Solvation Thermodynamics of Oligoglycine with Respect to Chain Length and Flexibility

Justin A. Drake,[1] Robert C. Harris,[2] and B. Montgomery Pettitt[1,*]

[1]Department of Biochemistry and Molecular Biology, Sealy Center for Structural Biology and Molecular Biophysics, University of Texas Medical Branch, Galveston, Texas; and [2]Department of Chemistry, Temple University, Philadelphia, Pennsylvania

ABSTRACT   Oligoglycine is a backbone mimic for all proteins and is prevalent in the sequences of intrinsically disordered proteins. We have computed the absolute chemical potential of glycine oligomers at infinite dilution by simulation with the CHARMM36 and Amber ff12SB force fields. We performed a thermodynamic decomposition of the solvation free energy (ΔG^{sol}) of Gly_{2-5} into enthalpic (ΔH^{sol}) and entropic (ΔS^{sol}) components as well as their van der Waals and electrostatic contributions. Gly_{2-5} was either constrained to a rigid/extended conformation or allowed to be completely flexible during simulations to assess the effects of flexibility on these thermodynamic quantities. For both rigid and flexible oligoglycine models, the decrease in ΔG^{sol} with chain length is enthalpically driven with only weak entropic compensation. However, the apparent rates of decrease of ΔG^{sol}, ΔH^{sol}, ΔS^{sol}, and their elec and vdw components differ for the rigid and flexible models. Thus, we find solvation entropy does not drive aggregation for this system and may not explain the collapse of long oligoglycines. Additionally, both force fields yield very similar thermodynamic scaling relationships with respect to chain length despite both force fields generating different conformational ensembles of various oligoglycine chains.

INTRODUCTION

The protein trinity hypothesis says that proteins may exist in an ordered, collapsed-disordered (molten globule), or extended-disordered (random coil) conformational state, and that these states give rise to particular, biological functions (1,2). Intrinsically disordered proteins (IDPs) or regions (IDRs) within proteins present a diverse ensemble of structures with a range of molecular dimensions. To capture this diversity, Uversky (3) proposed the addition of an intermediate premolten globule thermodynamic state characterized with a compactness and residual structure between that of the collapsed and extended states. An IDP or IDR's occupancy in and interconversion between these states depends, in part, on sequence composition and chain length (2,4), while other processes (e.g., binding, allostery, posttranslational modifications) shift the equilibrium between these states. Here we are interested in how the innate properties of a short, model IDP may affect the competition between intrapeptide and peptide-solvent interactions and thus result in extended or compact conformational ensembles of relevance to longer IDPs.

Extended IDPs tend to lack hydrophobic side chains and often have a large overall net charge (3–7). Many IDPs collapse or aggregate despite the absence of hydrophobic side chains and/or the presence of a large, net charge (8–11). Oligoglycine (i.e., the common protein backbone) is a particularly interesting example, as it is found in many IDRs (12,13), exhibits IDR qualities, and has been shown to collapse in a length-dependent manner (14–17). High glycine content has also been associated with compact IDRs (5). Evidence suggests that as chain length increases, favorable protein backbone interactions out-compete backbone-solvent interactions resulting in oligoglycine collapse (15,16). The fact that the solubility of oligoglycine decreases dramatically with chain length ($N = 1–5$) further supports this idea (18). However, whether favorable intrapeptide interactions are due to H-bonding, electrostatic (elec), and/or van der Waals (vdw) interactions, is a subject of debate (14–16,19). Interestingly, the solvation free energy of fixed conformations of oligoglycine in infinite dilution is negative and continues to decrease with chain length (20,21)—further suggesting that collapse may not be a consequence of unfavorable solvent interactions, but due to the availability of more potential intrapeptide interactions. Few studies have considered the entropic contributions to the solvation free energy of oligoglycines in particular—or IDPs in general—as a function of chain

CrossMark

length. While the hydrophobic effect (i.e., solvation entropic penalty) is known to play a key role in protein folding, its importance to aggregation and collapse of IDPs is less clear.

Using computational free energy methods, we performed a decomposition of the solvation thermodynamics for successively longer oligoglycine polypeptides (Gly$_2$-Gly$_5$). We use multistage free energy perturbation (FEP) to calculate solvation free energy (ΔG^{sol}) and its vdw (ΔG^{vdw}) and elec (ΔG^{elec}) components. We further decomposed ΔG^{sol}, ΔG^{vdw}, and ΔG^{elec} into their entropic and enthalpic components. FEP simulations were performed with oligoglycine models either constrained to a rigid-extended conformation or completely flexible (i.e., without positional constraints) using both the CHARMM36 (C36) [22] and AMBER ff12SB [23] force fields. We consider rigid and flexible oligoglycine models because we have previously shown how ΔG^{sol} depends on flexibility and geometry [20,24,25]. C36 and ff12SB are used to discriminate and control for force field effects and to investigate how the differences observed in the structural properties of oligoglycine predicted by these two force fields [17] affect the solvation thermodynamics.

We find that ΔG^{vdw} and ΔG^{elec} decrease with chain length but their magnitude and rate of decrease depends on conformational flexibility. Despite force-field-dependent structural properties of oligoglycine, both force fields yield very similar thermodynamic scaling profiles with respect to chain length. As chain length increases, $-T\Delta S^{\text{sol}}$ becomes more unfavorable but is offset by a more favorable ΔH^{sol} for both rigid and flexible oligoglycines. Differences in $-T\Delta S^{\text{sol}}$ and ΔH^{sol} between flexible and rigid oligoglycines is pronounced; however, due to their compensatory nature, they yield small differences in ΔG^{sol}. Significant compaction of oligoglycine is predicted for chains containing >10 glycine residues [14–16]. The question remains whether the solvation thermodynamics of short oligoglycines can predict (i.e., group-additivity) the solvation free energy of longer oligoglycine chains that can form complex, compact structures.

Both free energy error estimates and the necessary computational resources increase with chain length and limit this study to relatively short oligoglycines. However, our results help to provide mechanistic insight into the forces that may drive the collapse of longer oligoglycines.

MATERIALS AND METHODS

FEP simulations were performed for a rigid-extended conformation of Gly$_2$, Gly$_3$, and Gly$_4$ as well as for completely flexible Gly$_{2-5}$ using NAMD 2.10 [26] with the CHARMM36 (C36) [22] and AMBER ff12SB force fields [23] at 300 and 320 K. Solvation free energy, ΔG^{sol}, was calculated by first scaling the vdw interactions with the solvent followed by elec charging by means of a coupling parameter, λ. Below we describe the setup of the simulations, common simulation parameters, parameters specific to FEP, and the analysis methods.

System

For simulations with C36, oligoglycines were built in a fully extended state using the VMD plugin MoleFacture [27] and capped with neutral acetyl (ACE) and N-methylamide (NME) groups. They were similarly built using XLeap in AmberTools [23] for simulations with ff12SB. Systems were solvated with TIP3P water using either VMD's Solvate plugin or XLeap ensuring at least a 10 Å padding between oligoglycine and the sides of the simulation box.

General simulation parameters

All simulations were performed at constant temperature (either 300 or 320 K) and pressure (1 atm) using a Langevin thermostat and barostat. NAMD's default switching function at 1.0 nm was used to smoothly truncate nonbonded interactions at 1.2 nm. Electrostatic forces were calculated using particle mesh Ewald on a grid with a 1.0 Å spacing. The 1–4 scaling was set to 0.8333 for simulations with ff12SB to match AMBER's nonbonded exclusion convention. The velocity Verlet algorithm with a 2 fs time step was used to integrate the equations of motion. Coordinates and energies were saved every 1 ps.

Free energy perturbation simulations

To set up the FEP simulations, we first performed a steepest descent minimization for each system, allowing only water coordinates to change for simulations with fixed oligoglycine. Configurations for FEP simulations of flexible oligoglycines were taken either from a short equilibration run or from the final configurations from long molecular dynamics (MD) simulations from previous studies in our lab at constant temperature and pressure. These configurations were used to initiate the FEP simulations at 300 K.

The process of solvating each oligoglycine was decomposed into two pathways following a protocol similar to that of Kokubo et al. [24]. First, the vdw interactions between oligoglycine and the solvent are gradually turned on via a coupling parameter, λ^{vdw}, that takes on values between 0 (i.e., gas phase) and 1 (i.e., vdw interactions fully on). For intermediate λ^{vdw} values, the vdw interactions are scaled according to a soft-core potential [28]. Then, with $\lambda^{\text{vdw}} = 1$, the elec interactions are linearly scaled with λ^{elec} such that when $\lambda^{\text{elec}} = 1$, oligoglycine is fully solvated. The interactions between atoms in oligoglycine were not scaled. For the vdw pathway, a $\Delta\lambda^{\text{vdw}}$ value of 0.02 (50 windows) was used for rigid/fixed and flexible Gly$_2$, Gly$_3$, and Gly$_4$ simulations, while a spacing of 0.04 was used for flexible Gly$_5$. A $\Delta\lambda^{\text{elec}}$ value of 0.0625 (16 windows) was used for all electrostatic pathways. Independent, concurrent MD simulations at 300 K were performed at each λ for 50 ns for fixed and flexible Gly$_{2-4}$ and 80 ns for flexible Gly$_5$. The change in free energy was computed from these simulations at each λ^{vdw} and λ^{elec} and is discussed in the following section. The simulations were then repeated at 320 K, using the final configuration from the simulations at 300 K, to measure the change in solvation entropy [29,30]. One nanosecond at each λ was attributed to equilibration. In total, 1748 simulations were performed for an aggregate simulation time of 92.3 μs. While the overall solvation free energy is a state function, its vdw and elec components, as defined here, are path/protocol-dependent.

