# Distribution of Binding Energies of a Water Molecule in the Water Liquid−Vapor Interface†

**Shaji Chempath‡ and Lawrence R. Pratt*,§**

*Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, and Department of Chemical and Biomolecular Engineering, Tulane University, New Orleans, Louisiana 70118*

*Received: July 31, 2008*

Distributions of binding energies of a water molecule in the water liquid−vapor interface are obtained on the basis of molecular simulation with the SPC/E model of water. These binding energies together with the observed interfacial density profile are used to test a minimally conditioned Gaussian quasi-chemical statistical thermodynamic theory. Binding energy distributions for water molecules in that interfacial region clearly exhibit a composite structure. A minimally conditioned Gaussian quasi-chemical model that is accurate for the free energy of bulk liquid water breaks down for water molecules in the liquid−vapor interfacial region. This breakdown is associated with the fact that this minimally conditioned Gaussian model would be inaccurate for the statistical thermodynamics of a dilute gas. Aggressive conditioning greatly improves the performance of that Gaussian quasi-chemical model. The analogy between the Gaussian quasi-chemical model and dielectric models of hydration free energies suggests that naive dielectric models without the conditioning features of quasi-chemical theory will be unreliable for these interfacial problems. Multi-Gaussian models that address the composite nature of the binding energy distributions observed in the interfacial region might provide a mechanism for correcting dielectric models for practical applications.

## I. Introduction

The distribution, $P(\varepsilon)$, of the binding energy of a water molecule in *bulk* liquid water has recently been shown to be unimodal, in fact, normal (Gaussian) to a useful approximation.[1] This result was unanticipated and has practical consequences, some of which are identified in the following *Introduction*. This contribution obtains and analyzes comparative results for the distribution of the binding energy of a water molecule in the water liquid−vapor interface.

The previous[1−4] practical attention to $P(\varepsilon)$ followed from quasi-chemical theory which suggests conditioning on the basis of the linear size of a defined inner shell to force the conditioned $P(\varepsilon)$ toward a normal form. Convergence toward that limiting normal form can be slow, but intermediate levels of conditioning are already revealing.

More specifically, consider of a distinguished water molecule and define an inner shell as the spherical region of radius $\lambda$ centered on the O (oxygen) atom of that molecule. Then the assumption of a normal model for the conditional distribution $P(\varepsilon|n_\lambda=0)$ yields

$$\mu^{\mathrm{ex}} \approx \mu^{\mathrm{ex}}_{\mathrm{HS}}(\lambda) + RT \ln p(n_\lambda=0) + \langle \varepsilon|n_\lambda=0\rangle + \langle \delta\varepsilon^2|n_\lambda=0\rangle/2RT \quad (1)$$

$T$ is the temperature, and $R$ is the gas constant. The condition $(n_\lambda = 0)$ is that the inner shell characterized by $\lambda$ be *empty*. $p(n_\lambda=0)$ is the marginal probability of that event. The contribution $RT \ln p(n_\lambda=0)$ is the chemical contribution because it is expressed as

$$p(n_\lambda=0) = \frac{1}{1 + \sum_{m\geq 1} K_m \rho^m} \quad (2)$$

with the chemical equilibrium constants $K_n$ defined by

$$\frac{p(n_\lambda=n)}{p(n_\lambda=0)} = K_n \rho^n \quad (3)$$

This is a standard "product/reactants" form. A simplification of special interest occurs when one clustered species dominates all others so that, e.g., $p(n_\lambda=n) \approx 1$ for a specific $n$. Then, eq 3 becomes $\ln p(n_\lambda=0) \approx -\ln K_n \rho^n$. Thus, $RT \ln p(n_\lambda=0)$ of eq 1 characterizes the possibilities for binding of solvent molecules to the defined inner shell as they contribute to the free energy.

