

Efficient computation of absolute free energies of binding by computer simulations. Application to the methane dimer in water

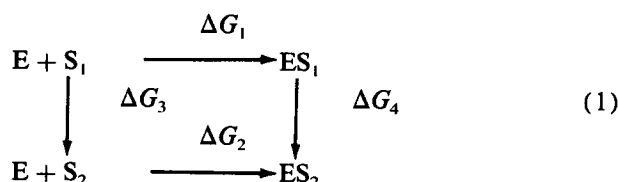
William L. Jorgensen, J. Kathleen Buckner, Stephane Boudon, and Julian Tirado-Rives
Department of Chemistry, Purdue University, West Lafayette, Indiana 47907

(Received 15 March 1988; accepted 13 May 1988)

An efficient procedure is noted for computing absolute free energies of binding for complexes in solution. Two series of computer simulations are required in which the substrate is annihilated in the solvent by itself and in the solvated complex. For illustration, the free energy of binding for two methane-like particles at their contact separation of 4 Å has been computed in TIP4P water. Though several alternatives are possible, in this case, Monte Carlo simulations were employed with statistical perturbation theory in the NPT ensemble at 25 °C and 1 atm. The results for the free energy of binding as well as for the potential of mean force are consistent with prior findings from the integral equation theory of Pratt and Chandler.

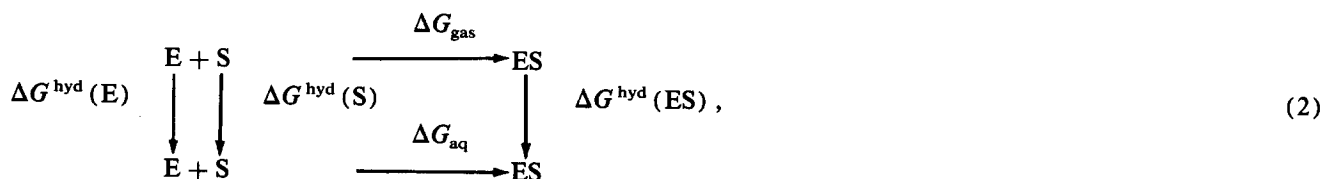
I. INTRODUCTION

In 1984, Tembe and McCammon pointed out that relative free energies of binding could be obtained from computer simulations by calculating the more accessible ΔG_3 and ΔG_4 in the illustrated thermodynamic cycle,



rather than ΔG_1 and ΔG_2 .¹ In the equations, E is a receptor, such as an enzyme, S_1 and S_2 are the substrates, and ES_1 and ES_2 are the corresponding complexes. They also noted that umbrella sampling methods or statistical perturbation theory (SPT) could be used to compute ΔG_3 and ΔG_4 for the mutations in solution. Though they illustrated the procedure for a simple Lennard-Jones model,¹ enthusiasm for the methodology was much enhanced by the observation that SPT could be used to obtain relative free energies of solution (ΔG_3) for organic molecules in water with high precision.² Since that time, the overall procedure, known as "thermodynamic cycle perturbation theory," has been used to compute relative free energies of binding for some impressive systems in aqueous solution including halide ions with a cryptand,³ benzamidine inhibitors with trypsin,⁴ phosphonate, and phosphoramidate inhibitors with thermolysin,⁵ a tripeptide with native and mutant subtilisin,⁶ and methotrexate with native and mutant dihydrofolate reductase.⁷ The accord with experimental binding data and the predictive abilities of the calculations have been remarkably good.³⁻⁷

