The Binomial Cell Model of Hydrophobic Solvation

V. V. Alexandrovsky,[†] M. V. Basilevsky,*[,†] I. V. Leontyev,[†] M. A. Mazo,[‡] and V. B. Sulimov*[,‡]

Karpov Institute of Physical Chemistry, ul. Vorontsovo Pole 10, Moscow 105064, Russia, and Department of Quantum Chemistry, Algodign, LLC, Bolshaya Sadovaya 8, 123379 Moscow, Russia

Received: October 27, 2003; In Final Form: May 25, 2004

On the basis of the cell model of dense fluids, we derive the binomial distribution law for a number of solvent particles occupying a given void of excluded volume (a cavity) which arises in a bulk solvent as a fluctuation. It is inserted as a default distribution in the information theory approach (Hummer, G.; Garde, S.; Garcia, A. E.; Paulaitis, M. E.; Pratt, L. R. *J. Phys. Chem. B* **1998**, *102*, 10469) for treating the thermodynamics of cavitation; the imaginary process is considered as a component of the total solvation process. Computations of cavitation free energies, entropies, and enthalpies for 11 hydrocarbons in water solvent are compared with results of the computer simulation of these properties reported by E. Gallicchio et al. (Gallicchio, E.; Kubo, M. M.; Levy, R. *J. Phys. Chem. B* **2000**, *104*, 6271). A similar analysis of cavitation for model spherical solutes in water in a wide range of cavity radii is also provided. Linear correlation between the cavitation free energy and the cavity volume is observed and justified. The present binomial approach for treating cavitation effects efficiently covers a large range of cavity volumes v's (10^2 Å^3 < v <10^3 Å^3).

1. Introduction

The theory of hydrophobic interaction in a water solution is the background for many applications in chemistry and biology. Customarily, it is associated with an idealized concept of cavitation (i.e., creating, as a first step of solvation, a cavity of excluded volume in the bulk solvent, in which the solute particle is accommodated at the next step).

Among the components constituting the total solvation phenomenon, the cavitation effect seems to be the most obscure one. Contrary to ordinary potential-like intermolecular interactions, reasonably well-described in terms of electrostatic and Lennard-Jones (LJ) forces, the cavitation free energy represents a complicated, many-particle collective effect that is mainly responsible for entropy changes accompanying solvation processes. Because of the nonlocal and nonpotential character of underlying structural changes in a solvent, physical understanding is not complete, and its computation represents a serious problem.

For a solute molecule occupying a cavity of excluded volume in a solvent, the cavitation free energy change is expressed as[1,2]

$$\Delta G_{\mathrm{cav}} = -k_{\mathrm{B}}T \ln P_0 \tag{1}$$

where P_0 is the probability of observing no solvent particles within the cavity. Earlier work was aimed at modeling this probability.[1−4] Recent progress was achieved[5−8] by considering a new idea of treating the whole probability distribution $\{P_m\}$ in which each P_m represents a probability of finding exactly m solvent particles in a cavity. Then, P_0 is obtained as P_m with $m = 0$.

The following prescription has been advanced for quantitatively describing $\{P_m\}$.[5−8] In the expression

$$P_m = P_m^{(0)} \mathrm{e}^{(-\lambda_0 - \lambda_1 m - \lambda_2 m^2)} \tag{2}$$

the primary (default) distribution $P_m^{(0)}$ is not exactly specified: It is supposed to represent simplistic models. The second exponential factor refines this simple description by adjusting three constants (λ_0, λ_1, λ_2) so that the three lowest moments of the true distribution $\{P_m\}$ are exactly fitted. Fitting is based on equations defining these three moments

$$\langle 1 \rangle = \sum_{m=0}^{N} P_m$$

$$\langle m \rangle = \sum_{m=0}^{N} m P_m$$

$$\langle m^2 \rangle = \sum_{m=0}^{N} m^2 P_m \tag{3}$$

where the left-hand sides are available either from experimentation or from a numerical molecular level simulation.

Although several default distributions $P_m^{(0)}$ have been discussed,[6−9] only the simplest flat model

$$P_m^{(0)} = \frac{1}{N+1} \quad (m \le m_{\max}) \qquad P_m^{(0)} = 0 \quad (m > m_{\max}) \tag{4}$$

was investigated systematically in a number of applications.[6,7,10−12] Usually, the upper bound $m_{\max}$ of m is not exactly specified in the literature, so it seems to be considered as a sufficiently large number providing convergence of sums in eq 3. Equation 2 then yields the discrete Gaussian distribution P_m, truncated from below ($m \ge 0$).

* Corresponding authors. E-mail: basil@cc.nifhi.ac.ru (M.V.B.), vladimir.sulimov@algodign.com (V.B.S.).
[†] Karpov Institute of Physical Chemistry.
[‡] Algodign, LLC.

10.1021/jp031189e CCC: $27.50
Published on Web 09/10/2004

Equations 2−4 were successfully implemented for the interpretation, at a qualitative level, of several effects accompanying solvation of hydrophobic solutes in water.[5−12] However, a systematic deficiency of the flat model (eq 4) has been observed.[6,8,9] This approach predicts an incorrect asymptotic behavior in the limit when the cavity size becomes much larger than the size of the solvent (water) particles. In this case, the flexibility of the procedure is exhausted by fitting exact values of a position and a section of the real distribution that is indeed insufficient when the desired point $m = 0$ corresponds to a far tail of $\{P_m\}$, remote from the peak maximum, such that $(\langle m \rangle/\sigma) \gg 1$, where σ^2 denotes the distribution variance. This drawback was clearly visualized by comparing the results of the flat default model with numerical simulations of cavitation free energies for water solutions for a series of hydrocarbons consisting of 1−6 carbon atoms.[13]

Creation of very large cavities (with radius $R \gg 10$ Å for the spherical case) obeys a macroscopic law, according to which the thermodynamics of cavitation can be treated in terms of a continuum (hydrostatic) solvent model.[14] This represents the extreme case opposite to the limit of small cavities as suggested by eqs 2−4. Several interpolation procedures connecting these two extremes have been developed recently.[14−16]

In the present work, we considered another analytical model of water fluid as a candidate for default distribution $P_m^{(0)}$. This is the binomial distribution which can be derived from a hypothesis of a cell structure of bulk water at ambient conditions. Within this approach, the simulation results for ΔG_{cav} reported in ref 13 are reasonably well reproduced. This success is gained, at least partly, from adjustable parameters inherent to the binomial default model. The additional flexibility, as compared to the parameterless flat model, allowed for successful fitting of the simulation data. The present method also provides a natural separation of the cavitation free energy into its enthalpy and entropy components. We believe that, with parameter values established in this study, our treatment can be extended for describing other hydrophobic solutes in water, even when they are too bulky to be involved in a straightforward simulation of the cavitation effect. The above-mentioned macroscopic limit cannot be, however, covered in terms of the primitive binomial approximation, and a more sophisticated default model is required to treat very large cavities.

2. The Binomial Distribution as a Default Model

The present treatment is based on the cell model of the liquid state.[17−22] Let us consider a cell as a maximum-sized element of space within a solute cavity that can accommodate, at most, one solvent particle. This concept, borrowed from the earlier theories,[17,18,22] is well motivated for simply shaped solvent molecules (spheres). We do not enter the detailed description of a cell here, assuming that few parameters determining its shape and size can be introduced empirically. In the most primitive description, the cell volume is the only such parameter.

Thereby, we can consider the probability y of finding a single water molecule in a cell; then, $1 - y$ is the probability that this cell is empty. The quantity y is assumed to be the same for all cells, irrespective of their location inside the cavity. Under such assumptions, a consideration of a single cell is a Bernoulli dual test,[23] with the probability of success being y. For a cavity containing n cells, n Bernoulli tests, provided they are statistically independent, generate the binomial distribution law:[23,24]

$$P_m^{(0)}(y, n) = \binom{n}{m} y^m (1 - y)^{n - m} \tag{5}$$

where

$$\binom{n}{m}$$

are binomial coefficients. The upper boundary of m can be defined as

$$n = \mathrm{Int}\left(\frac{v}{v_0}\right) + 1 \tag{6}$$

where v and v_0 are volumes of the cavity and the cell, respectively, and the symbol Int denotes the "integer part of". Because of shape effects, this definition is not fully consistent. Alternative prescriptions defining n in terms of the first and second distribution moments are also considered below. Finally, following the slightly modified prescription (eq 2) suggested by the information theory of the cavitation effect,[5−8] we consider the distribution

$$P(y, n) = \binom{n}{m} y^m (1 - y)^{n-m} \mathrm{e}^{(-\lambda_0 - \lambda_1 m)} \tag{7}$$

It will be called the BE (binomial/exponential) distribution in the following text. Constants λ_0 and λ_1 are found by fitting the BE distribution (eq 7) in terms of the first two lines of eq 3. For $m = 0$, the result

$$P_0(y, n) = (1 - y)^n \, \mathrm{e}^{-\lambda_0} \tag{8}$$

represents the probability expression to be substituted in eq 1.

