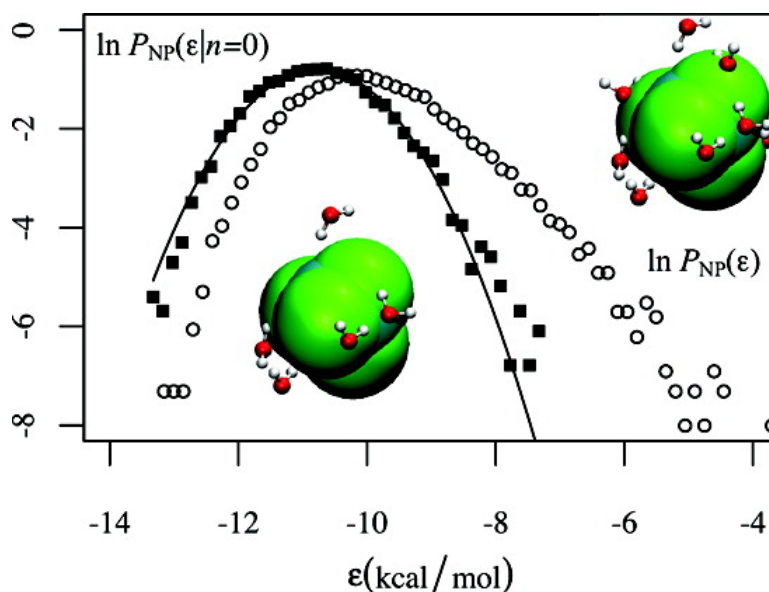Article

Non-van der Waals Treatment of the Hydrophobic Solubilities of CF

D. Asthagiri, H. S. Ashbaugh, A. Piryatinski, M. E. Paulaitis, and L. R. Pratt

More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Links to the 4 articles that cite this article, as of the time of this article download
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

Non-van der Waals Treatment of the Hydrophobic Solubilities of CF$_4$

D. Asthagiri,[†] H. S. Ashbaugh,[‡] A. Piryatinski,[§] M. E. Paulaitis,[||] and L. R. Pratt*,[§]

Contribution from the Department of Chemical & Biomolecular Engineering, Johns Hopkins University, Baltimore, Maryland 21218, Department of Chemical & Biomolecular Engineering, Tulane University, New Orleans, Louisiana 70118, Theoretical Division, Los Alamos National Laboratory, Los Alamos New Mexico 87545, and Department of Chemical & Biomolecular Engineering, The Ohio State University, Columbus, Ohio 43210

Received February 13, 2007; E-mail: lrp@lanl.gov

Abstract: A quasi-chemical theory implemented on the basis of molecular simulation is derived and tested for the hydrophobic hydration of CF$_4$(aq). The theory formulated here subsumes a van der Waals treatment of solvation and identifies contributions to the hydration free energy of CF$_4$(aq) that naturally arise from *chemical* contributions defined by quasi-chemical theory and *fluctuation* contributions analogous to Debye—Hückel or random phase approximations. The resulting Gaussian statistical thermodynamic model avoids consideration of hypothetical drying-then-rewetting problems and is physically reliable in these applications as judged by the size of the fluctuation contribution. The specific results here confirm that unfavorable tails of binding energy distributions reflect few-body close solute—solvent encounters. The solvent near-neighbors are pushed by the medium into unfavorable interactions with the solute, in contrast to the alternative view that a preformed interface is pulled by the solute—solvent attractive interactions into contact with the solute. The polyatomic model of CF$_4$(aq) studied gives a satisfactory description of the experimental solubilities including the temperature dependence. The proximal distributions evaluated here for polyatomic solutes accurately reconstruct the observed distributions of water near these molecules which are nonspherical. These results suggest that drying is not an essential consideration for the hydrophobic solubilities of CF$_4$, or of C(CH$_3$)$_4$ which is more soluble.

I. Introduction

It was famously noted many years ago that "no one has yet proposed a quantitative theory of aqueous solutions of nonelectrolytes, and such solutions will probably be the last to be understood fully."[1] Much has been done since that comment was offered first, and molecular simulation, as distinct from theory, is now routinely applied to aqueous solutions of nonelectrolytes. But quantitative molecular theory is still incomplete for these systems. Despite the fact that aqueous solutions are central to biochemistry, biophysics, and to nanotechnology invented by analogy to sophisticated biomolecular systems, molecular theory for these systems is still an outstanding challenge.

Motivated in the first place toward validation of simulation models,[2] this paper defines and solves several problems which are essential for quantitative theories of aqueous solutions. The results of the present topical application to CF$_4$(aq) shed new light on hydrophobic hydration in an interesting setting of solute size and interactions.

Specifically, we formulate a theory that permits direct evaluation of the hydration free energy of CF$_4$(aq) on the basis of primitive simulation data and without the reliance on the specialized perspective of van der Waals (vdW) theories in which physical interactions are separated into reference and perturbation interactions. Furthermore, these developments solve the difficulties of high-energy (unfavorable) fluctuations—expressed in tails of binding energy distributions—which are central to the statistical thermodynamic problem. Finally, these developments avoid consideration of unphysical intermediate states that might occur, for example, in coupling-parameter integrations.

A feature of coupling-parameter integrations for hydration free energies for CF$_4$(aq) that has been of specific interest recently[2,3] is *drying-then-rewetting* of the solute through intermediate levels of attractive couplings between solute and solvent. In this drying-then-rewetting scenario, it is argued that water pulls away from hydrophobic contact,[4] then the effect of attractive interactions is to pull a preformed interface structure back into contact with the hydrophobic solute. As discussed below, the insights obtained by considering the distribution of solute—solvent binding energies help to avoid drying of a surface when it is destined only to rewet with physical

---

[†] Johns Hopkins University.
[‡] Tulane University.
[§] Los Alamos National Laboratory.
[||] The Ohio State University.

(1) Rowlinson, J. S. *Liquids and Liquid Mixtures*; Plenum: New York, 1969; p 167.
(2) Li, X.; Li, J. Y.; Eleftheriou, M.; Zhou, R. H. *J. Am. Chem. Soc.* **2006**, *128*, 12439—12447.

