Revisiting the Carboxylic Acid Dimers in Aqueous Solution: Interplay of Hydrogen Bonding, Hydrophobic Interactions, and Entropy†

Jianhan Chen,‡ Charles L. Brooks, III,*,‡ and Harold A. Scheraga*,§

Department of Molecular Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, and Baker Laboratory of Chemistry and Chemical Biology, Cornell University, Ithaca, New York 14853-1301

Received: June 5, 2007; In Final Form: July 27, 2007

Carboxylic acid dimers are useful model systems for understanding the interplay of hydrogen bonding, hydrophobic effects, and entropy in self-association and assembly. Through extensive sampling with a classical force field and careful free energy analysis, it is demonstrated that both hydrogen bonding and hydrophobic interactions are indeed important for dimerization of carboxylic acids (except formic acid). The dimers are only weakly ordered, and the degree of ordering increases with stronger hydrophobic interactions between longer alkyl chains. Comparison of calculated and experimental dimerization constants reveals a systematic tendency for excessive self-aggregation in current classical force fields. Qualitative and quantitative information on the thermodynamics of hydrogen bonding and hydrophobic interactions derived from these simulations is in excellent agreement with existing results from experiment and theory. These results provide a verification from first principles of previous estimations based on two statistical mechanical hydrophobic theories. We also revisit and clarify the fundamental statistical thermodynamics formalism for calculating absolute binding constants, external entropy, and solvation entropy changes upon association from detailed free energy simulations. This analysis is believed to be useful for a wide range of applications including computational studies of protein−ligand and protein−protein binding.

1. Introduction

Hydrogen bonding and hydrophobic association are two of the principal interactions that determine biomolecular structures and assemblies.[1] Globular proteins are believed to fold by minimizing exposed hydrophobic surface, while simultaneously forming specific hydrogen bonds mainly in the form of secondary structures;[2] amphiphilic molecules such as lipids spontaneously form mesoscopic and macroscopic assemblies in water, which often exhibit a large degree of structural polymorphism because of slight shifts in the balance among hydrogen bonding, hydrophobic interactions, and entropy.[3] Despite significant progress in understanding these interactions from experiment as well as theory, it remains generally difficult to dissect individual contributions to the overall stability of structures, and controversies persist. For example, the net contribution of backbone hydrogen bonding to protein folding kinetics and thermodynamics has been under much debate.[4,5] There is also much uncertainty regarding the role of putative initial hydrophobic collapse in the early stage of protein folding.[6−8] Progress in clarifying these issues requires quantitative understanding of hydrogen bonding and hydrophobic effects as well as various entropic costs. For this purpose, small molecule model systems have played a very important role in the past, and will continue to be useful in extending and refining our understanding of the detailed balance among different interactions in aqueous solution.

The carboxylic acids represent a particularly useful series of model compounds for studying hydrogen bonding and hydrophobic interactions.[9−12] They are known to form dimers in the gas phase and in dilute solutions. In the gas phase and in nonpolar solvents, cyclic dimers with two strong C═O···HO hydrogen bonds dominate, and the dimerization constants, K_D, are similar regardless of the alkyl chain length. By contrast, K_D increases as a function of the alkyl chain length in aqueous solution. It has been proposed that open chain dimers with a single hydrogen bond are the dominant dimer conformations in aqueous solutions, and the chain length dependence of K_D could then be attributed to hydrophobic interactions between the alkyl groups.[9] The homologous series of carboxylic acids has proved to be particularly useful in examining the theoretical understanding of hydrophobic interactions,[13,14] as contributions from hydrogen bonding and external entropy changes (translational and orientational) can be subtracted out by examining the difference in binding free energies across the series. With development of modern molecular mechanics force fields and sampling techniques, we are now ready to revisit carboxylic acid dimers to obtain statistical mechanical descriptions of the dimerization process from basic physical principles. Furthermore, through careful free energy analysis, we can verify and extend our understanding of hydration of polar and nonpolar molecules, and understand how they interplay together with entropy to determine the self-association of carboxylic acids in water.

The rest of this paper is organized as follows. First, we describe the molecular force field and biased sampling technique used to compute the potentials of mean force (PMFs) of dimerization. Then, the statistical thermodynamic formalism for calculating absolute binding affinities and translational and

† Part of the "James T. (Casey) Hynes Festschrift".

*Corresponding authors. Phone: (858) 784-8035. Fax: (858) 784-8688. E-mail: brooks@scripps.edu (C.L.B.). Phone: (607) 255-4034. Fax: (607) 254-4700. E-mail: has5@cornell.edu (H.A.S.).

‡ The Scripps Research Institute.

§ Cornell University.

10.1021/jp074355h CCC: $40.75
Published on Web 09/19/2007

orientational entropy changes upon binding is reviewed. We note here that, while the fundamental relations have been clearly laid out in textbooks[15] as well as in published literature,[16,17] some confusion appears to persist,[18−20] which is often related to various approximations necessary for complex systems. In Results and Discussion, we will apply the above formalism to analyze the thermodynamics of hydrogen bonding and hydrophobic interactions and compare the results to existing experimental data and theoretical calculations. Accurate estimation of the total dimerization entropy, which is a sum of the external entropy (from changes in translation and rotation), the internal entropy (from changes in conformation), and the solvation entropy (from changes in solvation), allows a reliable estimation of the solvation entropy contribution to dimerization. Comparison with previous theoretical results thus provides verification of existing statistical mechanical theories of hydrophobic hydration and hydrophobic interactions from a first-principles consideration.

