with decreasing values for the fraction of Mo^{4+} and Mo^{3+}. From a comparison of parts a and b of Figure 7 it appears that both Mo metal and Mo^{2+} are active for benzene hydrogenation.

The onset of the catalytic activity for $T = 600$ °C coincides with significant formation of Mo^{2+}. The abrupt increase in the activity for $T = 700$ °C can be related to the formation of Mo metal. The catalytic activity correlates reasonably well with variation in the abundance of Mo metal. Figure 6 also shows taht for comparable abundance of Mo^{2+} and Mo metal (ca. 43% and 48% for catalysts reduced at 660 and 800 °C, respectively) the catalyst containing Mo metal is ca. 5-fold more active. This indicates a higher intrinsic activity of Mo metal compared to that of Mo^{2+}. One must note, however, that because of the high intrinsic activity of Mo metal for benzene hydrogenation, one cannot rule out the possibility that the observed activity ascribed to Mo^{2+} is due to a small amount of Mo metal which cannot be detected by ESCA.

The hydrogenation of benzene is a reaction typically catalyzed by metals.[37-42] This is, however, in variance with recent results obtained in this laboratory on reduced Mo/TiO_2 catalysts[16] which indicate that Mo^{2+} sites are more active for benzene hydrogenation than Mo metal. One possible interpretation of this finding postulated in the aforementioned study[16] is that the activity of Mo metal obtained on reduction of Mo/TiO_2 is inhibited by site blocking. The latter is caused by migration of reduced Ti moieties or impurities in the TiO_2 support on reduction. The results of the present study are consistent with this interpretation since, unlike TiO_2, Al_2O_3 is more resistant to reduction and therefore less likely to affect the surface structure of the Mo metal obtained at high-temperature reduction.

Conclusions

An ESCA study of the distribution of Mo oxidation states in Mo/Al_2O_3 catalysts reduced between 500 and 900 °C indicates the formation of mixed oxidation states ranging from Mo^{6+} to Mo metal. For catalysts reduced between 800 and 900 °C, Mo metal was the major species formed. The average Mo oxidation states in reduced Mo/Al_2O_3 catalysts estimated from ESCA results correlated well with those determined from O_2 consumption measurements.

The comparison of benzene hydrogenation activity of reduced Mo/Al_2O_3 catalysts with the distribution with Mo oxidation states indicates that Mo metal and (to a lesser extent) Mo^{2+} are the only active centers for benzene hydrogenation.

Acknowledgment. We thank Professor W. K. Hall for providing the Mo/Al_2O_3 catalyst and for stimulating discussions. This work was supported by the Department of Energy under Grant DE-AC02-79ER10485.

---

(37) Chappen, J.; Brenner, A. Presented at the 9th North American Meeting of the Catalysis Society, 1985, Houston, TX.

(38) Che, M.; Bennett, C. O. *Adv. Catal.* **1989**, *36*, 114.

(39) Martin, G. A.; Dalmon, J. A. *J. Catal.* **1982**, *73*, 233.

(40) Basset, J. M.; Dalmai-Imelik, G.; Primet, M.; Mutin, R. *J. Catal.* **1975**, *37*, 22.

(41) Moss, R. L.; Pope, D.; Dairs, B. J.; Edwards, D. H. *J. Catal.* **1979**, *58*, 206.

(42) Fuentes, S.; Figueras, F. *J. Catal.* **1980**, *61*, 443.

The Possibility of Obtaining an Effective Pairwise Additive Intermolecular Potential via an ab Initio Route by Fitting to a Cooperative Model of Condensed-Phase Configurations

Mihaly Mezei

Department of Chemistry and Center for Study in Gene Structure and Function, Hunter College and Graduate Center of CUNY, New York, New York 10021 *(Received: November 15, 1990)*

This paper evinces the feasibility of obtaining an effective pairwise additive intermolecular potential for liquid water at room temperature by fitting a pairwise additive function to the cooperatively calculated energies and virial sums of simulated liquid water configurations. The procedure requires iterative refinement of the fit. The importance of trying different functional forms has been demonstrated as well as the importance of including the virial sum into the fitting process. The cooperative energies and virial sums were calculated with the Campbell–Mezei model (derived from ab initio dimer energies) which includes a dipole polarization term.

Introduction

The description of the condensed phase of matter hinges upon an adequate representation of the intermolecular interactions. As the modeling of a condensed phase necessarily involves a large number of particles in a large number of configurations, the intermolecular energy has to be a simple expression of the particle coordinates. Such simple expressions may be obtained either by fitting an assumed functional form to (a necessarily limited number of) experimental properties of the substance under study (empirical potentials) or by fitting the same function to ab initio calculations performed on selected configurations of pairs of molecules (ab initio potentials).

The difficulty with the empirical potentials lies in the limited number of experimental data available and in particular in the hidden (or at least not well-known) correlations existing among them. For the ab initio potentials, the adequate sampling of the configuration space and the multibody effects cause problems (vide infra).

