorable and spans a range of $\sim$ 10–20 kcal/mol, corresponding to the most and least energetically favorable or probable states, respectively. The characteristic sigmoidal shape of ΔA_i^{oligo} shows that there are a large number of somewhat energetically degenerate conformational states with very similar values of ΔA_i^{oligo}. Conformational states with the lowest values of ΔA_i^{oligo} are characterized by a higher polyproline II content (Fig. 5, *right*), whereas those states with the highest values are composed of residues more equally proportioned across the six major regions of the ϕ-ψ map (Fig. 1). Interestingly, the confinement free energy is insensitive to α-helical content.

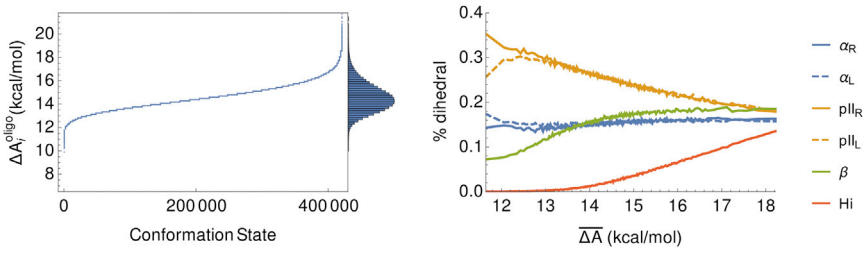

FIGURE 5 (*Left*) Gly_{15} confinement free energy as a function of conformational states (*i*). The vertical bar chart is a histogram of ΔA_i^{oligo} values. (*Right*) A fraction of residues falling within the six partitions of the ϕ-ψ map (see Fig. 1) are shown as a function of the average confinement free energy ($\overline{\Delta A}$) taken over a sliding, nonoverlapping window of 100 conformational states, ordered by increasing values of ΔA_i^{oligo}. A windowing average was taken to reduce noise and is different than the weighted average, $\langle \Delta A \rangle^{oligo}$, which is taken across all conformations.

DISCUSSION

The ability of an IDR to undergo structural transitions is key to disorder-mediated recognition and allostery (10,14,15). Toward providing bounds on the thermodynamics associated with such transitions, we consider the conformational entropy and free energy of oligoglycine, a model IDR and protein backbone model. We began first by calculating the absolute backbone conformational entropy of oligoglycine as a function of chain length over a biologically relevant range using the QHA, BQH, and MIE methods. The conformational entropy is significant, and all methods consistently report a linear scaling of backbone conformational entropy with chain length (Fig. 2), yet with different slopes. Recently, Towse et al. observed a linear scaling of backbone conformational entropy with chain length (up to ~400 residues), albeit with considerable spread, from a large-scale analysis of 807 structured proteins in the Dynameomics MD data set (72), suggesting that this linear relationship may be robust with respect to chain length and sequence space. We find that S_{res}^{MIE}, which scales at $\sim$ 1.4 kcal/mol/res, captures the most significant entropic contributions from pairwise dihedral correlations along the oligoglycine backbone.

Disorder-to-order transitions, commonly observed in short IDRs on the order of $\sim$ 10 amino acids that bind target proteins, necessitate a significant loss of conformational entropy (or gain of free energy). We estimate the loss of conformational entropy (ΔS) from the absolute conformational entropy scaling profiles in Fig. 2. The MIE estimates yield $T\Delta S$ = 1.16–1.42 kcal/mol/res at 300 K, which is consistent with the loss of entropy upon folding of well-structured proteins reported from a number of different experimental and computational studies (67,72,73). The ensemble confinement free energies, $\langle \Delta A \rangle^{res}$ and $\langle \Delta A \rangle^{oligo}$, which implicitly account for intrachain and chain-solvent enthalpic interactions, also scale linearly with chain length with slopes of 0.94 and 0.99 kcal/mol/res, respectively. Although of similar magnitude, these slopes are slightly less than those based on the absolute backbone entropy because, by definition, each conformational state retains internal entropy. From an analysis of 103 polypeptide-protein complexes, London et al. (74) found that polypeptide-protein interfaces use significantly more main-chain/main-chain hydrogen bonds than larger protein-protein interfaces. The protein backbone appears equipped to provide, at least partially, compensating enthalpic interactions to promote folding upon binding or recognition.