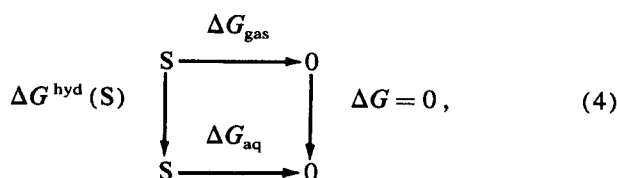
In comparison, little work has been done on the more difficult problem of obtaining absolute free energies of binding in solution. From the thermodynamic cycle below



where the labels are for aqueous solutions, the free energy of binding in water, ΔG_{aq} , can be expressed by

$$\Delta G_{\text{aq}} = \Delta G_{\text{gas}} + \Delta G^{\text{hyd}}(ES) - \Delta G^{\text{hyd}}(E) - \Delta G^{\text{hyd}}(S) \quad (3)$$

in terms of the free energy difference in the gas phase and the free energies of hydration for the reactants and complex. In turn, the free energies of hydration are related to hypothetical processes in which the substrate is made to vanish as expressed in the following thermodynamic cycle and Eq. (5):



$$\Delta G^{\text{hyd}}(S) = \Delta G_{\text{gas}}(S \rightarrow 0) - \Delta G_{\text{aq}}(S \rightarrow 0). \quad (5)$$

Consequently, the absolute binding energy for ES can be obtained via Eq. (3) upon computation of ΔG_{gas} and from simulations in which E, S, and ES are annihilated in the ideal gas phase and in solution. For the simple case where any of the species, $X = E, S$, or ES , has no internal degrees of freedom, then $\Delta G_{\text{gas}}(X \rightarrow 0) = 0$ and $\Delta G^{\text{hyd}}(X) = -\Delta G_{\text{aq}}(X \rightarrow 0)$. This procedure has recently been applied by Cieplak and Kollman to compute the absolute free energies of binding for nucleic acid base pairs, A-T and G-C, in aqueous solution.⁸ They used molecular dynamics simulations with SPT to compute all of the free energy changes on the right-hand side of Eq. (3) including the mutations of the base pairs and separated bases to nothing in water.

It is pointed out here that free energies of binding in solution can be computed efficiently in a way that avoids the

demanding calculations of $\Delta G_{\text{aq}}(\text{ES} \rightarrow 0)$ and $\Delta G_{\text{aq}}(\text{E} \rightarrow 0)$ and all of the gas-phase calculations. Combining Eqs. (3) and (5) yields Eqs. (6) and (7):

$$\Delta G_{\text{aq}} = \Delta G_{\text{gas}} + \Delta \Delta G_{\text{gas}}^0 + \Delta G_{\text{aq}}(\text{S} \rightarrow 0) + \Delta G_{\text{aq}}(\text{E} \rightarrow 0) - \Delta G_{\text{aq}}(\text{ES} \rightarrow 0), \quad (6)$$

$$\Delta \Delta G_{\text{gas}}^0 = \Delta G_{\text{gas}}(\text{ES} \rightarrow 0) - \Delta G_{\text{gas}}(\text{E} \rightarrow 0) - \Delta G_{\text{gas}}(\text{S} \rightarrow 0). \quad (7)$$

However, consideration of a trivial thermodynamic cycle shows $\Delta G_{\text{gas}} = -\Delta \Delta G_{\text{gas}}^0$. Furthermore, $\Delta G_{\text{aq}}(\text{ES} \rightarrow 0)$ can be computed in two stages by first mutating ES to E, as represented by

$$\Delta G_{\text{aq}}(\text{ES} \rightarrow 0) = \Delta G_{\text{aq}}(\text{ES} \rightarrow \text{E}) + \Delta G_{\text{aq}}(\text{E} \rightarrow 0). \quad (8)$$

Substituting into Eq. (6) then yields

$$\Delta G_{\text{aq}} = \Delta G_{\text{aq}}(\text{S} \rightarrow 0) - \Delta G_{\text{aq}}(\text{ES} \rightarrow \text{E}) \quad (9)$$

which shows that only the two simpler mutations are required. They correspond to the substrate disappearing in solution by itself and in the solvated complex. Equation (9) can also be readily derived in retrospect by considering the processes:

$$\begin{array}{ll} \text{S} \rightarrow 0 & \Delta G_{\text{aq}}(\text{S} \rightarrow 0) \\ \text{E} \rightarrow \text{ES} & -\Delta G_{\text{aq}}(\text{ES} \rightarrow \text{E}). \end{array} \quad (10)$$

$$\text{E} + \text{S} \rightarrow \text{ES} \quad \Delta G_{\text{aq}}$$

However, the connection to Eq. (3) is no longer obvious.