2. Methods

2.1. Force Field, Biased Sampling, and PMF Calculations. In this study, we investigate the dimerization of formic, acetic, propionic, and butyric acids in water at low pH where the carboxyl group is not ionized. The acids, described by the CHARMM22 all-atom force field,[21] were solvated in a truncated octahedral box of about 960 TIP3P[22] water molecules with periodic boundary conditions. The particle−mesh Ewald method[23] was used for long-range electrostatic interactions, and the van der Waals energy was smoothly switched off over the range of 10 to 11 Å.[24,25] The lengths of all bonds involving hydrogen atoms were constrained using the SHAKE algorithm[26] to allow a 2 fs time step. Biased molecular dynamics simulations along a reaction coordinate were carried out using the umbrella sampling technique.[27] The reaction coordinate during biased sampling was defined as the distance between the carbonyl carbon atoms of two monomers (r_{CC}). Harmonic restraint potentials with a force constant of 5 kcal/mol/Å^2 were used, and sampling was carried out every 0.5 Å along r_{CC} from 3 to 12 Å. Except for a weak flat-bottom harmonic restraint on the position of the center of mass of the whole dimer (with a force constant of 1 kcal/mol/Å^2 and a width of the flat bottom of 2 Å), both monomers were allowed to rotate freely during simulation. For each window, an equilibration simulation of 100 ps was followed by 4.0 ns of production sampling at constant pressure and temperature. Unbiased distance probability distributions, $P(r)$, were computed by using the weighted histogram analysis method (WHAM).[28,29] The final PMFs were then constructed as[30]

$$\omega(r) = -RT \ln g(r) = -RT \ln P(r) + 2RT \ln r \qquad (1)$$

where R is the ideal gas constant and T is the absolute temperature. $P(r)$ is related to the radial distribution function, $g(r)$, by $P(r) = 4\pi r^2 g(r)$, and the last term in eq 1 corresponds to an apparent centrifugal force correction. Multidimensional PMFs can be computed from unbiased multidimensional distribution functions in a similar fashion. However, since biased sampling is carried out in only one of the dimensions (r_{CC}), convergence in unrestrained dimensions might be expected to be limited and to depend on the nature of the corresponding free energy surfaces. However, all of the multidimensional PMFs examined in the current work are smooth in the unrestrained dimensions and are well converged.

2.2. Calculation of Absolute Binding Constants. As experiments report quantitative values of the absolute binding constants of dimerization, it is necessary to compute the same quantities from PMFs to allow a direct comparison. While extensive literature exists on the fundamental statistical thermodynamics formalism for such calculations,[15−17] some confusion in its practice appears to persist. Here, we briefly revisit and clarify the theoretical framework for deriving absolute binding constants from PMFs.

Considering a dimerization/binding reaction A + B ⇌ AB, the absolute binding constant is defined as[16]

$$K_D = \frac{C^0 C_{AB}}{C_A C_B} = C^0 K_{AB} \qquad (2)$$

where C_i is the concentration of species i and C^0 is the standard concentration in the same units as C_i. The subscript "D'' stands for dimerization. Since $C^0 = 1$ M, it is often dropped from eq 2. Nonetheless, it should be emphasized that K_D is dimensionless and the standard concentration is implied in the unit of concentration when C^0 is absent.

Given an arbitrary reaction coordinate, r, some geometric criterion for defining the bound states is necessary for computing K_{AB}. Without loss of generality, let us assume that r is some distance that describes the separation between A and B and that the bound states can be defined as those states with r no greater than some cutoff distance r_c. As $P(r) = 4\pi r^2 \, e^{-\beta\omega(r)}$, where $\beta = 1/RT$, the fraction of configurations in the bound state can be computed as

$$f_{AB} = \int_{r<r_c} P(r)\, dr / \int P(r)\, dr \qquad (3)$$

For a total solution volume of V, $C_i = f_i/V$ and

$$K_{AB} = \frac{f_{AB} V}{(1 - f_{AB})^2} \qquad (4)$$

Note that multiplication of K_{AB} by $C^0 = 1/V^0$ makes K_D dimensionless. The PMF obtained from the biased sampling protocol described above reflects the interactions of A and B at infinite dilution. At such a limit, $(1 - f_{AB}) \rightarrow 1$. Therefore,

$$K_{AB}|_{V\to\infty} = f_{AB} V|_{V\to\infty} = \int_{r<r_c} P(r)\, dr \qquad (5)$$

The above derivation makes use of the relationship $\int P(r)\, dr = V$ at infinite dilution, and the integral $\int_{r<r_c} P(r)\, dr$ defines a Boltzmann-weighted phase volume.[17] As the standard concentration $C^0 = 1$ M $= 1/1660$ molecule/Å^3, the absolute binding constant is then given by

$$K_D = \frac{1}{1660} \int_{r<r_c} 4\pi r^2 \, e^{-\beta\omega(r)}\, dr \qquad (6)$$

where the unit of distance is angstrom and the unit of free energy is that of RT.

2.3. External Entropy Changes of Association. Association of two monomers leads to restriction in both positional and orientational degrees of freedom (DOFs). Such a loss of translational/rotational entropy can be computed directly by estimating the corresponding Boltzmann-weighted configurational volumes in the bound state.[17] Such an external entropy change is often decomposed into translational and rotational contributions, even though it has been demonstrated that such partitioning is arbitrary to some extent.[20] Furthermore, such partitioning typically involves an implicit assumption of uncorrelated translational and rotational DOFs, such that the

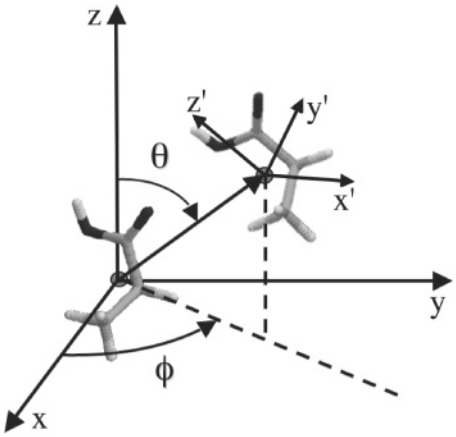

Figure 1. Coordinate systems for describing the orientational configuration of dimers. One of the monomers is fixed at the origin of the reference coordinates such that the center of mass (identified by filled circles) is located at the origin, and the principal axes of inertia align with the coordinate axes, *x*, *y*, and *z*. The polar coordinates of the center of mass of the second monomer are given by θ and ϕ; the orientation of the second monomer is described by the Euler angles that define the rotation between the reference coordinates and the fixed body coordinates, (x', y', z'). The latter are defined by the principal axes of interia of the second monomer.