The modification introduced in eq 7 eliminates the Gaussian factor from eq 2 and allows the fitting procedure to be performed analytically. The complete treatment of the probability distribution (eq 2) will be discussed later. It results in the same expression (eq 8) for P_0 but suggests a different interpretation for n and λ_0.

3. Interpretation of Equation 8: The NVT Ensemble

Let us now discriminate between the pure cell model producing the binomial distribution and the corrected BE version represented by eq 7 that contains the additional exponential factor. The above reasoning, defining a cell for which $P_0^{(0)} + P_1^{(0)} = 1$ and $P_m^{(0)} = 0$ ($m > 1$), is fully consistent only when all molecules considered, both solutes and solvents, are hard particles. The intermolecular interaction potentials vanish in this case for any nonzero separation of neighboring hard particles. Therefore, the binomial model (eq 5) represents an assembly of noninteracting particles contained in different cells.

We consider now the NVT ensemble with a total number N of hard solvent particles and total volume V. The solution is dilute, such that considering explicitly a single solute molecule is sufficient. Hence, the values V and N must be counted per single solute particle; the extension to M solute particles simply requires multiplying them by M. The external boundary of this system is kept fixed. A network of cells covers its volume. By considering the cavitation process as a transfer of particles between different cells, we become convinced that it is not accompanied by any energy changes and only entropy changes

are produced. The computation illustrating this reasoning is presented in Appendix A; it gives the first binomial factor of eq 8.

By this means, we conclude that the correcting exponential factor in this equation results from intercell interactions that arise when hard solvent (water) particles acquire their real soft interaction potentials, and the quantity λ_0 represents a measure of the interaction. The free (Gibbs) energy decomposition then follows from eqs 1 and 8:

$$\Delta G = -T\Delta S + \Delta H$$
$$-T\Delta S = -k_B T n \ln(1-y)$$
$$\Delta H = k_B T \lambda_0 \tag{9}$$

where ΔS and ΔH are entropy and enthalpy changes, respectively. (Actually, enthalpy changes reduce to internal energy changes in an NVT ensemble.)

Quantities y and n are parameters of the binomial distribution. As stated above, y denotes the probability of finding a hard particle in a given cell. The definition of n as the number of cells in the cavity, as given in eq 6, will now be modified for the case $n \gg 1$ as $n = (\langle m\rangle_0/y) = (\rho_0 v/y)$, where $\langle m\rangle_0 = \rho_0 v$ is the mean number of particles in the cavity, ρ_0 is the mean concentration (number density), and subscript 0 means that the hypothetical hard-particle fluid is considered in this reasoning. We can attain a continuous formulation without a significant loss of accuracy when $n \gg 1$ by allowing for noninteger values of n in eq 8.

Finally, we have to consider a real solvent with interacting particles. Its number density is ρ, and the mean number of soft particles in the cavity is $\langle m\rangle = \rho v$. With this notation, the scaling parameter ϵ is introduced: $\langle m\rangle_0 = \epsilon\langle m\rangle$, $\rho_0 = \epsilon\rho$; then, n is expressed as

$$n = \Theta\langle m\rangle \qquad \Theta = \frac{\epsilon}{y} \tag{10}$$

For the BE distribution (eq 7), the explicit expression for λ_0 based on eq 10 is obtained in Appendix B:

$$\lambda_0 = \Theta\langle m\rangle \ln\left[\frac{\Theta}{\Theta - 1}(1-y)\right] \tag{11}$$

Equation 9 is simplified to

$$\Delta G = -k_B T\langle m\rangle\, \Theta \ln\frac{\Theta - 1}{\Theta}$$
$$-T\Delta S = -k_B T\langle m\rangle\, \Theta \ln(1-y)$$
$$\Delta H = k_B T \lambda_0 \tag{12}$$

Most remarkable is the result that y has canceled in the free energy expession, giving a simple one-parameter formula. The full description in terms of eqs 11 and 12 is two-parametric. The parameters of the model must satisfy the following conditions: $0 \le y \le \epsilon \le 1$; $\Theta = \epsilon/y \ge 1$. The rationalization of this condition for ϵ is discussed below. This parameter has a clear meaning of density scaling, but in the final eqs 11 and 12, it is more convenient to substitute it by the combined parameter Θ. The values of y and Θ are supposed to fit the thermodynamic data obtained from a molecular simulation or from experimentation. According to the present cell model, they represent properties inherent to the bulk solvent, independent of a solute. This is why further extrapolation to other solutes in water becomes possible.

Equation 10 is an approximation establishing a purely linear dependence of n on the cavity volume. By this means, the thermodynamic characteristics of cavitation become proportional to v, and eq 12 represents the linearized BE approach.

4. Extension of the Cell Model for the Case of the NPT Ensemble

A modification of the simple transparent description of the cavitation effect as suggested in the preceding section is required when the NPT ensemble is considered. Let us introduce it first under NVT conditions. The total volume V of the solution system is covered by the network of identical cells with volume v_0; the number of cells is

$$L = \frac{V}{v_0} \tag{13}$$

The number of solvent particles is N, producing the concentration $\rho = N/V$. The background hypothesis is that the number of cells does not change during the solvation process. So, not only L but also the value of the probability y in eq 5 ($y = N/L < 1$) remains constant. Moreover, we have to assume that the values L and y are determined already at the primary level of reasoning when the hard-particle approximation (no interaction) is invoked, so that allowing for interaction at the next stage does not change them. Because the volume V decreases when the attractive interaction is switched on, we conclude that the density ρ_0 of the hypothetical hard-particle fluid is lower than the density ρ of real water. Therefore, the binomial approach, as well as another implicit condition that probability y and scaling parameter ϵ are independent of the cavity size and shape, arises as a natural consequence of the present formulation.

The NVT solvation process is treated as an insertion of a solute particle in the box of volume V containing N solvent particles at temperature T. The pressure in the resulting system fluctuates, and its mean value slightly differs from the external pressure P. We consider a dilute solution system, such that its thermodynamic properties are independent of N and V, which means that the volume fraction of the solute is extremely small:

$$\frac{v}{V} \approx \frac{n}{L} \to 0 \tag{14}$$

Considering further the NPT ensemble, we find that the constraint of constant V is relaxed. We have to account for the volume changes: $V \to V + \delta V$. The volume variations always exist as fluctuations, but the change of the average volume is not forbidden as well, and δV represents this average change corresponding to the NVT $\to$ NPT transition. We retain the basic assumption that the structure of the cell lattice is not destroyed, so values of L and y are not affected by the volume changes. However, the volume of cells must change under such conditions $v_0 \to v_0 + \delta v_0$, such that

$$L\delta v_0 = \delta V = \alpha v \tag{15}$$

The last equality postulates that δV is proportional to the cavity volume v with the proportionality factor α discussed below. Generally, α is expected to depend on (v/V), but it becomes constant under condition 14.

According to the general thermodynamic analysis,[25−27] the free energy does not change in a process of such type. On the other hand, entropy and enthalpy changes do occur, and they cancel one another. The entropy change δS can be estimated

explicitly in the frame of the cell model with noninteracting particles:

$$(-T\Delta S)_{\mathrm{NPT}} = (-T\Delta S)_{\mathrm{NVT}} - T\delta S$$

The superscripts label the relevant ensembles. The simple expression for δS follows when the solvent molecule in a cell is considered as a free structureless spherical particle contained in a box with volume $v_0 + \delta v_0$

$$T\delta S = Nk_{\mathrm{B}}T \ln\left(1 + \frac{\delta v_0}{v_0}\right) = k_{\mathrm{B}}T\alpha\epsilon\langle m\rangle \qquad (16)$$

Accepting a spherical shape for water particles is standard in molecular simulations, whereas their free motion inside the cell is possible until the intercell interaction is neglected. The result in eq 16 follows from the sequence of transformations of the value

$$\frac{\delta v_0}{v_0} = \frac{\alpha v}{Lv_0} = \frac{\alpha}{Nv_0}(yv) = \frac{1}{N}\alpha\epsilon\langle m\rangle$$

under condition 14.