To illustrate this procedure the results for a model calculation are reported here. Specifically, the absolute free energy of binding for two methane-like particles in TIP4P water has been computed at their contact separation of 4 Å. This distance was obtained from determination of the potential of mean force as a function of internuclear distance. The results support the viability of the procedure; however, for large substrates the method is still computationally demanding, though much less so than application of Eq. (3). In addition, the present results for $(\text{CH}_4)_2$ in water are consistent with findings from prior simulation work and the theory of the hydrophobic effect developed by Pratt and Chandler.⁹⁻¹¹

II. COMPUTATIONAL PROCEDURE

The potential functions for methane and water were taken from previous work that showed their appropriateness for describing the thermodynamic and structural properties of the pure liquids.^{12,13} Specifically, methane is represented as a single Lennard-Jones particle with $\sigma_M = 3.730$ Å and $\epsilon_M = 0.2940$ kcal/mol,¹² and the TIP4P model is used for water.¹³ The potential functions are pair-wise additive and the only interaction between methane and water is a Lennard-Jones term based on the C–O distance with $\sigma_{MW} = 3.430$ Å and $\epsilon_{MW} = 0.2135$ kcal/mol.

To begin, the free energy profile or potential of mean force (PMF) was determined for the separation of two methane particles in water. This was effected by a series of Monte Carlo (MC) statistical mechanics simulations in the isothermal–isobaric ensemble at 25 °C and 1 atm. The sys-

tem consisted of the two methanes plus 250 TIP4P water molecules in a periodic cell with dimensions of $\sim 17 \times 17 \times 26$ Å. Statistical perturbation theory was used to compute the free energy changes as the methanes were perturbed apart in increments of typically 0.25 Å. This was done by perturbing both methanes in and out by 0.125 Å with the interatomic axis kept parallel to the long axis of the periodic cell. Thus, 0.5 Å could be covered in one simulation. Each simulation consisted of an equilibration phase for at least 5×10^5 configurations followed by averaging over 2×10^6 configurations. The overall procedure was the same as used previously in the determination of the PMF for $(\text{CH}_3)_3\text{C}^+ \cdots \text{Cl}^-$ in water.¹⁴ The earlier study can be consulted for additional details. However, two points should be noted here. The intermolecular interactions were smoothly reduced to zero at 8.5 Å based on the C–O and O–O distances by quadratic feathering between 8.0 and 8.5 Å.¹⁴ Also, these simulations only provide the relative free energy as a function of the intermolecular distance. The PMF can be zeroed by computing the absolute free energy of interaction for one point, e.g., via Eq. (9).

The calculations for the PMF determined that the optimal separation for the contact methane dimer is 4 Å in TIP4P water. The absolute free energy of binding for this point was then determined by application of Eq. (9). Namely, a series of MC simulations was again carried out using SPT to calculate the free energy change as one of the two methanes in the contact dimer was gradually annihilated. The same system setup and procedures were employed in all respects as for the PMF computations. In this case, the Lennard-Jones parameters for the disappearing methane were scaled to zero as λ went from 1 to 0 in

$$\sigma_i = \lambda_i \sigma_M, \quad (11)$$

$$\epsilon_i = \lambda_i \epsilon_M. \quad (12)$$

It should also be noted that the disappearing methane was “reeled in” to the other methane from 4 Å at $\lambda = 1$ to 2.5 Å at $\lambda = 0$. In preliminary calculations, this procedure yielded more stable free energy changes in the SPT calculations than when the disappearing particle was left in place. The latter alternative may lead to some screening of the solvation of the remaining solute even for small λ that could produce slower convergence.