translational entropy is captured in the radial distribution function and the rotational entropy is captured in the orientationl distribution function. Such an assumption is often an oversimplification. Nonetheless, this general approach simplifies the estimation of external entropy changes and is quite useful in practice. With such a framework, one can compute the translational entropy directly from the PMF, $\omega(r)$. Given the probability, $P_S(r)$, of finding configurations at separation distance r in state S (i.e., dimerized or free), and using the Shannon definition of entropy,[31] the corresponding translational entropy is given by

$$S_{trans}^{(S)} = -R\int P_S(r) \ln P_S(r)\, dr \quad (7)$$

The probability distribution functions for the dimerized and free states are

$$P_{AB}(r) = P(r)/\int_{r<r_c} P(r)\, dr \quad (8)$$

and $P_{free}(r) = C^0$, respectively. Therefore, the translational entropy change is

$$\Delta S_{trans}/R = -\int_{r<r_c} P_{AB}(r) \ln P_{AB}(r)\, dr + \ln C^0 = -\langle \ln P(r)\rangle_{r<r_c} + \ln K_D \quad (9)$$

where $<\ln P(r)>_{r<r_c} = \int_{r<r_c} P(r) \ln P(r)\, dr/\int_{r<r_c} P(r)\, dr$. The same expression can also be derived by differentiating the absolute binding free energy with respect to temperature.[16] Also, its similarity to the standard thermodynamic relation, $\Delta S/R = \beta(\Delta H - \Delta G)$, should be noted.

In principle, similar expressions can be derived for the rotational entropy loss, ΔS_{rot}, from the orientational configuration distribution. Without loss of generality, we can define the coordinate system by placing one of the monomers in a fixed orientation at the origin, shown in Figure 1. The orientational configuration of the second monomer is then described by five angles: two polar angles for the center of mass and three Euler angles for the monomer orientation. Therefore, a five-dimensional probability function is required. Furthermore, if the orientational and positional DOFs are correlated (which is typically true), a full six-dimensional configuration distribution is required for estimating the total external entropy change.[17] As such, direct calculation of ΔS_{rot} is challenging in practice even for small systems like carboxylic acids. Nonetheless, as will be shown in Results and Discussion, ΔS_{rot} is much smaller than ΔS_{trans} for the weak association of carboxylic acids under study here. It is, therefore, suitable to obtain an approximate estimate by considering the orientational restrictions only in the dimerized state near the contact minimum. Furthermore, since a carboxylic acid is largely linear, one might describe the orientational configuration of the second monomer using only three variables: polar angles of the center of mass (θ,ϕ; see Figure 1) and the cross angle of the primary principal axes of inertia of the two monomers, $\theta_2 = \arccos(\mathbf{z}\cdot\mathbf{z}')$. In analogy to eq 9, the rotational entropy loss can then be estimated as

$$\Delta S_{rot}/R \approx -\int P_{AB}(\theta,\phi,\theta_2) \ln P_{AB}(\theta,\phi,\theta_2)\, d\theta\, d\phi\, d\theta_2 - \ln 8\pi \quad (10)$$

where $P_{AB}(\theta,\phi,\theta_2)$ is estimated from the ensemble sampled near the contact minimum, and 8π is the total (free) phase volume in the three angular coordinates.

For large systems such as protein−ligand complexes, extensive sampling of the configurational space is not typically practical, and approximations are necessary for estimation of both ΔS_{trans} and ΔS_{rot}. A particularly popular approximation is the quasiharmonic approach.[17] Alternatively, one might also obtain a reasonable estimate using an intuitive approach laid out by Steinberg and Scheraga.[32] In this approach, complete loss of the translational DOFs for the ligand (or one of the monomers) is compensated by gain of (i) a rotational DOF of the vector joining the centers of mass of the two monomers in the complex and (ii) a vibrational DOF along the same vector. We will estimate the translational entropy loss using this approach and demonstrate that the results are in quantitative agreement with the more rigorous calculations.

2.4. Solvation Entropy Changes. Solvation entropy is an important property for understanding complex formation, especially with large nonpolar ligands. To estimate the solvation entropy change upon association, one might first compute the total entropy change upon binding and then subtract from it the external and internal entropy contributions. As the change in the internal (conformational and vibrational) entropy is minimal for weak association of small molecules (see Results and Discussion for more details), the solvation entropy change can be estimated as

$$\Delta S_{solv} = \Delta S_D - \Delta S_{trans} - \Delta S_{rot} \quad (11)$$

In principle, the total entropy change upon association can be directly estimated as $\Delta S_D = (\Delta H_D - \Delta G_D)/T$. However, extremely precise averages of the total potential energies are required. Further complication comes from the need to unbias the sampling to construct the bound and unbound ensembles: large fluctuations in the total potential energy (proportional to the total number of atoms in the system), together with the high precision requirement, imply that a very large number of bins are required in WHAM analysis. Alternatively, by analogy to the experimental approach, one can compute the absolute binding constants at multiple temperatures and derive the total entropy and enthalpy changes from van't Hoff plots,