Solvation free energy

The vdw and elec contributions to solvation free energy are computed separately according to the exponential form of the FEP formula [31,32]:

$$\Delta G^{\text{vdw}} = -\frac{1}{\beta}\sum_{i=1}^{n^{\text{vdw}}} ln\langle \exp(-\beta\Delta U_i^{\text{vdw}})\rangle_i, \qquad (1)$$

$$\Delta G^{\text{elec}} = -\frac{1}{\beta}\sum_{i=1}^{n^{\text{elec}}} ln\left\langle \exp\left(-\beta\Delta U_i^{\text{elec}}\right)\right\rangle_i, \qquad (2)$$

$$\Delta G^{\text{sol}} = \Delta G^{\text{vdw}} + \Delta G^{\text{elec}}, \qquad (3)$$

where i is used to index a particular λ_i^{vdw} or λ_i^{elec} that take on values between 0 and 1; β is the inverse temperature; and $\Delta U_i = \Delta U(\lambda_{i+1}) - \Delta U(\lambda_i)$, which is the difference in either the vdw or elec potential energy for a configuration generated in the λ_i^{vdw} or λ_i^{elec} ensemble and those for the same configuration if it were in the $\lambda_{i+1}^{\text{vdw}}$ or $\lambda_{i+1}^{\text{elec}}$ ensemble, respectively. $\langle ... \rangle_i$ is an ensemble average and the summation is either over the λ^{vdw} or λ^{elec} values. Simulations at each λ yield trajectories of ΔU_i^{vdw} or ΔU_i^{elec} values. Note that the small contributions from pressure and volume are not included in the estimates of ΔG. For a detailed review on multistage FEP calculations, we refer the reader to Kokubo et al. (24) and Pohorille et al. (31). The total solvation free energy, ΔG^{sol}, is the sum of ΔG^{vdw} and ΔG^{elec}.

Solvation entropy and enthalpy

Two approaches were used to decompose ΔG^{sol} into its enthalpic and entropic components. In the first method, referred to as the finite difference (FD) approach, we further decompose ΔG^{vdw} and ΔG^{elec} into their enthalpic and entropic components. In this approach, the vdw and elec contributions to solvation entropy are given by a central finite difference approximation (29,30):

$$\Delta S^{\text{vdw}}(310\text{ K}) = -\frac{\partial \Delta G^{\text{vdw}}}{\partial T}\bigg|_{310\text{ K}}$$
$$\approx -\frac{\Delta G^{\text{vdw}}(310\text{ K} + \Delta T) - \Delta G^{\text{vdw}}(310\text{ K} - \Delta T)}{2\Delta T}, \qquad (4)$$

$$\Delta S^{\text{elec}}(310\text{ K}) = -\frac{\partial \Delta G^{\text{elec}}}{\partial T}\bigg|_{310\text{ K}}$$
$$\approx -\frac{\Delta G^{\text{elec}}(310\text{ K} + \Delta T) - \Delta G^{\text{elec}}(310\text{ K} - \Delta T)}{2\Delta T}, \qquad (5)$$

where ΔT is 10 K. Similarly, ΔS^{sol} is the sum of ΔS^{vdw} and ΔS^{elec}. To calculate solvation enthalpy and its components, we estimate ΔG^{sol}, ΔG^{vdw}, and ΔG^{elec} at 310 K as the average of those measured at 300 and 320 K, and use the fact that $\Delta G = \Delta H - T\Delta S$.

Alternatively, we directly calculate solvation enthalpy at 300 and 320 K as the difference in the average, total potential energy of the fully solvated state (aqueous), and the gas phase reference state (referred to as the end point energy (EP) approach):

$$\Delta H^{\text{sol}} \approx \Delta U^{\text{sol}} = \langle U^{\text{aq}} \rangle - \langle U^{\text{gas}} \rangle, \qquad (6)$$

$$\Delta H^{\text{sol}} \approx \left\langle U^{\text{total}}\left(\lambda^{\text{vdw}} = 1 \cap \lambda^{\text{elec}} = 1\right)\right\rangle$$
$$- \left\langle U^{\text{total}}\left(\lambda^{\text{vdw}} = 0 \cap \lambda^{\text{elec}} = 0\right)\right\rangle. \qquad (7)$$

Here, U^{total} is the sum of all bonded and nonbonded peptide-peptide, peptide-solvent, and solvent-solvent potential energy terms. Whereas the solvation free energy depends only on the peptide-solvent potential energy, the solvation enthalpy and entropy include contributions from all bonded and nonbonded interactions in the system (33). ΔS^{sol} at 300 and 320 K is calculated as $T\Delta S = \Delta H - \Delta G$. Note that ΔS^{sol} was not further decomposed into

its vdw and elec components as in the FD approach. To directly compare to the FD approach, ΔS^{sol} and ΔH^{sol} were approximated at 310 K as the average of those measured at 300 and 320 K. The change in the heat capacity of solvation is minor and indicates that both the FD approach and the averaging of the thermodynamic quantities to yield estimates at 310 K are reasonable approximations for the temperature range considered.

Error analysis

We use two independent methods to estimate errors (i.e., statistical uncertainties) in ΔG^{vdw} and ΔG^{elec} at 300 and 320 K. We briefly describe these methods but a more detailed description can be found in Appendix A. The first method involves estimating the correlation times from autocorrelation functions (ACFs) of $\exp(-\beta\Delta U_i(t))$ for each of the 1748 λ-simulations. We note that ACFs of $\Delta U_i(t)$ yielded very similar autocorrelation times. Error propagation, corrected for the number of independent observations using the correlation times, is applied to the exponential form of the FEP equation (31) to estimate errors, $\sigma_{\text{ACF}}^{\text{vdw}}$ and $\sigma_{\text{ACF}}^{\text{elec}}$, in ΔG^{vdw} and ΔG^{elec}, respectively. In the second method, we use a blocking approach (34,35) in which trajectories of ΔU_i^{vdw} and ΔU_i^{elec} are broken up into blocks that span λ-space. ΔG^{vdw} and ΔG^{elec} are calculated using Eq. 1 and Eq. 2, respectively, for each block and the variance of these block estimates is used to calculate the block standard error (BSE). The BSEs are calculated as a function of block length and the point at which the BSE stops varying with block length will yield error estimates, $\sigma_{\text{BSE}}^{\text{vdw}}$ and $\sigma_{\text{BSE}}^{\text{elec}}$, of ΔG^{vdw} and ΔG^{elec}, respectively. The errors measured by these two methods are then propagated, separately, to estimate errors in ΔG^{sol}.

For the FD approach, ACF and BSE errors in ΔG^{sol}, ΔG^{vdw}, and ΔG^{elec} are propagated to estimate errors in ΔS^{sol}, ΔS^{vdw}, ΔS^{elec}, ΔH^{sol}, ΔH^{vdw}, and ΔH^{elec}. For the EP approach, autocorrelation times of $U^{\text{aq}}(t)$ and $U^{\text{gas}}(t)$ were estimated from their respective ACFs similarly detailed in Appendix A. Errors in $\langle U^{\text{aq}} \rangle$ and $\langle U^{\text{gas}} \rangle$ were corrected for the number of independent observations and then propagated to estimate the error in ΔH^{sol}. Additionally, BSEs were calculated in a similar manner detailed above but with $\langle U^{\text{aq}} \rangle$ or $\langle U^{\text{gas}} \rangle$ estimated within each block. The error in the endpoint energy estimate of ΔH^{sol} calculated via the ACF or BSE approaches were combined with the corresponding errors in ΔG^{sol}, to estimate the errors in ΔS^{sol}.

RESULTS

Toward understanding the solvation thermodynamic mechanisms that dictate, in part, the structural properties of disordered oligoglycine chains, we have decomposed the solvation free energy (ΔG^{sol}) of oligoglycine into its vdw (ΔG^{vdw}) and elec (ΔG^{elec}) components as a function of chain length and force field using stratified FEP. The solvation free energy was further decomposed into its entropic (ΔS^{sol}) and enthalpic (ΔH^{sol}) contributions using two independent approaches (i.e., the FD and EP approaches as discussed in the Materials and Methods). The FD approach also yields the vdw and elec components of ΔS^{sol} and ΔH^{sol}. Because the FD approach provides a more accurate estimate of ΔS at the midpoint (i.e., 310 K) and to compare results from both approaches, ΔS^{sol} and ΔH^{sol} calculated via the EP approach were approximated at 310 K as the average of those measured at 300 and 320 K. We also investigate the effects of conformational flexibility on these various thermodynamic quantities. Below we present the results for each quantity at 310 K, and provide the results at 300 and 320 K in the Supporting Material.