The enthalpy change, as the counterpart of eq 16, is determined by invoking the entropy−enthalpy compensation effect[25−27] mentioned above: $\delta H = T\delta S$. It reflects implicitly the change of the interaction energy. We obtain therefore

$$\begin{aligned} \Delta G_{\mathrm{NPT}} &= \Delta G_{\mathrm{NVT}} \\ (-T\Delta S)_{\mathrm{NPT}} &= -k_{\mathrm{B}}T(\Theta\langle m\rangle)[\ln(1-y) + \alpha y] \\ (\Delta H)_{\mathrm{NPT}} &= k_{\mathrm{B}}T(\lambda_0 + \Theta\langle m\rangle\alpha y) \end{aligned} \qquad (17)$$

According to the first eq 17, the free energy can be treated in terms of the NVT formulation (eq 9). The expressions for entropy and enthalpy changes are modified. Again, as in the case of the NVT ensemble, y, Θ, and α should be considered as solute-independent adjustable parameters. They can be fitted, given simulation or experimental data for $(-T\Delta S)$ and ΔH.

5. Fitting the Simulation Data

5a. Treatment of the Simulation Results. We examined the binomial cell model numerically by comparing its results with extensive simulation data reported by Gallicchio et al.[13] for a series of hydrocarbons in water. In that study, partitioning ΔG into enthalpy and entropy components was explicitly performed under NPT conditions. The corresponding free energy contributions associated with cavity formation were separated. Actually, the decomposition of the cavitation free energy was performed as

$$\begin{aligned} \Delta G_{\mathrm{cav}} &= \Delta H_{\mathrm{cav}} - T\Delta S_{\mathrm{cav}} \\ \Delta H_{\mathrm{cav}} &= \Delta U_{vv} + \Delta U_{uv} \end{aligned} \qquad (18a)$$

where ΔU_{vv} is the solvent reorganization energy accompanying the cavity formation and ΔU_{uv} is the mean repulsion energy between solute and solvent particles.

The entropy term was found as a difference. The interaction term ΔU_{uv} was considered as a part of the cavitation energy due to technical reasons determined by conditions of the simulation. Equation 18a corresponds to the soft-core cavity model. We can assume alternatively

$$\begin{aligned} \Delta G_{\mathrm{cav}} &= \Delta H_{\mathrm{cav}} - T\Delta S_{\mathrm{cav}} \\ \Delta H_{\mathrm{cav}} &= \Delta U_{vv} \end{aligned} \qquad (18b)$$

This decomposition differs from that given in eq 18a because the interaction term ΔU_{uv} is withdrawn. Free energy ΔG_{cav} and enthalpy ΔH_{cav} change correspondingly, whereas the entropy remains the same. The interaction enthalpy term is then considered as a part of the total interaction free energy, $\Delta G_{\mathrm{int}} = \Delta G_{\mathrm{att}} + \Delta U_{uv}$, where ΔG_{att} denotes the average attraction energy generated by dispersion (R^{-6}) LJ interactions and small electrostatic terms.[13]

Neither of the approximations in eqs 18a and 18b for cavitation energy is fully consistent, and we tested both of them. As in ref 13, our computations were performed with the 11 hydrocarbons listed in Table 1. For their cavities, solvent-accessible surface areas (SASAs) and corresponding excluded volumes v's were found. Recommendations of ref 13 were followed in parametrization of the cavities: We accepted the same structures for hydrocarbon molecules, the thermal radii for carbon−water and hydrogen−water being 3.1095 and 2.568, respectively.[28] Mean values $\langle m\rangle$ and mean deviations $\sigma = \langle(m - \langle m\rangle)^2\rangle^{1/2}$ are also listed in Table 1 for the NPT ensemble corresponding to the TIP4P water model.[29] They were calculated from MD trajectories. A comparison of $\langle m\rangle$ and ρv values provides a test for checking the accuracy in the simulations of moments (see Appendix C for simulation details).

TABLE 1: Benchmark Hydrocarbon Systems

			TIP4P, NPT ensemble		
molecule	SASA, Å^2	v, Å^3	ρv	$\langle m\rangle$	σ
methane	141.26	153.30	5.113	5.113	1.180
ethane	178.66	211.19	7.044	7.047	1.326
propane	209.13	263.19	8.779	8.775	1.445
butane	239.85	315.66	10.529	10.528	1.560
pentane	270.53	368.12	12.279	12.278	1.655
hexane	300.60	419.25	13.984	13.976	1.751
isobutane	234.39	310.69	10.363	10.364	1.540
2-methylbutane	258.79	355.63	11.862	11.857	1.628
neopentane	254.33	352.76	11.766	11.769	1.620
cyclopentane	243.56	329.15	10.979	10.979	1.577
cyclohexane	262.40	365.94	12.206	12.208	1.640

5b. Linearized BE Scheme. Results obtained with the data based on eqs 18a and 18b are very similar. Compared to eq 18b, eq 18a produces fits that are slightly worse for ΔG and slightly better for ΔS and ΔH. It is expedient to rewrite eq 17 as

$$\begin{aligned} \Delta G &= \left(\Theta \ln\frac{\Theta}{\Theta - 1}\right)k_{\mathrm{B}}T(\rho v) \\ -T\Delta S &= (\Theta\xi)k_{\mathrm{B}}T(\rho v) \\ \Delta H &= \Theta\left(\ln\frac{\Theta}{\Theta - 1} - \xi\right)k_{\mathrm{B}}T(\rho v) \end{aligned} \qquad (19)$$

Here, eq 11 for λ_0 is applied. The combined parameter ξ is defined as

$$-\xi = \ln(1 - y) + \alpha y \qquad (20)$$

Equation 19 predicts a linear dependence of ΔG on the cavity volume v. For all 11 substrates, it reproduces the simulation results with accuracy within 0.2 kcal/mol. Figure 1 illustrates

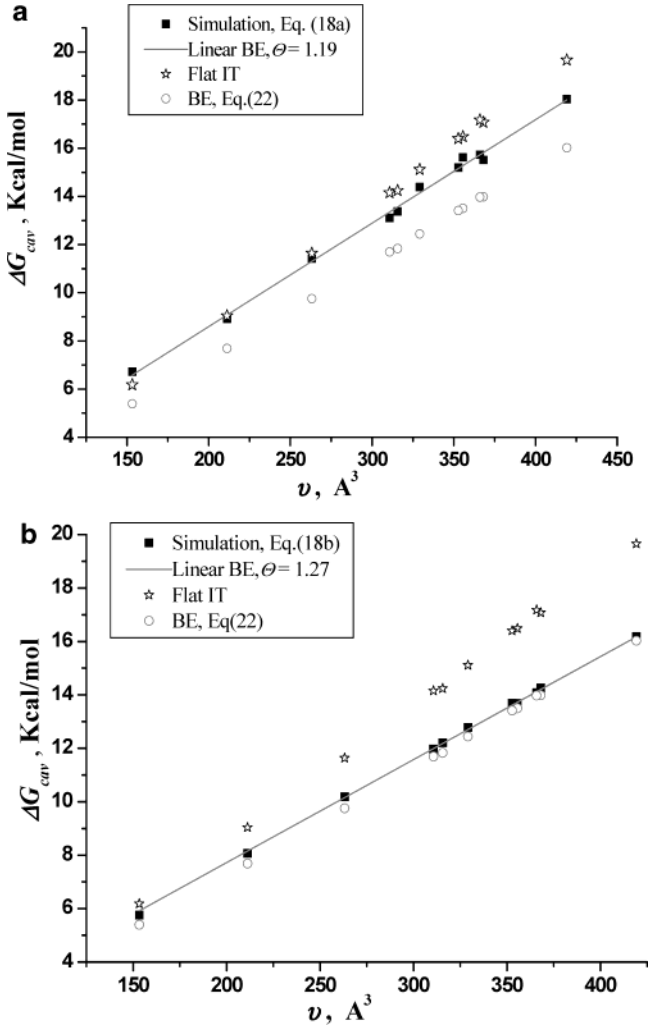

Figure 1. Correlations of the cavitation free energy ΔG_{cav} and the cavity volume v. Straight lines (—) correspond to the linear BE model, eq 19. The open circles (○) represent the full BE treatment, eqs 21 and 23. The stars (☆) are obtained from the flat distribution. Black boxes (■) are the simulation results: (a) eq 18a, $\Theta = 1.190$; (b) eq 18b, $\Theta = 1.272$.