(3) Weeks, J. D. *Annu. Rev. Phys. Chem.* **2002**, *53*, 533—562.
(4) Ashbaugh, H. S.; Pratt, L. R. *Rev. Mod. Phys.* **2006**, *78*, 159—78.

interactions fully expressed.[5,6] Because drying phenomena can be associated with a characteristic temperature dependence of the hydration free energy, an important quantitative outcome here is verification of a temperature dependence of the aqueous solubility of CF_4 that is characteristic of simple hydrophobic solutes; that is, the computed hydration free energy of CF_4 increases with increasing temperature through moderate temperatures.[7,8] The work reported here thus gives new results and analyses of the significance of drying to the hydration free energy of molecular solutes such as CF_4.[9]

The vdW theory is the well-established starting point for the theory of liquids generally.[10,11] A signature of a vdW theory is the separation of the physical interactions into reference interactions and a remainder that is treated perturbatively. Suitably identified hard-core model interactions are one possibility for the reference interactions.[12] The solution of the reference model problem is then the first task.[4] A vdW approach anticipates minimal changes in the local structural characteristics upon inclusion of perturbative interactions. Then final evaluation of thermodynamic properties can use first-order perturbation theory and is simple.[10,11]

Several general complications can arise with this approach. One complication is that a good identification of reference interactions is not available for physical intermolecular interactions, in general. If a conventional reference interaction model, for example a hard-core model, is merely assumed, another complication can arise if local structural characteristics observed for the reference case are not closely similar to those with the full, physical interactions. The latter complication is realized for hard-core models of hydrophobic solutes in water when the solute is substantially larger in size than a water molecule; for hard-core solutes a gradual dewetting with increasing solute size is expected. This has been clearly recognized for a long time[13] and accurately characterized recently.[4,14]

Thus, the CF_4(aq) case and many other liquid solutions eventually require an understanding beyond the concepts of vdW theories. A desideratum for such extensions is that they should be naturally embedded in the molecular simulations that are the way calculations on liquids are routinely done now. Another desideratum is that theoretical extensions should permit treatment of accurate molecular-scale potential-energy models, including chemical interactions where they are important.

Insightful extensions of vdW theories have been proposed.[15] Because these are difficult problems the historical theories typically start with modeling simplifications that are viewed as external to the theory. The *primitive models* that are standard
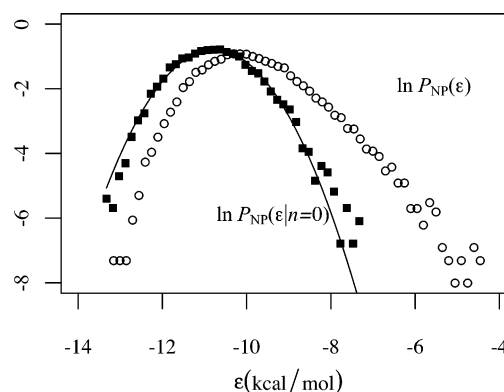
Figure 1. Distributions of binding energies of neopentane (NP = C(CH$_3$)$_4$) to water observed during simulation of the solution at $T = 298$ K, $p = 1$ bar, and infinite dilution. The open circles are the unconditional distribution $P_{NP}(\epsilon) = \langle \delta(\epsilon - \Delta U) \rangle$. The filled squares are the conditional distribution $P_{NP}(\epsilon|n = 0)$ obtained from the sample in which no water O-atoms occupy the volume defined by the distance $\lambda \leq 0.335$ nm of any C-atom. The curve is a Gaussian model for the conditional distribution.

to electrolyte solution theory are obvious examples.[16] To the extent that the preliminary modeling is important, ambitious theories often do not satisfy the desiderata noted above, and that is a limitation.

Here we discuss and test a theory that does satisfy those desiderata. This theory transparently subsumes vdW approaches and identifies additional natural contributions, in particular both *chemical* contributions as defined by quasi-chemical theories[17] and *fluctuation* contributions that would naturally arise with Debye—Hückel (or random-phase) approximations.[18] We test this new theory for a problem, CF_4 solubility in liquid water, where vdW approaches should be just fine. A reason for this application is that the physical arguments justifying a vdW approach are sometimes so specialized that it is helpful to see how the present approach naturally encompasses the physical issues of vdW problems.

A basic theoretical issue is exemplified by the results of Figure 1. Shown there by the open circles is the probability distribution of the binding energy[18,19] (ϵ) of a neopentane molecule C(CH$_3$)$_4$ in model liquid water. (Simulation details are given later.) Notice that this distribution displays a long, nearly exponential tail extending to high ϵ. Here we follow the idea that this tail reflects the influence of repulsive forces between the solute and solvent molecules. We rearrange the estimation of the excess chemical potential to acknowledge information on problems of packing a hard-core-model-C(CH$_3$)$_4$ into this solvent. A consequence of that rearrangement is an explicit identification of a *chemical* contribution to the desired free energy, and the theory is then recognized as an instance of *quasi-chemical* theory.[18,19] The modified distribution of binding energies that we then seek from the data—shown by the black squares of Figure 1—is Gaussian to sufficient accuracy. The result is then an accessible, approximation of the

(5) Ashbaugh, H. S.; Paulaitis, M. E. *J. Am. Chem. Soc.* **2001**, *123*, 10721—10728.

(6) Ashbaugh, H. S.; Pratt, L. R.; Paulaitis, M. E.; Clohecy, J.; Beck, T. L. *J. Am. Chem. Soc.* **2005**, *127*, 2808—2809.

(7) Graziano, G. *J. Phys. Chem. B* **2004**, *108*, 9371—9372.

(8) Huang, X.; Margulis, C. J.; Berne, B. J. *J. Phys. Chem. B* **2004**, *108*, 9373—9374.

(9) Huang, X.; Margulis, C. J.; Berne, B. J. *J. Phys. Chem. B* **2003**, *107*, 11742—11748.

(10) Widom, B. *Science* **1967**, *157*, 375—382.

(11) Widom, B. *Statistical Mechanics. A concise introduction for chemists*; Cambridge University Press: Cambridge, U.K., 2002.

