Multibody correlations in the hydrophobic solvation of glycine peptides

Robert C. Harris, Justin A. Drake, and B. Montgomery Pettitt[a)]
Sealy Center for Structural Biology and Molecular Biophysics, University of Texas Medical Branch, 301 University Blvd, Galveston, Texas 77555-0304, USA

(Received 31 August 2014; accepted 4 November 2014; published online 1 December 2014)

Protein collapse during folding is often assumed to be driven by a hydrophobic solvation energy (ΔG_{vdw}) that scales linearly with solvent-accessible surface area (A). In a previous study, we argued that ΔG_{vdw}, as well as its attractive (ΔG_{att}) and repulsive (ΔG_{rep}) components, was not simply a linear function of A. We found that the surface tensions, γ_{rep}, γ_{att}, and γ_{vdw}, gotten from ΔG_{rep}, ΔG_{att}, and ΔG_{vdw} against A for four configurations of deca-alanine differed from those obtained for a set of alkanes. In the present study, we extend our analysis to fifty decaglycine structures and atomic decompositions. We find that different configurations of decaglycine generate different estimates of γ_{rep}. Additionally, we considered the reconstruction of the solvation free energy from scaling the free energy of solvation of each atom type, free in solution. The free energy of the isolated atoms, scaled by the inverse surface area the atom would expose in the molecule does not reproduce the γ_{rep} for the intact decaglycines. Finally, γ_{att} for the decaglycine conformations is much larger in magnitude than those for deca-alanine or the alkanes, leading to large negative values of γ_{vdw} (-74 and -56 cal/mol/Å^2 for CHARMM27 and AMBER ff12sb force fields, respectively). These findings imply that ΔG_{vdw} favors extended rather than compact structures for decaglycine. We find that ΔG_{rep} and ΔG_{vdw} have complicated dependencies on multibody correlations between solute atoms, on the geometry of the molecular surface, and on the chemical identities of the atoms. © *2014 AIP Publishing LLC*. [http://dx.doi.org/10.1063/1.4901886]

I. INTRODUCTION

Computing accurate solvation energies (ΔG) is a major goal of computational chemistry. If such calculations could be made simple and routine, then they would allow more complicated chemical transformations to be computed in vacuum, where such calculations are often easier to perform.[1,2] One common approach to computing ΔG is to divide it into two components: the energy of inserting an uncharged molecule (cavity) into solution (ΔG_{vdw}) and the energy of turning on the atomic partial charges (ΔG_{el}).[3,4] Several theories have been developed to predict ΔG_{el}, including approaches based on the Poisson equation (Poisson-Boltzmann[5] and generalized Born[6] methods), integral equations,[7] and structured continuum approaches.[8–10] Here, we tested various theories for computing ΔG_{vdw}.

Many researchers have assumed[2,11–13] that $\Delta G_{vdw} = \gamma_{vdw} A$, where A is the solvent-accessible surface area of the molecule and γ_{vdw} is a positive surface tension that is independent of the properties of the molecule. This behavior is expected for macroscopic interfaces with and cavities in solvent.[14] Because a solvation free energy that increased with A would favor compact conformations, the notion that ΔG_{vdw} drives the initial collapse during protein folding and aggregation is widely held.[2,15,16]

Other researchers have, however, noted that the arguments that lead to the idea that $\Delta G_{vdw} = \gamma_{vdw} A$ only account for the energy of expelling the water from the molecular cavity and do not account for the formation of favorable dispersive interactions between the solute and solvent.[17–25] These studies have therefore proposed to split ΔG_{vdw} into purely repulsive (ΔG_{rep}) and attractive (ΔG_{att}) components. Also, several other studies[26–29] have claimed that ΔG_{rep} should be proportional to the solvent-accessible volume (V) rather than A for sufficiently small solutes. Following these ideas, several studies have attempted to compute ΔG_{vdw} by assuming that ΔG_{rep} is a linear function of A and V and that ΔG_{att} can be computed from other means.[20,21,24]

Two other methods that have been used to estimate ΔG_{vdw} or ΔG_{rep} are to assign separate γ_{vdw} or γ_{rep} to different atom types[30] and in an analogy with macroscopic interfaces to add a correction term to γ_{vdw} to account for the curvature of the molecular surface.[31,32]

Recent work seems to contradict some of the assumptions underlying these models.[25,33–35] These studies demonstrated that ΔG_{vdw} decreases with the number of monomers for glycine and alanine peptides and that it also decreases with increasing A and V for decaalanine.[34] These findings appear to contradict the common hypothesis[2,15,16] that ΔG_{vdw} drives aggregation and collapse during peptide aggregation and protein collapse and folding. By decomposing ΔG_{vdw} into ΔG_{rep} and ΔG_{att} according to a Weeks-Chandler-Andersen (WCA) decomposition,[17,18] we showed that ΔG_{vdw} decreased with A for deca-alanine because although ΔG_{rep} did increase with A, this increase was more than matched by increasingly favorable ΔG_{att}.[25] This finding appears to support studies[17–24] that argue that ΔG_{att} should be computed separately from ΔG_{rep}. However, γ_{vdw} for a series of alkanes differed from

a)E-mail: mpettitt@utmb.edu. Telephone: 409 7720723. Fax: 409 7720725.

that for four configurations of decaalanines, and the derivatives ($\partial \Delta G_{rep}/\partial x_i$) of ΔG_{rep} with respect to the coordinates (x_i) of the atomic centers were not linear in $\partial A/\partial x_i$, as would be expected if γ_{vdw} were a well-defined quantity. These findings may help explain such apparent anomalies as the observation that ΔG increases at different rates with A for cyclic or branched, rather than linear, alkanes.[36] They also indicate that γ_{rep} is probably not universally well defined and that how ΔG_{rep} changes when an atom is moved depends not just on how A changes but also on multibody interactions with neighboring atoms and the local structure of the molecular surface.