The other contributions to eq 1 are physically transparent. $\mu^{\mathrm{ex}}_{\mathrm{HS}}(\lambda)$ is the excess chemical potential of a hard-sphere (HS) solute that rigidly blocks the solvent from the inner shell of radius $\lambda$. Note, however, that the whole of the theory can be based upon a soft-cutoff definition of "inner shell",[5] so in fact "HS" needn't indicate a limitation to either *hard* or *spherical*. Similarly, conditional expectations $\langle \varepsilon|n_\lambda=0\rangle$ and $\langle \delta\varepsilon^2|n_\lambda=0\rangle$ can be obtained directly from a physical simulation record if the conditioning is not too aggressive, i.e., for $\lambda$ not too large.[1] For liquid water at 298 K and 1 atm, this physical theory with $\lambda \approx$ 2.8 Å accurately gives the numerically known free energy.[1]

Our physical argument here is that the quasi-chemical theory eq 1 transparently expresses all of the standard physical concepts that are typically used to rationalize exhaustive numerical results on the molecular thermodynamics of liquid solutions, namely, packing, chemical binding, mean interactions, and fluctuations. The latter two contributions, in combination, are precisely the effects that dielectric models attempt to capture.[4] Furthermore,

† Part of the special section "Aqueous Solutions and Their Interfaces".

* To whom correspondence should be addressed. E-mail: lpratt@tulane.edu.
‡ Los Alamos National Laboratory.
§ Tulane University.

10.1021/jp806858z CCC: $40.75 
Published on Web 11/11/2008

these various contributions are evaluated here most accurately for the specific case of interest, i.e., on the basis of the same data. Thus, the quasi-chemical theory eq 1 provides an important basis for testing the sufficiency of those traditional physical concepts.

Equation 1 is the natural simplification of the fully correct

$$\beta\mu^{\mathrm{ex}} = \beta\mu^{\mathrm{ex}}_{\mathrm{HS}}(\lambda) + \ln p(n_\lambda{=}0) + \ln \int \mathrm{e}^{\beta\varepsilon} P(\varepsilon|n_\lambda{=}0)\, \mathrm{d}\varepsilon \tag{4}$$

with $\beta = 1/RT$. The only assumption is that $P(\varepsilon|n_\lambda{=}0)$ is a general Gaussian distribution. The physical idea is that the conditioning focuses on a subensemble of the full data set, one for which the binding energies of the distinguished molecule are composed of numerous, weak, and weakly correlated contributions from solvent molecules that are not close neighbors. If the conditioning achieves that, then a Gaussian form would be expected for $P(\varepsilon|n_\lambda{=}0)$.

The general structure of this quasi-chemical theory eq 4, and its application to liquid water,[1] particularly with the simplification of eq 1, already permits striking conclusions for this introductory discussion. First, consider a model liquid in which intermolecular interactions vanish outside a finite range. Take $\lambda$ to be larger than that range. Then, the binding energies observed in the conditioned subensemble are always zero, and $\beta\mu^{\mathrm{ex}} = \beta\mu^{\mathrm{ex}}_{\mathrm{HS}}(\lambda) + \ln p(n_\lambda{=}0)$ exactly.

Second, notice that, for typical simulation models of liquid water with $\lambda$ in the range of the sizes of water molecules, the conditional mean $\langle\varepsilon|n_\lambda{=}0\rangle$ is negative and has a magnitude substantially larger than the free energy sought.[1] The conditional variance contribution $\langle\delta\varepsilon^2|n_\lambda{=}0\rangle/2RT$ then typically has a magnitude about half as large, and it is positive clearly. These terms compete with each other, and the net result is a negative magnitude typically about the size of the free energy sought. From these points, we conclude that, for typical simulation models of liquid water, a treatment that neglects long-range interactions would be inaccurate for the free energy. We further conclude that a treatment that includes long-ranged interactions as a mean field, neglecting fluctuations expressed by the variance, would be inaccurate also.

Third, the packing and chemical contributions have opposite signs. For typical simulation models of liquid water, there is a value of $\lambda \approx 3.3$ Å at which these contributions cancel each other precisely.[1] With that specific value of $\lambda$, the approximate physical theory eq 1 takes the form of a dielectric model generally expressed. The fact that the remaining terms have magnitudes roughly in the ratio 2:1 is a signature of linear-response models.[6] Of course, this specific value of $\lambda$ is generally not known without examination of the canceled terms, and the physical theory eq 1 is still approximate. However, we conclude that a dielectric model can be physically valid when packing and chemical contributions together are assessed to be negligible.