(12) Chandler, D.; Weeks, J. D.; Andersen, H. C. *Science* **1983**, *220*, 787.

(13) Stillinger, F. H. *J. Soln. Chem.* **1973**, *2*, 141.

(14) Hummer, G.; Garde, S. *Phys. Rev. Letts.* **1998**, *80*, 4193—4196.

(15) Stell, G. In *Statistical Mechanics. Part A: Equilibrium Techniques*; Berne, B. J., Ed.; Plenum: New York, 1977; pp 47—84.

(16) Friedman, H. L.; Dale, W. D. T. In *Statistical Mechanics. Part A: Equilibrium Techniques*; Berne, B. J., Ed.; Plenum: New York, 1977, pp 85—135.

(17) Pratt, L. R.; Asthagiri, D. In *Free Energy Calculations. Theory and Applications in Chemistry and Biology*; Chipot, C., Pohorille, A., Eds.; Springer: Berlin, 2006; Chapter 9.

(18) Beck, T. L.; Paulaitis, M. E.; Pratt, L. R. *The Potential Distribution Theorem and Models of Molecular Solutions*; Cambridge University Press: New York, 2006.

(19) Paulaitis, M. E.; Pratt, L. R. *Adv. Prot. Chem.* **2002**, *62*, 283—310.

desired interaction-contribution to the chemical potential of $C(CH_3)_4$(aq). The width of the Gaussian distribution provides a correction to a mean-field picture of this free energy and thus a check whether a mean-field treatment is sufficient.

II. Theory

The probability density $P_{CF_4}(\epsilon) = \langle \delta(\epsilon - \Delta U) \rangle$ of initial interest determines the excess chemical potential of CF_4(aq) through

$$\exp(\beta \mu_{CF_4}^{(ex)}) = \int P_{CF_4}(\epsilon) \exp(\beta\epsilon) \, d\epsilon \quad (2.1)$$

the potential distribution theorem in the *inverse* form.[18,19] $\delta(x)$ is the Dirac-delta function. The binding energy $\Delta U = U_{N+1} - U_N - U_1$ is the difference at a sampled configuration of the potential energy of the whole system from the sum of the decoupled energies of the solution and the solute, $\beta = 1/kT$ with T being the temperature, and $\mu_{CF_4}^{(ex)}$ is the interaction contribution to the chemical potential of CF_4(aq). Specifically,

$$\mu_{CF_4}^{(ex)} \equiv \mu_{CF_4} - \mu_{CF_4}^{(ideal)} \quad (2.2)$$

The right-most term of eq 2.2, $\mu_{CF_4}^{(ideal)}$, would be the chemical potential if intermolecular interactions were not present, so $\mu_{CF_4}^{(ex)}$ is the part of the chemical potential μ_{CF_4} in *excess* of the ideal gas result at the same density and temperature.

We discuss below results on $P_{CF_4}(\epsilon)$. But motivated by the introductory discussion, we also pursue an interest in hard-core (HC) model solutes. A HC model solute rigidly excludes the solvent water molecules from a molecular volume positioned on the solute molecule. That excess chemical potential $\mu_{HC}^{(ex)}$ is assumed to be known; there has been extensive study of that packing problem in modern times[4,20−31] so that is a realistic assumption. To calculate $\mu_{HC}^{(ex)}$ on the basis of the same simulation data, we would naturally use the difference formula[18]

$$\exp(-\beta(\mu_{HC}^{(ex)} - \mu_{CF_4}^{(ex)})) = \langle \exp(-\beta(\Delta U_{HC} - \Delta U_{CF_4})) \rangle \quad (2.3)$$

In this elaborated notation, ΔU_{CF_4} now stands for the binding energy for the CF_4 molecule, and ΔU_{HC} is the corresponding binding energy for the HC case. ΔU_{HC} is either zero (no solute−solvent overlaps) or infinite (overlapping). The simulation data being generated in the same way as before, the use of eq 2.3 merely requires eliminating configurations in which a solvent molecule penetrates the assigned solute exclusion volume. The

remaining configurations are a fraction $0 \leq p_{CF_4}(n=0) \leq 1$ of the original number. Then

$$\mu_{CF_4}^{(ex)} = \mu_{HC}^{(ex)} + kT \ln p_{CF_4}(n=0) + kT \ln \int P_{CF_4}(\epsilon|n=0) \exp(\beta\epsilon) \, d\epsilon \quad (2.4)$$

re-expresses the free-energy that is sought in eq 2.3. This introduces the notation that $P_{CF_4}(\epsilon|n=0)$ is the probability distribution of $\epsilon = \Delta U_{CF_4}$, conditional on the event that there are no overlaps; $p_{CF4}(n=0)$ is the associated marginal probability.

The intention is that $P_{CF_4}(\epsilon|n=0)$ should be more suitable in providing an accurate evaluation of the integral of eq 2.4 than is $P_{CF_4}(\epsilon)$ for the corresponding integral of eq 2.1. Both those integrals involve the exponential factor $\exp(\beta\epsilon)$ that makes their values crucially sensitive to high-ϵ behavior of those distributions. That sensitivity to low-probability features is the characteristic difficulty for evaluation of free energies, and is unavoidable as a general matter. The intention here then is that the conditioning is a device to exclude near-neighbor, high-ϵ contributions in the integral that must be evaluated; those near-neighbor, high-ϵ contributions are properly assessed by the packing and chemical contributions, the first two terms on the right-side of eq 2.4.

The remaining interactions in the conditioned sample are ones that are a sum of numerous contributions and less energetic than instances that are excluded. A Gaussian model then seems natural for the conditional distribution $P_{CF_4}(\epsilon|n=0)$. If a Gaussian approximation is assumed, eq 2.4 adopts the simple form

$$\mu_{CF_4}^{(ex)} \approx kT \ln p_{CF_4}(n=0) + \mu_{HC}^{(ex)} + \langle \Delta U_{CF_4}|n=0 \rangle + \frac{\beta}{2} \langle (\delta\Delta U_{CF_4})^2|n=0 \rangle \quad (2.5)$$

Of course[32]

$$\mu_{CF_4}^{(ex)} - kT \ln p_{CF_4}(n=0) - \mu_{HC}^{(ex)} = kT \ln \int P_{CF_4}(\epsilon|n=0) \exp(\beta\epsilon) \, d\epsilon \geq \langle \Delta U_{CF_4}|n=0 \rangle \quad (2.6)$$

In providing a *lower bound*, this mean-field approximation is different from traditional vdW treatments which typically provide upper bounds.