Conversely, order-to-disorder transitions, which occur in response to allosteric effector binding and environmental changes (13,16,17,23), permit the protein backbone access to the disordered ensemble with a concomitant, substantial increase in conformational entropy ($-\langle \Delta A \rangle^{oligo} \approx -14$ kcal/mol for Gly_{15}) and change in functional state. The entropic expansion of conformation space resulting from local protein unfolding has been proposed as a general allosteric mechanism (10,15,16,75) and a possible mode of targeting IDRs with small molecules (76). Lastly, protein/ligand binding and changes in cellular environment can remodel the disordered structural ensemble of an IDR (10) (e.g., extended versus collapsed disorder). We found that the conformational entropy of Gly_{15} is largely independent of end-to-end distance and radius of gyration, suggesting that backbone conformational entropy does not oppose ensemble remodeling—a property that may be necessary for certain types of disorder-mediated allostery.

IDRs as entropic reservoirs

Wand and co-workers proposed that the residual conformational entropy of proteins provides an entropic reservoir that may be energetically coupled to ligand binding (32,77). Evidence continues to mount that supports this idea (29–32,77). Changes in conformational entropy primarily attributed to changes in side-chain dynamics or fluctuations have been shown to promote or suppress the binding of structured proteins to their targets. The ability of zinc to allosterically inhibit homodimeric CzrA (chromosomal zinc-regulated repressor) binding to DNA is a particularly interesting example (30). Upon binding DNA, there is an increase in side-chain motion in CzrA that was estimated to be a significant contribution to the total favorable change in entropy. When bound, zinc appears to prevent this favorable change in entropy by preventing an increase in side-chain dynamics when binding to DNA, thus decreasing CzrA:DNA binding affinity.

In well-structured proteins, the protein backbone may not have the structural plasticity or capacity to alter its dynamics to the extent observed for side chains (31,77). That is, in well-structured proteins, side chains may represent an entropic (energetic) reservoir. Analogously, extending this concept to IDRs, we may consider the protein backbone as an entropic reservoir from which free energy may be extracted or deposited through IDR order-disorder transitions that alter the backbone conformational entropy.

To illustrate, we consider an idealized, disorder-mediated model of allostery (Fig. 6) that parallels the zinc-binding negative regulation of CzrA discussed above. In this example, a transcription factor (TF) composed of a DNA-binding domain and a regulatory domain (RD) binds to DNA, resulting in an order-to-disorder transition in the RD. The entropic expansion associated with the unfolding of RD contributes favorably to binding (ΔA^1). A cofactor (CF) can decrease the affinity of the transcription factor for DNA (i.e., $\Delta\Delta A = \Delta A^3 - \Delta A^1 = \Delta A^4 - \Delta A^2 > 0$) either by binding to and stabilizing the RD (ΔA^2), thus preventing the entropically favorable structural transition of RD, or by destabilizing the TF:DNA complex by reducing the conformational entropy of RD (ΔA^4). In either case, the increase in the free energy or chemical potential of the CF:TF:DNA complex shifts the equilibrium to favor the state in which the transcription factor is not bound to DNA (Fig. 6, *bottom left*) in a manner dependent on the concentration of CF. Taking Gly_{15} as a model for RD, we can approximate a bound on the decrease in the affinity across the possible confined states of Gly_{15}, as $0 < \Delta\Delta A < \langle A \rangle^{\text{oligo}} = TS^{\text{state}}_{\text{oligo}} \approx -14$ kcal/mol (Eq. 15). Although we have clearly neglected a number of (compensatory) thermodynamic contributions needed for a more physically representative model, our goal with this example is to illustrate the concept of the protein backbone as a free-energy reservoir and the potentially significant energetic/entropic contribution of IDR order-disorder transitions to binding.

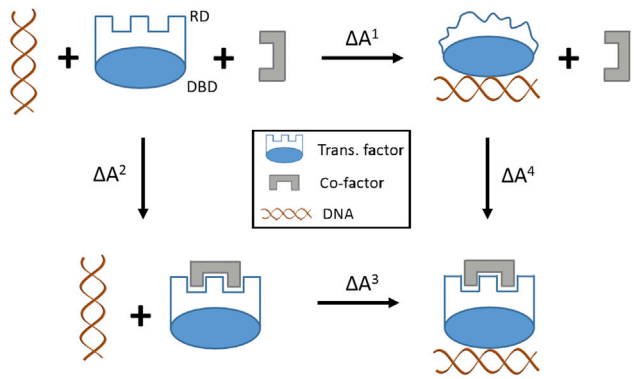

FIGURE 6 Thermodynamic cycle of an idealized transcription factor composed of a DNA-binding domain and regulatory domain (RD), exhibiting disorder-mediated, negative allostery. Upon binding DNA, the RD undergoes an order-to-disorder transition that entropically promotes binding. A cofactor may bind to the RD and prevent the favorable increase in conformational entropy (*bottom right*), thus decreasing the binding affinity of the transcription factor for its target DNA sequence. To see this figure in color, go online.