$$\ln K_D = \frac{\Delta S_D}{R} - \frac{\Delta H_D}{R}\frac{1}{T} \quad (12)$$

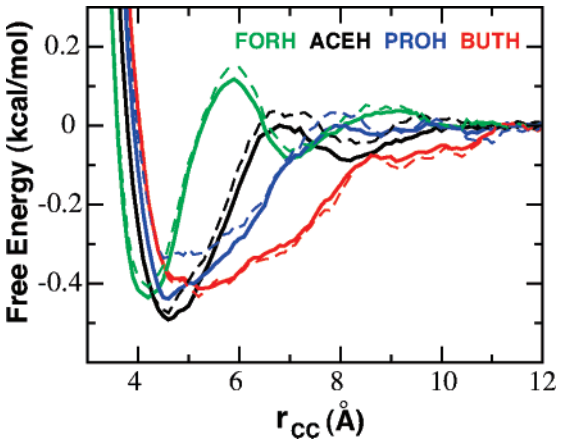

Figure 2. PMFs of carboxylic acid dimerization in aqueous solution. The color coding is as follows: green, formic acid (FORH); black, acetic acid (ACEH); blue, propionic acid (PROH); and red, butyric acid (BUTH). The solid curves were obtained by using all 4 ns production sampling, and the dashed ones were from the last 2 ns sampling.

TABLE 1: Calculated and Experimental Dimerization pK_D of Carboxylic Acids at 300 K

dimer	calcd[a] pK_D	exptl range[b]
FORH	0.25 ± 0.02	1.24−2.08
ACEH	0.038 ± 0.01	0.73−1.45
PROH	−0.064 ± 0.03	0.50−1.30
BUTH	−0.27 ± 0.01	0.26−1.00

[a]From eq 6. The convergence of the calculated pK_D is estimated from the difference between values computed from the 4 ns and the last 2 ns sampling. [b] Taken from ref 9. The lower bounds of the ranges were the values from the same group of investigators (see Table 3, column 6 of ref 9).

While computationally more expensive, this approach is often more tractable in practice and will be used in the current work. A formal statistical thermodynamics treatment of the solvent contributions to the entropy change upon dimerization has been provided in ref 32.

3. Results and Discussion

3.1. PMFs of Carboxylic Acid Dimerization in Aqueous Solution. In Figure 2, we examine the PMFs of dimerization in aqueous solution as a function of r_{CC}. As defined above, r_{CC} mainly captures hydrogen bond formation between the polar head groups. The PMFs were aligned on the basis of the flat regions at large separation, specifically, using the average free energy values from 11 to 12 Å. The convergence, estimated by comparing the results using all 4 ns sampling with those obtained using the last 2 ns sampling, appears to be very good except for the propionic acid dimer, where convergence is only fair. The desolvation peak is present only in the PMFs of formic and acetic acid dimers, but largely absent in those of propionic and butyric acids. Such differences can be attributed to greater interactions between longer nonpolar alkyl groups. For propionic and butyric acid dimers, the nonpolar tails can make hydrophobic contacts and contribute significantly to dimer stability even when the polar heads are substantially separated. This is a first, albeit indirect, evidence from detailed simulations to support the original hypothesis that both hydrogen bonding and hydrophobic interactions contribute to the dimerization of carboxylic acids in aqueous solutions.[9]

3.2. Absolute Dimerization Constants. Calculation of absolute dimerization constants (and related standard-state free energies there from) from the PMFs (eq 6) requires some geometric criteria for defining the bound states. For formic and acetic acid dimers, the distances at which the desolvation peaks maximize seem to be obvious choices. However, the choice is not as obvious for propionic and butyric acid dimers. To derive an objective estimate of cutoff distances, we examined the free energy as a function of both r_{CC} and the minimum heavy-atom distance, r_{min} (data not shown). The analysis indicates that appropriate values of r_c are 6.0, 6.5, 7.0, and 8.0 Å, respectively, for the dimers. With such choices, two monomers are indeed largely out of contact when $r_{CC} > r_c$, that is, with the minimum heavy-atom distances greater than 4.5 Å. It should be noted that the integral in eq 6 is dominated by contributions from

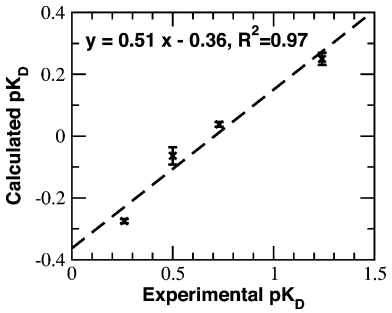

Figure 3. Correlation of experimental and calculated values of dimerization pK_D at 300 K. The experimental values correspond to the lower bound of the range summarized in Table 1.

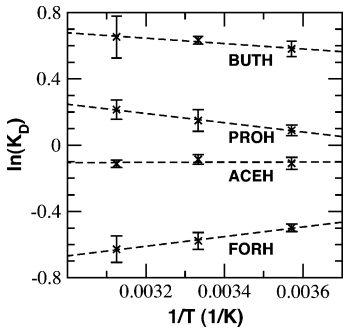

Figure 4. Van't Hoff plots of carboxylic acid dimerization constants. The dashed lines were obtained through least-square fitting. The residuals of fit (χ^2) are 6.0×10^{-5}, 4.0×10^{-4}, 3.6×10^{-5}, and 8.1×10^{-5} for FORH, ACEH, PROH, and BUTH, respectively.

TABLE 2: Calculated pK_D Values for Carboxylic Acid Dimers at Three Temperatures

T (K)	FORH	ACEH	PROH	BUTH
280	0.22 ± 0.01	0.048 ± 0.02	−0.039 ± 0.01	−0.25 ± 0.02
300	0.25 ± 0.02	0.038 ± 0.01	−0.064 ± 0.03	−0.27 ± 0.01
320	0.27 ± 0.03	0.049 ± 0.01	−0.093 ± 0.03	−0.28 ± 0.06

TABLE 3: Entropy and Enthalpy of Carboxylic Acid Dimerization[a]

dimer	ΔH_D	ΔS_D	ΔS_{trans}	ΔS_{rot}	ΔS_{solv}
FORH	−0.58 ± 0.05	−3.04 ± 0.16	−13.0	−0.2	10.2
ACEH	−0.02 ± 0.12	−0.24 ± 0.42	−12.8	−0.4	13.0
PROH	0.56 ± 0.04	2.16 ± 0.12	−12.5	−0.6	15.2
BUTH	0.32 ± 0.07	2.34 ± 0.24	−12.1	−0.4	14.8

[a]Total changes were obtained from least-squares fitting of the van't Hoff plots shown in Figure 4. Various entropic components were calculated by using the formalism outlined in Methods (eqs 9 and 10). The units are kilocalorie per mole for enthalpy and calorie per mole per Kelvin for entropy.