A further difference from vdW approximations is the contribution from the marginal probability $p_{CF_4}(n=0)$. The favorable contribution $kT \ln p_{CF_4}(n=0)$ corresponds to a free energy assessment for the solvent density that would have been excluded by the chosen HC model, but which is permitted in the problem of physical interest. This is the same as the contribution $kT \ln x_0$ in previous discussions of quasi-chemical theory. The present suggestion is thus a simple realization of quasi-chemical theory.[33] The challenging application of quasi-chemical methods to liquid water in ref 33 emphasized the important compensation between $\mu_{HC}^{(ex)}$ and $kT \ln p_{CF_4}(n=0)$, *packing* and *chemistry* contributions, respectively. In fact, the

(20) Pohorille, A.; Pratt, L. R. *J. Am. Chem. Soc.* **1990**, *112*, 5066−5074.
(21) Pratt, L. R.; Pohorille, A. *Proc. Nat. Acad. Sci. U.S.A.* **1992**, *89*, 2995−2999.
(22) Pratt, L. R.; Pohorille, A. *Conf. Proc.−Ital. Phys. Soc.* **1993**, *43*, 261−268.
(23) Hummer, G.; Garde, S.; García, A. E.; Pohorille, A.; Pratt, L. R. *Proc. Nat. Acad. Sci. U.S.A.* **1996**, *93*, 8951−8955.
(24) Gomez, M. A.; Pratt, L. R.; Hummer, G.; Garde, S. *J. Phys. Chem. B* **1999**, *103*, 3520−3523.
(25) Pratt, L. R.; Garde, S.; Hummer, G. *NATO Sci. Ser., Ser. C* **1999**, *529*, 407−420.
(26) Hummer, G.; Garde, S.; García, A. E.; Pratt, L. R. *Chem. Phys.* **2000**, *258*, 349−370.
(27) Pratt, L. R.; LaViolette, R. A.; Gomez, M. A.; Gentile, M. E. *J. Phys. Chem. B* **2001**, *105*, 11662−11668.
(28) Pratt, L. R.; Ashbaugh, H. S. *Phys. Rev. E* **2003**, *68*, 021505/1−021505/6.
(29) Pratt, L. R. *Annu. Rev. Phys. Chem.* **2002**, *53*, 409−436.
(30) Floris, F. M. *J. Phys. Chem. B* **2005**, *109*, 24061−24070.
(31) Ben-Amotz, D. *J. Chem. Phys.* **2005**, *123*, 184504.

(32) Hansen, J.-P.; McDonald, I. R. *Theory of Simple Liquids*; Academic Press: New York, 1976.
(33) Paliwal, A.; Asthagiri, D.; Pratt, L. R.; Ashbaugh, H. S.; Paulaitis, M. E. *J. Chem. Phys.* **2006**, *124*, 224502.
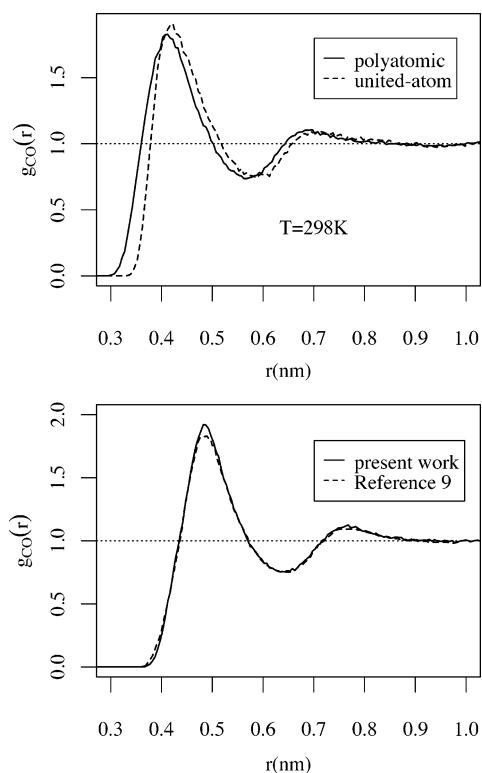
Figure 2. (Upper panel) Radial distribution function for oxygen (O) atoms relative to the carbon (C) atom of CF$_4$. (Lower panel) Corresponding results for C(CH$_3$)$_4$ except that the SPC model for water was used to permit comparison with results of previous work.[9]

present approach provides surprisingly realistic results for the free energy of liquid water.[34]

This theory is conceived on the basis of simulation data for the physical system, not for a reference model. This is an advantage because of the wide availability of simulation tools with realistic intermolecular interactions embedded. It is typically less easy to implement calculations for a reference model, perhaps even a HC model solute. Nevertheless, evaluation of the HC contribution of eq 2.4 is still required, and that might require a special purpose calculation.

Reliance on data for the physical system might be additionally advantageous for aqueous solutions at low pressure because some drying is expected for HC model solutes.[4,13] There are innocuous cases where the solute with realistic interactions does not exhibit noticeable drying, though drying is expected for the naturally corresponding HC model.[6,9] Where that difference is substantial, the present approach offers improved physical clarity by merely attempting to use statistical information parsimoniously, without attributing physical significance to intermediate results.