To avoid some of these difficulties, the present paper proposes and tests a procedure that generates an effective pair potential (instead of an empirical potential, vide infra) using an ab initio route to represent condensed-phase interactions at a particular thermodynamic state. The procedure starts out from an initial approximation to the potential (to be obtained from the existing literature or generating one from existing potentials describing related substances and assuming the transferability of the parameters) and generates a set of configurations (of the order of a few hundred) at that particular thermodynamic state. Next, the parameters of a pair potential are determined from fitting to the cooperatively calculated total energies and virial sums of a set of liquid configurations. These new parameters are next used to generate a new set of configurations. The procedure is iterated

0022-3654/91/2095-7042$02.50/0

until self-consistency. It is important to stress that, as is true for every effective pairwise potential, the parameters obtained would reflect the thermodynamic state(s) used for their derivation and their validity progressively worsens as one moves away from that state(s).

The procedure proposed above was tested on room temperature liquid water (at its experimental density). The cooperative calculations were performed with the ab initio water model of Campbell and Mezei.[1] That model includes multibody effects through dipole polarizability and was shown to work well for various ice forms[2] and was also able to reproduce[3] the energy of the recently discovered low-energy trifurcated water dimer structure.[4] The configurations used for the initial fitting were extracted from Monte Carlo simulations based on the ab initio MCY[5] and the empirical TIP2[6] and TIP4P[7] potentials.

Background

Classical statistical mechanical calculations required the energy $E(\mathbf{X}^N)$ of the assembly of N particles with coordinates $\mathbf{X}^N$. For the calculation of $E(\mathbf{X}^N)$ to be manageable, in general several approximations are made. In principle, the total energy $E(\mathbf{X}^N)$ can be written as the sum of two-body, three-body, etc., terms:

$$E(\mathbf{X}^N) = \sum_{i<j}^{N} e^{(2)}(X_i,X_j) + \sum_{i<j<k}^{N} e^{(3)}(X_i,X_j,X_k) + \ldots \qquad (1)$$

If one neglects all terms depending on more than two particles, then the total energy will be represented by a sum of pair energies. While the quantum mechanical calculation of these terms in a statistical mechanical computer simulation is still infeasible, the potential surface is usually amenable to a reasonable fit by a simple functional form. This approach has been pioneered by Clementi and his co-workers, who developed an extensive ab initio potential library for interactions between small molecules.[8,9] For water–water interactions, the expansion has been carried out to include three-body and four-body terms as well.[10]

An alternative treatment of the inclusion of higher order terms is through the physical mechanism of polarization. Here the two-body terms are supplemented by an N-body term that is computed by determining the mutual polarization energy of the system (usually at the induced dipole level). The induced moments are computed in an iterative fashion in most cases: compute the moments induced by the permanent field; update the field with the effect of the newly induced moments; repeat until self-consistency. However, Campbell has shown[11] that the induced dipoles can also be obtained directly by solving a system of equations—an approach that is particularly useful when the iteration scheme diverges. This formalism has been recently extended by Campbell and Mezei for higher order moments as well.[12] For Monte Carlo simulations this approach is very expensive since each step involves the move of only one molecule, but the cooperative energy calculations still have to involve all the others. In molecular dynamics, however, all molecules move at each time step and thus the additional work involved in calculating the cooperative contribution is manageable. For liquid water, the polarization approach has been pursued by Berendsen and Velde,[13] Campbell and Mezei,[1] Stillinger and David,[14] Barnes et al.,[15] Stillinger, Weber, and David,[16] and Campbell and Belford[17] and, more recently, by Lybrand and Kollman,[18] Sprik and Klein,[19] Rullman and van Duijnen,[20] Ahlstrom et al.,[21] and Kuwajima and Warshel.[22]

Stillinger laid the theoretical foundations for an improved two-body potential by mapping the contributions of the higher order terms to two-body terms in an average fashion[23] by selecting an $e^{\text{eff}}(X_i,X_j)$ in such a way as to minimize the quantity

$$\int \{\exp[-E(\mathbf{X}^N)/2kT] - \exp[-\sum_{i<j}^{N} e^{\text{eff}}(X_i,X_j)/2kT]\}^2 \, \mathrm{d}\mathbf{X}^N \qquad (2)$$

Such potentials are called effective pair potentials. Notice the presence of T in eq 2—the calculated parameters clearly would depend on them.

Normally, the parameters of an effective pair potential are obtained from fitting to experimental properties of the liquid. The first such water model (BNS) was derived by Ben Naim and Stillinger,[24] modified later (ST2) by Stillinger and Rahman.[25] Subsequently, several new effective pair potentials (often called empirical potentials) were derived for water: TIP,[26] SPC,[27] TIP2,[6] TIP4P,[7] to mention the ones used most frequently. They yielded reasonably satisfactory results when used in computer simulations.

However, a fundamental question still remains unanswered: How well is it possible to fit a pair potential to the full N-body energy? Comparison of the calculated liquid properties with the experimental data can, in principle, serve only as a source of possible negative answers since experimental data are relatively scarce and, more importantly, there may be (possibly hidden) correlations between them. With the development of a successful effective pairwise additive potential based on fitting to cooperatively calculated energies and virial sums of liquid configurations, this question can also be addressed quantitatively.