The other term needed for Eq. (9) is the free energy change for annihilating an isolated methane particle in TIP4P water. The results of this calculation have been reported elsewhere and involved the same procedures as described above.¹⁵ The computed free energy change of -2.27 ± 0.3 kcal/mol is in accord with the experimental value of -2.005 kcal/mol.^{15,16} In this prototypical case, the incremental free energy changes were computed in both directions, i.e., $\lambda_i \rightarrow \lambda_j$ and $\lambda_i \leftarrow \lambda_j$; the hysteresis from the two series of calculations then yielded the rough estimate of ± 0.3 kcal/mol for the statistical uncertainty. Overall, we have performed this mutation three times; based on the three independent results, the best estimate of the statistical uncertainty ($\pm 1\sigma$) for the computed free energy of hydration of methane is ± 0.24 kcal/mol. The earlier study also indicated that $\Delta\lambda$ values of 0.05–0.10 were satisfactory, so these

increments were used in the present calculations for $(\text{CH}_4)_2 \rightarrow \text{CH}_4$. However, in this case the perturbations were not checked in both directions for computational efficiency. Even so, seven separate simulations were performed and required ~ 14 days on a Gould 32/8750 computer which is 5–6 times faster than a VAX 11/780. The PMF calculation entailed ten more simulations for an additional 20 days on the Gould computer.

III. RESULTS AND DISCUSSION

A. Free energy of binding

The results for the mutations of $(\text{CH}_4)_2 \rightarrow \text{CH}_4$ and $\text{CH}_4 \rightarrow 0$ in TIP4P water are summarized in Fig. 1. The net free energy changes for the two processes are -1.85 and -2.27 kcal/mol, respectively. Combination via Eq. (9) then gives a net binding free energy of -0.42 kcal/mol at a separation of 4 \AA . Assuming that the statistical uncertainties for the two mutations are the same, the net uncertainty in the free energy of binding is ± 0.34 kcal/mol. It should be noted that the binding free energy includes the Lennard-Jones attraction of -0.28 kcal/mol for the two methanes separated by 4 \AA .

As seen in Fig. 1, variation in the free energy is relatively large near $\lambda = 1$. This is reasonable since free energies of hydration for hydrocarbons have been found to be proportional to the solvent-accessible surface area which varies with λ^2 [Eq. (11)].^{17,18} The enhanced decline in free energy for $(\text{CH}_4)_2 \rightarrow \text{CH}_4$ near $\lambda = 1$ and the diminished slope below $\lambda = 0.8$ in comparison to the $\text{CH}_4 \rightarrow 0$ results undoubtedly reflects the effects of the procedure by which the disappearing methane was simultaneously shrunk and pulled in to the other methane. In view of the consequent importance of the $\lambda = 1.0$ – 0.8 region, the free energy changes were computed in both 0.05 and 0.1 increments in this range. For the mutation of the dimer, the net free energy changes were -0.96 and -1.20 kcal/mol for $\lambda = 1.0$ – 0.8 with

$\Delta\lambda = 0.05$ and 0.10 , respectively. The average of these values was used in Fig. 1.