Solvation free energy

Fig. 1 depicts the scaling of ΔG^{sol}, ΔG^{vdw}, and ΔG^{elec} with chain length (N) for fixed, rigid-extended (*solid line*) and flexible (*dashed line*) oligoglycines from FEP simulations with C36 and ff12SB at 310 K. The subscripts "fix" and "flex" are used to denote solvation free energies from simulations in which oligoglycine is fixed in an extended, rigid conformation or allowed to freely explore conformation space, respectively. We used two independent approaches to estimate errors. These approaches are detailed in the Materials and Methods and Appendix A. Both methods yield errors ~1 kcal/mol or less for the ΔG values, suggesting our estimates have converged to sufficiently high precision to infer mechanism. Table S1 in the Supporting Material provides ΔG^{vdw}, ΔG^{elec}, and ΔG^{sol} for each oligoglycine at 300 and 320 K, which were averaged to yield estimates of those at 310 K.

Both force fields yield similar trends in the scaling of $\Delta G_{\text{fix}}^{\text{sol}}$, $\Delta G_{\text{flex}}^{\text{sol}}$, and their vdw and elec components with N. For example, $\Delta G_{\text{fix}}^{\text{sol}}$ is negative and continues to decrease linearly as the number of residues increase. The average difference, or slope, in $\Delta G_{\text{fix}}^{\text{sol}}$ between Gly_n and Gly_{n+1} is -5.12 and -5.44 kcal/mol per peptide unit for C36 and ff12SB, respectively. This scaling is consistent with the -5.00 and -5.48 kcal/mol per peptide unit reported by Tomar et al. (36) and Hu et al. (21) using different force fields and conformations of oligoglycine. Compared to $\Delta G_{\text{fix}}^{\text{sol}}$, $\Delta G_{\text{flex}}^{\text{sol}}$ decreases at a slower rate with respect to N with an average contribution per peptide of

-4.27 kcal/mol for C36 and -3.54 kcal/mol for ff12SB. There appears to be a transition occurring at Gly$_3$ that estimates the scaling of the longer oligoglycines. In terms of the free energy components, both ΔG^{elec} and ΔG^{vdw} are negative for rigid and flexible oligoglycines and also depend on conformational flexibility. $\Delta G_{\text{flex}}^{\text{vdw}}$ is consistently, slightly less favorable than $\Delta G_{\text{fix}}^{\text{vdw}}$, but their contributions to ΔG^{sol} are small in comparison to their dominant electrostatic counterparts. The differences in ΔG^{elec} between force fields (Fig. S1 in the Supporting Material) are ~2 kcal/mol or less with those for ΔG^{sol} at ~1 kcal/mol at both temperatures.

Solvation entropy

In the FD approach, solvation entropy was calculated from the numerical derivative with respect to temperature using the free energy at 300 and 320 K. Fig. 2 (*solid triangles*) shows the scaling of ΔS^{sol} and its elec (ΔS^{elec}) and vdw (ΔS^{vdw}) components for rigid and flexible oligoglycine. Data are provided in Table S2. Similar to the trends in solvation free energy, we find $\Delta S_{\text{fix}}^{\text{sol}}$, $\Delta S_{\text{fix}}^{\text{vdw}}$, and $\Delta S_{\text{fix}}^{\text{elec}}$ to be reasonably linear with respect to the number of residues with average contributions per peptide being -13.55, -8.68, and -4.90 cal/mol/K, respectively, for C36 and -13.48, -7.85, and -5.58 cal/mol/K for ff12SB. $\Delta S_{\text{flex}}^{\text{sol}}$, while still unfavorable, is more positive than $\Delta S_{\text{fix}}^{\text{sol}}$, indicating that conformational flexibility reduces the entropic penalty of solvating these short oligoglycines. The separation between $\Delta S_{\text{flex}}^{\text{sol}}$ and $\Delta S_{\text{fix}}^{\text{sol}}$ is largely due to
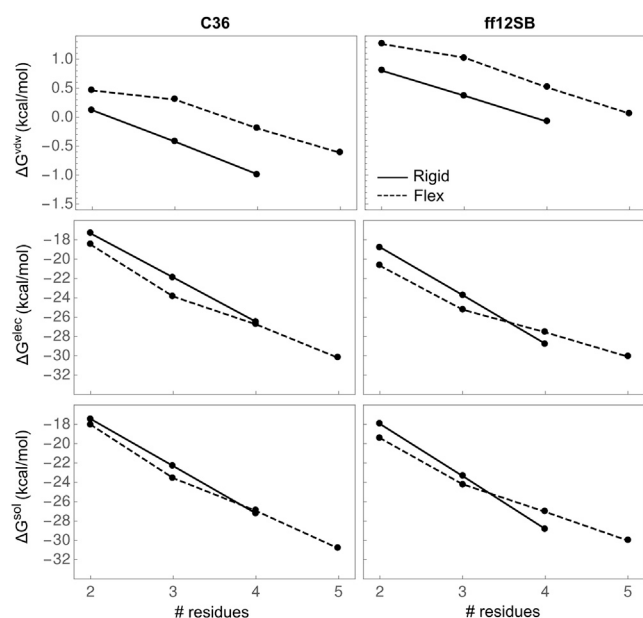

FIGURE 1 Solvation free energy of rigid/extended (*solid*) and flexible (*dashed*) oligoglycines at 310 K as a function of chain length from simulations with the C36 (*left*) and ff12SB (*right*) force fields. The vdw, elec, and overall solvation free energy are shown in order from top to bottom. Errors associated with each quantity can be found in Table S1.
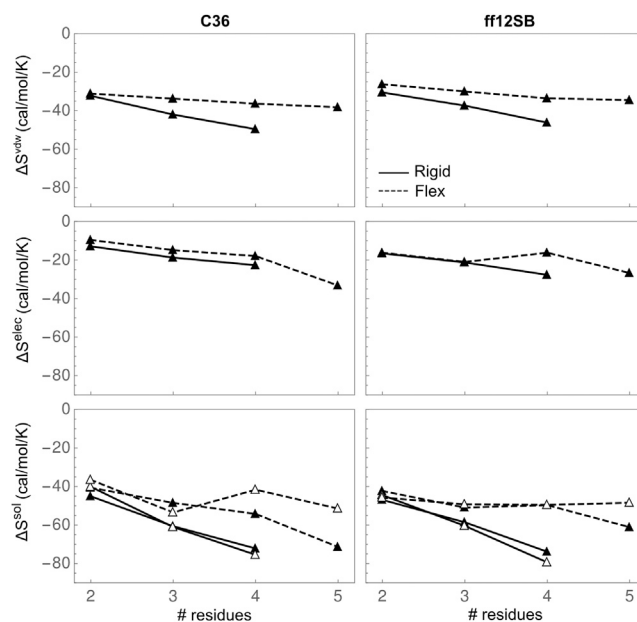

FIGURE 2 Solvation entropy of rigid/extended (*solid*) and flexible (*dashed*) oligoglycine as a function of chain from simulations with the C36 (*left*) and ff12SB (*right*) force fields using the FD (*solid triangle*) and EP (*open triangle*) approaches. The vdw, elec, and overall solvation entropy are shown in order from top to bottom. Error estimates are provided in Tables S2 and S3.

the differences in the scaling of the vdw component, which contributes more to the overall solvation entropy than the elec component.

In the EP approach, ΔS^{sol} at 300 and 320 K was calculated first by directly estimating ΔH^{sol} as the difference in the average potential energy of the system in the fully solvated state relative to the gas phase at both temperatures, then using $\Delta G^{\text{sol}} = \Delta H^{\text{sol}} - T\Delta S^{\text{sol}}$ to solve for ΔS^{sol} (Fig. S2; data provided in Table S3). To compare to the FD approach, ΔS^{sol} at 310 K was approximated as the average of ΔS^{sol} at 300 and 320 K. $\Delta S^{\text{sol}}_{\text{fix}}$ and $\Delta S^{\text{sol}}_{\text{flex}}$ calculated by the EP approach scale in a manner consistent with that observed from the FD approach (Fig. 2, *open triangle*). Again, both force fields yield a similar scaling of ΔS^{sol}, ΔS^{elec}, and ΔS^{vdw} for rigid and flexible oligoglycine with N (Fig. S3).