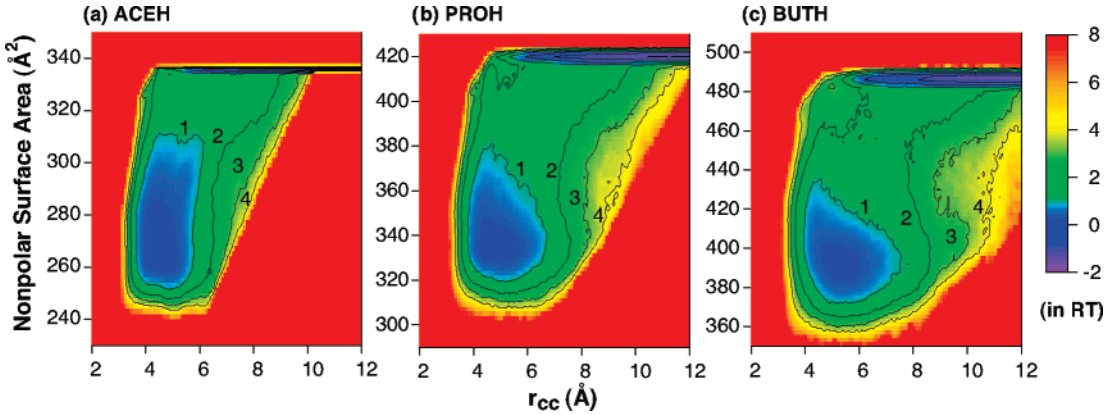

Figure 5. 2D PMFs as function of r_{CC} and solvent accessible nonpolar (alkyl chains) surface area at 300 K. The surface area was calculated by using default CHARMM22 van der Waals radii together with a 1.4 Å probe. The surfaces were shifted so that the free energy minimum in the bound basin ($r_{CC} \sim 5$Å) was zero. The contours are drawn at every RT, and levels higher than 4 RT are omitted from the plots for clarity.

low free energy regions and is influenced only slightly by the precise choice of r_c.

Table 1 compares the calculated pK_D with the experimental values at 300 K. Despite large uncertainties in the experimental values (as indicated by the wide ranges), it is clear that the calculated pK_D values are systematically lower than the experimental values. This indicates that the CHARMM22 force field employed in this study over stabilizes solute−solute interactions with respect to solute−water interactions. Such a systematic tendency of excessive self-association has also been observed in other studies for a range of classical force fields and model systems.[33,34] Nonetheless, the calculated pK_D values are well-correlated with the experimental results, as shown in Figure 3. A least-squares fit yields a slope (0.51) of less than one with a negative offset (−0.36). This indicates that the strengths of both hydrogen bonding and hydrophobic interactions are overestimated by the force field. The observed high correlation of experimental and theoretical values indicates that, despite the systematic deviation, the intrinsic trend within the homologous series is described well by the molecular model. Therefore, further analysis presented in the rest of this paper is likely relevant.

3.3. Total Entropy and Enthalpy of Dimerization. To estimate the total dimerization entropy and enthalpy, free energy calculations were carried out at three temperatures (280, 300, and 320 K). The resulting dimerization pK_D are summarized in Table 2. Uncertainties were estimated in the same way as in Table 1. These values were then fitted to the van't Hoff equation. As shown in Figure 4, the calculated pK_D values are well-described by a linear relationship with very small residuals. The derived total changes of entropy and enthalpy upon dimerization are summarized in Table 3. The uncertainties are estimated by considering the propagation of errors.[35] These results are largely in agreement with experimental values of $\Delta H_D \sim 0 \pm 1$ kcal/mol[9,36] and $\Delta S_D \sim 0 \pm 5$ cal/mol/K (note that the original references do not include an estimate of ΔS_D, and the above range is estimated here from the relation $T\Delta S_D = \Delta H_D - \Delta G_D$). In a more recent experimental study,[11] Yamamoto and Nishi used mass spectrometric analysis to estimate ΔH_D and ΔS_D of carboxylic acid dimerization in water clusters. Their measurements yielded ΔH_D of 0.0 to 0.5 kcal/mol and ΔS_D of −0.4 to 2.2 cal/mol/K for formic to butyric acids. We note some substantial difference in the experimental and simulation conditions. Nonetheless, their results are in excellent correlation with our calculations. Again, the existing deviation might also be related to the systematic tendency of excessive self-aggregation of current classical force fields, as noted above.

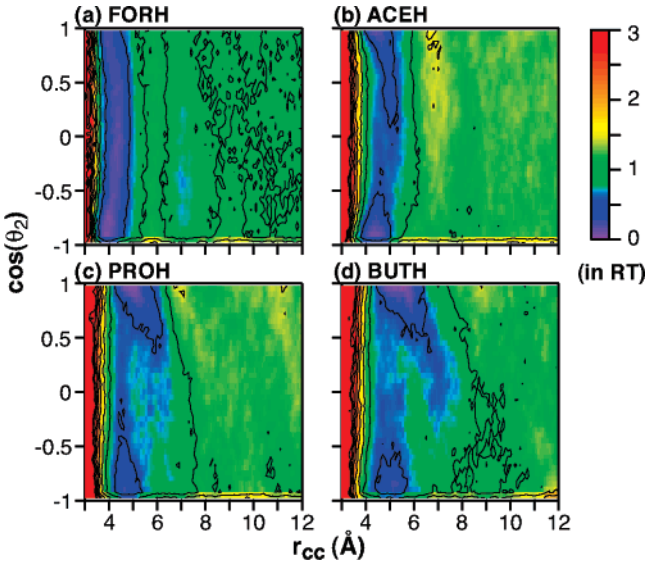

Figure 6. 2D PMFs as functions of r_{CC} and θ_2 (see the main text for definitions). The free energy contours are drawn every 0.5 RT (T = 300 K).