Calculations

The calculation of the total energy and the virial sum of each configuration was performed with the Campbell–Mezei model. This model contains an electrostatic term based on the field defined by the water monomer wave function and cooperatively induced dipoles, atom–atom repulsion terms obtained from fitting to Hartree–Fock dimer energies, and an empirical dispersion term:

$$E = E_{\mathrm{P}} + E_{\mathrm{I}} + E_{\mathrm{D}} + E_{\mathrm{R}} \qquad (3)$$

where E_{P} is the pairwise additive part of the electrostatic interactions of the permanent multipole moments, calculated for each pair as the sum of interaction of multipoles of various order. Each multipole interaction of order m and n, respectively, is calculated in the Maxwell formalism for spherical harmonics in the form

$$e_p^{ij}(n,m) = P_i^n P_j^m \sum_{p=1}^{n} \sum_{q=1}^{m} (\mathbf{s}_p^{i,n} \cdot \nabla)(\mathbf{s}_q^{j,m} \cdot \nabla)/r_{ij} \qquad (4)$$

where the constants P_i^n and P_j^m and vectors $\mathbf{s}_p^{i,n}$ and $\mathbf{s}_q^{j,m}$ are de-

---

(1) Campbell, E. S.; Mezei, M. *J. Chem. Phys.* **1977**, *67*, 2338.
(2) Campbell, E. S.; Mezei, M. *Mol. Phys.* **1980**, *41*, 883.
(3) Mezei, M.; Dannenberg, J. J. *J. Phys. Chem.* **1988**, *92*, 5860.
(4) Dannenberg, J. J. *J. Phys. Chem.* **1988**, *92*, 6869.
(5) Matsuoka, O.; Clementi, E.; Yoshemine, M. *J. Chem. Phys.* **1976**, *64*, 1351.
(6) Jorgensen, W. L. *J. Chem. Phys.* **1982**, *77*, 4156.
(7) Jorgensen, W. L.; Chandrashekar, J.; Madura, J. D.; Impey, R.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
(8) Clementi, E.; Cavallone, F.; Scordamaglia, R. *J. Am. Chem. Soc.* **1977**, *99*, 5531.
(9) Corongiu, G.; Clementi, E. *Gazz. Chim. Ital.* **1978**, *108*, 273.
(10) Detrich, J.; Corongiu, G.; Clementi, E. *Int. J. Quantum. Chem.* **1984**, *S18*, 701.
(11) Campbell, E. S. *Helv. Phys. Acta* **1967**, *40*, 388.
(12) Campbell, E. S.; Mezei, M. *Mol. Phys.* **1986**, *57*, 1201.
(13) Berendsen, H. J. C.; van der Velde, G. A. *CECAM Workshop Proceedings*; Centre Européen de calcul Atomique et Moléculaire: Orsay, France, 1972.
(14) Stillinger, F. H.; David, C. W. *J. Chem. Phys.* **1978**, *69*, 1473.
(15) Barnes, P.; Finney, J. L.; Nicholas, J. D.; Quinn, J. E. *Nature* **1979**, *282*, 459.
(16) Stillinger, F. H.; Weber, T.; David, C. W. *J. Chem. Phys.* **1982**, *76*, 3131.
(17) Belford, D.; Campbell, E. S. *J. Chem. Phys.* **1984**, *80*, 3288.
(18) Lybrand, T. P.; Kollmann, P. A. *J. Chem. Phys.* **1985**, *83*, 2923.
(19) Sprik, M.; Klein, M. L. *J. Chem. Phys.* **1988**, *89*, 7556.
(20) Rullmann, J. A. C.; van Duijnen, P. Th. *Mol. Phys.* **1988**, *63*, 451.
(21) Ahlstrom, P.; Wallqvist, A.; Engstrom, S.; Jonsson, B. *Mol. Phys.* **1989**, *68*, 563.
(22) Kuwajima, S.; Warshel, A. *J. Phys. Chem.* **1990**, *94*, 460.
(23) Stillinger, F. H. *J. Chem. Phys.* **1972**, *57*, 1780.
(24) Ben-Naim, A.; Stillinger, F. H. In *Structure and Transport Processes in Water and Aqueous Solution*; Horne, R. A., Ed.; Wiley: New York, 1972.
(25) Stillinger, F. H.; Rahman, A. *J. Chem. Phys.* **1974**, *60*, 1545.
(26) Jorgensen, W. L. *J. Am. Chem. Soc.* **1981**, *103*, 335.
(27) Berendsen, H. J. C.; Postma, J. M. P.; van Gunsteren, W. F.; Hermans, J. In *Jerusalem Symposia on Quantum Chemistry and Biochemistry*; Pullman, B., Ed.; Reidel: Dordrecht, Holland, 1981.

termined by the charge distribution.[28] The water charge density is split into atomic contributions: spherically symmetric terms on the hydrogen derived from the contributions of the basis functions centered on the hydrogens only and all other contributions are assigned to the oxygen, resulting in an exact second-order multipole representation of the densities on the hydrogen[29] and by using tenth-order multipole expansion for the oxygen.

The induced energy E_I is obtained from the field of the permanent moments and the induced dipoles μ_i:

$$E_I = -[\sum_{i=1}^{N}(\mu \cdot \mathbf{E}^\circ)]/2 \tag{5}$$

The calculation of induced dipoles used the polarizability values of Liebman and Moskowitz:[30] 9.79, 10.14, and 9.66 au along the HOH bisector, H–H line, and the normal of the molecular plane, respectively. The dispersion energy E_D is obtained as the sum of $1/r^6$ terms:

$$E_D = \sum_{i<j}^{N} C_6^{OO}/(r_{ij}^{OO})^6 \tag{6}$$

and the repulsion energy E_R is obtained as a sum of $1/r^9$ and $1/r^{12}$ terms:

$$E_R = \sum_{i<j}^{N} \sum_{p,q=1}^{3} C_9^{pq}/(r_{ij,pq})^9 + C_{12}^{pq}/(r_{ij,pq})^{12} \tag{7}$$

where the indices p and q run through the water atoms.

