THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

Conditional Solvation Thermodynamics of Isoleucine in Model Peptides and the Limitations of the Group-Transfer Model

Dheeraj S. Tomar,[†] Valéry Weber,[‡] B. Montgomery Pettitt,[§] and D. Asthagiri*[,†,§]

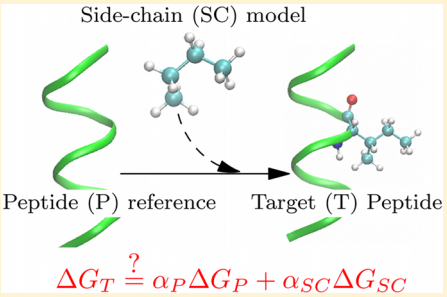
[†]Department of Chemical and Biomolecular Engineering, Johns Hopkins University, 3400 North Charles Street, Baltimore, Maryland 21218, United States

[‡]IBM Research, Zurich, Switzerland

[§]Sealy Center for Structural Biology and Molecular Biophysics, University of Texas Medical Branch, 301 University Boulevard, Galveston, Texas 77555, United States

Supporting Information

ABSTRACT: The hydration thermodynamics of the amino acid X relative to the reference G (glycine) or the hydration thermodynamics of a small-molecule analog of the side chain of X is often used to model the contribution of X to protein stability and solution thermodynamics. We consider the reasons for successes and limitations of this approach by calculating and comparing the conditional excess free energy, enthalpy, and entropy of hydration of the isoleucine side chain in zwitterionic isoleucine, in extended penta-peptides, and in helical deca-peptides. Butane in gauche conformation serves as a small-molecule analog for the isoleucine side chain. Parsing the hydrophobic and hydrophilic contributions to hydration for the side chain shows that both of these aspects of hydration are context-sensitive. Furthermore, analyzing the solute−solvent interaction contribution to the conditional excess enthalpy of the side chain shows that what is nominally considered a property of the side chain includes entirely nonobvious contributions of the background. The context-sensitivity of hydrophobic and hydrophilic hydration and the conflation of background contributions with energetics attributed to the side chain limit the ability of a single scaling factor, such as the fractional solvent exposure of the group in the protein, to map the component energetic contributions of the model-compound data to their value in the protein. But ignoring the origin of cancellations in the underlying components the group-transfer model may appear to provide a reasonable estimate of the free energy for a given error tolerance.



INTRODUCTION

The hydration thermodynamics of analogs of amino acid side chains (or of amino acid side chains in small-model peptides) has often been used to understand protein folding and protein−protein association. Examples of such approaches abound in the biochemical literature. A very small sampling of the many pioneering investigations where this approach has been used includes work identifying hydrophobicity as a dominant force in protein folding,[1] investigations of protein denaturation,[2−5] interpretation of calorimetric data on protein unfolding,[6−8] and, more recently, investigations on the role of osmolytes in protein folding.[9−11] Drawing upon insights attributed to Cohn and Edsall, Tanford[2,4] formulated this approach into a quantitative, predictive framework. In his approach, the free energy of unfolding is given as a sum of the free energy of transfer of "the small component groups of the molecule, from the environment they have in the native form, to the environment they have in the unfolded form".[2] Accounting for subsequent refinements that included corrections for the solvent-exposure of the "small component groups" (cf. ref 9), the generic equation of the unfolding free energy ($\Delta G_{N\to U}$) can be written as

$$\Delta G_{N\to U} = \sum_i \alpha_i \Delta g_i \tag{1}$$

where Δg_i is the free energy of transferring the group i from some reference phase to liquid water and α_i is a factor that corrects for the change in solvent-exposure of the group i in protein unfolding. In interpretation of calorimetric data,[8] for example, a gas-phase reference is natural, as is also the case in this article. The same form of equation can also be used to describe the effect of osmolytes on protein unfolding: identifying Δg_i with the free energy for transferring the group from water to the aqueous osmolyte solution and α_i with the change in the solvent exposure of the group upon unfolding in the osmolyte solution relative to water, the previous equation directly gives the so-called m value for 1 M osmolyte.[12]

At a time when theory, simulations, and experimental techniques were less developed than they are now, the group-additive approach was a pragmatic first step to understand the hydration thermodynamics of a complicated macromolecule. But it is also important to assess its limitations and probe if the

Received: January 21, 2014
Revised: March 20, 2014
Published: March 20, 2014

   dx.doi.org/10.1021/jp500727u | J. Phys. Chem. B 2014, 118, 4080−4087

physical conclusions based on this approach are valid. For example, the group-additive model appears to capture the effect of osmolytes rather well,[9] with the predicted and experimentally determined m values agreeing to within a couple kilocalories per mole for proteins with about 100 residues. On this basis, it has been suggested that some osmolytes largely act by tuning the solubility of the peptide backbone,[13] but it is unclear how to reconcile this observation with simulations that suggest that urea-induced denaturation is mediated by promiscuous urea–protein interactions tuned by typically nonspecific dispersion forces.[14,15] Furthermore, our previous study on the solvation of a peptide group[16] showed that in the water-to-osmolyte transfer solvent-mediated correlations between peptide units (in Gly_n) cancel, and this allows one to identify a peptide-group transfer free energy that is model-independent. But in the vapor-to-liquid transfer, a situation where solvent-mediated correlations are preserved, the identified group-additive contribution (Δg_i, eq 1) depends on the model used to define the peptide group.

Several recent studies have explored the limitations of the additive model.[16−19] While our work[16] emphasized solvent-mediated correlations, König et al.[18,19] have emphasized the role of "self-solvation"[20] in limiting additivity. In exploring context dependence of hydration, these authors also suggested how adding a methyl group to an amino acid changes the solvent-exposure of the peptide backbone, thereby influencing the excess free energy in a way that cannot be captured by side-chain analog data. Building on these efforts, here we study how well a model of the side chain describes the hydration of the side chain in the context of a protein. In particular, our aim is to better understand factors limiting additivity and the reasons why scaling model-compound data to describe its properties in a protein context may not always be satisfactory. To aid in parsing energetic differences, we study the conditional solvation of an isoleucine residue, often regarded as very hydrophobic,[21] in the context of model-extended and helical peptides. Butane in the gauche conformation, matching exactly the side-chain conformation of isoleucine in extended peptides, is used as a small-molecule analog of the side chain.