Interestingly, the changes in both the total entropy and enthalpy upon dimerization are negative for formic acid and become less negative/more positive with longer alkyl chains. As formic acid interacts mainly through hydrogen bonding, the observed trend of dimerization entropy and enthalpy indicates that (i) both hydrogen bonding and hydrophobic interactions contribute to dimerization of carboxylic acids with alkyl chains; (ii) changes of both entropy and enthalpy of hydrogen bonding of carboxyl groups are negative; and (iii) changes of both entropy and enthalpy of hydrophobic association of small nonpolar solutes are positive. These observations are in good agreement with the existing understanding of hydrogen bonding and hydrophobic interactions. For example, it is believed that, in hydrophobic hydration, small nonpolar solutes enhance the surrounding water structure, and these "structured'' waters are released upon hydrophobic association leading to increases in both entropy and enthalpy.[3,13,14] By examining the differences in ΔH_D and ΔS_D between the formic acid dimer and the rest of the homologous series, one can estimate that, for hydrophobic association of small nonpolar solutes like alkyl chains, $\Delta H_D \sim$ 0.6 to 1.0 kcal/mol and $\Delta S_D \sim 3$ to 5 cal/mol/K. These results are in excellent agreement with previous theoretical estimates of $\Delta H_D \sim 0.3$ to 1.8 kcal/mol and $\Delta S_D \sim 1.7$ to 11 cal/mol/K for hydrophobic association of amino acid side chains[13] and

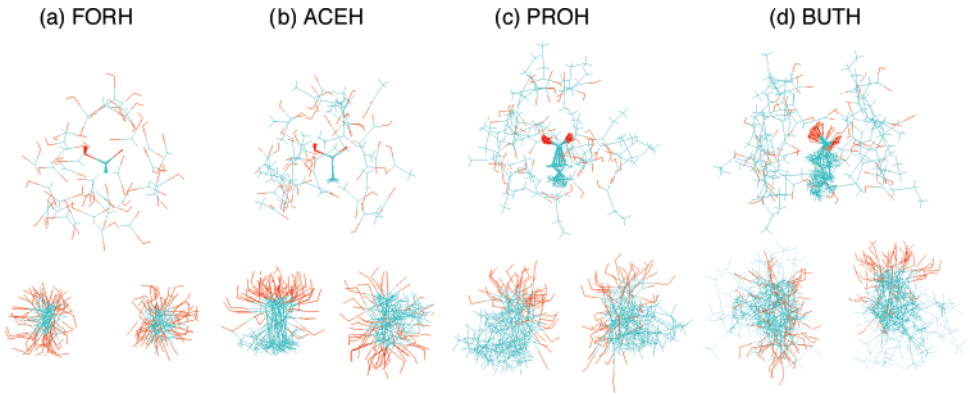

Figure 7. Representative snapshots of carboxylic acid dimers sampled with r_{CC} restrained near the corresponding contact minima. Charged oxygen atoms are plotted in red, and all carbon atoms are in blue. Hydrogen atoms are omitted for clarity. The snapshots were aligned using either all heavy atoms in one of the monomers (top) or selected atoms from both monomers (bottom). The selected atoms are oxygens for FORH and all carbons for the rest.

thus provide an interesting verification of previous estimates based on the hydrophobic interaction theory of Némethy and Scheraga[13] from first-principles. Similar ranges have also been obtained from experiments such as chromatography[14,37,38] and titration of cyclodextrin complexes.[39]

3.4. External Entropy Changes Upon Dimerization. Identifying the net solvent contribution to the total dimerization entropy is useful for improving the current understanding of solvation energetics of polar and nonpolar solutes. For this, one needs to first estimate the external and internal entropy changes upon dimerization. Because of the weak association of carboxylic acid dimers, the vibrational entropy (bonds and angles) is expected to remain constant. It can further be demonstrated that the conformational entropy change is also minimal, by examining the distribution of torsion angles in the bound and free states (data not shown). The external entropy change can be decomposed into translational and rotational entropy by assuming uncorrelated translational and orientational DOFs, using the formalism outlined in Methods. The results are summarized in Table 3. The rotational entropy loss is estimated from eq 10 using the orientational distributions sampled around the contact minima, which are 4.2, 4.6, 4.6, and 5.2 Å for formic, acetic, propionic, and butyric acid dimers, respectively. Clearly, the external entropy change is dominated by the translational entropy loss, and the approximation made in estimating the rotational entropy loss has little impact on the following analysis. We also estimated the loss of translational entropy using the scheme described by Steinberg and Scheraga.[32] Given the actual molecular weights, an average dimer separation of 5 Å, and average $\omega_{min}(r) \sim -0.6$ kcal/mol and $\delta r_{max} \sim 3$ Å, one can estimate the total translational entropy loss to be ~ -15 (eq 14 of ref 32) + 4 (eq 35 of ref 32) $= -11$ cal/mol/K. This is in excellent agreement with the values shown in Table 3, pointing to equivalence of different theoretical approaches. We further note that the calculated value of ~ -13 cal/mol/K for the total external entropy loss upon dimerization is in quantitative agreement with a recent experimental result of -10 ± 8 cal/mol/K, obtained from the dissociation/unfolding of dimeric proteins.[18]