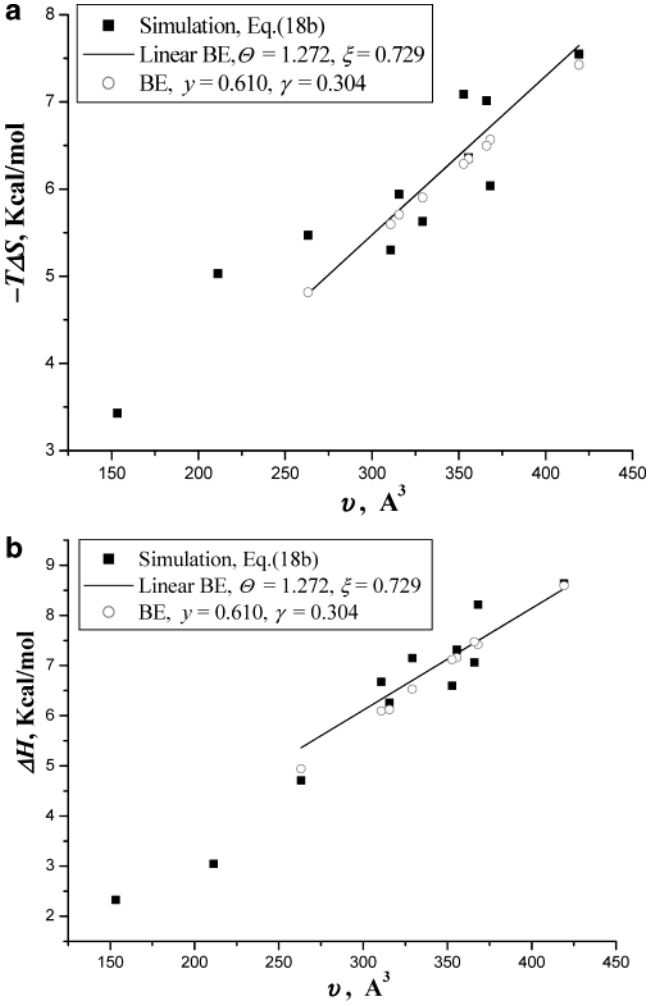

Figure 2. Free energy decomposition. Black boxes (■) are simulation results, eq 18b. Straight line (—) corresponds to eq 19: $\Theta = 1.272$, $\xi = 0.77$. Open circles (○) represent the BE treatment (eqs 21, 23) with $y = 0.610$, $\gamma = 0.304$. Fitting does not include the first two points (methane and ethane). (a) Correlation of the cavitation entropy ($-T\Delta S$) and the cavity volume v. (b) Correlation of the cavitation enthalpy ΔH and the cavity volume v.

this calculation: Straight lines represent predictions suggested by eq 19. The results found with the flat distribution[5−8] are also plotted. Open circles correspond to BE computations described below in section 5c.

Studies of the cavitation entropies and enthalpies include the second adjustable parameter ξ, eq 20. For a given Θ (determined from the fit for ΔG), we again obtain plots that are linear in v. They are shown in Figure 2a,b only for the data given by eq 18b with $\Theta = 1.272$. The two smallest substrates (methane and ethane) were excluded from this fitting. With the optimum value $\xi = 0.729$, the maximum discrepancy between eqs 18b and 19 for the entropy term does not exceed 0.7 kcal/mol. The discrepancy for ΔH does not exceed 0.6 kcal/mol. When methane and ethane were included, the discrepancy for ethane exceeded 1 kcal/mol for both $T\Delta S$ and ΔH. The results based on eq 18a are similar.

The accuracy of decomposing the cavitation free energy into entropic and enthalpic components has been estimated as 0.5−0.7 kcal/mol[13] (10% of the solvent reorganization energy U_{vv}). The effect of the soft-cavity approximation (having an order of magnitude compared with a discrepancy between the data of eqs 18a and 18b that is equal to ΔU_{uv}) is on the order of 1 kcal/mol. With this in mind, the linear fitting results seem to be acceptable, but they can be improved by advanced binomial procedures as discussed below.

5c. Beyond the Linearization Scheme. The conclusions of section 5b are based on the BE distribution (eq 7) and linearized expression (eq 10) for its parameter n, the number of cells in the cavity. An alternative approach is based on analytical expressions for the moments that are available for the BE model (Appendix B). The following result is obtained:

$$n = \frac{\langle m \rangle^2}{\langle m \rangle - \sigma^2} \tag{21}$$

This exact relation stands for the empirical prescription (eq 10) of the linear theory. It directly expresses n in terms of moments found by simulations (Table 1). For such a case, we obtain

$$e_0^{\lambda} = (1 - y)^n \left(\frac{\langle m \rangle}{\sigma^2}\right)^n \tag{22}$$

and finally

$$\Delta G = -k_B T n \ln\left(\frac{\sigma^2}{\langle m \rangle}\right) \tag{23}$$

with n given by eq 21. The parameter y has again dropped out from the expression for ΔG. As in the case of the flat distribution, eq 23 contains no adjustable parameters. Its fitting is illustrated by open circles in Figure 1. We compared the simulation data in both as suggested by eqs 18a (Figure 1a) and 18b (Figure 1b). The data from eq 18b are reproduced within 0.5 kcal/mol. As compared to the data of eq 18a, these BE results are systematically reduced (by 1.0−1.5 kcal/mol). The advantage of BE over the flat distribution is visible only in the case of eq 18b.

We have also studied the full scheme as formulated by eq 2 with binomial $P_m^{(0)}$. A rational guess about the functional form of parameter n is of primary importance in this refined test. Using linear eq 10 was unsuccessful, and the choice of scaling eq 21 as $n = \Theta[\langle m\rangle^2/(\langle m\rangle - \sigma^2)]$ also resulted in a large scattering of fitted points. A better approximation was found by applying the expression

$$n = \frac{1}{y}\left(\frac{\langle m\rangle\sigma^2}{1-y}\right)^{1/2} \tag{24}$$

It is obtained as a combination of two equations: $\langle m\rangle = ny$ and $\sigma^2 = ny(1 - y)$, which rigorously represent the moments of the binomial distribution. The following computations included, first, calculating the solution of eq 3 for λ_0, λ_1, and λ_2, and, next, fitting the parameter y in order to reproduce the simulation data. The procedure was totally numerical. It is denoted as the binomial information theoretical (BIT) model below. Figure 3a,b illustrates its results. It presents a satisfactory fit, approximately of the same quality as that for the nonlinear BE distribution in Figure 1b. For the data given by eq 18a, only a single point for methane produces the deviation of 1 kcal/mol. For the case of eq 18b, a significant deviation (0.7 kcal/mol) is observed only for propane. The merit of this refined nonlinear approach is that, because of the appearance of the adjustable parameter y, both data sets (suggested by eqs 18a and 18b) are reproduced with the correct average trends, such that the only fluctuations indicate that some internal inconsistencies of the procedure may still exist.

It should be noted that the results of the BIT scheme exhibited in Figure 3 were obtained with a continuous smeared version of the binomial distribution, where, in a computation of binomial coefficients, the factorials with noninteger n were substituted by Γ functions, instead of invoking rigorous eq 5 as a canonic prescription. The corresponding discrete version with true factorials produced more significant fluctuations.

5d. Nonlinear Treatment of the Free Energy Decomposition. We start from general NVT expressions (eq 9) in extending the nonlinear approach to perform the entropy/enthalpy decomposition of the cavitation free energy. The NPT ensemble is prepared, similarly to the linear case, by applying corrections $T\delta S$, eq 16. So, we obtain

$$-T\Delta S = -k_B T[n\ln(1-y) + \gamma\langle m\rangle]$$

$$\Delta H = k_B T[\lambda_0 + \gamma\langle m\rangle] \tag{25}$$

Here, $\gamma = \alpha\epsilon$ is a combined fitting parameter. In the BE model, we use eq 21 for n and eq 22 for λ_0. In the case of the full BIT scheme, eq 24 for n is used, whereas λ_0 is computed as described in section 5c.