In obtaining mechanistic insight from free-energy calculations, it is common to use decompositions of the free energy. Any decomposition into components that are not themselves state functions can lead to path dependencies in the analysis. However, with adequate care,[22,23] such components can yield insights into the molecular scale features underlying the free energy that are of first interest. Here we use the quasichemical approach[24−26] to separate the packing (steric) contributions from the hydrophilic contributions to both facilitate the calculation[27,28] and also provide insights into the limitations of additivity. These components (but not the net free energy) do depend on the specification of a hydration shell, but the approach has nevertheless provided important insights into the physics of hydration and in generating models of molecular solutions.[24−26,29,30]

We complement the quasichemical analysis with the traditional decomposition of the excess free energy into its enthalpic and entropic contributions. The quasichemical approach helps us better appreciate how the context influences the hydration of the solute, while the enthalpy entropy decomposition leads to the finding that nonobvious contributions from the reference (the context) get folded into the conditional contribution attributed to the side chain (the solute) that is of first interest in additive models.

The rest of the article is organized as follows. In the next section, we outline the Theory. Next, we outline the Methods, following which we present the Results and Discussion. The Concluding Discussion summarizes the main observations of this study.

THEORY

It is convenient, after choice of a standard state and concentration scale, to consider the chemical potential (the partial molar Gibbs free energy) as being composed of an ideal part with explicit dependence on the concentration and an excess part depending on solute–solvent interactions. The excess chemical potential (μ^{ex}) and its enthalpic (h^{ex}) and entropic (s^{ex}) components are calculated in this work. Formally,[26−31] $\mu^{ex} = \ln\langle e^{\beta\varepsilon}\rangle$, where the averaging $\langle ...\rangle$ is over the solute–solvent interaction energy distribution $P(\varepsilon)$ and $\beta = 1/k_BT$, where T is the temperature and k_B is the Boltzmann constant.

We compute μ^{ex} using the quasichemical decomposition of the potential distribution theorem.[24−26] To this end, we specify an inner hydration shell around the solute by a length parameter λ (Figure 1). The inner–outer demarcation serves

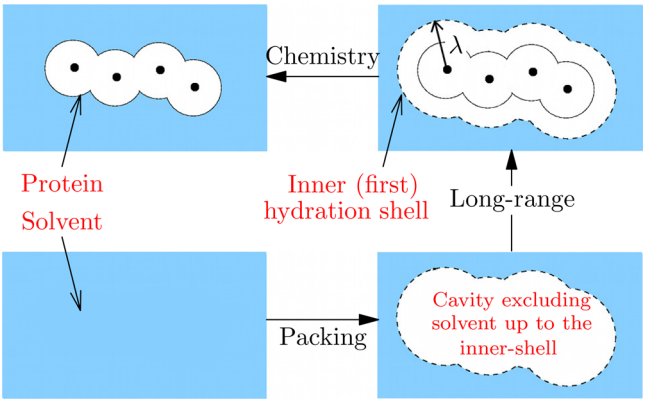


Figure 1. Schematic depicting the quasichemical organization of μ^{ex}. Schematic reproduced with permission from ref 16. Copyright 2013 Elsevier.

to separate strong, short-range solute–solvent interactions from their longer-range counterparts. With the definition of the inner shell, the excess chemical potential is given by[16,24−26,28]

$$\beta\mu^{ex} = \underbrace{\ln x_0[\phi_\lambda]}_{\text{local chemistry}} \underbrace{-\ln p_0[\phi_\lambda]}_{\text{packing}} + \underbrace{\beta\mu^{ex}[P(\varepsilon|\phi_\lambda)]}_{\text{long-range or outer}} \quad (2)$$

The constraint ϕ_λ is a field of range λ that serves to move the solvent away from the solute. Typically, for an inner shell extending up to the first hydration shell, $P(\varepsilon|\phi_\lambda)$, the (regularized) probability density of the solute–solvent binding energy is Gaussian.[16,27,28] Then

$$\mu^{ex}[P(\varepsilon|\phi_\lambda)] = \langle\varepsilon|\phi_\lambda\rangle + \frac{\beta}{2}\langle\delta\varepsilon^2|\phi_\lambda\rangle \quad (3)$$

where $\langle\varepsilon|\phi_\lambda\rangle$ and $\langle\delta\varepsilon^2|\phi_\lambda\rangle$ are the mean and variance of $P(\varepsilon|\phi_\lambda)$.

The remaining components in eq 2 have the following physical meaning. The free energy to empty the inner shell of solvent is $-k_BT \ln x_0$, where x_0 is the probability to find an empty inner shell. In terms of equilibrium solute–solvent clustering within the inner shell, $\ln x_0 = -\ln(1 + \sum_{i\geq 1} K_i\rho_w^i)$,

dx.doi.org/10.1021/jp500727u | J. Phys. Chem. B 2014, 118, 4080−4087

where K_i is the equilibrium constant to form a solute plus an i-water cluster and ρ_w is the bulk density of water.[24−26,29] We can consider an analogous process of solvent clustering within the same defined inner shell but in the absence of the solute: the quantity $-k_BT \ln p_0$ is the work done to create such a cavity in the liquid and is of fundamental interest in understanding liquid structure and primitive hydrophobic effects.[32,33] (The adjective primitive emphasizes that this is a property solely of the neat liquid.)

In simulations, we grow the external field ϕ_λ to calculate x_0 and p_0.[16,27,28] In practice, the field has a soft boundary at λ,[28] unlike the sharp-demarcation indicated in the schematic. It is straightforward to correct for this effect,[27] but we do not pursue that here. (Please note that the soft-cavity packing is a lower bound to the hard-cavity packing contribution and vice versa for the chemistry term.) We note in passing that for solutes such as nonpolar molecules,[34,35] water,[27,36,37] and ions,[29] including those with high charge density,[28] the quasichemical approach leads to free energies in good agreement with traditional free-energy perturbation or histogram overlap approaches, with the added advantages of the physical insights provided by the approach.