III. Methods

Reference 9 used the SPC model for water in studying hydration of neopentane, and we carried out that calculation for comparison. We used the same force field as ref 9 used for the realistic model of neopentane, and both SPC and SPC/E models for water. It might be anticipated that the SPC/E model predicts slightly more structured radial distribution functions and slightly better thermodynamic properties

(34) 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*; LA-UR-07−3494 Technical Report, Los Alamos National Laboratory: Los Alamos, NM, 2007.
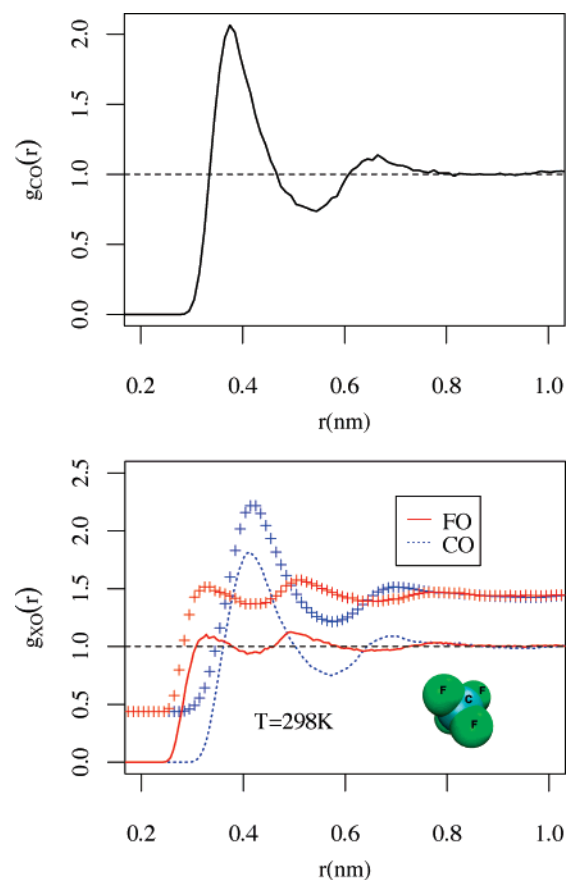
Figure 3. (Upper) Proximal FO radial distribution function extracted from the $T = 298$ K data for CF$_4$(aq). (Lower) Observed and reconstructed radial XO distribution functions (X = F or C). The crosses are the radial distribution functions reconstructed from the proximal distribution, elevated by 0.5 for clarity. The observed and reconstructed distribution functions are accurately the same.

sought here.[35] The polyatomic neopentane molecule was rigidly constrained in an optimum geometry. The united-atom potential energy model of CF$_4$ was that of ref 36; the CF$_4$ parameters for that all-atom model were those of ref 37.

We will focus principally on CF$_4$(aq) instead of C(CH$_3$)$_4$(aq) in what follows because of recent interest[2] but also because comparison of results with available thermodynamic data offers the possibility of helpful validation.[38] We note that C(CH$_3$)$_4$ is more soluble in water than is CF$_4$.[38] Neopentane is larger, but the contributions of attractive interactions are large as well.

The classical molecular dynamics simulations were performed with the NAMD 2.5[39] code. For the NPT simulations, the pressure was held constant at 1 bar using the Langevin piston method, and the temperature was controlled by Langevin dynamics applied to the oxygen atoms. The structure of each water molecule was constrained by the SHAKE[40] procedure.

For pressure control, the Langevin piston period was set at 200 fs and the decay constant was 100 fs. The temperature of the Langevin

(35) Paschek, D. *J. Chem. Phys.* **2004**, *120*, 6674−6690.
(36) Bonifácio, R. P.; Pádua, A. A. H.; Gomes, M. F. C. *J. Phys. Chem. B* **2001**, *105*, 8403−8409.
(37) Watkins, E. K.; Jorgensen, W. L. *J. Phys. Chem. A* **2001**, *105*, 4118−4125.
(38) Scharlin, P.; Battino, R.; Silla, E.; Tuñón, I.; Pascual-Ahuir, J. L. *Pure Appl. Chem.* **1998**, *70*, 1895−1904.
(39) Kale, L.; Skeel, R.; Bhandarkar, M.; Brunner, R.; Gursoy, A.; Krawetz, N.; Phillips, J.; Shinozaki, A.; Varadarajan, K.; Schulten, K. *J. Comp. Phys.* **1999**, *151*, 283−312.
(40) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comp. Phys.* **1977**, *23*, 327−341.
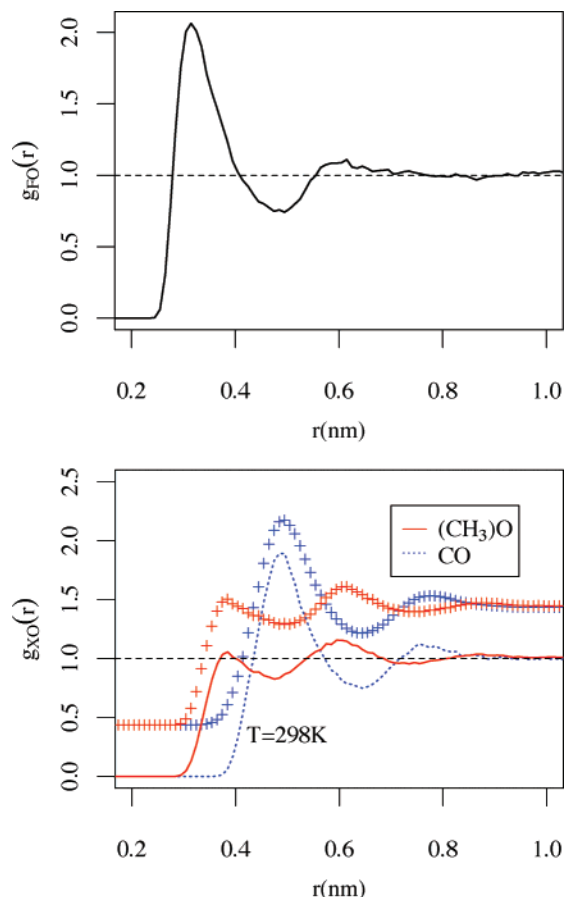
Figure 4. (Upper) Proximal C(methyl)O radial distribution function extracted from the C(CH₃)₄(aq) (neopentane) data at $T = 298$ K. (Lower) Observed and reconstructed radial XO distribution functions (X = C(methyl or C). The crosses are the radial distribution functions reconstructed from the proximal distribution, elevated by 0.5 for clarity. In this case, the observed CO distribution function has a slightly but perceptibly sharper principal peak than the CO distribution function reconstructed from the proximal distribution function.

piston was set equal to the temperature of the Langevin thermostat. The damping coefficient of the thermostat was set at 1 ps⁻¹.