CONCLUSIONS

In this article, our goal is to illustrate that the conformational entropy of the protein backbone represents a significant source of free energy that proteins may exploit through conformational transitions of IDRs as a means to regulate protein binding and, ultimately, function. We believe that our oligoglycine model provides an upper bound on the amount of conformational entropy potentially gained or lost as free energy when folding or unfolding of IDRs is coupled to binding. Lastly, we note that the development of force fields for simulations of intrinsically disordered proteins is an active field, as such systems continue to challenge the accuracy of existing force fields (78). Previously, we found that the solvation thermodynamics of Gly_{2-5} calculated from simulations with the CHARMM (Chemistry at Harvard Macromolecular Mechanics) 36 (79) and Amber ff12SB (60) force fields were very consistent (62) despite the two generating markedly different structural ensembles. In the future, we plan to investigate whether the conformational entropy estimates studied in this article are similarly insensitive to differences or perturbations in the structural ensemble of the same disordered polypeptide model.

SUPPORTING MATERIAL

One figure and three tables are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(18)30521-6.

AUTHOR CONTRIBUTIONS

J.A.D. performed all calculations and wrote the initial manuscript draft. Both authors conceptualized and planned the project and worked on the analysis of the results and the final manuscript.

ACKNOWLEDGMENTS

The authors thank Dr. Cheng Zhang for help with the analysis and a careful reading of the manuscript. The authors also acknowledge the Texas Advanced Computing Center at The University of Texas at Austin for providing high-performance computing resources that have contributed to the research results reported within this article. URL: http://www.tacc.utexas.edu.

We gratefully acknowledge the Robert A. Welch Foundation (H-0037), the National Institutes of Health (GM-037657), and the National Center for Supercomputing Applications Blue Waters Graduate Research Fellowship for partial support of this work. This work used the Extreme Science and Engineering Discovery Environment, which is supported by National Science Foundation grant number ACI-1548562. Additionally, this research is part of the Blue Waters sustained petascale computing project, which is supported by the National Science Foundation (award numbers OCI-0725070 and ACI-1238993).