In a recent study, we considered the decomposition of ΔG_{vdw} into ΔG_{rep} and ΔG_{att} for a series of alkanes and four conformers of deca-alanine.[25] The study raised questions about both variation with respect to chemistry and the sensitivity of the results to model parameters. In the present study, we extend our analysis to 50 sampled conformations of decaglycine taken over a few hundred nanoseconds and examine the sensitivity of our conclusions to the choice of force field by comparing two different force fields, CHARMM27[37,38] and AMBER ff12sb (models (1) and (2), respectively). Although the model force fields produce different quantitative results, the basic findings are consistent with our previous studies of decaalanines and alkanes. The values of γ_{rep}, γ_{att}, and γ_{vdw} differ significantly from those we found for deca-alanine and the alkanes, implying that these are not universally well-defined quantities, and that $\partial \Delta G_{rep}/\partial x_i$ is not correlated with $\partial A/\partial x_i$ for some atom types.

Additionally, here we consider ΔG_{rep} of each isolated atom type in the decaglycine to test atom-scaling models of hydrophobicity. We find the repulsive component of the free energy of the isolated atoms, scaled by the surface the atom would exposed in the molecule does not reproduce $\partial \Delta G_{rep}/\partial x_i$ versus $\partial A/\partial x_i$ for each atom type. This finding, combined with the poor correlations between $\partial \Delta G_{rep}/\partial x_i$ and $\partial A/\partial x_i$ for some atom types and the finding that the slopes of such plots for some atom types differed from the slopes obtained for the same atom types in decaalanine imply that using a separate surface tension for each atom type is not able to explain our results.

In the Results section, we obtain estimates of γ_{vdw} for decaglycines (-74 and -56 cal/mol/Å^2 for models (1) and (2)) that are larger and more negative than what we found for decaalanines (-3 cal/mol/Å^2) or alkanes (5 cal/mol/Å^2), implying that ΔG_{vdw} favors extended rather than compact structures for decaglycine.[25,34,35] We consider the implication of these results and the inability of area scaling the solvation free energy of isolated atoms to reproduce that for a molecule in the context of available models.

II. THEORY

Decomposing free energies into components can be challenging, as path dependencies in energy definitions can lead to difficulties in interpreting the results. Our chosen decomposition of ΔG_{vdw} into ΔG_{rep} and ΔG_{att} according to a WCA decomposition,[17,18] in contrast, is a well-defined decomposition with a strictly defined path. Because ΔG_{rep} is well-defined, we can examine its sensitivity to the molecular structure by computing its derivatives with respect to the atomic coordinates.[25]

A. Free energy definitions

We define ΔG_{vdw} to be the free energy required to go from a system where the solute and solvent do not interact to one where the interaction potential between solute and solvent atoms is given only by a Lennard-Jones potential,

$$U_{vdw}^{ij} = \varepsilon_{ij}\left[\left(r_{ij}^{min}/r_{ij}\right)^{12} - 2\left(r_{ij}^{min}/r_{ij}\right)^{6}\right], \tag{1}$$

where r_{ij} is the distance between the atoms, r_{ij}^{min} is the distance to the minimum of U_{vdw}, and ϵ_{ij} is the energy at the minimum of U_{vdw}. As we are computing ΔG_{vdw} and not ΔG, the electrostatic interaction between solute and solvent is not included. We define ΔG_{rep} to be the free energy required to go from a system where the solute and solvent did not interact to one where the interaction potential between solute and solvent atoms is given by the repulsive component of a Weeks-Chandler-Andersen[17,18] breakdown of ΔG_{vdw}

$$U_{rep}^{ij} = \varepsilon_{ij}\begin{cases}\left(\frac{r_{ij}^{min}}{r_{ij}}\right)^{12} - 2\left(\frac{r_{ij}^{min}}{r_{ij}}\right)^{6} + 1 & \text{if } r_{ij} < r_{ij}^{min}\\ 0 & \text{otherwise.}\end{cases} \tag{2}$$

We define $\Delta G_{att} \equiv \Delta G_{vdw} - \Delta G_{rep}$ and $U_{att}^{ij} \equiv U_{vdw}^{ij} - U_{rep}^{ij}$.

B. Computation of free energies and free energy derivatives

To compute the energy (ΔG_{rep}^{i}) of inserting a single atom of each atom type i in the force field, we used free energy perturbation (FEP).[39,40] A λ-dependent potential with a soft core was used,

$$U(\lambda) = \lambda\varepsilon_{ij}\sum\begin{cases}\left(\frac{r_{ij}^{min}}{r_{ij}(\lambda)}\right)^{12} - 2\left(\frac{r_{ij}^{min}}{r_{ij}(\lambda)}\right)^{6} + 1 & \text{if } r_{ij} < r_{ij}^{min}\\ 0 & \text{otherwise,}\end{cases} \tag{3}$$

where $r_{ij}(\lambda) = (r_{ij}^2 + (1-\lambda)(r_{ij}^{min})^2)^{1/2}$.