Equation 25 is governed by two parameters, y and γ, in both cases. Their fit to the simulation data (eq 18b) is illustrated by

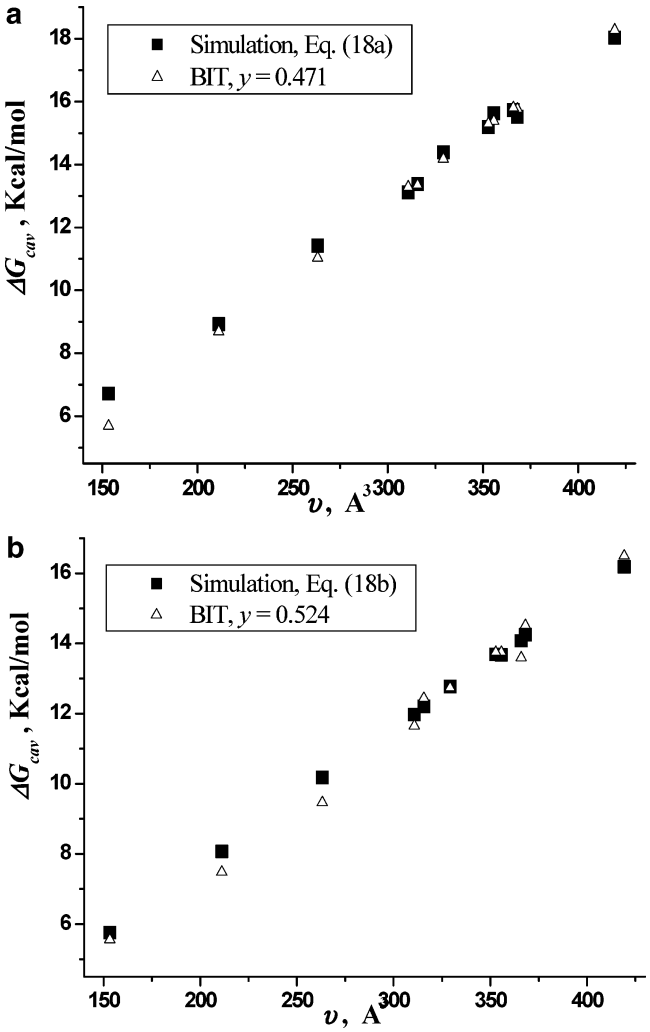

Figure 3. Treatment of cavitation free energies using numerical nonlinear binomial model, BIT (the open triangles, △). Black boxes (■) are the simulation results: (a) eq 18a, $y = 0.471$; (b) eq 18b, $y = 0.524$.

Figures 2 and 4. The data of eq 18a have resulted in a systematic error for ΔG when the BE model was used (Figure 1a); therefore, this case was not considered.

Applying nonlinear approaches improves the result of entropy/enthalpy decomposition. The change is small for the BE case (Figure 2, open circles). The quality of correlation is roughly the same for the BE and BIT models. Note, however, that all 11 substrates were considered for the BIT computations in Figure 4, contrasted to 9 in Figure 2. Including methane and ethane changes drastically the BE values of the fitting parameters.

The two models predict different values of fitting parameters y and γ. Those suggested by the BE model ($y = 0.610$, $\gamma = 0.304$) seem physically relevant. The BIT values are smaller ($y = 0.522$, $\gamma = 0.033$). For the BIT model, fitting the data of eq 18a produced results showing the same quality of correlation as that displayed in Figure 4a,b, with $y = 0.468$ and $\gamma = -0.064$. This computation, which gives a negative value for γ, although small in absolute value, seems unlikely for hydrophobic solutes.

Such estimates of y and γ are primarily tentative tests. Their dependence on the validity of the soft-core approximation for cavity/solvent interactions, which determines how reliable the underlying simulations are, is clearly seen. Further elaboration of this problem could modify the present values. The observed

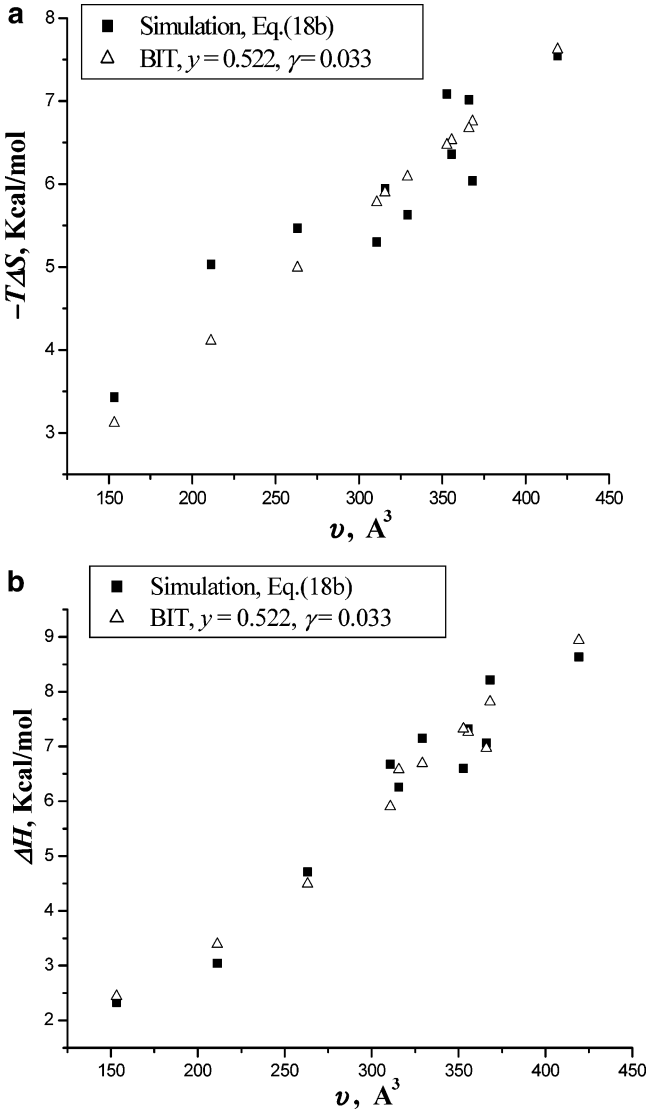

Figure 4. Free energy decomposition by the full nonlinear binomial scheme. Simulation results (■) correspond to eq 18b. Open triangles represent BIT ($y = 0.522$, $\gamma = 0.053$) computations. (a) Cavitation entropies versus cavity volumes. (b) Cavitation enthalpies versus cavity volumes.

empirical tendency is such that an increase in y is accompanied by an increase in γ. Generally, a larger value for γ (see the BE result) seems probable.[26] A larger value of y is also preferable according to the simplest versions of the cell theory of liquids.[19]

6. Spherical Solutes

Model spherical solutes have been studied in the literature for the cavity radius $R < 3.3$ Å ($v < 150$ Å^3).[7−11] The flat information theory (IT) model proved to be successful in this region. The range 2 Å $< R <$ 6.2 Å was studied in our simulations (33 Å^3 $< v <$ 1018 Å^3). Such variation covers a much larger range of the cavity size than the hydrocarbon set considered in section 5 (Table 1). The free energies, particle number mean values $\langle m \rangle$'s, and mean square deviations σ's were computed with the TIP4P water model (the NPT ensemble). As in ref 13, the soft repulsive core potential described the solute/solvent interaction. These interaction terms were included in the free energy as in eq 18a. The uncertainty similar to that discussed in section 5a is produced by this procedure. It can be relaxed by a systematic shift of the free energy curve $\Delta G(R)$ along the cavity radius (or volume) axis. The prescription for