Several extensions were made for the purpose of the present study. First, the calculations are supplemented with the same simple cubic periodic boundary conditions that were used in the simulation generating these configurations. Second, the multipole expansion at the oxygen was further truncated as a function of the distance between the oxygen atoms, resulting in large savings of computer time with negligible effect on the calculated energies. Finally, the calculation of the virial sum has also been implemented as follows. The virial sum V was used in two forms:[31]

$$V = \sum_{i=1}^{N}(\mathbf{R}_i \cdot \nabla_i E) = \sum_{i<j}^{N}(\mathbf{r}_{ij} \cdot \nabla_i e_{ij}) \tag{8}$$

The contribution of the permanent electrostatic energy terms can be easily computed from the second form since it is already in a form suitable for the Maxwell formalism and only requires the application of the operator $(r_{ij}\nabla)$ on each term calculated in the form of eq 5. Similarly, the virial contributions of a $1/r^x$ term (from E_D or E_R) are all in the form

$$-x(r_{ij}^0 r_{ij})/|r_{ij}|^{x-2} \tag{9}$$

where r_{ij}^0 is the intermolecular vector between the two molecules containing centers i and j at r_{ij}. An interesting implication of eq 9 is discussed in the Appendix. The contributions of the induced energy E_I were calculated by applying the first form of V, eq 8, to the expression:[21]

$$\nabla_k E_I = -\sum_{i=1}^{N}[\mu_i \cdot (\nabla_k E^\circ{}_i)] - \sum_{\substack{i=1\\ i\neq k}}^{N}[5(\mu_i \cdot \mathbf{r}_{ik}) \times (\mu_k \cdot \mathbf{r}_{ik})\mathbf{r}_{ik}/|\mathbf{r}_{ik}|^2 - (\mu_i \cdot \mu_k)\mathbf{r}_{ik} - (\mu_i \cdot \mathbf{r}_{ik})\mu_k - (\mu_k \cdot \mathbf{r}_{ik})\mu_i](3)/|\mathbf{r}_{ik}|^5 \tag{10}$$

The calculated viral sum was verified in two ways: first, by comparing the calculated V of a dimer with the value calculated by using finite difference approximations for the partial derivatives in eq 8, obtained by calculating the energies E_P, E_I, E_D, and E_R at systematically varied r_{ij}^0s; second, by a generating two conformations of 125 waters where the intermolecular distances were scaled up:

$$r_{ij}^0 \rightarrow (1 + \lambda)r_{ij}^0 \tag{11}$$

and using a finite difference approximation to the relation

$$\partial E/\partial\lambda = V = \Delta E/\Delta\lambda \tag{12}$$

to be compared with the value of V calculated from eq 8.

Once the cooperative energies of the selected configurations were calculated, the parameters of the effective pair potential were determined by minimizing the expression

$$\Delta^2 = w_E \Delta_E{}^2 + w_V \Delta_V{}^2 \tag{13}$$

where

$$\Delta_E{}^2 = \{\sum_{k}^{n}[E_k^c(\mathbf{X}^N) - \sum_{i<j}^{N} e^{\mathrm{eff}}(X_i,X_j)]^2\}/n \tag{14}$$

$$\Delta_V{}^2 = \{\sum_{k}^{n}[V_k^c(\mathbf{X}^N) - \sum_{i<j}^{N} v^{\mathrm{eff}}(X_i,X_j)]^2\}/n \tag{15}$$

Here the summation over k represents the summing over the n data points (configurations), $E_k^c(\mathbf{X}^N)$ and $V_k^c(\mathbf{X}^N)$ are the cooperatively calculated energy and viral sum, respectively, in the kth configuration, and $e^{\mathrm{eff}}(X_i,X_j)$ and $V^{\mathrm{eff}}(X_i,X_j)$ are the pairwise contributions to the energy and virial sum, respectively. They contain the potential parameters to be determined by the fitting procedure. The minization leads to a system of linear equations as long as the potential parameters are linear. The right-hand side of eq 15 is a variant of the Stillinger equation, eq 2. The exponential is not present explicitly, but since the configurations were selected from a simulation with a waterlike potential, it is included implicitly to a good approximation.

The quality of the fit can be conveniently characterized by the root mean square deviations Δ (a perfect fit would give zero), Δ_E, and Δ_V as well as the correlation coefficients

$$c_E = \langle(E^c - \langle E^c\rangle)(E^{\mathrm{eff}} - \langle E^{\mathrm{eff}}\rangle)\rangle/[\langle(E^c - \langle E^c\rangle)^2\rangle\langle(E^{\mathrm{eff}} - \langle E^{\mathrm{eff}}\rangle)^2\rangle]^{1/2} \tag{16}$$

$$c_V = \langle(V^c - \langle V^c\rangle)(V^{\mathrm{eff}} - \langle V^{\mathrm{eff}}\rangle)\rangle/[\langle(V^c - \langle V^c\rangle)^2\rangle\langle(V^{\mathrm{eff}} - \langle V^{\mathrm{eff}}\rangle)^2\rangle]^{1/2} \tag{17}$$

between the cooperatively calculated energies E^c and virial sums V^c considered to be "exact" and the fitted energies E^{eff} and virial sums V^{eff}, respectively. A "good" fit would give not only a small Δ but also large (i.e., approaching 1) correlation coefficients. Note that the parameters that minimize Δ are in general different from the ones that would maximize the analogous combination of the correlation coefficients. The adequateness of the new potential, however, can only be determined by applying the new potential in simulation although it is possible that after repeated application of the fitting procedure described here on different systems it will be possible to establish acceptability threshold values for the Δs and correlation coefficients.