REFERENCES

1. Dunker, A. K., C. J. Brown, …, Z. Obradović. 2002. Intrinsic disorder and protein function. *Biochemistry.* 41:6573–6582.
2. Oldfield, C. J., and A. K. Dunker. 2014. Intrinsically disordered proteins and intrinsically disordered protein regions. *Annu. Rev. Biochem.* 83:553–584.
3. van der Lee, R., M. Buljan, …, M. M. Babu. 2014. Classification of intrinsically disordered regions and proteins. *Chem. Rev.* 114:6589–6631.
4. Uversky, V. N., C. J. Oldfield, and A. K. Dunker. 2005. Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. *J. Mol. Recognit.* 18:343–384.
5. DeForte, S., and V. N. Uversky. 2016. Order, disorder, and everything in between. *Molecules.* 21:1–22.
6. Frankel, A. D., and P. S. Kim. 1991. Modular structure of transcription factors: implications for gene regulation. *Cell.* 65:717–719.
7. Liu, Y., K. S. Matthews, and S. E. Bondos. 2008. Multiple intrinsically disordered sequences alter DNA binding by the homeodomain of the Drosophila hox protein ultrabithorax. *J. Biol. Chem.* 283:20874–20887.
8. Liu, Y., K. S. Matthews, and S. E. Bondos. 2009. Internal regulatory interactions determine DNA binding specificity by a Hox transcription factor. *J. Mol. Biol.* 390:760–774.
9. Hilser, V. J., and E. B. Thompson. 2011. Structural dynamics, intrinsic disorder, and allostery in nuclear receptors as transcription factors. *J. Biol. Chem.* 286:39675–39682.
10. Tompa, P. 2014. Multisteric regulation by structural disorder in modular signaling proteins: an extension of the concept of allostery. *Chem. Rev.* 114:6715–6732.
11. Uversky, V. N., C. J. Oldfield, and A. K. Dunker. 2008. Intrinsically disordered proteins in human diseases: introducing the D2 concept. *Annu. Rev. Biophys.* 37:215–246.
12. Van Roey, K., B. Uyar, …, N. E. Davey. 2014. Short linear motifs: ubiquitous and functionally diverse protein interaction modules directing cell regulation. *Chem. Rev.* 114:6733–6778.
13. Uversky, V. N. 2013. Unusual biophysics of intrinsically disordered proteins. *Biochim. Biophys. Acta.* 1834:932–951.
14. Flock, T., R. J. Weatheritt, …, M. M. Babu. 2014. Controlling entropy to tune the functions of intrinsically disordered regions. *Curr. Opin. Struct. Biol.* 26:62–72.
15. Motlagh, H. N., J. O. Wrabl, …, V. J. Hilser. 2014. The ensemble nature of allostery. *Nature.* 508:331–339.
16. Jakob, U., R. Kriwacki, and V. N. Uversky. 2014. Conditionally and transiently disordered proteins: awakening cryptic disorder to regulate protein function. *Chem. Rev.* 114:6779–6805.
17. Zea, D. J., A. M. Monzon, …, G. Parisi. 2016. Disorder transitions and conformational diversity cooperatively modulate biological function in proteins. *Protein Sci.* 25:1138–1146.
18. Mohan, A., C. J. Oldfield, …, V. N. Uversky. 2006. Analysis of molecular recognition features (MoRFs). *J. Mol. Biol.* 362:1043–1059.
19. Davey, N. E., K. Van Roey, …, T. J. Gibson. 2012. Attributes of short linear motifs. *Mol. Biosyst.* 8:268–281.
20. Dinkel, H., K. Van Roey, …, T. J. Gibson. 2014. The eukaryotic linear motif resource ELM: 10 years and counting. *Nucleic Acids Res.* 42:D259–D266.
21. Staby, L., C. O'Shea, …, K. Skriver. 2017. Eukaryotic transcription factors: paradigms of protein intrinsic disorder. *Biochem. J.* 474:2509–2532.
22. Hsu, W. L., C. J. Oldfield, …, A. K. Dunker. 2013. Exploring the binding diversity of intrinsically disordered proteins involved in one-to-many binding. *Protein Sci.* 22:258–273.
23. Amemiya, T., R. Koike, …, A. Kidera. 2011. Classification and annotation of the relationship between protein structural change and ligand binding. *J. Mol. Biol.* 408:568–584.
24. Motlagh, H. N., J. A. Anderson, …, V. J. Hilser. 2015. Disordered allostery: lessons from glucocorticoid receptor. *Biophys. Rev.* 7:257–265.
25. Hilser, V. J., and E. B. Thompson. 2007. Intrinsic disorder as a mechanism to optimize allosteric coupling in proteins. *Proc. Natl. Acad. Sci. USA.* 104:8311–8315.
26. Nussinov, R. 2016. Introduction to protein ensembles and allostery. *Chem. Rev.* 116:6263–6266.
27. Marlow, M. S., J. Dogan, …, A. J. Wand. 2010. The role of conformational entropy in molecular recognition by calmodulin. *Nat. Chem. Biol.* 6:352–358.
28. Wand, A. J., V. R. Moorman, and K. W. Harpole. 2013. A surprising role for conformational entropy in protein function. *Top. Curr. Chem.* 337:69–94.
29. Tzeng, S. R., and C. G. Kalodimos. 2012. Protein activity regulation by conformational entropy. *Nature.* 488:236–240.
30. Capdevila, D. A., J. J. Braymer, …, D. P. Giedroc. 2017. Entropy redistribution controls allostery in a metalloregulatory protein. *Proc. Natl. Acad. Sci. USA.* 114:4424–4429.
31. Caro, J. A., K. W. Harpole, …, A. J. Wand. 2017. Entropy in molecular recognition by proteins. *Proc. Natl. Acad. Sci. USA.* 114:6563–6568.
32. Wand, A. J. 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.
33. Auton, M., and D. W. Bolen. 2004. Additive transfer free energies of the peptide backbone unit that are independent of the model compound and the choice of concentration scale. *Biochemistry.* 43:1329–1342.
34. Tran, H. T., A. Mao, and R. V. Pappu. 2008. Role of backbone-solvent interactions in determining conformational equilibria of intrinsically disordered proteins. *J. Am. Chem. Soc.* 130:7380–7392.
35. Hu, C. Y., H. Kokubo, …, B. M. Pettitt. 2010. Backbone additivity in the transfer model of protein solvation. *Protein Sci.* 19:1011–1022.
36. Drake, J. A., R. C. Harris, and B. M. Pettitt. 2016. Solvation thermodynamics of oligoglycine with respect to chain length and flexibility. *Biophys. J.* 111:756–767.
37. Uversky, V. N., and A. K. Dunker. 2010. Understanding protein nonfolding. *Biochim. Biophys. Acta.* 1804:1231–1264.
38. Marsh, J. A., and J. D. Forman-Kay. 2010. Sequence determinants of compaction in intrinsically disordered proteins. *Biophys. J.* 98:2383–2390.
39. Piovesan, D., F. Tabaro, …, S. C. E. Tosatto. 2017. DisProt 7.0: a major update of the database of disordered proteins. *Nucleic Acids Res.* 45:D219–D227.
40. Karplus, M., and J. N. Kushick. 1981. Method for estimating the configurational entropy of macromolecules. *Macromolecules.* 14:325–332.
41. Levy, R. M., M. Karplus, …, D. Perahia. 1984. Evaluation of the configurational entropy for proteins: application to molecular dynamics simulations of an α-helix. *Macromolecules.* 17:1370–1374.
42. Di Nola, A., H. J. Berendsen, and O. Edholm. 1984. Free energy determination of polypeptide conformations generated by molecular dynamics. *Macromolecules.* 17:2044–2050.
43. Harpole, K. W., and K. A. Sharp. 2011. Calculation of configurational entropy with a Boltzmann-quasiharmonic model: the origin of high-affinity protein-ligand binding. *J. Phys. Chem. B.* 115:9461–9472.
44. Hikiri, S., T. Yoshidome, and M. Ikeguchi. 2016. Computational methods for configurational entropy using internal and Cartesian coordinates. *J. Chem. Theory Comput.* 12:5990–6000.
45. Killian, B. J., J. Yundenfreund Kravitz, and M. K. Gilson. 2007. Extraction of configurational entropy from molecular simulations via an expansion approximation. *J. Chem. Phys.* 127:024107.
46. Hnizdo, V., and M. K. Gilson. 2010. Thermodynamic and differential entropy under a change of variables. *Entropy (Basel).* 12:578–590.
47. Gō, N., and H. A. Scheraga. 1976. On the use of classical statistical mechanics in the treatment of polymer chain conformation. *Macromolecules.* 9:535–542.