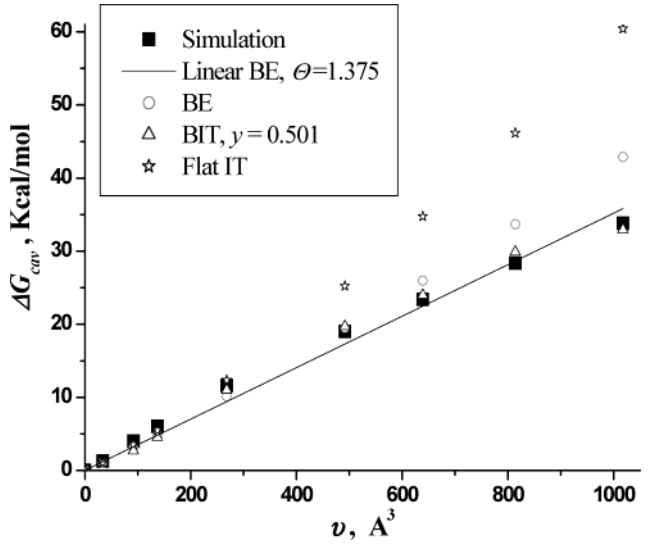

Figure 5. Correlations of the cavitation free energy and the cavity volume for model spherical cavities. The symbols are explained in Figure 1. For small solutes, the results of different models are close and cannot be clearly distinguished.

choosing R as described in Appendix C generated a free energy profile that proved to be practically identical to that reported[5,9] for $R < 3.3$ Å, and it was extrapolated to larger radii. Basic results obtained in this way are presented in Figure 5 for different cavitation models and can be summarized as follows.

1. A significant nonlinearity is visible in the set of simulation points over the whole range of the changing cavity size.

2. The linear BE model (straight line) corresponds to the slope with $\Theta = 1.375$, which is significantly higher than the value $\Theta = 1.19-1.27$ found in the hydrocarbon series. This change demonstrates how the cavity shape influences the cavitation free energy. For hydrocarbons, linear chains generated linear free energy dependences of ΔG versus v with high accuracy. Branched solutes displayed small irregular deviations from this straight line. (These deviations become notably larger when free energies are plotted against the cavity surface (SASA).) Now, with spherical solutes, a straight line with a different slope is observed, the quality of this linear correlation being significantly worse than for the case of hydrocarbons. The change of Θ corresponds to the 11−13% change of the correlation slope for the lower and upper boundaries of Θ corresponding to eqs 18a and 18b, respectively.

3. The nonlinear BIT model is better than the linear one adapted to reproducing the size and shape variation. Slightly changed values $y = 0.47-0.52$ and $y = 0.50$ were found for the hydrocarbon and the spherical series, respectively. Because of fitting y, even the largest solutes are well described in terms of the BIT scheme.

4. The flat IT model reproduces well the simulation results up to $R = 3.5$ Å, but for larger cavities, the error gradually increases, reaching 24 kcal/mol for $R = 6.24$ Å ($v = 1018$ Å^3). This behavior repeats on a larger scale the trend observed for the hydrocarbon series.

5. Finally, the nonlinear BE model predicts well the free energies below $v < 600$ Å^3 (as in the hydrocarbon set based on eq 18b), but the deviations increase gradually for larger volumes (8 kcal/mol for $v = 1018$ Å^3). With no adjustable parameters, its flexibility is insufficient to follow the observed nonlinearity trends in a wider range of cavity sizes.

7. Discussion

7a. Methodological Problems and Validity of Correlations. It is not only the merits of a particular model that determine the quality of correlations. Accuracy of the underlying simulations significantly affects the conclusions which can be inferred from a comparison between computational and model treatments. No less important is the computation of the second moments of fluctuations in the particle number in an observation volume, the quantities required as input data of the model theories.

The main ambiguity appears even at the level of the free energy simulation, due to the soft core accepted for the solute cavity. For hydrocarbons, the discrepancy between the results based on eqs 18a and 18b can serve as a measure of its significance. Because this factor produces a systematic error, the methods containing adjustable parameters (linear BE and BIT parameters) fit equally well to both sets of benchmark ΔG points. This is not so for two parameterless schemes (flat IT and BE schemes). It is seen from Figure 1a,b that the data of eq 18a are favorable for the flat IT method, whereas the BE method is more consistent with the data based on eq 18b. The pertaining systematic uncertainty depends on the definition of the cavity boundary (or the radius R for the spherical case) and can be corrected by its proper normalization. In simulations of spherical solutes, we used this normalization procedure to make these results compatible with earlier studies in a smaller range of cavity sizes.[7−11]

To check how sensitive the model computations are to the accuracy of the variances σ^2's, we compared, for spherical substrates, their values based on direct MD simulations (see section 5a) with those extracted from a computation of the double volume integral over the radial distribution function, as recommended in the literature.[5,6,13] The results practically coincide for the first five points (v < 400 Å), with the discrepancy for the largest cavity ($v = 1018$ Å^3) becoming 2.5%. This last misfit produced a discrepancy of 2 kcal/mol in the free energies of the flat IT model for the last point in Figure 5, a change which does not affect any conclusions based on this figure. For large solutes, the values of σ found in a simulation are smaller than those obtained from the integration of the radial distribution function, and this resulted in slightly increasing ΔG values obtained by means of the parameterless flat IT and BE methods. For the linear BE and the BIT schemes, the presence of adjustable parameters allows good fitting of the simulation data irrespective of the method invoked for a computation of σ. The results expressed in Figures 1−5 and used in the following discussion are all based on MD simulations of variances.

For hydrocarbons, the linear correlation between the cavitation free energy and the cavity volume is obeyed with high accuracy. The binomial cell model provides a rational background for this observation. The significance of $\Delta G-v$ correlations has already been advocated in the literature[8,30] on the basis of general statistical analysis. The same conclusion appears in recent statistical theories combining microscopic (small cavities) and continuum macroscopic (large cavities) solvent models of the cavitation effect.[15,16] We conclude from Figure 1a,b that the linear model is very satisfactory for ΔG despite an inaccuracy in the interpretation of the simulation data that originates, because of the soft-core approximation, from the discrepancy of approaches based on eqs 18a and 18b. This ambiguity is neutralized by a slight variation of parameter Θ.

7b. Entropy/Enthalpy Decomposition. The lower computational accuracy of entropies and enthalpies explains, at least partly, larger scattering of points in the corresponding plots, as compared to much better correlations observed for free energies. The BIT method, governed by parameters y and γ, is capable of describing the free energy decomposition into entropic and enthalpic components better than the linear BE model (with parameters Θ and ξ). This is seen from a comparison of Figures 2 and 4. The advantage of the BIT scheme also becomes visible in a combined consideration of hydrocarbons and spherical solutes. For the case of ΔG, with a single parameter (Θ for the linear BE and y for BIT), the BIT treatment covers more accurately both solute series with a smaller change of its governing parameter.

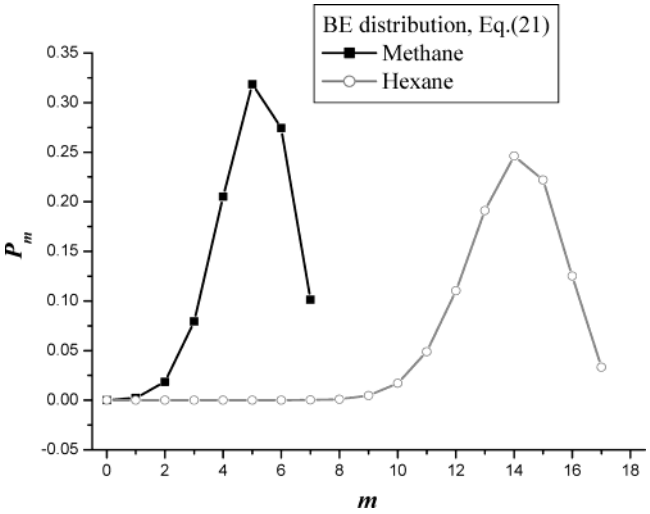

Figure 6. BE distributions for methane and hexane; $y = 0.6$.

On the other hand, the physical background of the decomposition procedure is also far from being fully satisfactory. It is indeed lower than that underlying the basic eq 8 which determines the total free energy analysis. The entropic term in eq 9, which pretends to go beyond the model of hard particles, is accepted to have the same form as in the most simple and transparent case with the interaction neglected. The interaction contribution to the entropy changes then appears only via a rescaling of the distribution parameters n and y, as in eqs 10−12. Such treatment of the interaction is actually the mean field approximation applied in the context of the lattice−gas model.[31] By this means, the nature of the normalization exponent λ_0 becomes entirely enthalpic. Even at this level, the inconsistency of neglecting the temperature dependence of the distribution parameters still remains. Nothing more can be gained within a formulation which avoids the explicit consideration of the interaction mechanism. Only for the parameterless BE model (eqs 21−23) could the alternative analysis be performed in terms of empirical temperature dependences of moments $\langle m \rangle$ and σ.