Three potential types were tried in the fitting procedure: the Lennard-Jones plus electrostatic (LJQ)

$$e^{\mathrm{eff}}(X_i,X_j) = C_{12}^{OO}/R_{OO}^{12} - C_6^{OO}/R_{OO}^6 + E_{es}(q_H,r_{OQ}) \tag{18}$$

a potential containing oxygen–oxygen and oxygen–hydrogen repulsion plus electrostatic (RRQ) terms

$$e^{\mathrm{eff}}(X_i,X_j) = C_{12}^{OO}/R_{OO}^{12} + \sum C_{12}^{OH}/R_{OH}^{12} + E_{es}(q_H,r_{OQ}) \tag{19}$$

(the summation is over all four O–H distances), and the Morse plus electrostatic (MQ) form

$$e^{\mathrm{eff}}(X_i,X_j) = D\exp[-2A(R_{OO} - R_{eq})] - 2D\exp[-A(R_{OO} - R_{eq})] + E_{es}(q_H,r_{OQ}) \tag{20}$$

In eqs 18–20 the electrostatic term E_{es} is calculated between q_H electron charges placed on the hydrogens and the $-2q_H$ charge placed on a center placed on the HOH bisector at r_{OQ}-Å distance from the oxygen atom. The subscripts O, H, and Q refer to the interaction site oxygen, hydrogen, and charge center, respectively.

The total energy expressions using eqs 18–20 are linear in C_{12}^{OO}, C_6^{OO}, C_{12}^{OH}, D, and $q_H{}^2$ and nonlinear in R_{OQ}, R_{eq}, and A. In general, when there are p linear parameters (C_i, $i = 1, ..., p$), the

---

(28) Mezei, M.; Campbell, E. S. *J. Comput. Phys.* **1976**, *20*, 110.
(29) Mezei, M.; Campbell, E. S. *Theor. Chim. Acta* **1977**, *43*, 227.
(30) Liebman, S. P.; Moskowitz, J. W. *J. Chem. Phys.* **1971**, *54*, 3622.
(31) Barker, J. A.; Henderson, D. *Rev. Mod. Phys.* **1976**, *48*, 587.

TABLE I: Characterization of the Monte Carlo Runs Used for the Initial Fittings[a]

	MCY	TIPS2	TIP4P
no. of MC steps	3.4×10^6	10^6	1.2×10^6
no. of configurations extracted	229	68	80
av pairwise energy	−8.679	−9.882	−10.053
min pairwise energy	−9.035	−10.106	−10.566
max pairwise energy	−8.262	−9.573	−9.485
av cooperative energy	−12.053	−12.535	−12.564
min cooperative energy	−12.229	−12.363	−13.252
max cooperative energy	−11.572	−12.055	−12.456
correlation between the pairwise and cooperative energies	0.690	0.769	0.538
av pairwise virial sum			1.504
min pairwise virial sum			−0.958
max pairwise virial sum			4.359
av cooperative virial sum	12.271	7.025	−0.434
min cooperative virial sum	7.582	3.177	−6.936
max cooperative virial sum	16.311	10.988	5.621
correlation between the pairwise and cooperative energies			0.752

[a] Energies are in kcal/mol.

approximating energy expression for the kth configuration can be written in the condensed form

$$E_k^{\mathrm{eff}} = \sum_{i=1}^{p} C_i a_{ik} \qquad (21)$$

where a_{ik} depend on the coordinates of the atoms in the kth configuration and the assumed values of the nonlinear parameters. Δ of eq 13 is minimized when the C_i's are the solution of the system of linear equations

$$\sum_{i=1}^{p} \sum_{k=1}^{n} C_i (w_{\mathrm{E}} a_{ik} a_{jk} + w_{\mathrm{V}} b_{ik} b_{jk}) = \sum_{k=1}^{n} w_{\mathrm{E}} E_k^{\mathrm{c}} a_{jk} + w_{\mathrm{V}} V_k^{\mathrm{c}} b_{jk} \qquad (j = 1, ..., p) \quad (22)$$

Here the b_{ik}'s are the virial sum contributions corresponding to the energy terms represented by the a_{ik}'s.

For the determination of the optimal values of the nonlinear parameters ideally one would use a nonlinear minimizer search procedure, with gradients calculated from straightforward differentiation of eqs 18–20. However, as the likely range of R_{eq} and R_{OQ} is rather limited, a simple grid search was chosen here. This strategy was justified a posteriori by the relative insensitivity of Δ to the nonlinear parameters in eqs 18–20.

The fitting procedure started from configurations selected from the history of Monte Carlo simulations of 125 water molecules under simple cubic periodic boundary conditions with a 7.75-Å spherical cutoff at room temperature using the MCY,[5] TIP2,[6] and TIP4P[7] potentials. Data characterizing these runs are given in Table I.