The computed free energy of binding, -0.4 ± 0.3 kcal/mol, is reasonable based on comparisons with results from the integral equation theory of Pratt and Chandler.^{9–11,19} The original theory,⁹ which did not include the attractive solute–solvent interactions, was applied by Pangali *et al.* to determine a PMF for two Lennard-Jones particles in water for comparison with their molecular dynamics results.¹⁹ It turns out that the present σ_{MW} and ϵ_{MW} (3.43 \AA and 0.21 kcal/mol) are close to their values of 3.43 \AA and 0.15 kcal/mol. With use of the experimental oxygen–oxygen radial distribution function (RDF) for water, the Pratt–Chandler theory then gave an absolute binding energy of ~ -0.45 kcal/mol at the contact separation.¹⁹ The theory was subsequently refined to include the attractive solute–solvent forces.^{10,11} This modification was shown to have little effect on the results of Pangali *et al.*; the first peak in the solute–solute RDF was lowered by less than 10% which corresponds to weakening the free energy of binding to ~ -0.40 kcal/mol, consistent with the present result. Pratt and Chandler also considered alternative Lennard-Jones parameters with a stronger solute–solvent attraction, $\sigma_{MW} = 3.2 \text{ \AA}$ and $\epsilon_{MW} = 0.33$ kcal/mol. In this case, inclusion of the attractive forces has a greater effect on the results and the net free energy of binding at the contact separation is near zero.¹¹ Though the results from the present simulations and the Pratt–Chandler theory are in accord, it should be noted that differences in the water models can also have some quantitative effect.²⁰

A more empirical approach to an estimate of the free energy of binding is also possible. Following the spirit of the correlation between solvent-accessible surface area (SASA) and free energy of hydration,^{17,18} the SASA is plotted vs intermolecular separation for the methane dimer in Fig. 2. For the purposes of this plot, the diameter of methane has been taken as $2^{1/6} \sigma_M$, which corresponds to the energy minimum

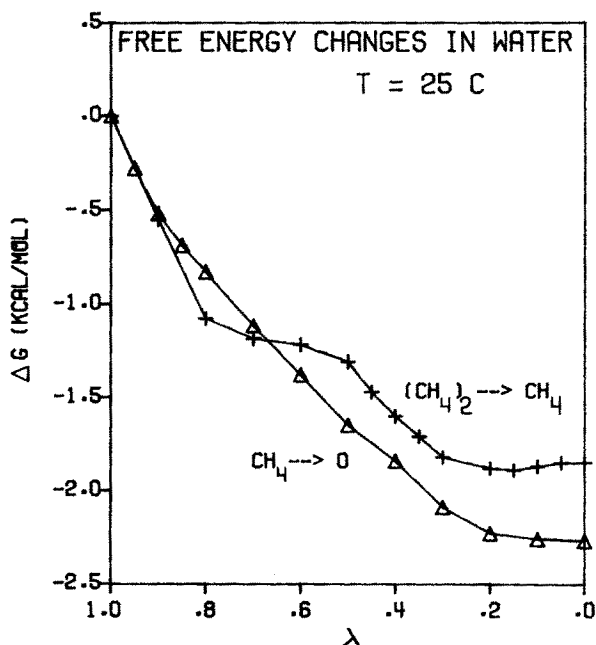


FIG. 1. Changes in free energy for the mutations of $(\text{CH}_4)_2 \rightarrow \text{CH}_4$ and $\text{CH}_4 \rightarrow 0$ in TIP4P water. One CH_4 disappears as λ goes from 1 to 0.

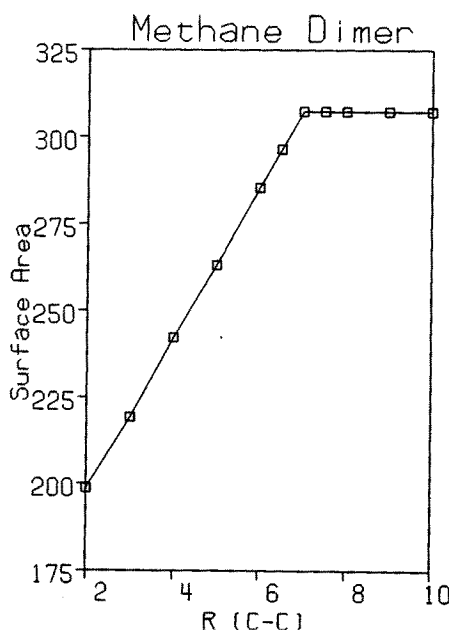


FIG. 2. Solvent-accessible surface area (\AA^2) for the methane dimer as a function of the C–C distance (\AA).