3.5. Solvation Entropy of Polar and Nonpolar Solutes. The solvation entropy contribution to the dimerization entropy was estimated by subtracting the external (and internal) entropic contributions from the total dimerization entropy (eq 11). The results, provided in Table 3, demonstrate that desolvation entropy is positive for both polar (formic acid) and small nonpolar solutes (alkyl chains) at room temperature, in agreement with previous qualitative experimental results.[40] The difference between the solvation entropy contributions of the formic acid dimer and the rest of the homologous series provides an estimation of the net solvation entropy contribution from hydrophobic association, which is approximately 3 to 5 cal/mol/K. Examples of hydrophobic entropies of association computed for maximum overlap in dimers of similar-sized side chains in proteins[13] are Ala−Ala (4.7 cal/mol/K) and Leu−Leu (6.0 cal/mol/K). Since the alkyl side chains of the carboxylic acids do not overlap maximally, the values cited above (3 to 5 cal/mol/K) are slightly lower. Furthermore, by examining the 2D PMF as a function of r_{CC} and the total solvent accessible nonpolar surface area, shown in Figure 5, it can be estimated that the average nonpolar surface area reduction is about 70 to 100 Å^2 (by considering the difference in the nonpolar surface areas between the bound, $r_{CC} \sim 5$ Å, and free, $r_{CC} > 8$Å, basins). Therefore, the hydration entropy per unit nonpolar surface area for alkyl chains can be estimated to be ~ -0.05 cal/mol/K/Å^2. Interestingly, this is similar to previous theoretical calculations of hydration of small spherical cavities using the Lum−Chandler−Weeks statistical thermodynamics theory of hydrophobic hydration[41] (based on Figure 2 of ref 41).

3.6. Interplay of Hydrogen Bonding, Hydrophobic Interactions, and Entropy in Dimerization. The above analysis of the PMFs, dimerization constants, total dimerization entropy, and enthalpy changes as well as the solvation entropy contributions all point to an interplay of hydrogen bonding, hydrophobic interaction, and entropy in carboxylic acid dimerization processes. To directly verify the hypothesis that both hydrogen bonding and hydrophobic association contribute to the stability of carboxylic acid dimers,[9] we examine the 2D free energy as a function of both r_{CC}, which captures the influence of hydrogen bonding, and solvent accessible nonpolar surface area, which reflects the contribution from hydrophobic association. The free energy basins that correspond to the dimer states can be outlined approximately by the 1 *RT* contour lines, shown in Figure 5. Clearly, both hydrogen bonding and hydrophobic interactions are involved in dimerization. The shapes of the free energy basins further reveal that, while hydrogen bonding appears to play a more dominate role for all dimers, hydrophobic interactions have increasing roles with longer alkyl chains: for acetic acid, the stability of the dimer decreases rapidly with larger r_{CC}, but much more slowly with larger nonpolar surface area (i.e., smaller reduction in the nonpolar surface area); however, for butyric acid, the dimer

stability decreases at similar rates with increase in either r_{CC} or nonpolar surface area.

Self-assembly of amphiphilic molecules into mesoscopic and macroscopic structures in water, a process fundamental to all living organisms, arises from the interplay of hydrogen bonding, hydrophobic interactions, and entropy. The essence of these complex phenomena is captured even in simple model systems such as the carboxylic acid dimers. In Figure 6, we examine the free energy as a function of both r_{CC} and θ_2. As defined in Methods, θ_2 measures the alignment of two largely linear monomers and captures the degree of ordering in dimerization. Not surprisingly, there is a lack of significant features along θ_2 beyond the 1.0 RT contour level, indicating weak degree of monomer alignment for the whole series. Nonetheless, interesting fine features exist at lower free energy levels. While the formic acid dimer free energy surface displays no feature even at the 0.5 RT level, some degree of ordering in monomer alignment is evident for the other three dimers, where both hydrogen bonding and hydrophobic interactions are involved in dimerization. The strongest degrees of alignment are observed for propionic and butyric acid dimers, reflected by dominant free energy basins around $\cos(\theta_2) \sim 1$ and $r_{CC} \sim 5$ Å. This indicates that propionic and butyric acid dimers are weakly ordered with parallel configurations being most probable. For the acetic acid dimer, even though there is a slight enrichment of side-to-side aligned configurations, that is, $\cos(\theta_2) \sim \pm 1$, hydrophobic interactions between the short alkyl chains do not appear to be strong enough to induce substantial bias toward fully aligned parallel configurations. Figure 7 shows some representative snapshots of carboxylic acid dimers, taken every 100 ps from the 4 ns production simulations with r_{CC} restrained near the contact minima. It illustrates the emergence of weak ordering with longer alkyl chains; for example, see the top panels of Figure 7. However, all dimers including butyric acid dimers remain highly disordered, as depicted in the bottom panels of Figure 7. Only through careful free energy analysis do the important underlying features described above start to emerge. This highlights the importance of extensive sampling and statistical analysis in deriving a reliable, quantitative understanding of systems with weak association and ordering.