Fourth, and finally for now, the approximation eq 1 will be *inaccurate* for a dilute gas when $\lambda \lesssim 5.5$ Å for the case of SPC/E water considered below. The application below to the water liquid−vapor interface, which naturally provides information on both high density and low density environments, thus provides testing information on that issue. A naive view of this issue is that it is analogous to the contrast between the ancient LIN and EXP approximations,[7] and the even more ancient Debye−Hückel and Poisson−Boltzmann approximations.[8]

A similar but less ancient observation derives from the point that

$$\varepsilon = \sum_j \varepsilon_j \tag{5}$$

Naively considered, a probability distribution of individual values $\varepsilon_j$ for the summand is not well defined in the following sense:[9,10] a histogram of observed $\varepsilon_j$'s for a finite-sized system will include a contribution straddling $\varepsilon_j = 0$ with *integrated* population proportional to the system size. This represents contributions from the large number of molecules that scarcely interact with the distinguished molecule being considered. Considering macroscopically large systems, the normalization of the distribution of $\varepsilon_j$ comes nearly entirely from that feature straddling $\varepsilon_j = 0$. Any other features of potential physical significance will be progressively diminished in the thermodynamic limit. The *shape* of the distribution will then be system-size dependent. Thus, definition of a distribution of $\varepsilon_j$ must be less naive.

Despite these complexities, the mean $\langle\varepsilon|n_\lambda{=}0\rangle$ is well defined, as is the variance $\langle\delta\varepsilon^2|n_\lambda{=}0\rangle$. The evidence that the Gaussian theory eq 1 is physically reliable for the bulk liquid suggests that there is substantial, correlated compensation between $\varepsilon_j$ of eq 5; furthermore, it suggests that understanding pair correlations $\langle\varepsilon_k\varepsilon_j\rangle$ would be a substantial step in understanding liquid water.

These points exemplify our view that the structure of the physical theory eq 1 conveys important information. In what follows, we obtain new results on the thermodynamic quantities involved above, as they vary through the interfacial region between water liquid and vapor phases. This paper discusses results for the excess chemical potential of water at the liquid−vapor interface. These results are relevant for understanding the hydration of ions and hydrophobic species near aqueous interfaces.[11]

## II. Methods

The present application to the water liquid−vapor interface has the advantage that the observation of the interfacial density profile permits direct inference of spatial variation of the excess free energy that is sought; specifically,

$$\mu^{\mathrm{ex}}(z) - \mu^{\mathrm{ex}} = -RT \ln[\rho(z)/\rho] \tag{6}$$

where $\rho$ is the density and $\mu^{\mathrm{ex}}$ the excess free energy of the bulk liquid. Of course, the excess free energy $\mu^{\mathrm{ex}}(z)$ is well defined on the basis of fundamental principles, explicit formulas are well established,[4] and all quantities in eqs 1 and 4 are straightforwardly understood as a functions of $z$, the perpendicular distance to the interface. Figure 1 shows the system considered. During postprocessing of the raw simulation data, we divide the simulation volume into 1 Å layers perpendicular to the $z$-axis. We then report the binding energy distributions and free energies for each layer. The equimolar surface is defined using the standard procedure so that the surface excess of water is zero. The surface $z = 0$ corresponds to the equimolar surface, $z > 0$ represents the vapor phase, and $z < 0$ represents the liquid phase. Note that there are two liquid−vapor interfaces in the simulation (top and bottom) and all properties reported here are averages of the quantities observed at these two interfaces.