Linear response theory (LRT) estimates of ΔG_{att} (ΔG_{att}^{lrt}) were computed by

$$\Delta G_{att}^{lrt} = 1/2[\langle U_{att}\rangle_0 + \langle U_{att}\rangle_1], \tag{4}$$

where $\langle\ldots\rangle_0$ signifies an average over an ensemble where the solute and solvent interaction potential is U_{rep}^{ij}, $\langle\ldots\rangle_1$ signifies an average over an ensemble where the potential is U_{vdw}^{ij}, and $U_{att} = \sum U_{att}^{ij}$, where this summation was taken over all solute-solvent atom pairs.

As in our previous study,[25] $\partial \Delta G_{rep}/\partial x_i$ was computed by

$$\partial \Delta G_{rep}/\partial x_i = \langle \partial U_{rep}/\partial x_i\rangle_0, \tag{5}$$

where $U_{rep} = \sum U_{rep}^{ij}$, again taken over all solute-solvent atom pairs.

Once estimates of $\partial \Delta G_{rep}/\partial x_i$ had been obtained, perturbative estimates (γ_{rep}^{der}) of γ_{der} could be obtained by computing the slopes of best-fit lines of plots of $\partial \Delta G_{rep}/\partial x_i$ against $\partial A/\partial x_i$.

III. METHODS

The simulations used to obtain the configurations of decaglycine in models (1) and (2) were run with NAMD 2.9.[41] To create the conformations of decaglycine in model (1), a simulation was run with the CHARMM27 force field.[37,38] First, an extended decaglycine was created with the MOLEFACTURE plugin in VMD[42] and placed in a 45×45×45 Å water box. It was then minimized for 10 000 steps and equilibrated for 20 ps. We then simulated for 300 ns, and configurations were saved every 10 ps. The structures with the largest and smallest A were taken from these configurations. The remaining 48 structures were selected to ensure an even spacing in A between the structures with the largest and smallest A. To create the conformations of decaglycine in model (2), the same protocol was followed except that the initial extended decaglycine was created with xleap[43] and the force field parameters were taken from the AMBER ff12sb force field.

To compute $\Delta G_{\rm att}^{\rm lrt}$ and $\partial G/\partial x_i$, each selected decaglycine structure was placed in a water box 20 Å larger than the molecule in each dimension, and these systems were minimized for 5000 steps. Two copies of each system were then created, one where the interaction potential between the solute and solvent was $U_{\rm rep}^{ij}$ and the other where it was $U_{\rm vdw}^{ij}$. The temperatures of these systems were then increased from 25 to 300 K in increments of 25 K, with 2 ps simulations at each temperature. Next, each system was simulated for 1 ns, snapshots were taken every 0.2 ps, and these frames were used to compute the averages in Eq. (4). We computed $\partial \Delta G_{\rm rep}/\partial x_i$ from the same simulations we used to compute $\Delta G_{\rm att}^{\rm lrt}$.

ALPHASURF in the PROGEOM package[44] was used to compute A and $\partial A/\partial x_i$. As in our previous work, A was defined to be the solvent-accessible surface area of the molecule,[45] defined with a probe radius of 1.7682 Å (the van der Waals radius of a water oxygen in the CHARMM27 force field) rather than the more normal 1.4 Å. The traditional choice of a radius of 1.4 Å is due to the OO peak in the water-water radial distribution function. However, this exceptionally close approach between water molecules (more than 0.5 Å closer than the minimum in their $U_{\rm vdw}^{ij}$) is caused by the strong favorable hydrogen-bonding energy between neighboring molecules. In our calculations, the solute molecules contained no charges, so there were no similar strong attractive forces between the solute and solvent to change the effective radius in $U_{\rm vdw}^{ij}$. For this reason and because it yielded slightly better correlations between $\Delta G_{\rm rep}$ and A in our previous study on decaalanines we made this area defining choice.[25,35]

Another approach that might be considered would be to use alternative surface definitions, and curvature corrections such as those explored by Honig and co-workers.[31] However, many alternative surfaces are often approximated with triangulated or other meshes. We did attempt to use other surface definitions in our previous work on decaalanines,[25] but we did not include the analysis because the resulting estimates of $\partial A/\partial x_i$ were frequently not consistent with finite-difference estimates of these derivatives. The type of analysis performed in the present study, where we compare $\partial A/\partial x_i$ to $\partial \Delta G_{\rm rep}/\partial x_i$ to test the proposition that $\Delta G_{\rm rep}$ is linear in A, would not converge well with such triangulated surface definitions.

To compute the repulsive component ($\Delta G_{\rm rep}^{\rm i}$) of the $\Delta G_{\rm vdw}$ of inserting a single atom of type i into solution, λ-space was divided into 10 equally spaced windows. A single solute atom of each type was placed in a water box 20 Å on a side, and this structure underwent 50 000 steps of minimization. From this minimized structure, initial structures for each λ value were created by increasing the temperature of this box from 25 to 300 K in 25 K increments. At each temperature, the system was simulated for 2 ps. Each of the resulting systems was then run for 1 ns, configurations were taken every 0.2 ps, and these frames were used in the FEP calculations. Although FEP is more accurate when a particle is being inserted rather than removed from solution,[46] the differences between the estimates obtained from forward FEP and those from backward FEP provided rough estimates of the errors in these calculations.