7c. Binomial Versus Flat Distributions. The advantage of using the binomial distribution for large solutes, revealed from the inspection of Figures 1 and 5, can be explained by Figure 6. It compares BE distributions for methane and hexane. For hexane, the case $m = 0$, responsible for the cavitation effects, corresponds to the remote tail of the distribution ($(\langle m \rangle/\sigma) \gg 1$), where the influence of the second moment (i.e., the peak width σ) is suppressed. This is why fitting the exact σ value is not so important. On the other hand, the presence of adjustable parameter Θ in the linear BE approach or y in the BIT method allows for properly tuning the tail effect. Such flexibility is lacking in the flat model. This method works well for methane

where point $m = 0$ lies closer to the main body of the distribution peak, and the correct result can be attained by fitting σ. It, however, provides a poor approximation for hexane. Because both solutes are well described by the nonlinear BE model (Figure 1b) that is also parameterless, one could consider this fact as an indication of a real physical background that indeed underlies the BE approach. However, such a conclusion cannot be made on the basis of the data of eq 18a.

7d. Interpretation of the Parameter Values. We can better understand some distinctive features of the binomial model by considering parameters found for its nonlinear versions, which are more closely related to the physical background than the combined parameters of the linear BE scheme. The values obtained by fitting for the basic parameter y ($y = 0.5-0.6$) are in reasonable agreement with its interpretation, in terms of the cell theory of fluids,[17,19−21] as the degree of occupation of the cell network. We also expect that parameter γ, governing the degree of volume expansion under NPT conditions, must be positive for hydrophobic solutes. From this point of view, the values of γ which were found for the nonlinear BE model look most reliable, whereas its evaluation from the BIT model based on eq 18a, giving a small negative value, can be to a larger extent determined by accidental details of the underlying simulation. (For reasons responsible for this result, see sections 5d and 7a,b.)

As a further example, let us consider the temperature dependence of the solvation entropy of water, the phenomenon intensively discussed in the literature.[6−8] It is known that the hydration entropy, in which the cavitation contribution dominates, tends to disappear when the temperature is increased, whereas the temperature change of the enthalpy term is not so significant. A possible mechanism for this effect is contained in eq 25. For the entropy, the two terms in brackets have opposite signs (provided $\gamma > 0$) with the first NVT term dominating and the second term being a correction for the volume expansion under NPT conditions. As y decreases and γ increases with the temperature increase, the absolute values of the two contributions are balanced, leading to a cancellation of the total entropy.

8. Conclusion

The present work borrowed the basic ideas of the cell theories of dense fluids in order to derive the binomial distribution law as a background for estimating the probabilities of occupation by solvent particles of a given excluded volume inside a pure solvent. For practical applications, this approach is further elaborated in a phenomenological manner rather than by refining the details of cell models. Invoking the information theory (IT) technique leads to several versions of the advanced binomial scheme which contain adjustable parameters and are sufficiently flexible for describing cavitation effects in a large range of variation of the cavity size and shape. The earlier parameterless flat IT model[6−12] worked only for voids available for small solute particles, like methane (with cavity volume <150 Å^3). The present binomial approach successfully extends the information theory treatment to much larger solutes and covers excluded volumes in the range 10^2-10^3 Å^3. The possibility of extrapolation beyond the upper boundary of this range requires further investigation.

This range of cavity sizes covers many applications, where, according to the binomial approach, the thermodynamic characteristics of the excluded cavity formation should be correlated to the volume of the cavity rather than to the surface of its boundary. However, strong correlation between cavity surfaces and volumes always exists within restricted series of solutes. Moreover, in the treatment of association processes such as ligand binding, volume changes are roughly proportional to the interface surface inside the contact region between the complex components, which becomes then a true factor determining the cavitation free energy contribution. Thus, a correlation with the boundary surface is not excluded, although it is not considered as a general primary origin of the cavitation effect.

A comparison with simulation results for the water solvent seems to support the validity of the binomial distribution as a starting default approximation to be inserted in the information theory treatment. The cell model underlying this binomial law implies that the total volume of a fluid is not available to each of its molecules. This is a reasonable assumption for such a dense and well-structured liquid like water under ambient conditions. Its validity for the case of other solvents is possible, but it must be tested for each particular case.

In regard to the problem of large cavities, the present approach, which extends significantly the scope of microscopic treatment and covers the lower part of the cavity range, is quite appropriate for inclusion in the interpolation procedures designed to cover the whole range of cavity size.[14−16] As a challenging target to treat uniformly, the whole problem would require a significant modification of the default (binomial) approximation which deserves special studies.

Acknowledgment. V.V.A., M.V.B., and I.V.L. thank the Russian Foundation of Fundamental Research (Project 02-03-33049) for financial support.

Appendix A

Probability of Cavity Formation in Terms of the Cell Model.[32] Let us consider a large (macroscopic) volume V containing N hard solvent particles and covered by a network of equivalent cells. The degree of cell population is

$$y = N/L < 1 \tag{A1}$$

where L is the number of cells in V, such that $L > N$. The number of empty cells (holes) is $L - N$. The excluded solute volume (the cavity) contains n cells, such that

$$\frac{n}{N}, \frac{n}{L}, \frac{n}{L-N} \ll 1 \tag{A2}$$

Its formation is viewed as a condensation of n holes within a single volume. The statistical weight of such a fluctuation is

$$Z(n) = \frac{(L-n)!}{N!(L-N-n)!}$$

The total (equilibrium) statistical weight is

$$Z(n=0) = \frac{L!}{N!(L-N)!}$$

The probability of cavity formation is

$$P_0(n) = \frac{Z(n)}{Z(n=0)} = \frac{(L-n)!(L-N)!}{L!(L-N-n)!} \tag{A3}$$

Asymptotically, under conditions in eq A2, eq A3 reduces to

$$P_0(n) = (1-y)^n \tag{A4}$$

Appendix B

Moments of the BE Distribution. We consider the BE distribution

$$P_m = \tilde{P}_m e^{-\lambda_0}$$

$$\tilde{P}_m = P_m^{(0)} \exp(-\lambda_1 m) \quad \text{(B1)}$$

where $P_m^{(0)}$ represents the binomial distribution. In terms of the unnormalized distribution $\tilde{P}_m$, the BE moments can be computed, given the following sums

$$S_0(t) = \sum_{m=0}^{n} \tilde{P}_m = \sum_{m=0}^{n} \binom{n}{m} (yt)^m (1-y)^{n-m}$$

$$S_1(t) = \sum_{m=0}^{n} m\tilde{P}_m = t\frac{d}{dt}(S_0(t))$$

$$S_2(t) = \sum_{m=0}^{n} m^2\tilde{P}_m$$

where we used the notation $e^{-\lambda_1} = t$.

Because $S_0(t)$ is the generating function for the binomial distribution,[23] we can write

$$S_0(t) = (1 - y + yt)^n \quad \text{(B2)}$$

and derive $S_1(t)$. The first moment of distribution (eq B1) equals $\langle m \rangle = (S_1(t)/S_0(t))$, and this can be considered as the equation defining t; a simple calculation gives

$$t = \frac{\langle m \rangle \quad (1-y)}{(n - \langle m \rangle) \quad y}$$

By substituting this value in eq B2, noting that $S_0(t = e^{-\lambda_1}) = e^{\lambda_0}$ and applying $n = \Theta\langle m \rangle$, we come out with the expression (eq 11) for λ_0.

In a similar way, the sum $S_2(t)$ is calculated, and the second moment equals $\langle m^2 \rangle = (S_2(t)/S_0(t))$. This yields the variance

$$\sigma^2 = \langle m^2 \rangle - \langle m \rangle^2 = \langle m \rangle - \frac{\langle m \rangle^2}{n} \quad \text{(B3)}$$

Expressions 21 for n and 22 for λ_0 follow from this result.