All calculations were performed with the MUSIC software[12,13] which is a set of FORTRAN 90 modules for MC and MD simulations. 534 water molecules were initially placed in a cubic simulation cell of sides 20 Å × 20 Å × 40 Å under periodic boundary conditions. After 100 ps of equilibration at 300 K

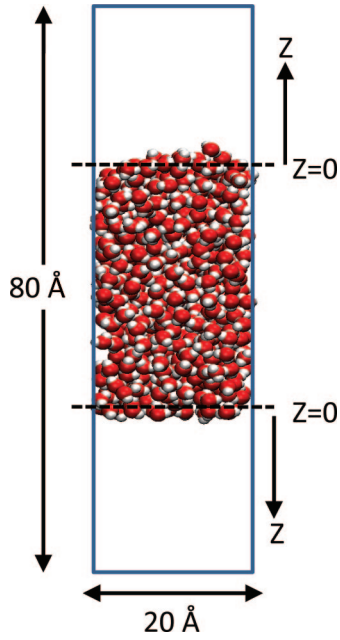


**Figure 1.** Snapshot of the simulation cell at 300 K. The dashed lines represent the equimolar surfaces and define the planes $z = 0$ at the top and bottom interfaces.

with a velocity rescaling algorithm, the simulation cell size was changed to add vacuum regions of 20 Å on the top and bottom of the simulation cell, making the total simulation cell 20 Å × 20 Å × 80 Å in size. 1.0 ns of equilibration was performed at 300 K with velocity rescaling. This system was then simulated for 8.0 ns further by the NVE MD technique. The average temperature estimated from the kinetic energy remained at 302 K. Ewald summation with a smooth-particle mesh algorithm was used in all calculations. MD simulations used the velocity Verlet algorithm with a 2 fs time step, and the molecular structure of the SPC/E water molecules was rigidly constrained with the RATTLE algorithm and three distance constraints. 80 000 snapshots of the simulation were stored, and in the subsequent postprocessing, we again used Ewald summation to calculate the electrostatic contributions to binding energies of individual molecules. In order to reduce the numerical sampling errors for the case of $\lambda = 3.4$ Å, we conducted 10 independent simulations of 4.0 ns and stored and analyzed a total of 800 000 snapshots at 300 K.

## III. Results and Discussion

Density profiles observed from our simulations, and the relative excess chemical potential $\mu^{ex}(z) - \mu^{ex}$, are shown in Figure 2. These profile results are similar to those obtained previously,[14] particularly in exhibiting weak oscillations for the lowest temperature considered (275 K). We note that the tension of this interface for the SPC/E model has been reevaluated recently.[15]

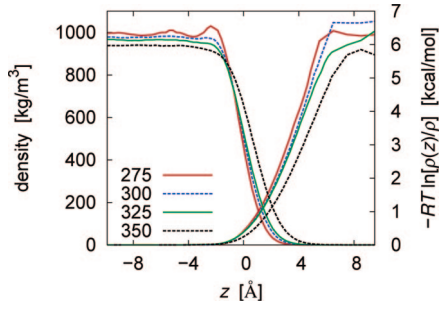


**Figure 2.** Density profiles (left $y$-axis) and excess chemical potentials (right $y$-axis) at the water−vapor interface. The excess chemical potential is reported with respect to its value in the bulk liquid water.

Figure 3 presents the binding energy distributions of a water molecule as observed in the bulk and in a layer of 1 Å thickness just above the equimolar surface; remember that $z = 0$ indicates the equimolar surface. The results for the bulk liquid (upper panel) are roughly Gaussian, while the distribution for the outer interfacial layer is qualitatively composite. To make this composite characteristic more evident, we fit these distributions to a sum of Gaussians according to

$$P(\varepsilon) = \sum_{i=1}^{n} \frac{f_i}{\sqrt{2\pi\sigma_i^2}} \exp\left[-\frac{(\varepsilon - \bar{\varepsilon}_i)^2}{2\sigma_i^2}\right] \qquad (7)$$

with $n = 2$. With $f_1 + f_2 = 1$, a double-Gaussian function will have five adjustable parameters. Figure 3 (upper panel) then shows that the results for the bulk liquid are roughly Gaussian; on the all-important high energy side (see eq 4), an overall Gaussian model matches closely one of the Gaussians of the composite fit, and the data are *slightly* super-Gaussian there. This agrees with previous results.[1] In contrast, for the outer interfacial region (lower panel), the composite nature of the observed distribution is definite, an overall Gaussian matches one of the individual Gaussians less well on the high energy side, and the observed distribution is there *less* than predicted by a single-Gaussian model.