The excess entropy of hydration of the solute is given approximately by

$$Ts^{ex} \approx E_{sw} + E_{reorg} - \mu^{ex} \quad (4)$$

where the average solute−solvent interaction contribution to the excess enthalpy is E_{sw} and the average solvent reorganization contribution is E_{reorg}. The excess enthalpy $h^{ex} \approx E_{sw} + E_{reorg}$, where we have neglected small corrections due to pressure−volume effects and the finite isothermal compressibility of the pure liquid. In the previous equation we additionally ignore a small correction due to the finite thermal expansion coefficient of the pure liquid.

METHODS

The pentapeptides GGGGG, GGIGG, and IGGGG are modeled in the extended configuration with the long axis aligned with the diagonal of the simulation cell and the center of the peptide placed at the center of the simulation cell. Helical deca-glycine (G_9G) and a helical peptide with nine alanine and one glycine at position 6 (A_9G) serve as references for the helical peptides, G_9I and A_9I, respectively. All peptides had an acetylated (ACE) N-terminus and an n-methyl-amide (NME) C-terminus. Butane in the gauche conformation was built using the isoleucine conformation in the extended pentapeptide.

The solvent was modeled by the TIP3P[38,39] model, and the CHARMM[40] force field with CMAP correction terms for dihedral angles[41] was used for the peptide. A total of 2006 water molecules solvated the pentapeptide; 3500 water molecules were used for the helical peptides. The Lennard-Jones parameters for the isoleucine side chain were used for g-butane. Because C_β in isoleucine becomes a CH_2 group in g-butane, the partial charges of that center were slightly adjusted to account for the presence of a capping H atom. We note that CHARMM does have a parameter set for butane. In particular, the parameter we use for C_β is very slightly shifted from the parameter values for the corresponding carbon atom in butane. Our parametrization was motivated by our desire to be as close to isoleucine as possible, but this minor shift in parameters is not expected to change any conclusions.

The G_9G and A_9G peptides were built from α helical deca-alanine. These structures were energy-minimized with weak restraints on heavy atoms to prevent large distortions of the helix. (After energy minimization, the solute atoms are held fixed for the remainder of the simulation.) The G_9I and A_9I helices were built by grafting the conformation of isoleucine in the GGIGG system onto position 6. Thus the internal conformation of the isoleucine is the same in g-butane, GGIGG, and the helical peptides.

With minor changes, the free-energy calculations and error analysis follow the procedure previously described.[16] In brief, ϕ_λ is applied such that λ varies from 0 to 5 Å. For every unit angstrom, a five-point Gauss-Legendre quadrature rule defines the λ points sampled. The work to apply the field is then obtained by quadratures. At each gauss-point, the system was equilibrated for 0.5 ns, and the data were collected over the subsequent 0.5 ns. For the extended peptides, the long-range contribution was obtained by performing particle-insertion calculations in the appropriate molecular-shaped cavity (Figure 1). Water with the appropriate cavity was simulated for 1 ns and 1250 frames from the last 0.625 ns used for analysis. Confirming the Gaussian distribution of binding energies, we found that particle extraction calculations (eq 3) agree with the more robust[27,42] particle-insertion procedure.[16,43] For the helical peptides, given their high dipole moment we made the conservative choice of obtaining the electrostatic contribution to the long-range interaction using a two-point Gauss-Legendre quadrature;[44] the van der Waals (vdW) contribution was obtained using particle insertion in the molecular cavity. Electrostatic self-interaction corrections (of ∼0.5 kcal/mol) were applied.[16,45,46] (For the helices, the sum of the vdW and quadrature-based electrostatic contribution deviates by ∼1 kcal/mol from the Gaussian model, but this deviation is significant at the statistical resolution of the chemistry and packing contributions.)

For the zwitterionic isoleucine and glycine, we obtained the free energy in two stages. First, the excess free energy of the completely discharged amino acid was obtained using the quasichemical procedure. Then, the work required to turn-on the charges was obtained using a three-point Gauss-Legendre quadrature. The three-point rule gave the same answer (within 0.1 kcal/mol) as the two-point rule, but these estimates deviate by over 10 kcal/mol from the linear-response result.

The excess energy was obtained by adapting the shell-wise calculation procedure previously used for studying the hydration of methane.[34] For the peptides, water molecules in the first two hydration shells were considered, where the hydration shell is defined by the union of shells of radius λ centered on the heavy atoms: $\lambda \leq 5$ Å defined the first, $5.0 < \lambda \leq 8$ Å defined the second, and $8.0 < \lambda \leq 11.0$ Å defined the third hydration shell. Within statistical uncertainties, the reorganization contribution from the third shell is close to zero, justifying our use of just the first two shells for calculating h^{ex}. For calculating the excess energy, we equilibrated the solvated peptide system an additional 1.5 ns (beyond what was used in the free-energy calculation) and propagated the trajectory for an additional 3 ns, collecting data every 500 ps for a total of 6000 frames. Entropies obtained using eq 4 agree within statistical uncertainty with entropy calculated using $-\partial\mu^{ex}/\partial T$ (manuscript in preparation).

Conditional Contributions. The conditional hydration free energy of the side-chain X in a peptide MX relative to the reference MH (the corresponding amino acid is glycine) is

defined by $\mu^{ex}[X|MX] = \mu^{ex}[MX] - \mu^{ex}[MH]$.[47] Similar definitions apply to the excess enthalpy and excess entropy of hydration. Provided that additivity holds, $\mu^{ex}[X|MX]$ will be proportional to $\mu^{ex}[g\text{-butane}]$, the side-chain analog. (The same proportionality, by definition, must then also hold for h^{ex} and s^{ex}.) To better understand how the reference influences the hydration of the side chain, we use the quasichemical approach to study the packing, chemistry, and long-range contributions of the side chain in the presence of the reference. (We term these the conditional quasichemical components.) For the packing calculation, we open a cavity at the appropriate position next to MH to accommodate the side chain, with MH fully coupled to the solvent. Likewise, for the chemistry contribution, we expel solvent only around the side chain X in MX. With solvent thus excluded, the long-range contribution was obtained using eq 3.