IV. RESULTS

Panels (a) and (c) of Figure 1 show $\gamma_{\rm rep}^{\rm der}$ as a function of A for models (1) and (2). Rather than being a constant, $\gamma_{\rm rep}^{\rm der}$ ranged from 19.7 to 50.3 and 15.8 to 41.4 cal/mol/Å^2 for models (1) and (2), and the correlations between $\partial \Delta G_{\rm rep}^{\rm der}/\partial x_i$ and $\partial A/\partial x_i$ were weak for many of the configurations of decaglycine. (The Pearson's correlation coefficients (R^2) between these two quantities ranged from 0.53 to 0.84 and 0.33 to 0.78 for models (1) and (2), respectively.) The average values of $\gamma_{\rm rep}^{\rm der}$ from models (1) and (2) (40.4 and 32.2 cal/mol/Å^2) differed from what we found for decaalanine (43 cal/mol/Å^2), but the range of $\gamma_{\rm rep}^{\rm der}$ in model (1) did at least contain the value of $\gamma_{\rm rep}$ we found for decaalanine. The values of $\gamma_{\rm rep}$ obtained here also differed significantly from what we found for a series of alkanes (69 cal/mol/Å^2), implying that $\gamma_{\rm rep}$ is not a well-defined quantity.

To verify that the estimates of $\gamma_{\rm rep}^{\rm der}$ had converged, we also computed these quantities from estimates of $\partial \Delta G_{\rm rep}/\partial x_i$ obtained from the first halves of the trajectories used to generate our estimates of $\gamma_{\rm rep}^{\rm der}$ and estimates of $\partial A/\partial x_i$ obtained from finite difference derivatives computed by moving each atom 0.001 Å in each direction. These coarser estimates of $\gamma_{\rm rep}^{\rm der}$ differed from those in Figure 1 by less than 1.1 cal/mol/Å^2.

Panels (b) and (d) of Figure 1 show $\Delta G_{\rm att}^{\rm lrt}$ plotted as a function of A for models (1) and (2). As in our previous study, this energy was roughly linear in A, although here the correlations between $\Delta G_{\rm att}^{\rm lrt}$ and A ($R^2 = 0.92$ and 0.98 for models (1) and (2)) were weaker than that between $\Delta G_{\rm att}$ and A for decaalanine in our previous study ($R^2 = 0.99$). The weaker correlations in the present study may be a consequence of the larger number of diverse configurations, which allowed us to better test the proposition that $\Delta G_{\rm att}$ should be linear in A. Furthermore, the estimates of $\gamma_{\rm att}^{\rm lrt}$ obtained from this data (−114 and −88 cal/mol/Å^2 for models (1) and (2)) differed from either what we observed for the decaalanine peptides ($\gamma_{\rm att} = -48$ cal/mol/Å^2) or what we observed for the alkanes ($\gamma_{\rm att} = -64$ cal/mol/Å^2).

As in our previous study, whether $\partial \Delta G_{\rm rep}/\partial x_i$ was proportional to $\partial A/\partial x_i$ depended on the atom type considered.