In Figure 4, we compare the excess chemical potentials predicted by the Gaussian model, eq 1 with $\lambda = 2.8$ Å, and a multi-Gaussian against the exact values derived with eq 6. An overall observation of great interest is that a water molecule entering the liquid ($z < 0$) phase from the vapor phase ($z > 0$)

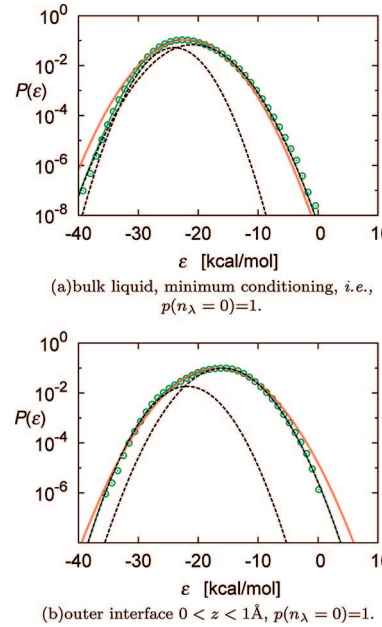


**Figure 3.** Probability distribution of binding energy from the simulations at 300 K (green circles). The red lines are single-Gaussian fits, and the green lines are double-Gaussian fits. The dashed black lines represent the component Gaussians contributing to the double-Gaussian fit. Note that the green lines are obscured due to overlap with the dashed black lines.

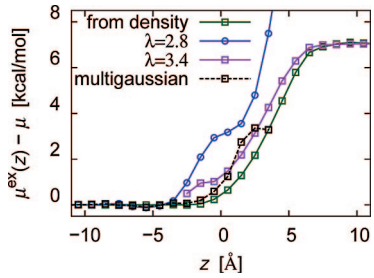


**Figure 4.** Comparison of excess chemical potential from density profile with various models.

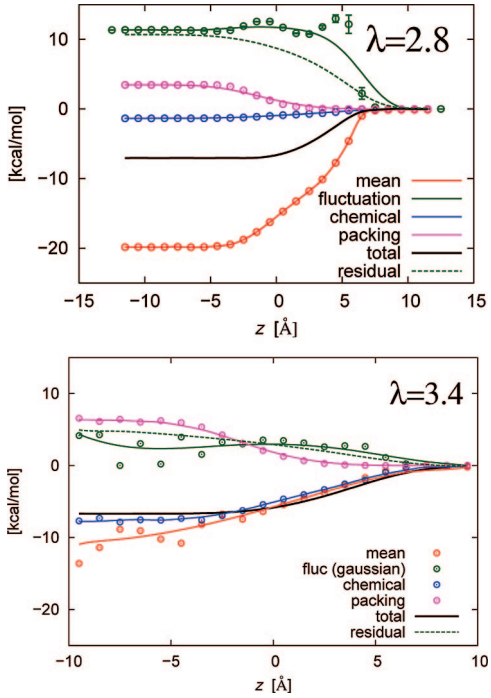


**Figure 5.** Individual contributions to excess chemical potential from the Gaussian model eq 1 as a function of distance from the interface. The green curve is the fluctuation contribution, in magnitude roughly half the mean contribution (red). Upper panel: $\lambda = 2.8$ Å. Lower panel: $\lambda = 3.4$ Å.

acquires more than 90% of its hydration free energy *before* entering the liquid phase, i.e., for $z > 0$.