Solvent-Exposure-Weighted Additive Model. In additivity-based approaches, a protein is partitioned down to single peptide units[9] (or even lower[8]), and the thermodynamics of the macromolecule is obtained using equations of the form of eq 1. To better understand the limitations of additivity, here we partition the helical peptide S into just two groups, the isoleucine side chain (sc) and the rest, generically termed the background (back); then

$$\mu_S^{ex} = \alpha_{sc}\cdot\mu_{sc\text{-}model}^{ex} + \alpha_{back}\cdot\mu_{back}^{ex} \quad (5)$$

where $\mu_{sc\text{-}model}^{ex}$ refers to the conditional hydration contribution of the side chain in a model compound.

Typically, the conditional quantity based on the zwitterionic amino acid[3,9,48] is combined with α factors based on the solvent exposure of the side-chain X in the GXG tripeptide. Here we consistently use the conditional quantity and the solvent exposure from GGIGG solely. (Our results make it apparent that using conditional quantities from the zwitterionic model will only worsen predictions of thermodynamic quantities.) The α factors here are based on solvent-accessible surface area (SASA) obtained using a standard code[49] and atomic radii.[50] As an example, let A_t be the SASA of G_9I and A_{sc} be the contribution due to the isoleucine side chain. Furthermore, let A_{sc}^0 be the SASA of the isoleucine side chain in GGIGG and A_{back}^0 be the SASA of G_9G. Then, $\alpha_{sc} = A_{sc}/A_{sc}^0$ and $\alpha_{back} = (A_t - A_{sc})/A_{back}^0$.

RESULTS AND DISCUSSION

In evaluating the results later, it serves to first define the error tolerance for an additive model to be satisfactory in predicting energetics of protein folding. Dill[51] has suggested that the allowable error per group for a modest 100 amino-acid protein is ~10 cal/mol/group (or ~0.02 k_BT/group at 298 K). This estimate sets a stringent precision level; our results show that the additive approach at best agrees at a couple of kilocalories per mole level.

Hydration of Zwitterionic Amino Acids. Table 1 summarizes the results on the hydration thermodynamics of the zwitterionic forms of isoleucine and glycine. The difference of these quantities (denoted by Δ in Table 1) is typically identified as the group contribution of the amino acid side chain.

Table 1 shows that whereas the conditional free energy of the isoleucine side chain is in good agreement with the excess free energy of g-butane, comparing the discharged and charged analogs of the amino acids shows that slightly over half of the 2.5 kcal/mol free energy comes from differences in the free energy of charging the amino acids. This is suggestive of screening of the zwitterion-dipole by the side chain (cf. ref 19), but this physics associated with charging is not one envisioned when considering the hydration of g-butane, an alkane. Furthermore, examining E_{sw} and E_{reorg} reveals large deviations between g-butane and the side chain. Thus, these results suggest physical differences in the conditional hydration of the side chain and g-butane, its small molecule analog.

Table 1. Conditional Hydration of Isoleucine Side Chain in Zwitterionic Amino Acid[a]

	μ^{ex}	h^{ex}	E_{sw}	E_{reorg}	Ts^{ex}
$\Delta[I^{(0)}]$	1.2 ± 0.1	−1.3 ± 0.7	−5.7 ± 0.03	4.3	−2.5
$\Delta[I]$	2.5 ± 0.1	−0.9 ± 1.8	−1.8 ± 0.1	1.0	−3.4
g-butane	2.5 ± 0.1	−3.4 ± 0.5	−9.4 ± 0.02	6.0	−5.9

[a] $\Delta[x]$, where $x = I^{(0)}$ or I is value relative to $G^{(0)}$ or G, respectively. The superscript 0 indicates that the partial charges are set to zero. The solute−solvent interaction contribution (E_{sw}) and solvent reorganization contribution (E_{reorg}) to h^{ex} are also given. The quantities E_{reorg}, Ts^{ex}, and h^{ex} have similar statistical uncertainties. The statistical uncertainty in E_{sw} is less than 1/10th that in h^{ex}. Standard error of the mean is given at 1σ. All thermodynamic quantities are in units of kilocalories per mole.

Hydration of Isoleucine in Pentapeptides. Table 2 summarizes the results on the hydration of the extended pentapeptides. It is evident that hydration thermodynamics of GGIGG and IGGGG are different, suggesting that the conditional free energy of isoleucine will depend on where it is placed along the backbone. The difference in μ^{ex} is ~1 kcal/mol, and it is all in the enthalpy of solution.

Table 2. Hydration of Isoleucine Side Chain in Extended Pentapeptides[a]

	μ^{ex}	h^{ex}	E_{sw}	E_{reorg}	Ts^{ex}
Δ[GGIGG]	2.9 ± 0.4	0.6 ± 2.0	−1.8	2.4	−2.3
$\alpha_{sc}\cdot$[g-butane]	1.5	−2.1	−5.8	3.7	−3.6
Δ[IGGGG]	2.1 ± 0.4	−0.5 ± 2.0	−2.1	1.7	−2.6
$\alpha_{sc}\cdot$[g-butane]	1.5	−2.1	−5.8	3.7	−3.6

[a] $\Delta[x]$, where x = GGIGG or IGGGG is the value relative to GGGGG. α_{sc} = 0.619 is the ratio of the solvent-accessible surface area of the isoleucine side chain in GGIGG to that for g-butane; α_{sc} = 0.616 for IGGGG. Rest as in Table 1.

Note that relative to g-butane the fractional solvent exposure of isoleucine in GGIGG and IGGGG is nearly the same, and hence the g-butane data corrected by fractional-solvent exposure are only modestly successful in modeling the positional dependence in this case. Relative to g-butane, the conditional h^{ex} is more positive (less favorable), and this effect primarily arises from changes in E_{sw}. This result is intuitively reasonable given that the isoleucine side chain in the pentapeptide is less solvent-exposed than g-butane, but the expected solvent exposure turns out to be physically unrealistic. One would need to scale the g-butane E_{sw} by 0.2 to match the corresponding value for the side chain, but the observed fractional exposure is ~0.6. Furthermore, the scaling factors are different for μ^{ex}, the subcomponents of h^{ex}, and hence also s^{ex}.