Results

Results of Fitting the Cooperative Energies Only. In the first series of calculations $w_{\mathrm{V}} = 0$ was used; i.e., only the energy was fitted. The initial fittings were performed with the LJQ and RRQ forms and for different R_{OQ}'s. As the TIP2 and TIP4P potentials used 0.15 Å and the MCY potential 0.268 Å, it was expected that similar values will give the best fit. However, the cooperative energies were best fitted by R_{OQ} values around 0. Tables II and III give the results of the fits using the three different sets of configurations individually as well as their combinations with R_{OQ} = 0.15 and 0.0 Å and with the best (within 0.005 Å) R_{OQ} value. The combinations of the configuration sets were prepared in two ways: in the first version all configurations were used, and in the second version the MCY set was reduced by taking only every third of the 229 configurations. The two combinations were labeled MT24A and MT24B in the tables, respectively. The fits on these two sets were very similar, indicating that the number of configurations sampled was adequate.

As the different sets of configurations gave different but close to zero R_{OQ} values and the quality of the fit was not much affected by small changes in R_{OQ}, it was set to zero. The small value found optimal for R_{OQ} can be understood by remembering that the charge distribution used to develop the Campbell–Mezei model has a higher dipole moment than the experimental value. For the MQ form, additional optimization was performed to get the

TABLE II: Results of the Fitting Potentials of the Form LJQ to the Total Energy Only[a]

data	R_{QQ}	C_{12}^{OO}	C_6^{OO}	σ	ϵ	q_{H}	Δ_{E}	c_{E}
MCY	0.15	722586	1438.89	2.82	0.716	0.49631	0.0779	0.857
MCY	0.00	676321	1250.09	2.85	0.578	0.41897	0.0674	0.896
MCY	−0.01	672834	1246.44	2.85	0.577	0.41370	0.0674	0.896
TIPS2	0.15	866549	1130.90	3.03	0.369	0.56152	0.0973	0.867
TIPS2	0.00	807343	1065.36	3.02	0.351	0.45348	0.0755	0.923
TIPS2	−0.02	793626	1080.40	3.00	0.367	0.43896	0.0750	0.924
TIP4P	0.15	802151	1484.98	2.85	0.687	0.50566	0.1017	0.772
TIP4P	0.00	818939	1421.44	2.88	0.614	0.42033	0.0849	0.852
TIP4P	−0.09	809752	1485.24	2.86	0.681	0.37010	0.0823	0.862
MT24A	0.15	777694	1364.73	2.88	0.599	0.51651	0.0886	0.951
MT24A	0.00	781596	1600.13	2.81	0.819	0.39485	0.0814	0.959
MT24A	0.05	789032	1525.34	2.83	0.737	0.43114	0.0808	0.960
MT24B	0.15	792864	1311.14	2.91	0.542	0.52629	0.0945	0.944
MT24B	0.00	780623	1574.31	2.81	0.794	0.39743	0.0869	0.952*
MT24B	0.04	789597	1508.80	2.84	0.721	0.42764	0.0863	0.953
LJQ	0.00	702162	1489.46	2.79	0.790	0.39246	0.1084	0.845
LJQ	−0.11	705524	1493.93	2.79	0.791	0.34087	0.1062	0.851*
RRQ	0.00	307997	889.01	2.65	0.647	0.40843	0.1041	0.733
RRQ	−0.16	334032	940.13	2.66	0.661	0.33720	0.0991	0.759
MQ	0.00	732351	1612.90	2.77	0.888	0.38334	0.1220	0.791
MQ	−0.09	724013	1609.75	2.77	0.895	0.34183	0.1198	0.798
LJQ1	0.00	730650	1222.74	2.90	0.512	0.43851	0.1096	0.830
LJQ1	−0.05	727782	1222.53	2.90	0.513	0.40891	0.1090	0.833
RRQ1	0.00	508519	1735.52	2.58	1.481	0.35001	0.1049	0.693
RRQ1	−0.25	494728	1446.63	2.64	1.058	0.28118	0.0981	0.739
MQ1	0.00	613696	1394.35	2.76	0.792	0.38271	0.1668	0.516
MQ1	−0.07	605868	1370.49	2.76	0.775	0.35080	0.1658	0.523

[a] C_{12}^{OO} and C_6^{OO} are in kcal/(mol·Å^{12}) and kcal/(mol·Å^6), respectively; σ and ϵ are the Lennard-Jones parameters equivalent to C_{12}^{OO} and C_6^{OO}; the hydrogen charge q_{H} is in electrons; Δ_{E} (defined by eq 14) is in kcal/mol; c_{E} is the correlation defined by eq 16; and R_{OQ} is in Å. An asterisk marks the fits chosen for new simulations. Data labels refer to the potential function (see Results) that generated the dataset used in the fit.