Considering the physical models, the Gaussian model, eq 1 with $\lambda = 2.8$ Å, performs poorly near the interface ($-4$ Å $< z <$ 4 Å), as anticipated above. Figure 5 examines the several contributions to that Gaussian theory, and also provides an empirically determined residual

$$RT \ln \int e^{\beta\delta\varepsilon} P(\delta\varepsilon|n_\lambda{=}0)\, d\delta\varepsilon = \mu^{ex} - \mu^{ex}_{HS}(\lambda) - RT \ln p(n_\lambda{=}0) - \langle\varepsilon|n_\lambda{=}0\rangle \quad (8)$$

Note (Figure 5) that the chemical contribution is particularly small for $\lambda = 2.8$ Å. For that case, the fluctuation contribution that is associated with dielectric models executes a liquid−vapor transition only *after* decisively exiting the liquid phase at $z = 0$. Otherwise, these distinct contributions change smoothly through the interfacial region. The individual contributions are typically as large as (or larger than) the net result, and thus, the balance of the various contributions is an essential aspect of the theory. The empirically determined residual is close to the

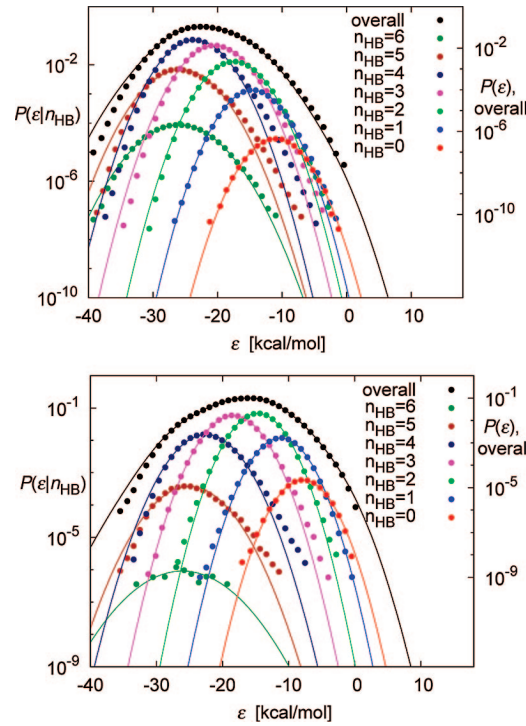


**Figure 6.** Binding energy distributions of water molecules based on their hydration environment. Colored symbols and lines are the observed probabilities and Gaussian fits, respectively. The overall $P(\varepsilon)$ (plotted on right $y$-axis) is obtained by adding the $P(\varepsilon|n_{HB})$ curves. Top: bulk water. Bottom: $0 < z < 1$ Å layer.

Gaussian fluctuation contribution in the bulk but smoothly deviates from it in the interfacial region.

We can go further with the multi-Gaussian analysis by defining structural substates that permit a decomposition[4,16]

$$P(\varepsilon) = \sum_{k=0}^{n} \frac{p_k}{\sqrt{2\pi\sigma_k^2}} \exp\left[-\frac{(\varepsilon - \bar{\varepsilon}_k)^2}{2\sigma_k^2}\right] \quad (9)$$

The $p_k$ are the observed probabilities for the various substates $k$, and the contrast to the previous fitting exercise is that parameters here are obtained directly from the simulation record and not by subsequent fitting to the net $P(\varepsilon)$. The important physical assumption is that the conditional probabilities $P(\varepsilon|k)$ are Gaussian. Figure 6 provides the distributions of a water molecule with minimal conditioning so that $p(n_\lambda{=}0) = 0$, with a multi-Gaussian model based on the number of hydrogen bonds a water molecule forms with its neighbors. The hydrogen bonds are defined in a conventional way. Specifically, we defined two water molecules to be hydrogen bonded if their OO separation was less than 3.3 Å, the smallest nonbonded OH separation as less than 2.3 Å, and the O−OH angle was less than 45°.[17] The free energy model is then specified by

$$\int e^{\beta\varepsilon} P(\varepsilon)\, d\varepsilon = \sum_{k=0}^{n} p_k \exp\left[\beta\bar{\varepsilon}_k + \frac{\beta^2\sigma_k^2}{2}\right] \quad (10)$$

Remarkably, we find that the $n_{HB} = 0$ and $n_{HB} = 1$ terms are the largest terms in the sum eq 10. This can be checked by noting that the red ($n_{HB} = 0$) curve in Figure 6 dominates when $\varepsilon > -5$ kcal/mol. This is precisely the region where the integrand of the integral of eq 4 is maximal. Though the contributions

from $n_{HB} = 3$ and $n_{HB} = 4$ are essential for defining the overall shape of the distribution $P(\varepsilon)$, for the purpose of predicting the free energy, all that matters is $P(\varepsilon)$ for $n_{HB} = 0$ and $n_{HB} = 1$ and their overall proportions.