Solvation enthalpy

Next, for the FD approach, we calculated solvation enthalpy (ΔH^{sol}) and its components at 310 K using $\Delta H = \Delta G + T\Delta S$. We selected a temperature of 310 K because the finite difference used to approximate ΔS is more precise at the midpoint than at 300 or 320 K. ΔG at 310 K was estimated as the average of ΔG at 300 and 320 K. For the EP approach, ΔH^{sol} was directly calculated at 300 and 320 K, then averaged to give ΔH^{sol} at 310 K. Fig. 3 shows ΔH^{vdw}, ΔH^{elec}, and ΔH^{sol} as a function of chain length, force field, and approach for rigid and flexible oligogly-
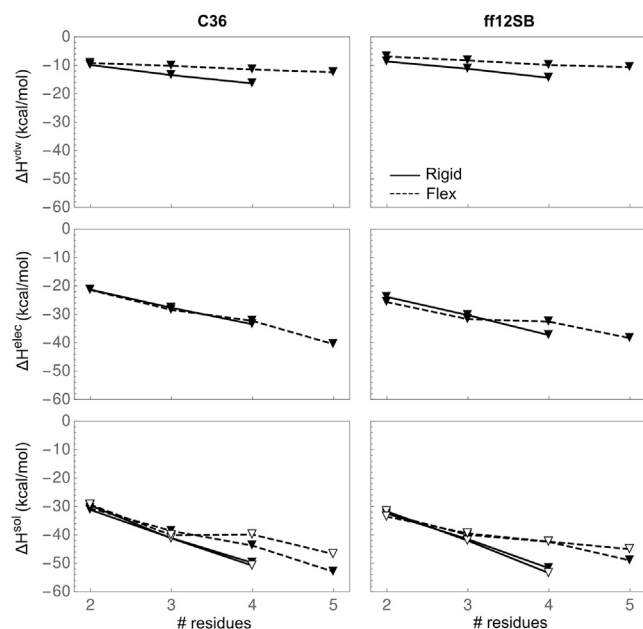

FIGURE 3 Solvation enthalpy of rigid/extended (*solid*) and flexible (*dashed*) oligoglycine at 310 K as a function of chain from simulations with the C36 (*left*) and ff12SB (*right*) force fields calculated by the FD (*solid triangle*) and EP (*open triangle*) approaches. The vdw, elec, and overall solvation enthalpy are shown in order from top to bottom.

cines. Data are provided in Tables S4 and S5. Both $\Delta H^{\text{sol}}_{\text{fix}}$ and $\Delta H^{\text{sol}}_{\text{flex}}$ are large and negative, but the rate of decrease with respect to length of $\Delta H^{\text{sol}}_{\text{flex}}$ is slightly less. The scaling of ΔH^{sol} with chain length is predominantly determined by ΔH^{elec}. The average rate of decrease of $\Delta H^{\text{vdw}}_{\text{flex}}$ is -1.09 and -1.25 kcal/mol/unit for C36 and ff12SB, respectively, which is less than half of those observed for $\Delta H^{\text{vdw}}_{\text{fix}}$. Differences between force fields were slightly larger for ΔH than ΔS and ΔG, with differences ranging between -4 and 4 kcal/mol (Fig. S5).

Whereas ΔG^{sol} depends only on peptide-solvent interaction energy, ΔH^{sol} and ΔS^{sol} include bonded and nonbonded intrapeptide contributions as well as a solvent reorganization energy (33). To parse out these contributions, we decomposed ΔU^{sol} (i.e., the EP approximation of ΔH^{sol}) into its various peptide-peptide ($\Delta U^{\text{vdw}}_{u,u}$, $\Delta U^{\text{elec}}_{u,u}$), peptide-solvent ($\Delta U^{\text{vdw}}_{u,v}$, $\Delta U^{\text{elec}}_{u,v}$), solvent-solvent ($\Delta U^{\text{vdw}}_{v,v}$, $\Delta U^{\text{elec}}_{v,v}$), and intrapeptide-bonded (ΔU^b) average energies for flexible Gly_2 and Gly_5. Fig. 4 depicts this energetic breakdown. We find that ΔH^{sol} is favorable and continues to decrease with chain length due primarily to a favorable, decreasing change in the elec peptide-solvent interaction energy. As expected, this is slightly offset by a positive $\Delta U^{\text{vdw}}_{v,v}$ and much more so by $\Delta U^{\text{elec}}_{v,v}$, or a solvent reorganization penalty. For rigid oligoglycines, bonded and nonbonded intrapeptide energies cancel and favorable peptide-peptide interactions yield a favorable, decreasing ΔH^{sol} with chain length. However, for flexible oligoglycines, solvation has the effect of reducing intrapeptide electrostatic energy, resulting in a positive $\Delta U^{\text{elec}}_{u,u}$, and, to a lesser extent, $\Delta U^{\text{vdw}}_{u,u}$, in a length-dependent manner. The bonded, intrapeptide potential energy contributes little to the solvation enthalpy. That the electrostatics dictates the scaling of ΔH^{sol} with N is consistent with what we observe by the FD approach.

Error analysis

Calculation of entropy differences from free energies requires strict control of errors. Accurately estimating errors in free energy calculations continues to be a challenging problem (31). Here, we used two independent methods to estimate errors, the details of which can be found in the Materials and Methods and Appendix A. Table S1 provides errors in the free energy and its components from the autocorrelation function (ACF) and the BSE approaches. Both approaches demonstrate that the error in $\Delta G^{\text{sol}}_{\text{flex}}$ is largely due to the error in $\Delta G^{\text{elec}}_{\text{flex}}$ whereas the errors in $\Delta G^{\text{vdw}}_{\text{fix}}$ and $\Delta G^{\text{elec}}_{\text{fix}}$ are of similar magnitudes. Errors also increase with chain length. However, the ACF errors are systematically larger than those calculated using the BSE approach. Two reasons may explain, in part, this observation.

First, the procedure used to fit the ACFs to estimate the correlation times appears to be sensitive to weak,
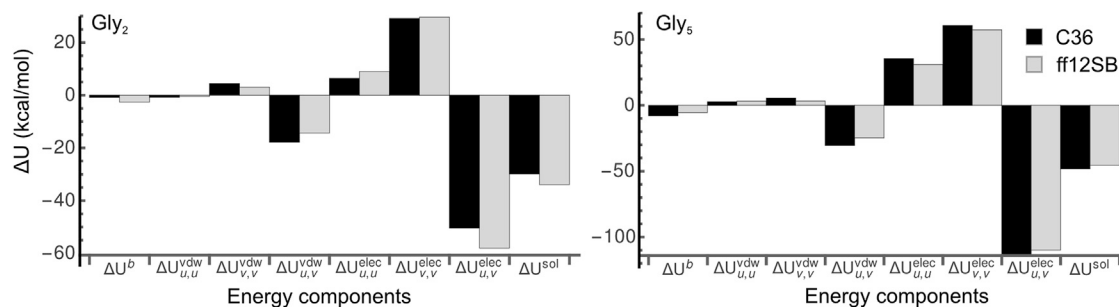
FIGURE 4 Differences in the average energies of the components contributing to $\Delta H^{\text{sol}} \approx \Delta U^{\text{sol}}$ for flexible Gly$_2$ (*left*) and Gly$_5$ (*right*). The subscript u denotes the solute/peptide and v denotes the solvent. ΔU^b is the difference between the average bonded intrapeptide energies (e.g., dihedral, bond, angle energies) of oligoglycine in the final, solvated state and the gas phase.

long-time correlations (<5% self-similarity and oscillation), which are predominantly present in simulations of flexible oligoglycines in which the electrostatic interactions are turned on. The correlation times across the λ^{elec} simulations for flexible oligoglycines ranged from 1 ps to 34 ns and increased with chain length. The BSE approach appears to be less sensitive to these weak, long-time correlations. To test this, we calculated the BSE at each λ^{vdw} and λ^{elec} at 300 and 320 K for flexible Gly$_5$ and calculated the correlation times using $BSE = \sigma\sqrt{\tau/N}$, where σ is the variance, τ is the correlation time, and N is the number of samples. While the profiles of correlation times across λ^{elec} were similar for both approaches, the correlation times calculated with the BSE approach were orders-of-magnitude (data not shown) less than those reported by the ACF approach. Others have noted that this BSE approach typically underestimates the true correlation time of dynamic observables (34,37). In contrast, the vdw correlation times differed, at most, by a factor of 2.

Second, we took a conservative approach in estimating the correlation times as four times the decay constant from a best fit of the ACFs to an exponentially decaying function. For these reasons, we interpret the errors from the BSE and ACF approaches as loose lower and upper bounds, respectively, on the true error. The ACF errors, although larger, still indicate that our estimates of ΔG^{vdw}, ΔG^{elec}, and ΔG^{sol} have converged to a precision <1 kcal/mol for rigid and flexible Gly$_{2-4}$ (Table S1) and slightly larger for flexible Gly$_5$. Due to the propagation of uncertainty and the fact that we are taking a finite derivative in the FD approach, the ACF and BSE errors are larger for ΔS^{sol}, ΔH^{sol} and their vdw and elec components (Tables S2 and S4). Interestingly, persistent, weak correlations were not present in the ACFs of $U^{\text{aq}}(t)$ and $U^{\text{gas}}(t)$ as compared to the ACFs of $\exp(-\beta\Delta U_i(t))$. This led to errors (Tables S3 and S5) in ΔH^{sol} and ΔS^{sol} that were systematically less than those calculated via the FD approach. The BSEs reported by both approaches are more similar. We also do not observe as strong of a length dependence of the errors estimated by the EP approach.