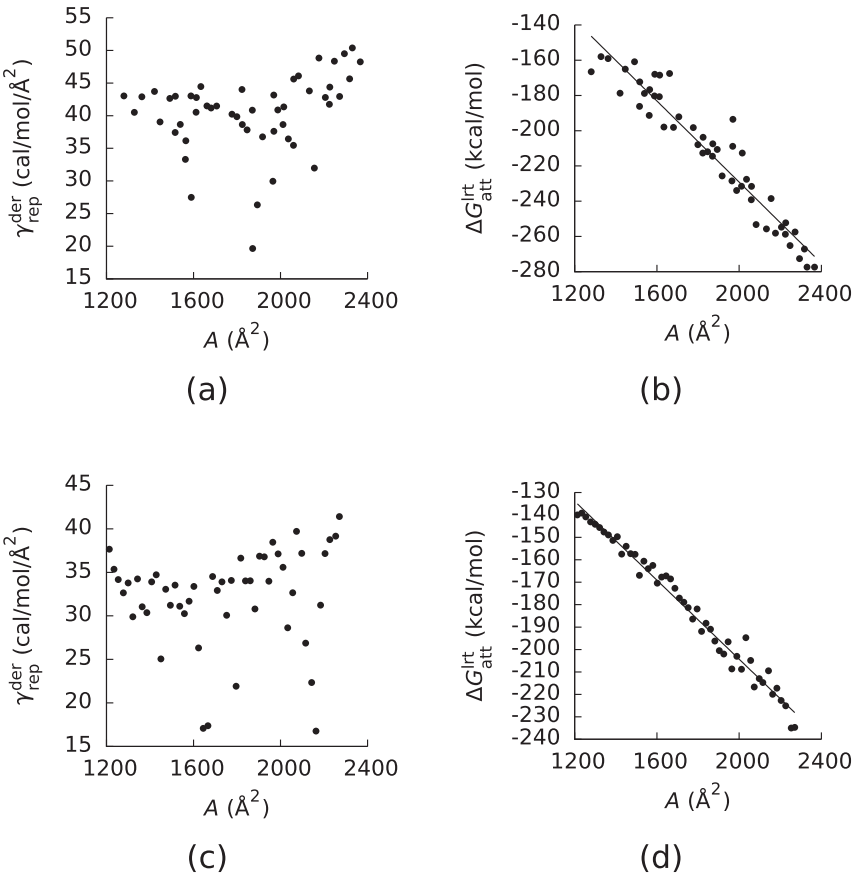

FIG. 1. (a) The slope (γ_{rep}^{der}) of the least-squares lines between the derivatives ($\partial \Delta G_{rep}^{der}/\partial x_i$) of the repulsive component (ΔG_{rep}) of the van der Waals component (ΔG_{vdw}) of the solvation free energy with respect to the coordinates (x_i) of the atomic centers and the derivative ($\partial A/\partial x_i$) of the solvent-accessible surface area (A) with respect to the x_i as a function of A. (b) Linear response theory estimates (ΔG_{att}^{lrt}) of the attractive component of ΔG_{vdw} as a function of A. The values in (a) and (b) were computed with the CHARMM27 force field and configurations taken from a simulation run with CHARMM27. (c) and (d) The same as (a) and (b), respectively, but the values in these plots were computed with AMBER ff12sb and configurations taken from a simulation run with AMBER ff12sb. The slopes of the least-squares lines in (c) and (d) were -114 and -88 cal/mol/Å^2, and the Pearson's correlation coefficients of these lines were $R^2 = 0.92$ and 0.98.

In panels (a) and (c) of Figure 2, $\partial \Delta G_{rep}/\partial x_i$ is plotted as a function of $\partial A/\partial x_i$ for models (1) and (2), and the points are colored by their atom types in the CHARMM27 force field. Atoms of type C were the carbonyl carbons, atoms of type CT2 were the α-carbons, atoms of type CT3 were the terminal carbons, atoms of type HA were the hydrogens bound to the terminal carbons, atoms of type HB were the hydrogens bound to the α-carbons, atoms of type NH1 were the nitrogens, and atoms of type O were the oxygens. The atom types in AMBER ff12sb do not exactly match those in the CHARMM27 force field (AMBER ff12sb groups the α- and terminal carbons into the same atom type (CT), it groups the hydrogens bound to the α-carbons with those that bind to the N-terminal carbons into the same atom type (H), and it gives a separate atom type (HC) to hydrogens that bind to the C-terminal carbons), but because the α- and terminal carbons had different relationships between $\partial \Delta G_{rep}/\partial x_i$ and $\partial A/\partial x_i$ in both models (Figure 2 and Table I), we classified the atoms according to the CHARMM27 force field. Given this difference in classification, the estimates of $\partial \Delta G_{rep}/\partial x_i$ and $\partial A/\partial x_i$ in model (2) were computed with the parameters from the AMBER ff12sb force field.

Table I contains the slopes ($\gamma_{rep}^{der,i}$) of plots of $\partial \Delta G_{rep}/\partial x_i$ against $\partial A/\partial x_i$ for each atom type in the CHARMM27 force field, computed on both models (1) and (2), along with the R^2 of these plots. Also shown in Table I are ΔG_{rep}^{i} of each atom type i in the CHARMM force field divided by A. If γ_{rep} did not depend on the chemical environment of the atom, then $\Delta G_{rep}^{i}/A$ would equal $\gamma_{rep}^{der,i}$. Clearly, $\Delta G_{rep}/A$ differs significantly from $\gamma_{rep}^{der,i}$. This observation further indicates that γ_{rep} depends on the molecular environment of the atom in question and is therefore not really a well-defined quantity.

As can be seen in Figure 2, the value of γ_{rep}^{der} is determined primarily by atoms of types HA, HB, and O, both because these atom types had the largest $\partial \Delta G_{rep}/\partial x_i$ and because their $\partial \Delta G_{rep}/\partial x_i$ were roughly proportional to $\partial A/\partial x_i$. Table I shows that the correlations between $\partial \Delta G_{rep}/\partial x_i$ and $\partial A/\partial x_i$ were significant for these three atom types in each model. Additionally, $\partial \Delta G_{rep}/\partial x_i$ was significantly correlated with $\partial A/\partial x_i$ for atoms of type CT2 and CT3 in model (1), but not in model (2).

Some of the statistics in Table I differ from those in our previous work on decaalanine and various alkanes.[25] In that study, $\partial \Delta G_{rep}/\partial x_i$ was correlated with $\partial A/\partial x_i$ for the oxygens and hydrogens, as in the present study. These two quantities were significantly correlated for the α-carbons in model (1), whereas $\partial \Delta G_{rep}/\partial x_i$ was not significantly correlated with $\partial A/\partial x_i$ for the α-carbons in the decaalanines. Additionally,