TABLE III: Results of Fitting Potential of the Form RRQ to the Total Energy Only[a]

data	R_{OQ}	C_{12}^{OO}	C_{12}^{OH}	q_H	Δ_E	c_E
MCY	0.15	100933	7248.82	0.63155	0.0656	0.911
MCY	0.00	132228	3178.85	0.50905	0.0504	0.950
MCY	0.00	132228	3178.85	0.50905	0.0504	0.950
TIPS2	0.15	205228	4883.23	0.63511	0.0955	0.900
TIPS2	0.00	4422.69	3180.10	0.49498	0.0631	0.950
TIPS2	−0.02	464.061	3032.98	0.48789	0.0628	0.950
TIP4P	0.15	109400	4691.63	0.61803	0.1085	0.785
TIP4P	0.00	70172.1	2616.59	0.49497	0.0849	0.855
TIP4P	0.00	70172.1	2616.59	0.49497	0.0849	0.855
MT24A	0.15	366095	2900.09	0.63846	0.1040	0.936
MT24A	0.00	114361	2612.40	0.50077	0.0658	0.974
MT24A	−0.02	91305.3	2486.98	0.48671	0.0655	0.975
MT24B	0.15	382010	2858.24	0.64062	0.1096	0.928
MT24B	0.00	79154.6	2678.55	0.49722	0.0722	0.968*
MT24B	−0.01	65993.2	2615.69	0.48997	0.0721	0.969
LJQ	0.00	−82819	3405.12	0.49123	0.0946	0.876
LJQ	−0.01	−75573	3188.28	0.48469	0.0946	0.876
RRQ	0.00	75143.5	2778.64	0.49324	0.0727	0.883
RRQ	−0.08	108038	1708.73	0.45069	0.0708	0.889*
MQ	0.00	−80224	2371.81	0.48544	0.1078	0.838
MQ	−0.04	−78279	1877.91	0.45957	0.1071	0.841
LJQ1	0.00	45136.9	2624.55	0.50373	0.1025	0.848
LJQ1	−0.07	80893.7	1659.37	0.45955	0.1008	0.854
RRQ1	0.00	73114.4	2212.17	0.49289	0.0894	0.789
RRQ1	−0.04	95002.7	1675.70	0.46704	0.0887	0.793
MQ1	0.00	82176.7	817.25	0.46023	0.1657	0.569
MQ1	0.02	68562.4	964.67	0.47798	0.1655	0.569

[a] C_{12}^{OO} and C_{12}^{OH} are in kcal/(mol·Å^{12}); the hydrogen charge q_H is in electrons; Δ_E (defined by eq 14) is in kcal/mol; c_E is the correlation defined by eq 16; and R_{OQ} is in Å. An asterisk marks the fits chosen for new simulations. Data labels refer to the potential function (see Results) that generated the dataset used in the fit.

TABLE IV: Results of Fitting Potentials of the Form MQ to the Total Energy Only[a]

data	R_{OQ}	R_{eq}	A	D	q_H	Δ_E	c_E
MT24B	0.15	3.1	2.1	1.0788	0.4435	0.0975	0.942
MT24B	0.05	3.1	2.1	1.0873	0.4254	0.0886	0.951
MT24B	0.04	3.1	2.1	1.0897	0.4196	0.0886	0.951
MT24B	0.03	3.1	2.1	1.0919	0.4141	0.0886	0.951
MT24B	0.02	3.1	2.1	1.0958	0.4086	0.0890	0.950
MT24B	0.00	3.1	2.0	1.1557	0.3879	0.0907	0.948*
MT24B	0.00	3.1	2.2	1.0327	0.4079	0.0897	0.950
MT24B	0.00	3.0	2.1	1.0561	0.3581	0.0971	0.941
MT24B	0.00	3.2	2.1	0.7588	0.4319	0.0924	0.949
MT24B	0.00	3.2	2.2	0.6889	0.4409	0.0961	0.948
MT24B	0.00	3.2	2.0	0.8283	0.4218	0.0902	0.950
MT24B	0.00	3.0	2.2	1.5236	0.3670	0.0902	0.950
MT24B	0.00	3.0	2.0	1.5856	0.3492	0.1011	0.936
MQ	−0.05	2.9	2.9	1.3279	0.3561	0.1242	0.782
MQ	−0.05	3.0	2.9	0.8551	0.3947	0.1125	0.819
MQ	−0.05	3.1	2.9	0.4584	0.4281	0.1210	0.788
MQ	−0.05	2.9	3.0	1.2917	0.3607	0.1221	0.790
MQ	−0.05	3.0	3.0	0.7908	0.4003	0.1123	0.820
MQ	−0.05	3.1	3.0	0.4075	0.4316	0.1219	0.786
MQ	−0.05	2.9	3.1	1.2505	0.3654	0.1203	0.796
MQ	−0.05	3.0	3.1	0.7280	0.4053	0.1126	0.819
MQ	−0.05	3.1	3.1	0.3618	0.4346	0.1229	0.783
MQ	−0.01	3.0	3.0	0.7935	0.4208	0.1139	0.814
MQ	−0.07	3.0	3.0	0.7902	0.3907	0.1121	0.821
MQ	−0.08	3.0	3.0	0.7900	0.3861	0.1121	0.821*
MQ	−0.09	3.0	3.0	0.7900	0.3815	0.1121	0.821

[a] R_{OQ} and R_{eq}, and A are in Å; D is in kcal/mol; the hydrogen charge q_H is in electrons; Δ_E (defined by eq 14) is in kcal/mol; and c_E is the correlation defined by eq 16. An asterisk marks the fits chosen for new simulations. Data labels refer to the potential function (see Results) that generated the dataset used in the fit.

best value of A and R_{eq}. Tables II–IV contain the results of the fits.