DISCUSSION

Scaling of solvation free energy, enthalpy, and entropy with chain length and the effects of flexibility

Computational free energy methods can be used to measure the thermodynamic properties of disordered polypeptides that are otherwise difficult to access with experiment. Here, we have performed a complete thermodynamic decomposition of the solvation free energy of oligoglycine as a function of chain length (N), force field, and positional constraints. Few studies have considered the scaling of solvation entropy of the protein backbone with the number of residues and the effects of flexibility on this scaling relationship. Accurately estimating the solvation free energy and entropy even for these short oligoglycine models required extensive sampling, particularly for flexible oligoglycines. This study required a total of 1748 simulations for an aggregate simulation time of 92.3 μs.

We find that both force fields yield very similar trends in the scaling of ΔG^{sol}, ΔH^{sol}, and ΔS^{sol} and their vdw and elec components with N (Figs. 1–3). Fig. 5 shows ΔG^{sol}, ΔH^{sol}, and $-T\Delta S^{\text{sol}}$ for rigid-extended and flexible oligoglycines as a function of chain length and force field at 310 K (Tables S6 and S7) calculated by the FD and EP approaches. ΔG^{sol} is negative and continues to decrease at a rate similar to that observed in previous studies of oligoglycine fixed in various conformations (21,36,38). Unlike for successively larger alkanes (39), the scaling of ΔG^{sol} is enthalpically driven and only moderately compensated by unfavorable entropy changes. This is consistent with the transfer free energy of N-methylacetamide (NMA), a peptide group analog, from the gas to solvent phase (14,40). However, ΔG^{sol} of NMA is −10 kcal/mol at 298 K, which is nearly twice as large as the glycine peptide contribution per monomer we find for rigid oligoglycines and larger yet for flexible oligoglycines.

When considering the effects of flexibility, we find that $-T\Delta S^{\text{sol}}$ increases and ΔH^{sol} decreases at a slower rate than their rigid counterparts. However, the increase in entropy associated with solvating flexible oligoglycines offsets
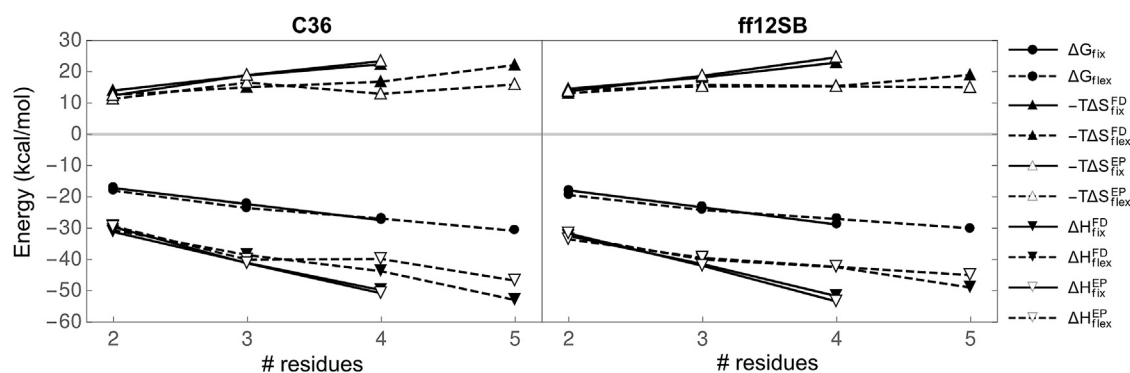
FIGURE 5 Thermodynamic decomposition of solvation free energy for rigid/extended (*solid*) and flexible (*dashed*) oligoglycines as a function of chain length and force field at 310 K calculated by the FD and EP approaches. (*Solid circle*) ΔG^{sol} calculated by free energy perturbation; (*solid and open triangles*) ΔH^{sol} and $-T\Delta S^{sol}$ calculated by the FD and EP approaches, respectively. Note that ΔH^{sol} and $-T\Delta S^{sol}$ at 310 K was approximated as their averages at 300 and 320 K.

the reduction in enthalpic interactions relative to that observed for rigid oligoglycines, and, as a result, differences in ΔG_{flex}^{sol} and ΔG_{fix}^{sol} are less pronounced. The FD and EP approaches predict very similar thermodynamic scaling profiles for the rigid and flexible oligoglycine models. However, the EP approach yields smaller statistical uncertainties and requires half as many FEP simulations to obtain ΔH^{sol} and ΔS^{sol} as the FD approach.

Given the trends in the scaling relationships established here for short oligoglycines, we expect the differences in ΔG_{flex}^{sol}, and their entropic and enthalpic components, to be amplified for longer oligoglycines. As we will discuss below, this will have implications for the validity of group-additive models to predict the thermodynamics of long, disordered polypeptides from the thermodynamic properties of its isolated chemical groups. Future work will address the predictive capabilities of these scaling relationships for longer oligoglycines.

Decomposing ΔG^{sol}, ΔH^{sol}, and ΔS^{sol} into vdw and elec components provides insight into the mechanisms that yield their scaling relationships with N. ΔG^{elec} dominates ΔG^{vdw} and largely dictates the scaling of ΔG^{sol}, as has been previously noted for various oligoglycine and oligoalanine chains (21,24). On the other hand, ΔS^{vdw} is of the same order of magnitude of ΔS^{elec} but contributes more to ΔS^{sol}. This is consistent with the idea that inserting a large vdw cavity in water reduces the translational entropy of the solvent and disrupts the water-water interaction network more so than the subsequent charging of the cavity (41,42). The inverse relationship between enthalpy and entropy is also apparent in their vdw and elec components with ΔH^{elec} being ~1.5–3 times more favorable than ΔH^{vdw} for rigid and flexible oligoglycines. In general, both the absolute value and the rate of decrease of the vdw and elec components of ΔG, ΔH, and ΔS change when positional constraints are not imposed on the oligoglycine chain. Most notably, ΔG_{flex}^{vdw}, ΔH_{flex}^{vdw}, and ΔS_{flex}^{vdw} are consistently more positive than their fixed counterparts. We note that in a previous

study in which oligomers of alanine were decomposed in a similar fashion, vdw solvation free energies were found to be consistently more positive for flexible alanine polypeptides than their rigid-extended counterparts (24). Flexibility has the consequence of giving less instantaneous vdw exposure to the solvent for most conformations as favorable peptide-solvent interactions may be screened. This is understandable, given the range of vdw attractions to solvent. Note that, for flexible alanine chains, the slope of ΔG_{flex}^{vdw} turned positive. But just as in this study, the overall solvation free energy was dominated by the elec component.

Force-field comparison

Figs. S1, S3, and S5 show the differences in ΔG^{sol}, ΔH^{sol}, and ΔS^{sol} and their components between force fields. Differences in ΔG^{sol} between C36 and ff12SB are within 1–2 kcal/mol for rigid and flexible Gly$_{2-5}$. Interestingly, there appears to be a compensatory relationship between force fields in that ff12SB consistently underestimates ΔG^{vdw} relative to C36, but overestimates ΔG^{elec} for Gly$_{2-4}$, resulting in smaller differences in ΔG^{sol}. These cancellations are also observed for flexible Gly$_5$, but do not follow the aforementioned trend. Hu et al. (21) previously measured ΔG^{sol} for various fixed conformations of Gly$_{2-5}$ using the CHARMM27 (C27) force field and found an average contribution per peptide of −5.48 kcal/mol, which is consistent with the −5.26 and −5.57 kcal/mol/unit we find with C36 and ff12SB, respectively. However, the absolute values of ΔG^{sol} were different. For example, they reported a ΔG^{sol} of Gly$_5$ that was roughly 10 kcal/mol less than what we find with C36 and ff12SB, which further highlights the effects of flexibility on solvation free energy.

While employing different strategies, C36 and ff12SB were parameterized to match experimental and theoretical structural and thermodynamic data (22,23). We have previously shown that C36 generates structural ensembles of

Gly$_3$ and Gly$_{10}$ to a greater extent, which are more extended and less compact than ff12SB with significant differences in individual dihedral angle populations (17). The small differences we observe in ΔG^{sol} between C36 and ff12SB may suggest either that the solvation thermodynamic properties of oligoglycine are robust to differences in force-field structural parameters or that the chain lengths considered here are too short to realize significant differences in structural ensembles generated by C36 and ff12SB. The former implies a degenerate relationship between the structural manifold of oligoglycine and its solvation thermodynamics. Even for Gly$_5$, differences in force fields are pronounced, with ff12SB sampling more compact conformations than C36 (data not shown). Given the striking differences in the structural ensemble of Gly$_{10}$ produced by these two force fields (17), we expect differences in the scaling of ΔG^{sol} to become more apparent for larger N. While this clearly needs to be tested, we note that caution should be exercised when attempting to connect structural ensembles to biological mechanisms from such potential models (i.e., force fields) even when those models are in thermodynamic agreement.

Implications for group-additivity

The notion of group-additivity posits that the solvation free energy of a macromolecule is the result of additive contributions of the chemical groups that comprise it. Additive principles have been applied to and observed in numerous studies of hydrocarbon and protein systems (18,21,43–49). However, evidence continues to mount that additivity is context-specific, and depends on local chemical environment and molecular conformation/geometry. While rough estimates are possible, the method may not accurately predict the solvation free energy of amino acids and proteins (20,25,36,38,50–55). From a study of 50 conformations of decaglycine, we have shown that even an individual atom's contribution to ΔG^{vdw} and its repulsive component are not well defined and vary with respect to conformation; the result depends on local chemical environment and force field, which is contrary to what one would expect if additivity held (20).