## IV. Conclusions

Binding energy distributions for water molecules in the water liquid−vapor interfacial region exhibit a composite structure (Figure 3). A minimally conditioned Gaussian quasi-chemical model that works surprisingly accurately for the free energy of bulk liquid water breaks down for water molecules in the liquid−vapor interfacial region. That breakdown can be associated with the fact that this minimally conditioned Gaussian quasi-chemical model would be an inaccurate model of a dilute gas. Aggressive conditioning, e.g., increasing the radius of the defined inner shell region to $\lambda = 3.4$ Å, greatly improves the performance of that Gaussian quasi-chemical model. The analogy between the Gaussian quasi-chemical model and dielectric models of hydration free energies suggests (Figure 5) that naive dielectric models without the conditioning features of quasi-chemical theory will be unreliable for these interfacial problems. Multi-Gaussian models that address the composite nature of the binding energy distributions observed in the interfacial region might provide a mechanism for correcting dielectric models for practical applications.

**Acknowledgment.** This work was carried out under the auspices of the National Nuclear Security Administration of the U.S. Department of Energy at Los Alamos National Laboratory under Contract No. DE-AC52-06N A25396. This project was funded by the Office of Basic Energy Sciences under the U.S. Department of Energy. We thank H. S. Ashbaugh for helpful conversations on this research.

## References and Notes

(1) Shah, J. K.; Asthagiri, D.; Pratt, L. R.; Paulaitis, M. E. *J. Chem. Phys.* **2007**, *127*, 144508.
(2) Asthagiri, D.; Ashbaugh, H. S.; Piryatinski, A.; Paulaitis, M. E.; Pratt, L. R. *J. Am. Chem. Soc.* **2007**, *129*, 10133.
(3) Asthagiri, D.; Merchant, S.; Pratt, L. R. *J. Chem. Phys.* **2008**, *128*, 244512.
(4) Beck, T. L.; Paulaitis, M. E.; Pratt, L. R. *The potential distribution theorem and models of molecular solutions*; Cambridge University Press: 2006.
(5) Chempath, S.; Pratt, L. R.; Paulaitis M. E. (2008),Quasi-chemical theory with a soft cutoff.
(6) Ben-Amotz, D.; Underwood, R. *Acc. Chem. Rev.* **2008**, *41*, 957.
(7) Stell, G. In *Statistical Mechanics. Part A: Equilibrium Techniques*; Berne, B. J., Ed.; Plenum: New York, 1977; pp 47−84.
(8) Onsager, L. *Chem. Rev.* **1933**, *13*, 73.
(9) Rahman, A.; Stillinger, F. H. *J. Chem. Phys.* **1971**, *55*, 3336.
(10) Stillinger, F. *Science* **1980**, *209*, 451.
(11) Pratt, L. R.; Pohorille, A. *Chem. Rev.* **2002**, *102*, 2671.
(12) Web site for obtaining source code given in this paper and other potentially useful modules: http://zeolites.cqe.northwestern.edu/Music/music.html.
(13) Gupta, A.; Chempath, S.; Sanborn, M. J.; Clark, L. A.; Snurr, R. Q. *Mol. Simul.* **2003**, *29*, 29.
(14) Sokhan, V. P.; Tildesley, D. J. *Mol. Phys.* **1997**, *92*, 625.
(15) Chen, F.; Smith, P. *J. Chem. Phys.* **2007**, *126*, 221101.
(16) Hummer, G.; Pratt, L.; García, A. *J. Am. Chem. Soc.* **1997**, *119*, 8523.
(17) Luzar, A.; Chandler, D. *J. Chem. Phys.* **1993**, *98*, 8160.

JP806858Z