48. Herschbach, D. R., H. S. Johnston, and D. Rapp. 1959. Molecular partition functions in terms of local properties. *J. Chem. Phys.* 31:1652–1661.

49. Chang, C.-E., M. J. Potter, and M. K. Gilson. 2003. Calculation of molecular configuration integrals. *J. Phys. Chem. B.* 107:1048–1055.

50. Li, D.-W., and R. Brüschweiler. 2009. *In silico* relationship between configurational entropy and soft degrees of freedom in proteins and peptides. *Phys. Rev. Lett.* 102:118108.

51. Kassem, S., M. Ahmed, …, K. H. Barakat. 2015. Entropy in bimolecular simulations: a comprehensive review of atomic fluctuations-based methods. *J. Mol. Graph. Model.* 62:105–117.

52. Chang, C. E., W. Chen, and M. K. Gilson. 2005. Evaluating the accuracy of the quasiharmonic approximation. *J. Chem. Theory Comput.* 1:1017–1028.

53. Karplus, M., T. Ichiye, and B. M. Pettitt. 1987. Configurational entropy of native proteins. *Biophys. J.* 52:1083–1085.

54. Suárez, E., N. Díaz, and D. Suárez. 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.* 7:2638–2653.

55. Cukier, R. I. 2015. Dihedral angle entropy measures for intrinsically disordered proteins. *J. Phys. Chem. B.* 119:3621–3634.

56. Teufel, D. P., C. M. Johnson, …, H. Neuweiler. 2011. Backbone-driven collapse in unfolded protein chains. *J. Mol. Biol.* 409:250–262.