Hydration of Isoleucine in Helical Peptides. Table 3 summarizes the results on the conditional hydration of isoleucine in the helical peptides G_9I and A_9I. Once again, we find large deviations between the conditional contribution and

dx.doi.org/10.1021/jp500727u | J. Phys. Chem. B 2014, 118, 4080−4087

Table 3. Conditional Hydration of Isoleucine in Helical Deca-Peptides[a]

	μ^{ex}	h^{ex}	E_{sw}	E_{reorg}	Ts^{ex}
$\Delta[G_9I]$	3.0 ± 1.1	1.1 ± 2.7	−2.3	3.4	−1.9
α_{sc}·[*g*-butane]	1.5	−2.0	−5.5	3.5	−3.5
$\alpha_{sc}\cdot\Delta_{GGIGG}$	2.7	0.6	−1.7	2.3	−2.4
$\Delta[A_9I]$	3.2 ± 1.2	3.3 ± 4.0	1.0	2.3	0.1
α_{sc}·[*g*-butane]	1.3	−1.9	−5.3	3.4	−3.3
$\alpha_{sc}\cdot\Delta_{GGIGG}$	2.5	0.5	−1.6	2.2	−2.3

[a]$\Delta[x]$, where x = G_9I or A_9I, gives value relative to G_9G or A_9G, respectively. $(Gly)_9$-Ile is indicated as G_9I, where isoleucine occupies the sixth position. A similar notation is used for the other peptides. The fractional solvent exposure of isoleucine in G_9I is α_{sc} = 0.59, and in A_9I it is α_{sc} = 0.56. Relative to isoleucine in GGIGG, α_{sc} = 0.95 (G_91) and α_{sc} = 0.90 (A_9I). Rest as in Table 2.

the value from the side-chain model appropriately scaled by the solvent exposure. The deviations can be dramatic. The conditional E_{sw} contribution to h^{ex} in A_9I is positive, whereas E_{sw} for *g*-butane and the conditional E_{sw} in GGIGG are both negative.

Equation 5, with a protein partitioned down to a single peptide unit, for example, ref 9 (or sometimes even lower, for example, ref 8), is often used in group-additivity-based interpretation of experimental data. Following this precedent but limiting the partitioning of the peptide into only two groups − the side chain and the rest − we use eq 5 to reconstruct the free energy of the helical peptides using the conditional hydration thermodynamics of isoleucine in GGIGG together with the appropriate reference for the helix (Table 4). It is evident that the error based on the additivity-based approach (eq 5) can be several k_BT values (and is itself comparable to typical folding free energies).

Table 4. Predicted Thermodynamics of G_9I and A_91 Helical Peptides Using Equation 5[a]

	μ^{ex}	h^{ex}	E_{sw}	E_{reorg}	Ts^{ex}
G_9I	−46.7	−82.5	−155.9	73.4	−35.8
$[G_9I]_A$	−43.7	−76.7	−143.7	67.0	−33.0
error	3.0	5.8	12.2	−6.4	2.8
A_9I	−37.4	−74.8	−151.1	76.3	−37.4
$[A_9I]_A$	−35.9	−72.8	−144.3	71.5	−36.9
error	1.5	2.0	6.8	−4.8	0.2

[a]Subscript A denotes the predicted properties obtained by combining Δ[GGIGG] (Table 2) and $\Delta[x]$ (Table 3, x = G_9I or A_9I) using eq 5. Error in the predicted value relative to the simulated value is also noted. All values are in kilocalories per mole.

Role of the Reference in Conditional Hydration. The results above show that at the precision level suggested by Dill[51] the conditional contribution of amino acid X from a small-length-scale model or side-chain analog is of limited utility in modeling the contribution of X in model helices (and by induction in more complicated proteins). To understand the reasons for this, we look at two interrelated aspects of the problem: (1) hydration of the side chain in the presence of the background and how this compares with the side-chain analog and (2) how the side chain changes the hydration of the background and how this effect is conflated with the conditional contribution attributed to the side chain.

Table 5 gives the conditional quasichemical components for isoleucine side chain in the context of extended and helical

Table 5. Conditional Quasichemical Components for Isoleucine Hydration[a]

	$k_BT \ln x_0$	$-k_BT \ln p_0$	$\mu^{ex}_{long-range}$	μ^{ex}
g-butane	−16.4	24.0	−5.1	2.5 ± 0.1
GGIGG	−13.4	20.2	−4.4	2.4 ± 0.1
IGGGG	−13.2	19.7	−4.4	2.1 ± 0.1
G_9I	−13.0	18.6	−4.0	1.6 ± 0.1
A_9I	−11.5	17.7	−3.5	2.7 ± 0.1

[a]Statistical uncertainties in chemistry and packing are ∼0.1 kcal/mol, while that in the long-range contribution is an order of magnitude lower. All values are in kilocalories per mole.

peptides. Except for isoleucine in G_9I, the agreement in the conditional μ^{ex} with values obtained as $\mu^{ex}[MX] - \mu^{ex}[MH]$ (Tables 2 and 3) is excellent. (For G_9I, the agreement is good only at 2σ.)

Table 5 shows that primitive hydrophobic effects (as assessed here by the soft-cavity packing contribution) is context-sensitive. Comparing A_9I with G_9I and IGGGG with GGIGG, we find, as is expected,[52] that opening a cavity near hydrophobic groups is more facile than opening it next to a more polar surface. This observation is fundamentally related to how water density fluctuations vary depending on the polarity of the surface.[53] Furthermore, relative to *g*-butane, the primitive hydrophobic effect is lower in the context of these peptide models, but the fractional exposed area does not provide the correct proportionality. Finally, the free energy to open a cavity, for cavity sizes of interest in modeling amino acid side chains, is expected to scale with the volume and not the area.[54−56]

Mirroring the trend in the packing contribution, the chemistry contribution is less negative when the side chain is near other hydrophobic groups. The long-range contribution also follows a trend similar to the chemistry contribution. The compensating variation in the packing and hydrophilic contributions tends to balance, and (excluding the G_9I case) the net conditional free energy of the side chain is not significantly different from that for *g*-butane. Like the packing contribution, the short-range and long-range hydrophilic contributions are not expected to have a simple dependency on the solvent-exposed surface area. Because different physical effects are implicit in the net μ^{ex}, no single scale factor can be expected to map the *g*-butane data to the isoleucine side chain in a given model, and, in general, properties of a small-scale model to its value in a protein.