The Lennard-Jones interaction was terminated at 11 Å by smoothly switching to zero starting at 10 Å. Electrostatic interactions were treated with the particle mesh Ewald method with a grid spacing of approximately 0.75 Å. The SPC/E water model[41] was used for both the united-atom and the polyatomic CF₄ models. Our choice of SPC/E for the united-atom model is consistent with the choice made by ref 36. It is known that these water models are among the more successful ones in describing the low-pressure equation of state for liquid water and also for studies of hydrophobic hydration.[35]

We constructed the system by placing a CF₄ molecule at the center of a box containing 512 water molecules. For convenience in analysis, we fixed the carbon atom of the CF₄ molecule to the center of the simulation cell. The starting system was equilibrated for 2 ns. During this time, fluctuations of instantaneous temperature, pressure, and density were monitored. After equilibration, a 2 ns production run was used for the united-atom CF₄ model. Configurations were stored every 200 fs, and the same period was used to accumulate statistics on the temperature, the pressure, and the volume. The density of water was calculated from the mean volume of the simulation cell. For the rigid polyatomic model of CF₄, a 4 ns production run was carried out, with all other data collection and analysis procedures identical to that for the united-atom model.

(41) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91*, 6269–6271.
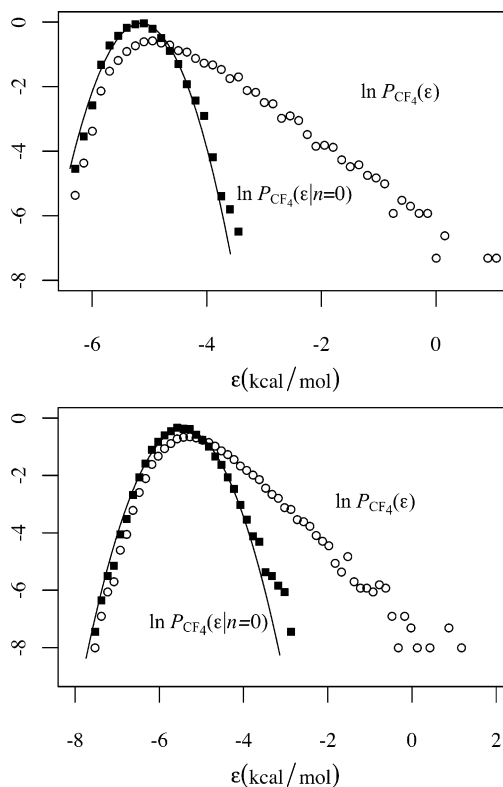
Figure 5. (Upper panel) Binding energy distributions for the united-atom model CF₄ in water at $T = 298$ K and $p = 1$ bar. The open circles are $P_{CF_4}(\epsilon)$ and the filled squares are $P_{CF_4}(\epsilon | n = 0)$ with $\lambda = 0.3789$ nm, the smallest r at which $g_{CO}(r) = 1$. (Lower panel) Corresponding results for the polyatomic model of CF₄. Here $n = 0$ indicates that in the conditional sample all water O-atoms are located a distance greater than 0.278 nm from any F-atom.
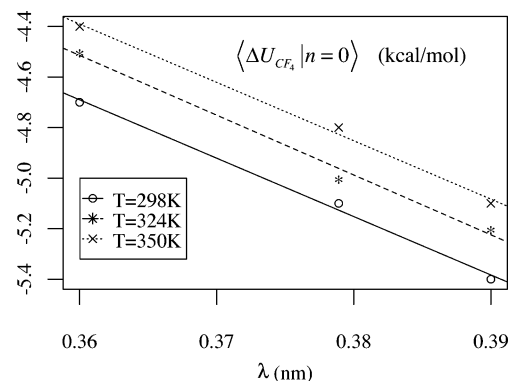
Figure 6. For the united-atom model of CF₄(aq), dependence of conditional mean binding energy with radius λ of the conditioning sphere. More aggressive conditioning, eliminating more near-neighbors, lowers the conditional mean binding energies, i.e., makes binding energies more favorable. This expected behavior is not consistent with attractive solute−solvent interactions drawing a preformed interface into closer contact.

For the polyatomic model, electrostatic and nonelectrostatic contributions to the binding energy need to be assessed. For electrostatic interactions, we implemented the generalized reaction field method (GRF),[42−44] and also used our earlier[45] Ewald simulation approach. As

(42) Hummer, G.; Soumpasis, D. M.; Neumann, M. *Mol. Phys.* **1992**, *77*, 769–785.
(43) Hummer, G.; Soumpasis, D. M.; Neumann, M. *J. Phys. Condens. Matter* **1994**, A141–A144.
(44) Hummer, G.; Pratt, L. R.; García, A. E. *J. Phys. Chem. A* **1998**, *102*, 7885–7895.
(45) Asthagiri, D.; Pratt, L. R.; Ashbaugh, H. S. *J. Chem. Phys.* **2003**, *119*, 2702–2708.

Table 1. Results (eq 2.5) for the United-Atom Model of CF$_4$[a]

T (K)	λ (nm)	$\langle \Delta U_{CF_4} \| n = 0 \rangle$	$+ \beta/2 \langle (\delta \Delta U_{CF_4})^2 \| n = 0 \rangle$	$+ \mu_{HC}^{(ex)}$	$+ kT \ln p_{CF_4} (n = 0)$	$= \mu_{CF_4}^{(ex)}$
298	0.36	−4.7	+0.4	+8.2	−0.1	= 3.9
	0.3789	−5.1	+0.1	+9.6	−0.5	= 4.1
	0.39	−5.4	+0.1	+10.4	−1.0	= 4.1
324	0.36	−4.5	+0.4	+8.5	−0.1	= 4.3
	0.3789	−5.0	+0.2	+9.8	−0.5	= 4.5
	0.39	−5.2	+0.1	+10.7	−1.1	= 4.5
350	0.36	−4.4	+0.4	+8.8	−0.1	= 4.7
	0.3789	−4.8	+0.1	+10.2	−0.6	= 4.9
	0.39	−5.1	+0.1	+11.1	−1.2	= 5.0