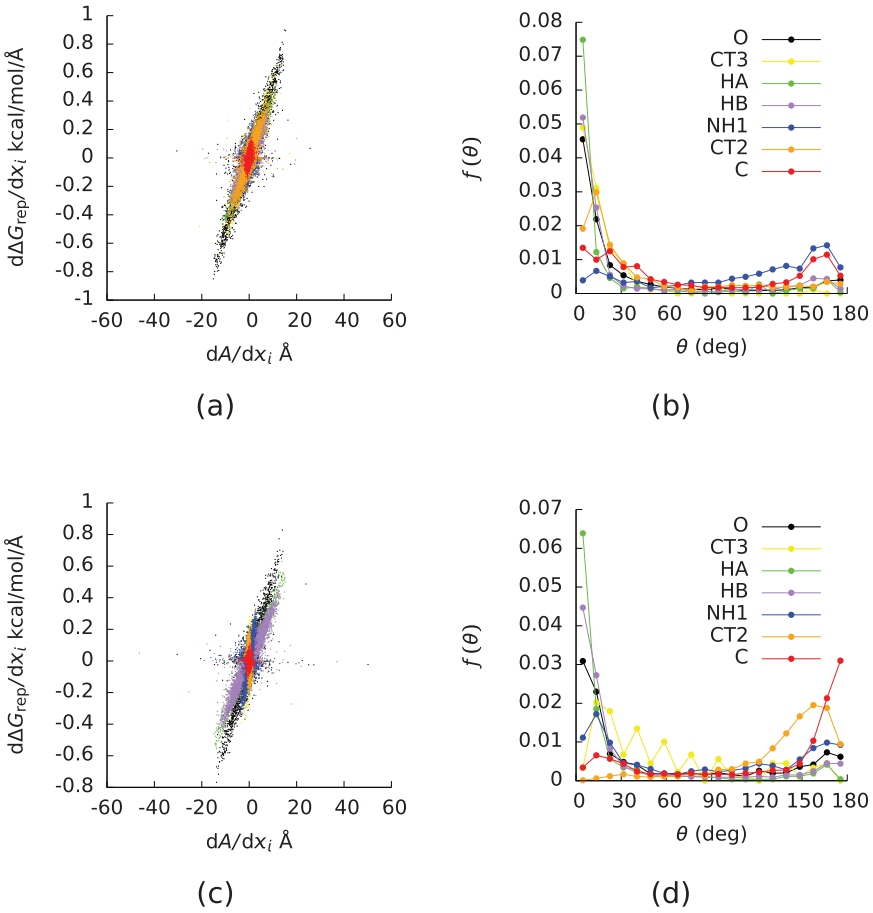

FIG. 2. (a) The derivative ($\partial\Delta G_{\text{rep}}/\partial x_i$) of the repulsive ($\Delta G_{\text{rep}}$) component of the van der Waals component (ΔG_{vdw}) of the solvation free energy with respect to the coordinates (x_i) of the atomic centers as a function of the derivative ($\partial A/\partial x_i$) of the solvent-accessible surface area (A) with respect to the x_i. (b) The probability densities (f) of the angle (θ) between $\nabla_i A$ and $\nabla_i \Delta G_{\text{rep}}$ where $\nabla_i = (\partial/\partial x_i, \partial/\partial y_i, \partial/\partial z_i)$ and (x_i, y_i, z_i) were the coordinates of the center of atom i. The points in (a) and the curves in (b) are colored by the atom type in the CHARMM27 force field. The values in (a) and (b) were computed with the CHARMM27 force field and configurations taken from a simulation run with CHARMM27. (c) and (d) The same as (a) and (b), respectively, but the values in these plots were computed with AMBER ff12sb and configurations taken from a simulation run with AMBER ff12sb.

in our previous study $\partial\Delta G_{\text{rep}}/\partial x_i$ was correlated with $\partial A/\partial x_i$ for the nitrogens and carbonyl carbons, but the $\gamma_{\text{rep}}^{\text{der},i}$ were negative. In contrast, $\partial\Delta G_{\text{rep}}/\partial x_i$ and $\partial A/\partial x_i$ were not correlated for these atom types in the present study. These findings demonstrate how the relationship between $\partial\Delta G_{\text{rep}}/\partial x_i$ and $\partial A/\partial x_i$ for an atom type can change when the surrounding atoms in the solute molecule change.

TABLE I. FEP estimates of $\gamma_{\text{rep}}^{\text{der},i}$ for each atom type, Pearson's correlation coefficients (R^2) of the corresponding least-squares lines of plots of $\partial\Delta G_{\text{rep}}/\partial x_i$ versus $\partial A/\partial x_i$, and estimates of $\Delta G_{\text{rep}}^{i}/A$ obtained by inserting a single atom of each type into solvent and dividing by the solvent-accessible surface areas of these atoms. All $\gamma_{\text{rep}}^{\text{der}}$ and $\Delta G_{\text{rep}}/A$ were in units of (cal/mol/Å^2).