Turning now to the effect of the side chain on the background, we see that in grafting a *g*-butane onto a zwitterionic glycine to construct the zwitterionic isoleucine the background (here the peptide backbone) experiences an unfavorable change in E_{sw} of 6.3 (= 120.7−114.4) kcal/mol and the side chain contributes −8.1 kcal/mol for a net change of −1.8 kcal/mol. Thus in assigning the conditional E_{sw} for the side chain as $E_{sw}[I] - E_{sw}[G]$, we ascribe 6.3 kcal/mol of backbone contributions to the side chain. The effects of this problem are seen to varying extents in the other peptide systems studied here (Table 6).

In the previous analysis, we have chosen a specific, but illustrative, example of hydration of isoleucine in model peptide systems. We have also considered the hydration of the aromatic phenylalanine in the GGFGG system and find that all qualitative conclusions of the isoleucine study apply to this system as well. (These results are collected in the Supporting Information.) A more extensive investigation of sequence and

dx.doi.org/10.1021/jp500727u | J. Phys. Chem. B 2014, 118, 4080−4087

Table 6. Decomposition of Peptide-Solvent Mean Binding Energy, E_{sw}, into Contributions from the Background and Side Chain[a]

	$\alpha_{back}\cdot E_{sw}$[back-model]	E_{sw}[back]	E_{sw}[side chain]	$\alpha_{sc}\cdot E_{sw}$[sc-model]
g-butane			−9.4	
G		−120.7		
I		−114.4	−8.1	
GGGGG		−105.5		
GGIGG	−95.3	−100.2	−7.1	−5.8
IGGGG	−95.4	−100.4	−7.2	−5.8
G_9G		−153.6		
G_91	−142.0	−149.1	−6.8	−6.7
A_9G		−152.0		
A_9I	−142.7	−144.9	−6.2	−6.4

[a]For GGIGG and IGGGG, the scaled-E_{sw} of *g*-butane, the side-chain model (sc-model), is given under $\alpha_{sc}\cdot E_{sw}$ [sc-model]; for the helices, the scaled-E_{sw} of the side chain in GGIGG is given. $\alpha_{back}\cdot E_{sw}$[back] is the scaled value of the background reference (back-model). For example, G_9G is the background reference for G_9I. All values are in kilocalories per mole.

structure space may be desirable to better understand the success and limitation of the additivity approach, but within the context of the potential model chosen, the examples considered here do serve to indicate the size of deviations from additivity and, importantly, the compensating deviations in entropy and enthalpy and the conflation of backbone-specific changes with effects attributed to the side chain.

In summary, the interdependence of background hydration and side-chain hydration previously indicated suggests that using Y[GXG]−Y[GGG] to model Y[X], where Y is some thermodynamic state function of interest, in proteins has important limitations. Acknowledging the changes in the solvent exposure of both the background and the side chain in the peptide under study is an intuitively reasonable step, but no single scaling factor proves to be satisfactory in capturing both the hydrophilic and hydrophobic aspects of hydration, especially if one uses the stringent precision levels suggested by Dill.[51]

CONCLUDING DISCUSSION

Our results show that there are nontrivial correlations between the side chain, backbone, and water that do not allow a perfect pairwise decomposition. In essence, the protein context and its induced correlations in water change the hydration of objects within that correlation range. Conversely, the object thus affected itself changes the hydration of the context. Thus composing the enthalpy, entropy, and free energy of a macromolecule using a group-additive approach will necessarily include reference and group contributions in entirely non-transparent ways.

Constructing the free energy of a protein by scaling appropriate model compound data by the fractional solvent-exposure of that group in the protein is intuitively reasonable, but this approach attempts to map simultaneously both packing and hydrophilic aspects of hydration, but each of these aspects addresses inherently different physical features of solvent behavior and solute−solvent interaction. Moreover, the derived conditional energetics of the side chain are seen to contain nonobvious contributions from the changes in the hydration of the background. Thus no single scale factor can be expected to describe the conditional μ^{ex}, h^{ex}, and s^{ex} of the side chain relative to the property of the model compound. Furthermore, relative to a model compound, the identification of greater or lesser hydrophobicity or hydrophilicity will itself be context-dependent, raising questions about the utility of small-molecule data to understand finely detailed features of protein hydration. This could be of concern in assessing contributions of hydrophilic versus hydrophobic mutations in proteins.

If we take the stringent Dill error criteria,[51] our analysis shows that the hydration of *g*-butane is inadequate to model the conditional hydration of GGIGG, and the conditional hydration of isoleucine in GGIGG is inadequate in modeling the hydration of isoleucine in model helices. The successes in using small-molecule data in modeling protein hydration should thus be considered with respect to the errors required in a given context. If an additive description of free energy predicts the numerical values with reasonable error expectations in accord with experiments, it is likely that this success arises due to some compensation in the underlying components.

ASSOCIATED CONTENT

Supporting Information

Hydration of phenylalanine side chain in capped GGFGG pentapeptide. Conditional quasichemical components for phenylalanine hydration. Decomposition of peptide-solvent mean binding energy, E_{sw}, into contributions from the background and side chain. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: Dilip.Asthagiri@utmb.edu.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Dr. Gillian Lynch (UTMB) for providing the initial structure of a deca-alanine in the helical conformation. We thank Arieh Ben-Naim for helpful comments on an early draft of the manuscript. D.A. and B.M.P. gratefully acknowledge the financial support of the National Institutes of Health (GM 037657), the National Science Foundation (CHE-1152876), and the Robert A. Welch Foundation (H-0037). This research used resources of the National Energy Research Scientific Computing Center, which is supported by the Office of Science of the U.S. Department of Energy under contract no. DE-AC02-05CH11231. We also acknowledge the Texas Advanced Computing Center (TACC) at The University of Texas at Austin for providing HPC resources that have contributed to the research results reported within this paper.