Appendix C

Simulation Details. MD trajectories were computed with 2215 TIP4P water molecules[32] and one solute particle in a cubic box with periodic boundary conditions. The electrostatics was treated by the Particle Mesh Ewald technique[33] with the low range cutoff at 10 Å. After the system relaxed at temperature 300 K and pressure 1 bar, the calculations were carried out using a Berendsen thermostat and barostat[34] with time constants of 0.5 ps for temperature and pressure. The equations of motion were integrated with a time step of 2 fs. A run for a calculation of the distribution moments had a duration of 1 ns. The calculated water density was equal to 0.997 g/cm^3.

In the studies of spherical solutes, we considered them as van der Waals spheres interacting with water oxygen atoms by the LJ potential:

$$U(r) = 4\epsilon_{ws}\cdot\left[\left(\frac{\sigma_{ws}}{r}\right)^{12} - \left(\frac{\sigma_{ws}}{r}\right)^{6}\right] \quad \text{(C1)}$$

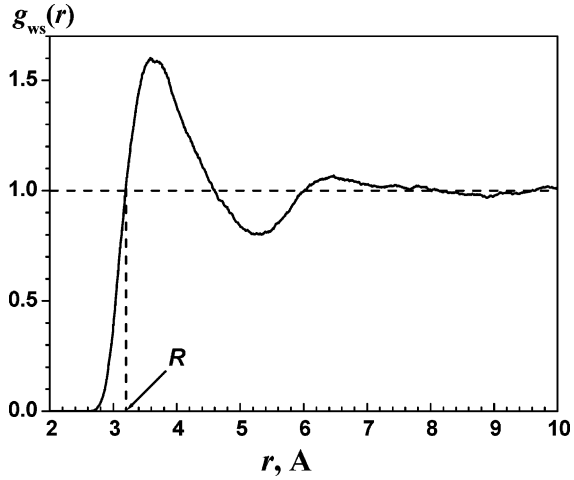

Figure 7. 7. Radial distribution function, $g_{ws}(r)$, as a function of distance r between centers of water oxygen atoms and a solute sphere. The solute with diameter $\sigma_s = 5$ Å is considered here. The effective radius R is determined as described in the text.

TABLE 2: Effective Cavity Radii *R*'s and Thermal Radii R_T's for Several Values of the Solute LJ Parameter σ_s

σ_s, Å	2.0	2.5	3.0	4.0	5.0	7.0	9.0	10.0	11.0	12.0
R, Å	2.0	2.2	2.4	2.8	3.2	4.0	4.9	5.34	5.79	6.24
R_T, Å	1.93	2.12	2.308	2.68	3.06	3.81	4.558	4.93	5.308	5.68

Here, r is the distance between the centers of interacting particles, $\epsilon_{ws} = (\epsilon_w\epsilon_s)^{1/2}$, $\sigma_{ws} = 1/2(\sigma_w + \sigma_s)$, and ϵ_w, σ_w are parameters of TIP4P water molecules (0.649 kJ/mol and 3.154 Å, respectively). For the solute sphere, $\epsilon_s = 0.002$ kJ/mol, whereas σ_s was varied in the range 2−12 Å. (With such a value of ϵ_s, the potential is almost repulsive.) The following procedure was accepted to determine the effective cavity radius R: The radial distribution function $g_{ws}(r)$ for (water oxygen)/solute centers was computed (see Figure 7 as a typical example), and R was extracted as the minimum value of r at which $g_{ws}(R) = 1$. A correlation between σ_s and R is illustrated by Table 2. The thermal radii R_T's, obtained as those values of r where $U(R_T) = k_BT$ (see eq C1), are also shown for the sake of comparison.

The solvation free energy was computed by the slow-growth technique with linear change of scaling parameter λ. A systematic error, estimated by the calculations in the forward and opposite directions (exhibited as a hysteresis), was no more than 5.5%. All MD calculations were carried out using the MD program package *GROMACS*.[35] The present prescription reproduced the energy profile for cavitation free energies reported earlier[5,9] for $R < 3.3$ Å and extended it to the larger range as shown in Figure 5. When using R_T instead of R (Table 2), the energy profile goes higher.

References and Notes

(1) Reiss, H. *Adv. Chem. Phys.* **1965**, *9*, 1.
(2) Reiss, H.; Frisch, H. L.; Helfand, F.; Lebowitz, J. L. *J. Chem. Phys.* **1960**, *32*, 119.
(3) Pierotti, R. A. *Chem. Rev.* **1976**, *76*, 717.
(4) Tomasi, J.; Persico, M. *Chem. Rev.* **1994**, *94*, 2027.
(5) Hummer, G.; Garde, S.; Garcia, A.; Pohorille, A.; Pratt, L. R. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 8951.
(6) Hummer, G.; Garde, S.; Garcia, A. E.; Paulaitis, M. E.; Pratt, L. R. *J. Phys. Chem. B* **1998**, *102*, 10469.
(7) Hummer, G.; Garde, S.; Garcia, A. E.; Pratt, L. R. *Chem. Phys.* **2000**, *258*, 349.
(8) Pratt, L. R. *Annu. Rev. Phys. Chem.* **2002**, *53*, 409.
(9) Gomez, M. A.; Pratt, L. R.; Hummer, G.; Garde, S. *J. Phys. Chem. B* **1999**, *103*, 3520.
(10) Hummer, G.; Garde, S.; Paulaitis, A. E. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 1552.

(11) Garde, S.; Garcia, A. E.; Pratt, L. R.; Hummer, G. *Biophys. Chem.* **1999**, *78*, 21.
(12) Ashbaugh, H. S.; Garde, S.; Hummer, G.; Kaler, E. W.; Paulaitis, M. E. *Biophys. J.* **1999**, *77*, 645.
(13) Gallicchio, E.; Kubo, M. M.; Levy, R. *J. Phys. Chem. B* **2000**, *104*, 6271.
(14) Lum, K.; Chandler, D.; Weeks, J. D. *J. Phys. Chem. B* **1999**, *103*, 4570.
(15) Sun, S. X. *Phys. Rev. E* **2001**, *64*, 021512.
(16) Rein ten Wolde, P.; Sun, S. X.; Chandler, D. *Phys. Rev. E* **2001**, *65*, 011201.
(17) Cernuschi, F.; Eyring, H. *J. Chem. Phys.* **1939**, *7*, 547.
(18) Lennard-Jones, J. E.; Devonshire, A. F. *Proc. R. Soc. London, Ser. A:* **1937**, *163A*, 53; **1938**, *165A*, 1.
(19) Rowlinson, J. S.; Curtiss, C. F. *J. Chem. Phys.* **1951**, *19*, 1519.
(20) Peek, R. M.; Hill, T. L. *J. Chem. Phys.* **1950**, *18*, 1252.
(21) Hirshfelder, J. O.; Bird, R. B.; Curtiss, C. F. *Molecular Theory of Gases and Liquids*; Wiley: New York, 1954.
(22) Frenkel, IA. I. *Kinetic Theory of Liquids*; Moscow, 1950.
(23) Feller, W. *An Introduction to Probability Theory and Its Applications*; Wiley: New York, 1970; Vol. 1.
(24) Reichl, L. E. *A Modern Course in Statistical Physics*; Wiley: New York, 1998.
(25) Levy, R. M.; Gallicchio, E. *Annu. Rev. Phys. Chem.* **1998**, *49*, 531.
(26) Kubo, M. M.; Gallicchio, E.; Levy, R. M. *J. Phys. Chem. B* **1997**, *101*, 10527.
(27) Qian, H.; Hopfield, J. J. *J. Chem. Phys.* **1996**, *105*, 9292.
(28) Professor Gallicchio provided us with the thermal radii listed in the present text. They actually correspond to the cavities studied in ref 13.
(29) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926
(30) Hummer, G. *J. Am. Chem. Soc.* **1999**, *121*, 6299.
(31) Rowlinson, J. S.; Widom, B. *Molecular theory of capillarity*; Dover: New York, 2002; Chapter 5.
(32) This derivation was considered earlier in the graduate thesis of I. V. Rostov (Rostov, I. V.; Basilevsky, M. V. Unpublished). .
(33) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *Chem. Phys.* **1995**, *103*, 8577.
(34) Berendsen, H. J. C.; Postma, J. P. M.; Gunsteren, W. F.; Di-Nola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3686.
(35) Lindahl, E.; Hess, B.; van der Spoel, D. *J. Mol. Model.* **2001**, *7*, 306.