for the Lennard-Jones potential, and the diameter of the probe sphere is 2.8 \AA , a standard assignment for water. In passing from infinite separation to 4 \AA , the SASA is reduced by 64 \AA^2 . Since the SASA for an isolated methane is 153 \AA^2 , the free energy benefit for reducing the SASA at 4 \AA should be $(64/153) \times 2.005 \text{ kcal/mol} = 0.84 \text{ kcal/mol}$. The methane-methane interaction of 0.28 kcal/mol needs to be added to this to yield an estimate of -1.1 kcal/mol for the free energy of binding. A more general correlation for surface areas and experimental free energies of hydration for *n*-alkanes yields a slope of $0.009 \text{ kcal/mol \AA}^2$ for methyl groups.²¹ Extension to the present problem would then give a net free energy of binding for $(\text{CH}_4)_2$ of $-64 \times 0.009 - 0.28 = -0.86 \text{ kcal/mol}$. These results are clearly crude and affected by several assumptions including the choice of particle diameters and the linear dependence of the free energy of hydration on the surface area. Nevertheless, they are consistent with the true free energy of binding not being lower than $\sim -1 \text{ kcal/mol}$.

B. Potential of mean force

The results for the PMF for the two methane-like particles in TIP4P water are summarized in Fig. 3. Separations from 3.5 to 7.5 \AA have been covered; extension much beyond the upper limit might yield artifactual results owing to the finite system size and cutoffs of the potential functions. The curve has been shifted so that the value at $R(\text{C-C}) = 4 \text{ \AA}$ is -0.42 kcal/mol , as computed above. The statistical uncertainty ($\pm 1\sigma$) for each point relative to the adjacent points is ± 0.05 – $\pm 0.10 \text{ kcal/mol}$ based on the fluctuations in the individual Monte Carlo simulations. This means that the uncertainties at the ends of the curve starting in the middle are ± 0.15 – $\pm 0.30 \text{ kcal/mol}$, though these are probably lower bounds. Nevertheless, the present results are in essentially perfect agreement with the findings of Pangali *et al.* from their molecular dynamics simulations with importance sampling in ST2 water, and with the results from the Pratt-Chandler theory.¹⁹ All studies show the two minima corre-

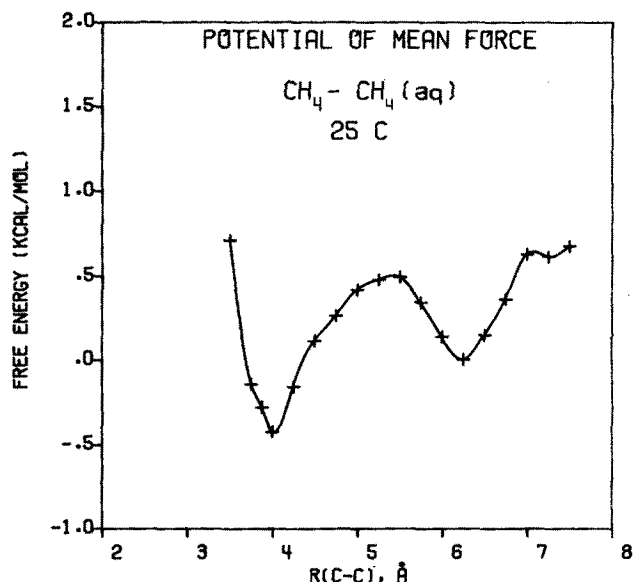


FIG. 3. Computed potential of mean force for $(\text{CH}_4)_2$ in TIP4P water.