While chemical context and conformation matter, the apparent linear scaling of $\Delta G_{\text{fix}}^{\text{sol}}$ with N and the relatively small differences between $\Delta G_{\text{flex}}^{\text{sol}}$ and $\Delta G_{\text{fix}}^{\text{sol}}$ observed here are consistent with an additive contribution of the glycine peptide unit to the overall solvation free energy. However, our data also indicate different rates of decrease of $\Delta G_{\text{flex}}^{\text{sol}}$ and $\Delta G_{\text{fix}}^{\text{sol}}$ with N. A recent study (36) has shown that the binding energy of a glycine peptide with the solvent depends on or is correlated with that of its neighboring peptides for a series of extended, fixed oligoglycines. They note that correlated fluctuations can change the free energy of the solute and this effect occurs over spatially distant moieties. For longer, flexible oligoglycines, structural fluctuations and long-range peptide-peptide interactions alter the local sol-

vent structure around any one glycine peptide in the chain from that of the linear peptide. As a result, the contribution of each glycine peptide to the overall solvation free energy of the entire chain is conditioned by both its near and spatially more distant neighbors. These correlations or cooperative effects could not have been predicted from the solvation free energy of an isolated glycine and, thus, leads to a departure from group-additivity. This may explain why $\Delta G_{\text{flex}}^{\text{sol}}$ and $\Delta G_{\text{fix}}^{\text{sol}}$ decrease at different rates initially for short oligoglycines and oligoalanines (24) and suggests that longer chain lengths will be needed to better resolve the effects of flexibility and structural correlations on the solvation free energy of each peptide unit.

Implications for aggregation of short oligoglycines and collapse of long oligoglycines

If the scaling relationships we establish here hold, then favorable enthalpic interactions with the solvent will continue to drive the decrease in ΔG^{sol} for long oligoglycines. However, long oligoglycines ($N \geq$ ~15) collapse in relatively dilute or infinitely dilute solutions as observed in both experiment and simulation (14–16). In addition, short oligoglycines become exceedingly insoluble with visible aggregates forming for Gly$_5$ at finite (millimolar) concentrations (18,19). Thus, solvation does not drive the aggregation of short oligoglycines and the collapse of long oligoglycines.

We previously investigated the hydrophobic effect as a potential force driving the collapse of decaglycine (20). ΔG^{vdw} is often defined as the hydrophobic solvation free energy, and, as seen for macromolecules and various alkanes (39), it is expected to increase roughly linearly with solvent-exposed surface area $\Delta G^{\text{vdw}} = \gamma^{\text{vdw}} A$, where γ^{vdw} is the implied surface tension, independent of the chemical properties of the solute and A is the exposed surface area (56,57). Others have suggested that ΔG^{vdw} should be split into an attractive (ΔG^{att}) and repulsive (ΔG^{rep}) component and that ΔG^{rep} should increase linearly with $A(\Delta G^{\text{rep}} = \gamma^{\text{rep}} A)$ (58,59). For a detailed review on various theories of hydrophobicity, we refer to Harris and Pettitt (54) and Ben-Amotz (60). A solvation free energy that increases with A would favor the initial folding or collapse of proteins by limiting the surface area exposed to solvent. Interestingly, from our study of 50 conformations of decaglycine, we found that ΔG^{vdw} actually favors extended conformations with $\gamma^{\text{vdw}} < 0$ because $-\gamma^{\text{att}} > \gamma^{\text{rep}}$ and that γ^{rep} was weakly correlated, if at all, with A and depended on the conformation of decaglycine (20). Taking N as a proxy for A, in this study we also find that $\gamma^{\text{vdw}} < 0$, but it depends both on force field and flexibility. While we did not further decompose ΔG^{vdw} here, it is reasonable to assume that the attractive component outweighs the repulsive component. Tomar et al. (61) similarly find that attractive peptide-solvent dispersion interactions can prevent the collapse of

deca-alanine. This seemingly counterintuitive observation that water-mediated, hydrophobic interactions promotes chain expansion rather than collapse as the surface area (i.e., solute size or chain length) increases is discussed in detail in Ben-Amotz (62).

A reasonable explanation for the observed collapse of long oligoglycines and their insolubility is that peptide-peptide interactions outcompete the still favorable peptide-solvent interactions—an observation that could not have been predicted from the solvation free energies of short, isolated oligoglycines. Using single molecule fluorescence correlation spectroscopy, Teufel et al. (15) found the hydrodynamic radius of gyration (R_h) of Gly_{20} to be significantly less than that of NME-Gly_{20} (wherein the amide hydrogens of Gly_{20} were replaced with methyl groups). They conjectured that the methyl groups preclude the formation of backbone H-bonds resulting in the observed increase in solubility and R_h of NME-Gly_{20} relative to Gly_{20}. In another study using MD simulations, Karandur et al. (16) found that the collapse of Gly_{15} and Gly_{25} may be explained by nonhydrogen-bonding, intrapeptide dipole-dipole interactions. In a similar manner, these dipole-dipole correlations and charge layerings were also observed to stabilize clusters or aggregates of Gly_5 (19). It appears that as chain length or concentration increase, and subsequently the number of available intra- or interpeptide interactions increase, these interactions are favored over peptide-solvent interactions. Uversky et al. (4) has noted that the extent of collapse or compaction of IDPs is, in part, a length-dependent property and that "…weak, intramolecular interactions can be unspecifically amplified by an increase in protein size."

Alternatively, if the thermodynamic scaling relationships presented here do not hold, then at some longer chain length, $T\Delta S^{sol}$ may overtake ΔH^{sol} and the initial collapse of long oligoglycines would be entropically driven, as others have noted (14). In this situation, collapsed states may be further favored via backbone interactions. So group-additive models would fail to predict the solvation thermodynamics of long oligoglycines from short ones. However, the scaling of ΔG^{sol}, ΔH^{sol}, and ΔS^{sol} with chain length for shorter oligoglycines may provide upper bounds on these quantities for longer oligoglycines by explicitly neglecting structural correlations between short segments that comprise the larger chain. Further work will need to assess the predictive capabilities of the scaling relationships we established, and whether a thermodynamic transition exists in which a negative entropic component to the free energy drives collapse of longer oligoglycines.

CONCLUSION

The structural ensemble of an IDP/IDR is the result of a complex competition between intrapeptide and peptide-solvent interactions (63–65). A number of determinants, like chain length (4), sequence composition (5), and charge patterning

(66), influence the propensity of an IDP/IDR to adopt compact or extended structures. The protein trinity hypothesis (1,2), and its variants (3), have been useful in classifying or characterizing the structural properties of IDPs/IDRs and relating those properties to biological function. Here, we sought to gain insight into the solvation thermodynamics of long chains by considering isolated, short oligoglycine chains more amenable to computational free energy methods. IDRs, especially those found in the disordered regulatory domains of nuclear transcription factors (13), are enriched with oligoglycine tracts and glycine content (5,12).

We have performed a complete thermodynamic decomposition of the solvation free energy into its enthalpic and entropic components as well as each of their vdw and elec contributions as functions of chain length. For both rigid and flexible oligoglycine models, the decrease in ΔG^{sol} with chain length is enthalpically driven with only a weak entropic compensation. Solvation free energies of alkanes and polypeptides are often measured by imposing positional constraints on their structures. Here we have shown that the rate of decrease of ΔG^{sol}, ΔS^{sol}, ΔH^{sol}, and their vdw and elec components with N depends on conformational flexibility; however, longer chain lengths may be needed to better resolve differences between ΔG_{flex}^{sol} and ΔG_{fix}^{sol}. The differences in the initial rate of decrease of ΔG_{flex}^{sol} and ΔG_{fix}^{sol} for shorter oligoglycines may be due to local solvent reorganization around individual peptide groups brought on by correlated, structural fluctuations (i.e., near and spatially distant neighbor effects) as proposed by Tomar et al. (36). For these reasons, we anticipate differences ΔG_{flex}^{sol} and ΔG_{fix}^{sol} to become more pronounced for longer chain lengths where long-range interactions (e.g. excluded volume) are more prominent and weak peptide-peptide interactions are amplified.

Previously, we have shown that the C36 and ff12SB fields yield considerably different conformational ensembles of Gly_3 and Gly_{10} with C36 sampling more extended oligoglycine structures (17). Even for Gly_5 we find ff12SB to sample more compact structures than C36 (data not published). Interestingly, these force fields, both of which were parameterized with different structural and thermodynamic target data, yield remarkably similar scaling relationships of ΔG_{flex}^{sol}, ΔS_{flex}^{sol}, and ΔH_{flex}^{sol} with chain length. This seems to imply a degenerate relationship between conformational ensemble and solvation thermodynamics, and complicates the ability to make consistent mechanistic inferences.