	CHARMM27			AMBER ff12sb	
Atom type	$\Delta G_{\text{rep}}^{i}/A$	$\gamma_{\text{rep}}^{\text{der},i}$	R^2	$\gamma_{\text{rep}}^{\text{der},i}$	R^2
C	32	18	0.11	−4	0.01
CT2	32	38	0.69	−54	0.15
CT3	32	48	0.86	12	0.44
HA	21	38	0.89	32	0.89
HB	22	34	0.78	28	0.83
NH1	33	8	0.02	22	0.14
O	30	45	0.85	38	0.69

Panels (b) and (d) of Figure 2 show the probability densities ($f(\theta)$) of the angle (θ) between $\nabla_i \Delta G_{\text{rep}}$ and $\nabla_i A$, where $\nabla_i = (\partial/\partial x_i, \partial/\partial y_i, \partial/\partial z_i)$ and (x_i, y_i, z_i) were the coordinates of the center of the atom i. If ΔG_{rep} were perfectly linear in A, then θ would equal 0°. For atom types HA, HB, and O, $f(\theta)$ was peaked at 0°, and these were the atom types noted above that determined $\gamma_{\text{rep}}^{\text{der}}$. Interestingly, some of the other atom types had significantly different $f(\theta)$ in the two models; (θ) was peaked at 180° for atom types C and CT2 in model (2) but not in model (1). In our previous study, we suggested that atom types for which $f(\theta)$ were peaked at 180° might tend to lie at the bottom of valleys in the solvent-accessible surface.[25] Additionally, some of these $f(\theta)$ differ from those we found for decaalanine. In that case, $f(\theta)$ was peaked at 180° for atoms of type NH1 and C, whereas in the present study $f(\theta)$ showed no clear peak for atoms of type NH1 and was only peaked at 180° for atoms of type C in model (2). These findings further emphasize the effects of the local molecular geometry and chemical environment on solvation free energy.

V. CONCLUSIONS

By examining 50 conformations of decaglycine, we were able to test the propositions that ΔG_{rep}, ΔG_{att}, and ΔG_{vdw} are linear functions of A more extensively than in our previous works.[25,33–35] The different configurations of decaglycine produced γ_{rep}^{der} that covered a large range of values, implying that γ_{rep} may not be a well defined quantity at the molecular scale we are investigating. Additionally, the correlations between ΔG_{att}^{lrt} and A in this study were weaker than those between ΔG_{att} and A for either decaalanine or alkanes, and the magnitude of γ_{att} was significantly larger than those we observed for either decaalanine or alkanes. As a result, the estimates of γ_{vdw} obtained in the present study (−74 and −56 cal/mol/Å^2 for CHARMM27 and AMBER ff12sb) were larger than what we observed previously. These findings, in combination with our previous findings, seem to imply that none of ΔG_{rep}, ΔG_{att}, and ΔG_{vdw} is a simple linear function of A for a wide range of molecules and molecular shapes.

That the full γ_{vdw} for decaglycine was large and negative confirmed previous findings on shorter glycine oligomers.[33,34] Previously, we found that γ_{vdw} was negative for decaalanine because $-\gamma_{att} > \gamma_{rep}$ and $\gamma_{att} < 0$, but in that case, the magnitude of γ_{vdw} was fairly small.[25] One could have argued that ΔG_{vdw} could simply be neglected for that system. In contrast, the large γ_{vdw} for these decaglycines is probably not negligible and that it is negative implies that ΔG_{vdw} favors extended over compact states for these peptides. Apparently, whether ΔG_{vdw} favors extended or compact conformations and how strongly will have to be determined on a case-by-case basis.

We found that the relationship between $\partial \Delta G_{rep}$ and $\partial A/\partial x_i$ was different for different types of atoms. Overall γ_{rep}^{der} was apparently largely determined by the well exposed hydrogen and oxygen atoms in the peptides. Additionally, the relationships between $\partial \Delta G_{rep}/\partial x_i$ and $\partial A/\partial x_i$ for some of these atom types differed from what we saw for decaalanines. In our previous study, we found that $f(\theta)$ was peaked at 180° for the carbonyl carbons and nitrogens in decaalanine, whereas in these decaglycines $f(\theta)$ for the nitrogens had no clear peak and it was only peaked for the carbonyl carbons when the AMBER ff12sb force field was used. Furthermore, for several atom types $\partial \Delta G_{rep}/\partial x_i$ was not correlated with $\partial A/\partial x_i$. These observations could not be explained by assigning different surface tensions to different atom types.

Collectively, these findings imply that none of ΔG_{rep}, ΔG_{att}, and ΔG_{vdw} can be assumed to be linear in A with well-defined surface tensions (γ_{rep}, γ_{att}, and γ_{vdw}). Instead, each atom's contribution to these free energies appears to depend on its chemical identity and multibody interactions with neighboring solute atoms. These contributions therefore appear to contain nontrivial dependencies on the local structure of the molecular surface. Successful hydrophobic theories will have to account for such interactions.

ACKNOWLEDGMENTS

The Robert A. Welch Foundation (H-0037), the National Science Foundation (NSF) (CHE-1152876) and the National Institutes of Health (NIH) (GM-037657) are thanked for partial support of this work. This research was performed in part using the Kraken and Stampede systems, part of the National Science Foundation XSEDE resources.