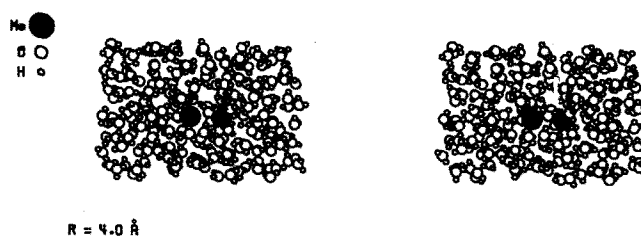


FIG. 4. Stereo view of one configuration of the contact methane dimer at 4 \AA separation in water. The periodic cell is shown, though water molecules more than 4.5 \AA in front of the solutes have been removed for clarity.

sponding to the contact (Fig. 4) and solvent-separated (Fig. 5) dimers. Furthermore, the barrier between the contact and solvent-separated forms is $\sim 1 \text{ kcal/mol}$ in each case and the solvent-separated species is found to be at higher free energy by 0.25 – 0.4 kcal/mol . However, upon integration over the volume element, the solvent-separated form is actually predicted to have the greater population. The only differences in the results are that the positions of the minima are shifted out by 0.2 and 0.7 \AA in the prior results¹⁹ relative to Fig. 3. This largely reflects the use of $\sigma_M = 4.12 \text{ \AA}$ in the earlier work vs 3.73 \AA here.

A PMF for $(\text{CH}_4)_2$ in water has also been determined by Ravishanker *et al.* from Monte Carlo simulations with importance sampling.²² They used potential functions derived from quantum mechanical calculations for all of the intermolecular interactions including the MCY model for water. Their results also show the two minima, though the intervening barrier is smaller and the solvent-separated form is now lower in free energy than the contact dimer by $\sim 0.2 \text{ kcal/mol}$. The latter discrepancy with the other calculations is traceable to a significantly more attractive, optimal methane-water interaction from the potential functions of Ravishanker *et al.*²² Recently, Watanabe and Andersen carried out lengthy molecular dynamics calculations for krypton ($\sigma = 3.6 \text{ \AA}$, $\epsilon = 0.401 \text{ kcal/mol}$) in water and again concluded that solvent-separated pairs are more probable than the contact dimer.²³ Thus, a major point of agreement among the calculations is that the solvent-separated structures may play a more significant role in the hydration of nonpolar groups than previously expected based on traditional notions about hydrophobic effects. Interestingly, Wallqvist and Berne have now found a preference for an analogous solvent-separated structure in molecular dynamics simulations for methane near a hydrophobic wall.²⁴

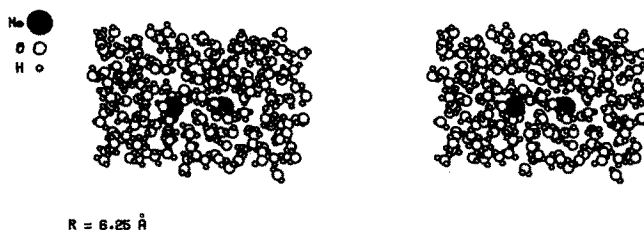


FIG. 5. Stereo view of one configuration of the solvent separated methane dimer at a C-C distance of 6.25 \AA . Note that a sheet of water separates the solutes rather than featuring one particularly unique water molecule as in some ion pairs (cf. Ref. 14). Other details as in Fig. 4.

Thus, the phenomenon is not restricted to the simplest case of the methane dimer in water.

IV. CONCLUSION

A straightforward procedure has been presented for obtaining absolute free energies of binding for complexes in solution. The method can be used in conjunction with Monte Carlo or molecular dynamics simulations and the requisite free energy changes can be calculated via several means including statistical perturbation theory and importance sampling. The procedure was applied in the present MC-SPT study to compute the absolute free energy of binding for two methane-like particles at their contact separation in the TIP4P model of water. The estimated statistical uncertainty in this case of ± 0.3 – ± 0.4 kcal/mol indicates the level of precision that may be expected. Of course, in cases where the binding is stronger, the statistical uncertainty should be a smaller fraction of the free energy of binding. However, the treatment of large solutes may result in greater error limits unless longer simulations are undertaken.