[a] λ is the conditioning radius. After the marginal probability p_{CF_4} ($n = 0$) is estimated, all configurations with an O-atom within λ of the C-atom are excluded from subsequent analysis. Free energy contributions are in kcal/mol.
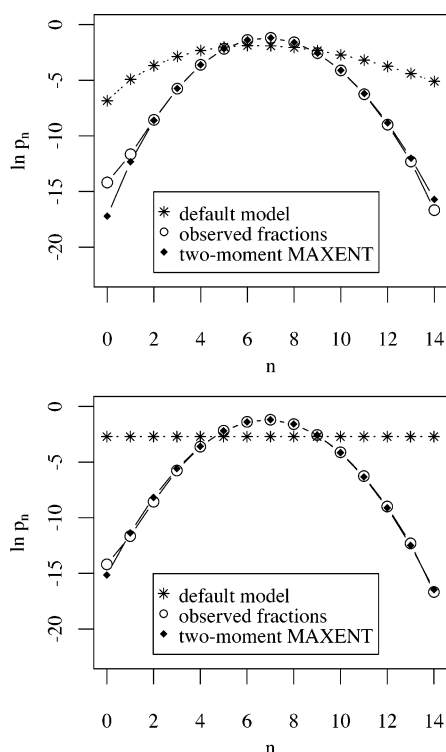
Figure 7. Observed distribution of occupancies n of the stencil shaped for the polyatomic hard-core model of CF$_4$; $T = 300$ K. Here $-\mu_{HC}^{ex}/kT = \ln p_0 \approx -14.2$. A similar direct approach would not be practical for neopentane; more sophisticated efforts will be required for larger solutes, and that deserves subsequent work. The diamonds give the results for the two-moment maximum entropy model which uses the probabilities indicated by the stars as the default model:[25] (upper panel) Poisson default model; (lower panel) uniform probability default model.

expected the results were imperceptibly different for the handful of configurations that were tested. Thus we consider here only results with the more economical GRF approach.

The required hydration free energies for hard-sphere solutes were taken from ref 4. HC polyatomic CF$_4$ models were evaluated by trial insertion into liquid water;[4] details of those results are noted as they arise in the discussion that follows.

As an alternative to the radial distribution function, the water structure near our polyatomic solutes was determined using *proximal* radial distribution functions.[5,6,46,47] In this characterization of structure, water density distortions are measured relative to the solute interaction site to which the individual water molecules are closest. The methods for evaluating the proximal radial distribution functions for a solute

composed of one type of interaction site—the present model neopentane is an example—are detailed in the references. In the case of a solute with more than one type of interaction site, such as CF$_4$, the size of the constituent sites must be taken into account.[48] The carbon of CF$_4$, however, is buried within fluorine atoms, so that no water molecules are closer to the carbon than to any other interaction sites. CF$_4$ can be effectively treated as a solute with just one type of proximal radial distribution function.

Traditional radial distribution functions can be reconstructed approximately as follows: the three-dimensional density of water sites of type α, denoted by αW, near an n-site solute in position $\mathscr{R}^n$ is assumed to be

$$\rho_{\alpha W}(\mathbf{r}|\mathscr{R}^n) \approx \rho_{\alpha W} g_{\alpha prox}(\min_i(|\mathbf{r} - \mathbf{r}_i|)) \qquad (3.7)$$

Here $\min_i(|\mathbf{r} - \mathbf{r}_i|)$ is the distance from $\mathbf{r}$ to the *nearest*, or *proximal*, site of the solute, $\mathbf{r}_i \in \mathscr{R}^n$. Traditional radial distribution functions can then be determined by contraction, taking the appropriate spherical average about the site of interest.

IV. Results and Discussion

Typical results for $g_{CO}(r)$—the radial distribution of oxygen atoms from the central carbon atom—are shown in Figure 2. For the case of CF$_4$, the differences in this characteristic between united-atom and polyatomic models are slight. These CF$_4$ results are qualitatively similar to the C(CH$_3$)$_4$ results, from both the present and the previous work.[9] These distribution functions are not qualitatively different from the measured Kr−O distribution functions for Kr(aq).[49] Judged by the location of the maximum of the measured radial distribution function,[49] Kr is about the same size as are the CF$_4$ model solutes studied here, and the solubility of Kr(aq) displays classic hydrophobic temperature dependence.[4] The maximum values of the distribution functions obtained here are slightly smaller than that measured for Kr−O, but the differences might not be significant. Since those data for the Kr(aq) system are not sensitive to pressure[50] it is reasonable to conclude that drying is not playing a significant role in the hydration structure of Kr(aq). This view is consistent with accurate theoretical results for hard-sphere solutes in water which indicate that maximum values of radial distribution in the neighborhood of 2 are distinct from the cases of larger-sized hard-sphere solutes where drying is well developed.[4]

(46) Ashbaugh, H. S.; Paulaitis, M. E. *J. Phys. Chem.* **1996**, *100*, 1900−1913.
(47) Garde, S.; Hummer, G.; García, A. E.; Pratt, L. R.; Paulaitis, M. E. *Phys. Rev. E* **1996**, *53*, R4310−R4313.

(48) Ashbaugh, H. S.; Paulaitis, M. E. *Ind. Eng. Chem. Res.* **2006**, *45*, 5531−5537.
(49) Filipponi, A.; Bowron, D. T.; Lobban, C.; Finney, J. L. *Phys. Rev. Lett.* **1997**, *79*, 1293−1296.
(50) Bowron, D. T.; Weigel, R.; Filipponi, A.; Roberts, M. A.; Finney, J. L. *Mol. Phys.* **2001**, *99*, 761−765.