4. Conclusion

To understand the interplay of hydrogen bonding, hydrophobic interactions, and entropy, we revisited carboxylic acid dimers and carried out extensive biased molecular dynamics simulations using an all-atom classical force field. Through careful free energy analysis, the current study provides unambiguous support for the original hypothesis that both hydrogen bonding and hydrophobic association are involved in dimerization of carboxylic acids[9] and provides a quantitative description of the distributions of the mutual orientations of the monomers in the dimers. Quantitative agreement in the thermodynamics of hydrogen bonding and hydrophobic interactions is demonstrated with existing experimental data as well as theoretical calculations. This illustrates the usefulness of carboxylic acid dimers as a model system for understanding hydration phenomena of both polar and nonpolar solutes and for verifying statistical mechanical theories of hydrophobic association. It is particularly interesting that such detailed, first-principles simulations can reproduce and support a previous qualitative and semiquantitative understanding of hydration energetics of different species. For example, the simulations correctly demonstrate that both polar and nonpolar solutes have positive dehydration entropy and that the enthalpy of hydrophobic association is positive. The calculations further estimate that the solvation entropy contribution to the hydrophobic association of the alkyl groups of carboxylic acids is 3 to 5 cal/mol/K, similar to that of nonpolar side chains of amino acids,[13] and that the hydration entropy per unit nonpolar area is ~ -0.05 cal/mol/K/Å^2 for small solutes, also in agreement with previous calculations.[41] We also revisit and clarify the underlying statistical thermodynamics formalism for calculation of absolute binding constants and external entropy of dimerization, as well as estimation of solvation entropy contributions to hydrogen bonding and hydrophobic interactions. These general formalisms are expected to be very useful for a wide range of applications including studies of protein−ligand and protein−protein interactions.

Acknowledgment. This work was supported by the National Institutes of Health (RR-12255 to C.L.B. and GM-14312 to H.A.S.) and the National Science Foundation (MCB05-41633 to H.A.S.). J.C. also thanks David Bostick for useful discussions regarding the calculation of absolute binding constants.

References and Notes

(1) Brooks, C. L., III; Karplus, M.; Pettitt, B. M. *Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics;* John Wiley and Sons: New York, 1987.
(2) Dyson, H. J.; Wright, P. E.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 13057−13061.
(3) Chandler, D. *Nature* **2005**, *437*, 640−647.
(4) Deechongkit, S.; Nguyen, H.; Powers, E. T.; Dawson, P. E.; Gruebele, M.; Kelly, J. W. *Nature* **2004**, *430*, 101−105.
(5) Rose, G. D.; Fleming, P. J.; Banavar, J. R. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 16623−16633.
(6) Ghosh, A.; Elber, R.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A* **2002**, *99*, 10394−10398.
(7) Jacob, J.; Krantz, B.; Dothager, R. S.; Thiyagarajan, P.; Sosnick, T. R. *J. Mol. Biol.* **2004**, *338*, 369−382.
(8) Jagielska, A.; Scheraga, H. A. *J. Comput. Chem.* **2007**, *28*, 1068−1082.
(9) Schrier, E. E.; Pottle, M.; Scheraga, H. A. *J. Am. Chem. Soc.* **1964**, *86*, 3444−3449.
(10) Suzuki, E.; Taniguchi, Y.; Watanabe, T. *J. Phys. Chem.* **1973**, *77*, 1918−1922.
(11) Yamamoto, K.; Nishi, N. *J. Am. Chem. Soc.* **1990**, *112*, 549−558.
(12) Chocholousova, J.; Vacek, J.; Hobza, P. *J. Phys. Chem. A* **2003**, *107*, 3086−3092.
(13) Némethy, G.; Scheraga, H. A. *J. Phys. Chem.* **1962**, *66*, 1773−1789.
(14) Scheraga, H. A. *J. Biomol. Struct. Dyn.* **1998**, *16*, 447−460.
(15) Hill, T. L. *Statistical Mechanics. Principles and Selected Applications;* McGraw-Hill: New York, 1956.
(16) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. *Biophys. J.* **1997**, *72*, 1047−1069.
(17) Luo, H.; Sharp, K. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, 99, 10399−10404.
(18) Yu, Y. B.; Privalov, P. L.; Hodges, R. S. *Biophys. J.* **2001**, *81*, 1632−1642.
(19) Lazaridis, T. *Curr. Org. Chem.* **2002**, *6*, 1319−1332.
(20) Carlsson, J.; Aqvist, J. *Phys. Chem. Chem. Phys.* **2006**, *8*, 5385−5395.
(21) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E., III; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(22) Joregensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.
(23) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577−8593.
(24) Brooks, C. L., III; Pettitt, B. M.; Karplus, M. *J. Chem. Phys.* **1985**, *83*, 5897−5908.
(25) Steinbach, P. J.; Brooks, B. R. *J. Comput. Chem.* **1994**, *15*, 667−683.

(26) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327−341.
(27) Torrie, G. M.; Valleau, J. P. *J. Comput. Phys.* **1977**, *23*, 187−199.
(28) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011−1021.
(29) Roux, B. *Comput. Phys. Commun.* **1995**, *91*, 275−282.
(30) Paci, E.; Ciccotti, G.; Ferrano, D.; Kapral, R. *Chem. Phys. Lett.* **1991**, *176*, 581−587.
(31) Shannon, C. E. *Bell System Tech. J.* **1948**, *27*, 379−423,623−656.
(32) Steinberg, I. Z.; Scheraga, H. A. *J. Biol. Chem.* **1963**, *238*, 172−181.
(33) Perera, A.; Sokolic, F. *J. Chem. Phys.* **2004**, *121*, 11272−11282.
(34) Kang, D.; Smith, P. E. *J. Comput. Chem.* **2006**, *27*, 1477−1485.
(35) Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. Numerical Recipies. *C: The Art of Scientific Computing*, 2nd ed.; Cambridge University Press: Cambridge, U.K., 1992.
(36) Nash, G. R.; Monk, C. B. *J. Chem. Soc.* **1957**, 4274−4280.
(37) Steinberg, I. Z.; Scheraga, H. A. *J. Am. Chem. Soc.* **1962**, *84*, 2890−2892.
(38) de Araujo, A. F. P.; Pochapsky, T. C.; Joughin, B. *Biophys. J.* **1999**, *76*, 2319−2328.
(39) Ross, P. D.; Rekharsky, M. V. *Biophys. J.* **1996**, *71*, 2144−2154.
(40) Privalov, P. L.; Makhatadze, G. I. *J. Mol. Biol.* **1993**, *232*, 660−679.
(41) Huang, D. M.; Chandler, D. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 8324−8327.