A favorable, enthalpically driven solvation free energy that continues to decrease with chain length (or surface area) does not account for the observed insolubility of Gly_5 (18,19) or the preferred collapse of longer oligoglycines (14–16). Alternatively, our work supports the idea that peptide-peptide interactions outcompete favorable peptide-solvent interactions. Whether solvation entropy promotes the collapse of longer oligoglycines outside this scope of work remains to be seen. Attempting to predict the solvation thermodynamics of longer, disordered

polypeptides from that observed for shorter chain segments where intrapeptide interactions are limited may be problematic and render group-additive models less useful.

While there is a tendency for long oligoglycines to collapse, they adopt a diverse set of conformations (14,16,17,21,67) with shorter chain segments assuming extended conformations more exposed to the solvent. This observation, in combination with the fact that we find short oligoglycine tracts to favorably interact with solvent, may suggest that the conformational ensemble of long oligoglycine chains could be shifted to favor more extended or collapsed states without the need to overcome excessively high free energy barriers. In this way, the protein backbone provides a scaffold amenable to both collapsed-disorder and extended-disorder and that evolutionary forces have selected unique combinations of side chains to alter the distribution of these two states.

APPENDIX A

The free energy change at a particular λ, ΔG_i, is a nonlinear function of ΔU_i measured from an MD simulation and is given by Eq. A.1 below. We drop the vdw and elec superscripts as the following applies to both components. While the simulations were performed in the NPT ensemble, we are neglecting the small contributions of pressure and volume to the free energy.

$$\Delta G_i = -\frac{1}{\beta}\ln\left[\frac{1}{N_i}\sum_{j=1}^{N_i}\exp\left(-\beta\Delta U_{i,j}\right)\right], \qquad (A.1)$$

where N_i is the number of observations or frames from a simulation with λ_i, j indexes each frame, and $\Delta U_{i,j} = U(\lambda_{i+1})_j - U(\lambda_i)_j$. ΔG is then the sum of the contributions ΔG_i:

$$\Delta G = \sum_{i=1}^{n}\Delta G_i, \qquad (A.2)$$

where n is the number of λ-values and will differ for the vdw and elec components.

Time correlation analysis

For a function $f(x_1, x_2...)$, that is a linear combination of independent, random variables $x_1, x_2...$, the variance of $f(x_1, x_2...)$ is given by Eq. A.3 (i.e. error propagation):

$$\sigma_f^2 = \left(\frac{\partial f}{\partial x_1}\right)^2\sigma_{x_1}^2 + \left(\frac{\partial f}{\partial x_2}\right)^2\sigma_{x_2}^2 + \dots \qquad (A.3)$$

To calculate the variance of ΔG_i at λ_i ($\sigma_{\Delta G_i}^2$), we apply Eq. A.3 but with only one independent variable (ΔU_i), to Eq. A.1, which gives

$$\sigma_{\Delta G_i}^2 = \frac{1}{N_i^{\text{eff}}\beta^2}\frac{\langle\exp(-2\beta\Delta U_i)\rangle_i}{\langle\exp(-\beta\Delta U_i)\rangle_i^2} - \frac{1}{N_i^{\text{eff}}\beta^2}, \qquad (A.4)$$

$$N_i^{\text{eff}} = \frac{N_i}{\frac{2\tau_i}{\Delta t} + 1}. \qquad (A.5)$$

N_i^{eff} is the effective number of independent observations from a simulation, τ_i is the correlation time, and Δt is the time interval at which the ΔU_i

values were saved. To determine τ_i, we first generated the time ACF, $C(t)$, of $\exp(-\beta\Delta U_i(t))$ for each simulation. Visual inspection of the ACFs showed an initial, rapid decay in correlations, followed in many cases by a second, slower decay as the ACF approaches zero. Due to the large number of simulations and to capture the two-phase decay, we fit the ACFs with a double exponential function:

$$C_{\text{fit}}(t; \gamma_1, \gamma_2) = A_1\exp\left(\frac{-t}{\gamma_1}\right) + A_2\exp\left(\frac{-t}{\gamma_2}\right), \qquad (A.6)$$

where γ_1 and γ_2 are the decay rate constants and $\gamma_1 < \gamma_2$. If the double exponential fit failed or $A_2 < 0.01$, the ACFs were fit with a single exponential. To improve the likelihood of capturing the true correlation time, the correlation times were conservatively estimated as $\tau_i = 4\gamma_2$ or $\tau_i = 4\gamma_1$ for the double or single exponential fits, respectively. In the situations where the fits yielded $\tau_i < 0$, we took a conservative approach and set τ_i to 1 ps (i.e., Δt). These correlation times were used with Eq. A.4 to estimate the errors, $\sigma_{\text{ACF}}^{\text{vdw}}$ and $\sigma_{\text{ACF}}^{\text{elec}}$ (square root of the variance), in ΔG^{vdw} and ΔG^{elec} at 300 and 320 K. For the FD approach, Eq. A.3 was used to estimate errors in ΔG^{sol} at both temperatures, ΔS^{sol}, ΔH^{sol}, and their vdw and elec components. For the EP approach, ACFs of $U^{\text{aq}}(t)$ and $U^{\text{gas}}(t)$ were similarly fit to Eq. A.6 and their autocorrelation times estimated as described above. Variances of $\langle U^{\text{aq}}\rangle$ and $\langle U^{\text{gas}}\rangle$ were adjusted for the effective number of independent observations (Eq. A.5), and Eq. A.3 was used to estimate the error in ΔH^{sol} at 300 and 320 K. Equation A.3 was then used to estimate the error in ΔS^{sol} at both temperatures.

BSE analysis

Separately for the vdw and elec components, we construct a matrix of ΔU_i-values in which the columns are particular λ-values that span either λ^{vdw} or λ^{elec} and the rows are time points. The matrix has dimensions $N \times n$, where N is the total number of frames and n is the number of λ-values—subscripts have been dropped because all simulations for the vdw and elec pathways, respectively, were simulated for the same length of time and number of λ-values. The matrix is then broken into blocks of a particular length such that $N = M \times b$, where M is the number of submatrices of dimensions $b \times n$ and b is the block length. For a given b, ΔG values are calculated using Eq. 1 for each of the M blocks. The BSE is the standard deviation of these block free energies divided by the square root of the number of blocks, M. A series of BSE values are computed for a range of block lengths and the error is estimated at the point in which the BSE curve plateaus or BSE stops varying with block length. A maximum block length of 8 and 12 ns were used for fixed and flexible Gly$_{2-4}$ and flexible Gly$_5$, respectively. Taking a conservative approach, we estimate the errors $\sigma_{\text{BSE}}^{\text{vdw}}$ and $\sigma_{\text{BSE}}^{\text{elec}}$ in ΔG^{vdw} and ΔG^{elec} by taking the maximum BSE across the range of block lengths. Equation A.3 was used to estimate the errors in solvation free energy at 300 and 320 K, solvation entropy, solvation enthalpy, and their vdw and elec components. For the EP approach, BSEs were calculated in a similar manner except with $\langle U^{\text{aq}}\rangle$ or $\langle U^{\text{gas}}\rangle$ estimated within each block. Equation A.3 was then used to estimate the BSE of $\Delta H^{\text{sol}} \approx \langle U^{\text{aq}}\rangle - \langle U^{\text{gas}}\rangle$, which was further combined with the BSE of ΔG^{sol} to yield the BSE of ΔS^{sol} at both temperatures.

SUPPORTING MATERIAL

Five figures and seven tables are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(16)30545-8.

AUTHOR CONTRIBUTIONS

J.A.D. was responsible for performing the free energy calculations, error analysis, interpretation, and cowriting the article. R.C.H. was responsible

REFERENCES

1. Dunker, A. K., J. D. Lawson, …, Z. Obradovic. 2001. Intrinsically disordered protein. *J. Mol. Graph. Model.* 19:26–59.

2. Dunker, A. K., C. J. Brown, …, Z. Obradovic. 2002. Intrinsic disorder and protein function. *Biochemistry.* 41:6573–6582.

3. Uversky, V. N. 2002. Natively unfolded proteins: a point where biology waits for physics. *Protein Sci.* 11:739–756.

4. Uversky, V. N., C. Santambrogio, …, R. Grandori. 2012. Length-dependent compaction of intrinsically disordered proteins. *FEBS Lett.* 586:70–73.

5. Marsh, J. A., and J. D. Forman-Kay. 2010. Sequence determinants of compaction in intrinsically disordered proteins. *Biophys. J.* 98:2383–2390.

6. Oldfield, C. J., and A. K. Dunker. 2014. Intrinsically disordered proteins and intrinsically disordered protein regions. *Annu. Rev. Biochem.* 83:553–584.

7. Weathers, E. A., M. E. Paulaitis, …, J. H. Hoh. 2004. Reduced amino acid alphabet is sufficient to accurately recognize intrinsically disordered protein. *FEBS Lett.* 576:348–352.

8. Crick, S. L., M. Jayaraman, …, R. V. Pappu. 2006. Fluorescence correlation spectroscopy shows that monomeric polyglutamine molecules form collapsed structures in aqueous solutions. *Proc. Natl. Acad. Sci. USA.* 103:16764–16769.