REFERENCES

(1) Kauzmann, W. Some Factors in the Interpretation of Protein Denaturation. *Adv. Protein Chem.* **1959**, *14*, 1−63.
(2) Tanford, C. Contribution of Hydrophobic Interactions to the Stability of the Globular Conformation of Proteins. *J. Am. Chem. Soc.* **1962**, *84*, 4240−4247.
(3) Nozaki, Y.; Tanford, C. The Solubility of Amino Acids and Related Compounds in Aqueous Urea Solutions. *J. Biol. Chem.* **1963**, *238*, 4074−4081.
(4) Tanford, C. Isothermal Unfolding of Globular Proteins in Aqueous Urea Solutions. *J. Am. Chem. Soc.* **1964**, *86*, 2050−2059.
(5) Tanford, C. Protein Denaturation. *Adv. Protein Chem.* **1970**, *24*, 1−95.
(6) Makhatadze, G. I.; Privalov, P. L. Contribution of Hydration to Protein Folding Thermodynamics: I. The Enthalpy of Hydration. *J. Mol. Biol.* **1993**, *232*, 639−659.
(7) Privalov, P. L.; Makhatadze, G. I. Contribution of Hydration to Protein Folding Thermodynamics: II. The Entropy and Gibbs Energy of Hydration. *J. Mol. Biol.* **1993**, *232*, 660−679.
(8) Makhatadze, G. I.; Privalov, P. L. Energetics of Protein Structure. *Adv. Protein Chem.* **1995**, *47*, 307−425.
(9) Auton, M.; Rösgen, J.; Sinev, M.; Holthauzen, L. M. F.; Bolen, D. W. Osmolyte Effects on Protein Stability and Solubility: A Balancing Act Between Backbone and Side-Chains. *Biophys. Chem.* **2011**, *159*, 90−99.
(10) Auton, M.; Holthauzen, L. M. F.; Bolen, D. W. Anatomy of Energetic Changes Accompanying Urea-induced Protein Denaturation. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 15317−15322.
(11) Auton, M.; Bolen, D. W. Predicting the Energetics of Osmolyte-induced Protein Folding/unfolding. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 15065−15068.
(12) Greene, R. F.; Pace, C. N. Urea and Guanidine Hydrochloride Denaturation of Ribonuclease, Lysozyme, Alpha-chymotrypsin, and Beta-lactoglobulin. *J. Biol. Chem.* **1974**, *249*, 5388−5393.
(13) Bolen, D. W.; Rose, G. D. Structure and Energetics of the Hydrogen-bonded Backbone in Protein Folding. *Annu. Rev. Biochem.* **2008**, *77*, 339−362.
(14) Hua, L.; Zhou, R.; Thirumalai, D.; Berne, B. J. Urea Denaturation by Stronger Dispersion Interactions with Proteins than Water Implies a 2-stage unfolding. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 16928−16933.
(15) Kokubo, H.; Hu, C. Y.; Pettitt, B. M. Peptide Conformational Preferences in Osmolyte Solutions: Transfer Free energies of Decaalanine. *J. Am. Chem. Soc.* **2011**, *133*, 1849−1858.
(16) Tomar, D. S.; Weber, V.; Asthagiri, D. Solvation Free Energy of the Peptide Group: its Model Dependence and Implications for the Additive Transfer Free Energy Model. *Biophys. J.* **2013**, *105*, 1482−1490.
(17) Staritzbichler, R.; Gu, W.; Helms, V. Are Solvation Free Energies of Homogeneous Helical Peptides Additive? *J. Phys. Chem. B* **2005**, *109*, 19000−19007.
(18) König, G.; Boresch, S. Hydration Free Energies of Amino Acids: Why Side Chain Analog Data are not Enough. *J. Phys. Chem. B* **2009**, *113*, 8967−8974.
(19) König, G.; Bruckner, S.; Boresch, S. Absolute Hydration Free Energies of Blocked Amino Acids: Implications for Protein Solvation and Stability. *Biophys. J.* **2013**, *104*, 453−462.
(20) Roseman, M. A. Hydrophilicity of Polar Amino Side-chains is Markedly Reduced by Flanking Peptide Bonds. *J. Mol. Biol.* **1988**, *200*, 513−522.
(21) Cornette, J. L.; Cease, K. B.; Margalit, H.; Spouge, J. L.; Berzofsky, J. A.; DeLisi, C. Hydrophobicity Scales and Computational Techniques for Detecting Amphipathic Structures in Proteins. *J. Mol. Biol.* **1987**, *195*, 659−685.
(22) Boresch, S.; Archontis, G.; Karplus, M. Free Energy Simulations: The Meaning of the Individual Contributions from a Component Analysis. *Proteins: Struct., Funct., Bioinf.* **1994**, *20*, 25−33.
(23) Boresch, S.; Karplus, M. The Meaning of Component Analysis: Decomposition of the Free Energy in Terms of Specific Interactions. *J. Mol. Biol.* **1995**, *254*, 801−807.
(24) Paulaitis, M. E.; Pratt, L. R. Hydration Theory for Molecular Biophysics. *Adv. Protein Chem.* **2002**, *62*, 283−310.
(25) Pratt, L. R.; Asthagiri, D. In *Free Energy Calculations: Theory and Applications in Chemistry and Biology*; Chipot, C., Pohorille, A., Eds.; Springer Series in Chemical Physics *86*; Springer: Berlin, 2007; Chapter 9, pp 323−351.
(26) Beck, T. L.; Paulaitis, M. E.; Pratt, L. R. *The Potential Distribution Theorem and Models of Molecular Solutions*; Cambridge University Press: Cambridge, U.K., 2006.
(27) Weber, V.; Merchant, S.; Asthagiri, D. Regularizing Binding Energy Distributions and Thermodynamics of Hydration: Theory and Application to Water Modeled with Classical and *Ab Initio* Simulations. *J. Chem. Phys.* **2011**, *135*, 181101.
(28) Weber, V.; Asthagiri, D. Regularizing Binding Energy Distributions and the Hydration Free Energy of Protein Cytochrome C from All-Atom Simulations. *J. Chem. Theory Comput.* **2012**, *8*, 3409−3415.
(29) Merchant, S.; Asthagiri, D. Thermodynamically Dominant Hydration Structures of Aqueous Ions. *J. Chem. Phys.* **2009**, *130*, 195102.
(30) Merchant, S. Regularizing Free Energy Calculations to Study Ion Specific Effects in Biology. Ph.D. Thesis, Johns Hopkins University, Baltimore, 2011.
(31) Widom, B. Potential-distribution Theory and the Statistical Mechanics of Fluids. *J. Phys. Chem.* **1982**, *86*, 869−872.
(32) Pratt, L. R.; Pohorille, A. Theory of Hydrophobicity: Transient Cavities in Molecular Liquids. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 2995−2999.
(33) Pratt, L. R. Molecular Theory of Hydrophobic Effects: "She is too Mean to have Her Name Repeated. *Annu. Rev. Phys. Chem.* **2002**, *53*, 409−436.
(34) Asthagiri, D.; Merchant, S.; Pratt, L. R. Role of Attractive Methane-water Interactions in the Potential of Mean Force between Methane Molecules in Water. *J. Chem. Phys.* **2008**, *128*, 244512.
(35) Asthagiri, D.; Ashbaugh, H. S.; Piryatinski, A.; Paulaitis, M. E.; Pratt, L. R. Non-van der Waals Treatment of the Hydrophobic Solubilities of CF_4. *J. Am. Chem. Soc.* **2007**, *129*, 10133−10140.
(36) Paliwal, A.; Asthagiri, D.; Pratt, L. R.; Ashbaugh, H. S.; Paulaitis, M. E. An Analysis of Molecular Packing and Chemical Association in Liquid Water Using Quasichemical Theory. *J. Chem. Phys.* **2006**, *124*, 224502.
(37) Shah, J. K.; Asthagiri, D.; Pratt, L. R.; Paulaitis, M. E. Balancing Local Order and Long-ranged Interactions in the Molecular Theory of Liquid Water. *J. Chem. Phys.* **2007**, *127*, 144508.
(38) Jorgensen, W.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(39) Neria, E.; Fischer, S.; Karplus, M. Simulation of Activation Free Energies in Molecular Systems. *J. Chem. Phys.* **1996**, *105*, 1902−1921.
(40) MacKerell, A. D., Jr.; Bashford, D.; Bellott; Dunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; et al. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(41) MacKerell, A. D., Jr.; Feig, M.; Brooks, C. L., III. Extending the Treatment of Backbone Energetics in Protein Force Fields: Limitations of Gas-phase Quantum Mechanics in Reproducing Protein Conformational Distributions in Molecular Dynamics Simulations. *J. Comput. Chem.* **2004**, *25*, 1400−1415.
(42) Lu, N.; Kofke, D. A. Accuracy of Free-energy Perturbation Calculations in Molecular Simulations. I. Modeling. *J. Chem. Phys.* **2001**, *114*, 7303−7311.
(43) Rogers, D. M.; Beck, T. L. Modeling Molecular and Ionic Absolute Solvation Free Energies with Quasichemical Theory Bounds. *J. Chem. Phys.* **2008**, *129*, 134505.