The free energy of binding for $(\text{CH}_4)_2$ in water and the potential of mean force computed here were found to be in remarkable agreement with previous results from a molecular dynamics study and the Pratt–Chandler theory.^{9–11,19} In view of the proven quality of the potential functions used in the present study,^{12,13,15} the results provide further support for the exceptional predictive abilities of the Pratt–Chandler theory in treating problems associated with the energetics and structure of non-polar solutes in aqueous solution.^{9–11,19,25}

ACKNOWLEDGMENTS

Gratitude is expressed to the National Institutes of Health and the National Science Foundation for support of this research.

- ¹B. L. Tempe and J. A. McCammon, *Comput. Chem.* **8**, 281 (1984).
- ²W. L. Jorgensen and C. Ravimohan, *J. Chem. Phys.* **83**, 3050 (1985).
- ³T. P. Lybrand, J. A. McCammon, and G. Wipff, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 833 (1986).
- ⁴C. F. Wong and J. A. McCammon, *J. Am. Chem. Soc.* **108**, 3830 (1986).
- ⁵P. A. Bash, U. C. Singh, F. K. Brown, R. Langridge, and P. A. Kollman, *Science* **235**, 574 (1987).
- ⁶S. N. Rao, U. C. Singh, P. A. Bash, and P. A. Kollman, *Nature* **328**, 551 (1987).
- ⁷U. C. Singh (to be published).
- ⁸P. Cieplak and P. A. Kollman, *J. Am. Chem. Soc.* **110**, 3734 (1988).
- ⁹L. R. Pratt and D. Chandler, *J. Chem. Phys.* **67**, 3683 (1977).
- ¹⁰L. R. Pratt and D. Chandler, *J. Chem. Phys.* **73**, 3430 (1980).
- ¹¹L. R. Pratt and D. Chandler, *J. Chem. Phys.* **73**, 3434 (1980).
- ¹²W. L. Jorgensen, J. D. Madura, and C. J. Swenson, *J. Am. Chem. Soc.* **106**, 6638 (1984).
- ¹³W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, *J. Chem. Phys.* **79**, 926 (1983); W. L. Jorgensen and J. D. Madura, *Mol. Phys.* **56**, 1381 (1985).
- ¹⁴W. L. Jorgensen, J. K. Buckner, S. E. Huston, and P. J. Rossky, *J. Am. Chem. Soc.* **109**, 1891 (1987).
- ¹⁵W. L. Jorgensen, J. F. Blake, and J. K. Buckner, *Chem. Phys.* (in press).
- ¹⁶A. Ben-Naim, *J. Chem. Phys.* **81**, 2016 (1984).
- ¹⁷R. B. Hermann, *J. Phys. Chem.* **76**, 2754 (1972).
- ¹⁸W. L. Jorgensen, J. Gao, and C. Ravimohan, *J. Phys. Chem.* **89**, 3470 (1985).
- ¹⁹C. Pangali, M. Rao, and B. J. Berne, *J. Chem. Phys.* **71**, 2975 (1979).
- ²⁰A. Tani, *Mol. Phys.* **51**, 161 (1984).
- ²¹S. Boudon, J. Tirado-Rives, and W. L. Jorgensen (to be published).
- ²²G. Ravishanker, M. Mezei, and D. L. Beveridge, *Faraday Symp. Chem. Soc.* **17**, 79 (1982).
- ²³K. Watanabe and H. C. Andersen, *J. Phys. Chem.* **90**, 795 (1986).
- ²⁴A. Wallqvist and B. J. Berne, *Chem. Phys. Lett.* **145**, 26 (1988).
- ²⁵W. L. Jorgensen, *J. Phys. Chem.* **87**, 5304 (1983); W. L. Jorgensen and J. K. Buckner, *ibid.* **91**, 6083 (1987).