9. Möglich, A., K. Joder, and T. Kiefhaber. 2006. End-to-end distance distributions and intrachain diffusion constants in unfolded polypeptide chains indicate intramolecular hydrogen bond formation. *Proc. Natl. Acad. Sci. USA.* 103:12394–12399.

10. Walters, R. H., and R. M. Murphy. 2009. Examining polyglutamine peptide length: a connection between collapsed conformations and increased aggregation. *J. Mol. Biol.* 393:978–992.

11. Vitalis, A., X. Wang, and R. V. Pappu. 2007. Quantitative characterization of intrinsic disorder in polyglutamine: insights from analysis based on polymer theories. *Biophys. J.* 93:1923–1937.

12. 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.

13. 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.

14. 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.

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

16. 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.

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

18. 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.

19. Karandur, D., K.-Y. Wong, and B. M. Pettitt. 2014. Solubility and aggregation of Gly$_5$ in water. *J. Phys. Chem. B.* 118:9565–9572.

20. Harris, R. C., J. A. Drake, and B. M. Pettitt. 2014. Multibody correlations in the hydrophobic solvation of glycine peptides. *J. Chem. Phys.* 141:22D525.

21. Hu, C. Y., H. Kokubo, …, B. M. Pettitt. 2010. Backbone additivity in the transfer model of protein solvation. *Protein Sci.* 19:1011–1022.

22. 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 χ^1 and χ^2 dihedral angles. *J. Chem. Theory Comput.* 8:3257–3273.

23. Case, D. A., J. T. Berryman, …, P. A. Kollman. 2012. AMBER 12. University of California-San Francisco, San Francisco, CA.

24. Kokubo, H., R. C. Harris, …, B. M. Pettitt. 2013. Solvation free energies of alanine peptides: the effect of flexibility. *J. Phys. Chem. B.* 117:16428–16435.

25. Harris, R. C., and B. M. Pettitt. 2014. Effects of geometry and chemistry on hydrophobic solvation. *Proc. Natl. Acad. Sci. USA.* 111:14681–14686.

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

27. Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: visual molecular dynamics. *J. Mol. Graph.* 14:33–38.

28. Beutler, T. C., A. E. Mark, …, W. F. van Gunsteren. 1994. Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. *Chem. Phys. Lett.* 222:529–539.

29. Choudhury, N., and B. M. Pettitt. 2006. Enthalpy-entropy contributions to the potential of mean force of nanoscopic hydrophobic solutes. *J. Phys. Chem. B.* 110:8459–8463.

30. Pettitt, B. M., and P. J. Rossky. 1986. Alkali halides in water: ion-solvent correlations and ion-ion potentials of mean force at infinite dilution. *J. Chem. Phys.* 84:5836–5844.

31. Pohorille, A., C. Jarzynski, and C. Chipot. 2010. Good practices in free-energy calculations. *J. Phys. Chem. B.* 114:10235–10253.

32. Zwanzig, R. W. 1954. High-temperature equation of state by a perturbation method. I. Nonpolar gases. *J. Chem. Phys.* 22:1420–1426.

33. Yu, H.-A., and M. Karplus. 1988. A thermodynamic analysis of solvation. *J. Chem. Phys.* 89:2366–2379.

34. Grossfield, A., and D. M. Zuckerman. 2009. Quantifying uncertainty and sampling quality in biomolecular simulations. *Annu. Rep. Comput. Chem.* 5:23–48.

35. Flyvbjerg, H., and H. G. Petersen. 1989. Error estimates on averages of correlated data. *J. Chem. Phys.* 91:461.

36. Tomar, D. S., D. Asthagiri, and V. Weber. 2013. Solvation free energy of the peptide group: its model dependence and implications for the additive-transfer free-energy model of protein stability. *Biophys. J.* 105:1482–1490.

37. Berg, B. A., and R. C. Harris. 2008. From data to probability densities without histograms. *Comput. Phys. Commun.* 179:443–448.

38. Gu, W., S. J. Rahi, and V. Helms. 2004. Solvation free energies and transfer free energies for amino acids from hydrophobic solution to water solution from a very simple residue model. *J. Phys. Chem. B.* 108:5806–5814.

39. Gallicchio, E., M. M. Kubo, and R. M. Levy. 2000. Enthalpy-entropy and cavity decomposition of alkane hydration free energies: numerical

results and implications for theories of hydrophobic solvation. *J. Phys. Chem. B.* 104:6271–6285.

40. Makhatadze, G. I., M. M. Lopez, and P. L. Privalov. 1997. Heat capacities of protein functional groups. *Biophys. Chem.* 64:93–101.

41. Pratt, L. R. 2002. Molecular theory of hydrophobic effects: "She is too mean to have her name repeated.". *Annu. Rev. Phys. Chem.* 53:409–436.

42. Hummer, G., S. Garde, …, L. R. Pratt. 2000. New perspectives on hydrophobic effects. *Chem. Phys.* 258:349–370.

43. Tanford, C. 1962. Contribution of hydrophobic interactions to the stability of the globular conformation of proteins. *J. Am. Chem. Soc.* 84:4240–4247.

44. Rytting, J. H., L. P. Huston, and T. Higuchi. 1978. Thermodynamic group contributions for hydroxyl, amino, and methylene groups. *J. Pharm. Sci.* 67:615–618.

45. Wolfenden, R., L. Andersson, …, C. C. B. Southgate. 1981. Affinities of amino acid side chains for solvent water. *Biochemistry.* 20:849–855.

46. Eisenberg, D., and A. D. McLachlan. 1986. Solvation energy in protein folding and binding. *Nature.* 319:199–203.

47. Canchi, D. R., and A. E. García. 2011. Backbone and side-chain contributions in protein denaturation by urea. *Biophys. J.* 100:1526–1533.

48. Dill, K. A. 1997. Additivity principles in biochemistry. *J. Biol. Chem.* 272:701–704.

49. Pace, C. N., H. Fu, …, G. R. Grimsley. 2011. Contribution of hydrophobic interactions to protein stability. *J. Mol. Biol.* 408:514–528.

50. König, G., S. Bruckner, and S. Boresch. 2013. Absolute hydration free energies of blocked amino acids: implications for protein solvation and stability. *Biophys. J.* 104:453–462.

51. Staritzbichler, R., W. Gu, and V. Helms. 2005. Are solvation free energies of homogeneous helical peptides additive? *J. Phys. Chem. B.* 109:19000–19007.

52. Chang, J., A. M. Lenhoff, and S. I. Sandler. 2007. Solvation free energy of amino acids and side-chain analogues. *J. Phys. Chem. B.* 111:2098–2106.

53. Avbelj, F., and R. L. Baldwin. 2006. Limited validity of group additivity for the folding energetics of the peptide group. *Proteins.* 63:283–289.

54. Harris, R. C., and B. M. Pettitt. 2016. Reconciling the understanding of 'hydrophobicity' with physics-based models of proteins. *J. Phys. Condens. Matter.* 28:083003.

55. Della Gatta, G., T. Usacheva, …, D. Ichim. 2006. Thermodynamics of solvation of some small peptides in water at $T = 298.15$ K. *J. Chem. Thermodyn.* 38:1054–1061.

56. Baldwin, R. L. 2007. Energetics of protein folding. *J. Mol. Biol.* 371:283–301.

57. Meyer, E. E., K. J. Rosenberg, and J. Israelachvili. 2006. Recent progress in understanding hydrophobic interactions. *Proc. Natl. Acad. Sci. USA.* 103:15739–15746.

58. Weeks, J. D., D. Chandler, and H. C. Andersen. 1971. Role of repulsive forces in determining the equilibrium structure of simple liquids. *J. Chem. Phys.* 54:5237–5247.

59. Gallicchio, E., L. Y. Zhang, and R. M. Levy. 2002. The SGB/NP hydration free energy model based on the surface generalized Born solvent reaction field and novel nonpolar hydration free energy estimators. *J. Comput. Chem.* 23:517–529.

60. Ben-Amotz, D. 2016. Water-mediated hydrophobic interactions. *Annu. Rev. Phys. Chem.* 67:617–638.

61. Tomar, D. S., V. Weber, …, D. Asthagiri. 2016. Importance of hydrophilic hydration and intramolecular interactions in the thermodynamics of helix-coil transition and helix-helix assembly in a deca-alanine peptide. *J. Phys. Chem. B.* 120:69–76.

62. Ben-Amotz, D. 2015. Hydrophobic ambivalence: teetering on the edge of randomness. *J. Phys. Chem. Lett.* 6:1696–1701.

63. Anfinsen, C. B. 1972. The formation and stabilization of protein structure. *Biochem. J.* 128:737–749.

64. Anfinsen, C. B. 1973. Principles that govern the folding of protein chains. *Science.* 181:223–230.

65. Flory, P. J. 1989. Statistical Mechanics of Chain Molecules. Hanser-Gardner, Cincinnati, OH.

66. Das, R. K., and R. V. Pappu. 2013. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proc. Natl. Acad. Sci. USA.* 110:13392–13397.

67. Hu, C. Y., G. C. Lynch, …, B. M. Pettitt. 2010. Trimethylamine N-oxide influence on the backbone of proteins: an oligoglycine model. *Proteins.* 78:695–704.