57. Auton, M., J. Rösgen, …, D. W. Bolen. 2011. Osmolyte effects on protein stability and solubility: a balancing act between backbone and side-chains. *Biophys. Chem.* 159:90–99.

58. Karandur, D., R. C. Harris, and B. M. Pettitt. 2016. Protein collapse driven against solvation free energy without H-bonds. *Protein Sci.* 25:103–110.

59. Bondos, S. E., X.-X. Tan, and K. S. Matthews. 2006. Physical and genetic interactions link hox function with diverse transcription factors and cell signaling proteins. *Mol. Cell. Proteomics.* 5:824–834.

60. Case, D., J. Berryman, …, P. Kollman. 2012. Amber 12. University of California, San Francisco.

61. Phillips, J. C., R. Braun, …, K. Schulten. 2005. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* 26:1781–1802.

62. Drake, J. A., and B. M. Pettitt. 2015. Force field-dependent solution properties of glycine oligomers. *J. Comput. Chem.* 36:1275–1285.

63. Wang, J., and R. Brüschweiler. 2006. 2D entropy of discrete molecular ensembles. *J. Chem. Theory Comput.* 2:18–24.

64. Hnizdo, V., E. Darian, …, H. Singh. 2007. Nearest-neighbor nonparametric method for estimating the configurational entropy of complex molecules. *J. Comput. Chem.* 28:655–668.

65. Onufriev, A., D. Bashford, and D. A. Case. 2000. Modification of the generalized born model suitable for macromolecules. *J. Phys. Chem. B.* 104:3712–3720.

66. Onufriev, A., D. Bashford, and D. A. Case. 2004. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins.* 55:383–394.

67. Baxa, M. C., E. J. Haddadian, …, T. R. Sosnick. 2014. Loss of conformational entropy in protein folding calculated using realistic ensembles and its implications for NMR-based calculations. *Proc. Natl. Acad. Sci. USA.* 111:15396–15401.

68. Hofmann, H., A. Soranno, …, B. Schuler. 2012. Polymer scaling laws of unfolded and intrinsically disordered proteins quantified with single-molecule spectroscopy. *Proc. Natl. Acad. Sci. USA.* 109:16155–16160.

69. Zhou, H. X. 2004. Polymer models of protein stability, folding, and interactions. *Biochemistry.* 43:2141–2154.

70. Stirnemann, G., D. Giganti, …, B. J. Berne. 2013. Elasticity, structure, and relaxation of extended proteins under force. *Proc. Natl. Acad. Sci. USA.* 110:3847–3852.

71. Suárez, E., N. Díaz, J. Méndez, and D. Suárez.ba. 2013. CENCALC: a computational tool for conformational entropy calculations from molecular simulations. *J. Computat. Chem.* 34:2041–2054.

72. Towse, C. L., M. Akke, and V. Daggett. 2017. The dynameomics entropy dictionary: a large-scale assessment of conformational entropy across protein fold space. *J. Phys. Chem. B.* 121:3933–3945.

73. Thompson, J. B., H. G. Hansma, …, K. W. Plaxco. 2002. The backbone conformational entropy of protein folding: experimental measures from atomic force microscopy. *J. Mol. Biol.* 322:645–652.

74. London, N., D. Movshovitz-Attias, and O. Schueler-Furman. 2010. The structural basis of peptide-protein binding strategies. *Structure.* 18:188–199.

75. Mitrea, D. M., and R. W. Kriwacki. 2013. Regulated unfolding of proteins in signaling. *FEBS Lett.* 587:1081–1088.

76. Heller, G. T., P. Sormanni, and M. Vendruscolo. 2015. Targeting disordered proteins with small molecules using entropy. *Trends Biochem. Sci.* 40:491–496.

77. Sharp, K. A., E. O'Brien, …, A. J. Wand. 2015. On the relationship between NMR-derived amide order parameters and protein backbone entropy changes. *Proteins.* 83:922–930.

78. Huang, J., and A. D. MacKerell, Jr. 2018. Force field development and simulations of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* 48:40–48.

79. Best, R. B., X. Zhu, …, A. D. Mackerell, Jr. 2012. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone φ, ψ and side-chain $\chi(1)$ and $\chi(2)$ dihedral angles. *J. Chem. Theory Comput.* 8:3257–3273.