Table 2. Results for the Polyatomic Model of CF₄, as in Table 1[a]

T (K)	$\langle \Delta U_{CF_4} | n = 0 \rangle$	$+ \beta/2 \langle (\delta \Delta U_{CF_4})^2 | n = 0 \rangle$	$+ \mu_{HC}^{(ex)}$	$+ kT \ln p_{CF_4}\ (n = 0)$	$= \mu_{CF_4}^{(ex)}$
298	−5.4	+0.3	+8.4	−0.3	= 3.0
325	−5.2	+0.3	+8.6	−0.4	= 3.3
350	−5.1	+0.3	+8.7	−0.4	= 3.5

[a] The experimental result at $T = 298$K is 3.1 kcal/mol.[38] For this model, the conditioning requires that all oxygen atoms be outside the volume defined by the distance $\lambda = 0.278$ nm from each F atom.
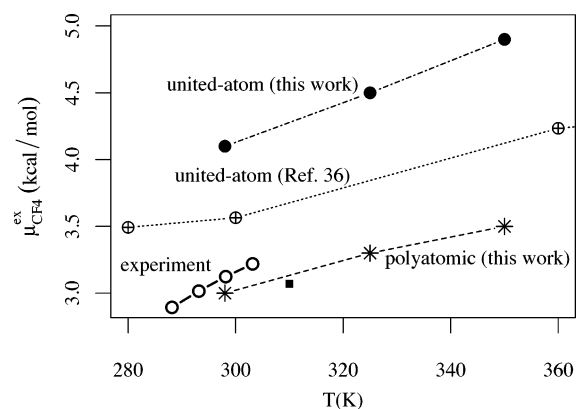
Figures 3 and 4 show the atom−atom radial distributions associated with the solute−water structures. The reconstructions of the traditional radial distribution functions on the basis of the proximal distribution functions are remarkably successful. The proximal distribution functions have a maximum value of about 2.0 and give no indication of drying. Indeed, the only noticeable differences between the observed and reconstructed radial distribution functions shown is associated with the central carbon of the neopentane solute. In that case, the observed principal maximum of the radial distribution function is slightly *sharper* than the principal maximum reconstructed on the basis of the proximal distribution function.

The conditioned and unconditioned binding-energy distributions for the CF₄ models considered are shown in Figure 5. The conditioning that eliminates close neighbors also predominantly eliminates unfavorable (high) binding energies, as was anticipated with Figure 1. This is consistent with the view that the solvent near-neighbors are pushed into repulsive interactions with the solute, that is, the solvent squeezes the solute. That is a standard view of hydrophobic effects[51] and contrasts with the idea that a preformed interface is pulled into contact with the solute by solute−solvent attractive interactions. This point is reinforced by examination (Figure 6) of the dependence of the conditional means for the united-atom case with increase of the conditioning radius λ: more aggressive conditioning, eliminating more near-neighbors, lowers the conditional mean binding energies, that is, makes binding energies more favorable. Again, this expected behavior contrasts with the idea that attractive solute−solvent interactions drawing a preformed interface into closer contact.

The observed linear trend in the conditional mean binding energies is in contrast to the net free energies, which vary marginally in the opposite direction, and become insensitive to the conditioning radius for the larger values (Table 1). This is the behavior that is sought for a satisfactory numerical result.

Evaluation of μ_{HC}^{ex} for the polyatomic HC models of CF₄ is not yet trivial. Here we used trial insertions of the tetrahedral five-hard-sphere structure into simulation configurations of liquid water, according to standard concepts and procedures.[4] Figure 7 shows some results of that effort.

The various results are accumulated in Table 2. The hard-core contribution is the largest of the several terms, and the Gaussian fluctuation contribution is small. We are justified therefore in concluding that this Gaussian model is physically accurate. These results thus support a mean-field picture of the hydration characteristics. The comparison with experiment (Figure 8) is satisfactory. The experimental results display conventional hydrophobic behavior in which the solubility decreases with increasing temperature for these moderate temperatures.

(51) Richards, F. M. *Sci. Am.* **1991**, *264*, 54.
(52) Scharlin, P.; Battino, R. *J. Chem. Eng. Data* **1995**, *40*, 167−169.
(53) Pohorille, A.; Wilson, M. A. *J. Chem. Phys.* **1996**, *104*, 3760−73.

Figure 8. Hydration free energies for CF₄(aq) as a function of temperature. The experimental results[52] are the open circles; The united-atom results used the solute model of ref 36, and circle-and-cross symbols plot the results from Table 5 of that reference. The filled square is the calculation of ref 53. The present results for the united-atom model are for $\lambda = 0.3789$ nm, the smallest distance at which $g_{CO}(\lambda) = 1$. The curvature of these graphs provides a heat capacity of hydration, but here three points are insufficient to support further consideration of that.

Figure 8 plots the results for the hydration free energies and compares with the available experimental results. The previous computational results for the united-atom model[36] are below the lower bound (eq 2.6) obtained here. The difference is not large. But in our experience working on this problem, this situation should not be considered unusual. Ambitious *full* calculations of such free energies can have statistical errors and systematic biases comparable to approximation errors of simple models. In fact, an alternative evaluation of this free energy on the basis of overlap of $P_{CF_4}(\epsilon)$ with the uncoupled (test) particle distribution $P_{CF_4}^{(0)}(\epsilon) = \langle \delta(\epsilon - \Delta U) \rangle_0$[18] was not successful.

V. Conclusions

Specialized consideration of nearest neighbors of CF₄(aq) or C(CH₃)₄(aq) changes the awkward binding energy distributions of Figures 1 and 5, with high-ϵ tails, into ones that are Gaussian to a useful approximation. This confirms that those unfavorable-binding-energy tails reflect few-body close solute−solvent encounters. Thus, the solvent nearest-neighbors are pushed by the medium into unfavorable interactions with the solute, in contrast to the alternative view that a preformed interface is pulled by the solute−solvent attractive interactions into contact with the solute.

The specialized consideration noted here, being based upon introduction of marginal and conditional distributions, is simple and general. The Gaussian statistical thermodynamic model, eq 2.5, avoids consideration of hypothetical drying-then-rewetting problems, and is physically reliable in these applications as judged by the size of the fluctuation contribution. The poly-atomic model of CF₄(aq) gives a satisfactory description of the hydrophobic solubility of that species. The proximal distributions evaluated here for polyatomic solutes accurately reconstruct the

observed distributions of water near these molecules which are nonspherical, of course. These results suggest that drying is not an essential consideration for the hydrophobic solubilities of CF$_4$, or of C(CH$_3$)$_4$ which is more soluble.