[1] C. S. Babu and B. L. Tembe, J. Chem. Sci. **98**, 235 (1987).
[2] R. Baldwin, J. Mol. Biol. **371**, 283 (2007).
[3] K. A. Sharp and B. Honig, Annu. Rev. Biophys. Biophys. Chem. **19**, 301 (1990).
[4] C. J. Cramer and D. G. Truhlar, Chem. Rev. **99**, 2161 (1999).
[5] P. Grochowski and J. Trylska, Biopolymers **89**, 93 (2008).
[6] D. Bashford and D. A. Case, Annu. Rev. Phys. Chem. **51**, 129 (2000).
[7] J.-F. Truchon, B. M. Pettitt, and P. Labute, J. Chem. Theory Comput. **10**, 934 (2014).
[8] V. Lounnas, B. M. Pettitt, and G. N. Phillips, Jr., Biophys. J. **66**, 601 (1994).
[9] V. Makarov, B. M. Pettitt, and M. Feig, Acc. Chem. Res. **35**, 376 (2002).
[10] B. Lin and B. M. Pettitt, J. Comput. Chem. **32**, 878 (2011).
[11] F. H. Stillinger, J. Sol. Chem. **2**, 141 (1973).
[12] R. A. Pierotti, Chem. Rev. **76**, 717 (1976).
[13] D. Sitkoff, K. A. Sharp, and B. Honig, J. Phys. Chem. **98**, 1978 (1994).
[14] T. Young, Philos. Trans. R. Soc. **95**, 65 (1805).
[15] E. E. Meyer, K. J. Rosenberg, and J. Israelachvili, Proc. Natl. Acad. Sci. U.S.A. **103**, 15739 (2006).
[16] P. Ball, Chem. Rev. **108**, 74 (2008).
[17] J. D. Weeks, D. Chandler, and H. C. Andersen, J. Chem. Phys. **54**, 5237 (2003).
[18] D. Chandler, J. D. Weeks, and H. C. Andersen, Science **220**, 787 (1983).
[19] E. Gallicchio, M. M. Kubo, and R. M. Levy, J. Phys. Chem. B **104**, 6271 (2000).
[20] E. Gallicchio, L. Y. Zhang, and R. M. Levy, J. Comput. Chem. **23**, 517 (2002).
[21] M. Zacharias, J. Phys. Chem. A **107**, 3000 (2003).
[22] N. Choudhury and B. M. Pettitt, J. Am. Chem. Soc. **127**, 3556 (2005).
[23] N. Choudhury and B. M. Pettitt, Mol. Simul. **31**, 457 (2005).
[24] J. A. Wagoner and N. A. Baker, Proc. Natl. Acad. Sci. U.S.A. **103**, 8331 (2006).
[25] R. C. Harris and B. M. Pettitt, Proc. Natl. Acad. Sci. U.S.A. **111**, 14681 (2014).
[26] K. Lum, D. Chandler, and J. D. Weeks, J. Phys. Chem. B **103**, 4570 (1999).
[27] D. M. Huang and D. Chandler, Proc. Natl. Acad. Sci. U.S.A. **97**, 8324 (2000).
[28] G. Hummer, S. Garde, A. E. Garcıa, and L. R. Pratt, Chem. Phys. **258**, 349 (2000).
[29] S. Rajamani, T. M. Truskett, and S. Garde, Proc. Natl. Acad. Sci. U.S.A. **102**, 9475 (2005).
[30] D. Eisenberg and A. D. McLachlan, Nature (London) **319**, 199 (1986).
[31] A. Nicholls, K. A. Sharp, and B. Honig, Prot.: Struct., Funct., Bioinf. **11**, 281 (1991).
[32] K. A. Sharp, A. Nicholls, R. F. Fine, and B. Honig, Science **252**, 106 (1991).
[33] C. Y. Hu, H. Kokubo, G. C. Lynch, D. W. Bolen, and B. M. Pettitt, Prot. Sci. **19**, 1011 (2010).
[34] H. Kokubo, C. Y. Hu, and B. M. Pettitt, J. Am. Chem. Soc. **133**, 1849 (2011).
[35] H. Kokubo, R. C. Harris, D. Asthagiri, and B. M. Pettitt, J. Phys. Chem. B **117**, 16428 (2013).
[36] T. Simonson and A. T. Bruenger, J. Phys. Chem. **98**, 4683 (1994).
[37] A. D. MacKerell, Jr., D. Bashford, M. Bellott, R. L. Dunbrack, Jr., J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher III, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wieórkiewicz-Kuczera, D. Yin, and M. Karplus, J. Phys. Chem. B **102**, 3586 (1998).
[38] A. D. MacKerell, Jr., M. Feig, and C. L. Brooks III, J. Comput. Chem. **25**, 1400 (2004).
[39] D. L. Beveridge and F. M. DiCapua, Annu. Rev. Biophys. Biophys. Chem. **18**, 431 (1989).
[40] T. P. Straatsma and J. A. McCammon, Annu. Rev. Phys. Chem. **43**, 407 (1992).
[41] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, and K. Schulten, J. Comput. Chem. **26**, 1781 (2005).

[42]W. Humphrey, A. Dalke, and K. Schulten, J. Mol. Graph. **14**, 33 (1996).
[43]D. A. Case, T. A. Darden, T. E. Cheatham III, C. L. Simmerling, J. Wang, R. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A. W. Götz, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, R. M. Wolf, J. Liu, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D. R. Roe, D. H. Mathews, M. G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and P. A. Kollman, *Amber 12* (University of California, San Francisco, 2012).
[44]H. Edelsbrunner and P. Koehl, Proc. Natl. Acad. Sci. U.S.A. **100**, 2203 (2003).
[45]B. Lee and F. M. Richards, J. Mol. Biol. **55**, 379 (1971).
[46]A. Pohorille, C. Jarzynski, and C. Chipot, J. Phys. Chem. B **114**, 10235 (2010).