(44) Hummer, G.; Szabo, A. Calculation of Free-energy Differences from Computer Simulations of Initial and Final States. *J. Chem. Phys.* **1996**, *105*, 2004−2010.
(45) Hummer, G.; Pratt, L. R.; Garcia, A. E. Free Energy of Ionic Hydration. *J. Phys. Chem.* **1996**, 1206−1215.
(46) Hummer, G.; Pratt, L. R.; Garcia, A. E. Molecular Theories and Simulation of Ions and Polar Molecules in Water. *J. Phys. Chem. A* **1998**, *102*, 7885−7895.
(47) Ben-Naim, A. *Molecular Theory of Water and Aqueous Solutions: The Role of Water in Protein Folding and Self-Assembly and Molecular Recognition*; World Scientific: Singapore, 2011; Vol. *2*.
(48) Tanford, C. Contribution of Hydrophobic Interactions to the Stability of the Globular Conformation of Proteins. *J. Am. Chem. Soc.* **1962**, *84*, 4240−4247.
(49) Sanner, M.; Olson, A. J.; Spehner, J. C. Reduced Surface: an Efficient Way to Compute Molecular Surfaces. *Biopolymers* **1996**, *38*, 305−320.
(50) Bondi, A. Van der Waals Volumes and Radii. *J. Phys. Chem.* **1964**, *68*, 441−451.
(51) Dill, K. A. Additivity Principles in Biochemistry. *J. Biol. Chem.* **1997**, *272*, 701−704.
(52) Hummer, G.; Garde, S.; Garcia, A. E.; Paulaitis, M. E.; Pratt, L. R. Hydrophobic Effects on a Molecular Scale. *J. Phys. Chem. B* **1998**, *102*, 10469−10482.
(53) Godawat, R.; Jamadagni, S. N.; Garde, S. Characterizing Hydrophobicity of Interfaces by Using Cavity Formation, Solute Binding, and Water Correlations. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 15119−15124.
(54) Stillinger, F. H. Structure in Aqueous Solutions of Nonpolar Solutes from the Standpoint of Scaled-particle Theory. *J. Solution Chem.* **1973**, *2*, 141−158.
(55) Chandler, D. Interfaces and the Driving Force of Hydrophobic Assembly. *Nature* **2005**, *437*, 640−647.
(56) Ashbaugh, H. S.; Pratt, L. R. Colloquium: Scaled Particle Theory and the Length Scales of Hydrophobicity. *Rev. Mod. Phys.* **2006**, *78*, 159−178.