In the first iteration, three force-biased[32] Metropolis[33] Monte Carlo simulations (1.8×10^6 MC steps each after equilibration) were performed by using the parameters from the initial fits in the LJQ, RRQ, and MQ forms. One hundred twenty configurations were extracted from each of the simulations, and their cooperative energies were calculated. These sets will be referred to as LJQ, RRQ, and MQ. As the calculations of the cooperative energies and the fitting procedure used the minimum image convention (i.e., no cutoffs), while the Monte Carlo calculations used a spherical cutoff, the calculation with the LJQ fit was repeated under the minimum image convention (labeled LJQM) to assess the possible effect of this inconsistency.

---

[32] Metropolis, M.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087.

[33] Rao, M.; Pangali, C.; Berne, B. J. *Mol. Phys.* **1979**, *37*, 1773.

TABLE V: Results of Simulations with the Effective Potentials Fitted to the Energies Only[a]

data	$-E^{\mathrm{eff}}$	p	C'_v	K	r_{M1}	g_{M1}	r_{M2}	g_{M2}	$-E^{\mathrm{c}}$	c_{E}
LJQ	13.18	−3969	13.1	6.3	2.80	3.70	5.50	1.30	13.04	0.79
LJQM	13.82	−5897	12.1	6.2	2.80	3.64	5.50	1.29	14.89	0.79
RRQ	14.30	3530	9.6	4.7	2.70	4.20	4.90	1.18	13.02	0.82
MQ	14.49	−3780	17.3	7.1	2.80	3.97	5.50	1.37	14.23	0.67
LJQ1	13.33	6530	11.5	6.7	2.80	4.15	5.30	1.67	13.36	0.78
RRQ1	16.06	−3592	12.2	4.6	2.65	4.44	4.80	1.20	14.63	0.73
MQ1	17.47	−3835	11.2	6.4	2.60	4.03	5.60	1.21	14.40	0.45

[a] E^{eff} and E^{c} are the calculated average internal energy and the calculated cooperative energy of 120 configurations generated by the new effective potential in kcal/mol, respectively; p is the liquid pressure in atm; C'_v is the configurational contribution to the constant-volume heat capacity; K is the coordination number defined as the average number of neighbors within 3.3 Å; r_{M1} and r_{M2} are the locations of the first and second peaks in the water–water radial distribution function $g(r)$, respectively; and c_{E} is the correlation between E^{eff} and E^{c}. Data labels refer to the potential function (see Results) that generated the dataset used in the fit.

In the second iteration, the three different functional forms were again fitted to the various datasets. Three new simulations were run, using the parameters fitted to the corresponding datasets. As at this stage the optimal value of R_{OQ} seemed to differ more from zero than at the initial stage, the optimal value was used. The datasets generated in the second iteration are referred to as LJQ1, RRQ1, and MQ1.

Tables II–IV also contain the results of the fits in the first and second iterations. The simulations performed in the first and second iterations are described in Table V for the LJQ, RRQ, and MQ forms. The table gives the calculated energy, pressure, the configurational contribution to the constant-volume heat capacity, the coordination number, and the locations and heights of the first two peaks of the water–water radial distribution function $g(r)$. "Waterlike" behavior requires a coordination number between 4 and 5 as well as a weak second peak that is located at about 1.4 times the first peak's distance. The average of the cooperatively calculated energies of the configurations selected and its correlation with the effective potential energies calculated during the simulation is also given to assess the convergence of the potential generation process. The LJQM calculation using the minimum image convention instead of the spherical cutoff showed no significant difference in the calculated structural properties although the waters outside the cutoff spheres contributed significantly to the energy and pressure—this is not surprising since the cutoff radius used was relatively small (dictated by the number of molecules).

Following the best fit parameters for the LJQ and RRQ forms from the initial stage through the two iterations, the calculated parameters show a tendency toward convergence although they cannot be called converged yet when the $R_{\mathrm{OQ}} = 0$ fits are compared. (For the MQ form the three nonlinear parameters make such comparison more difficult.)

Comparing the simulation results, the two MC simulations using the LJQ form gave reasonably similar structural and energetic results except that the pressures from the two simulations are very different. The RRQ simulations show some drift toward lower energy and shorter water–water distances but no drastic jump in the pressure. Additionally, it gives the most waterlike structure among the three in terms of coordination number and $g(r)$ second peak position. However, the first peak's position is significantly smaller than the experimental value, 2.8 Å.[34] The MQ simulations also show a drift toward lower energies, but the pressure and the structural indicators show reasonable convergence. The significant differences between the simulated liquids using the different functional forms are also highlighted by the large differences in the fitting parameters obtained for a given functional form from the fits on the datasets of the simulations with the different functional forms.

An unexpected result from the fitting procedure was the near singularity of the linear equation system's matrix. For the LJQ form, the rows corresponding to the r^{-6} and to the r^{-1} terms were close to being linearly dependent, indicating that the inclusion of the r^{-6} term is very nearly superfluous. For the RRQ form, the rows corresponding to the O–O and O–H repulsions behaved similarly, giving rise to analogous conclusions.

Thus the results suggest that the proposed procedure is likely to converge, even though it may take more than the two iterations performed here. The energies are consistently fitted with ca. 1% error, irrespective of the functional form used. The ranking of the functional forms varied from iteration to iteration; thus, on this basis no clear winner emerged. However, the deep sensitivity of the calculated structure to the functional form chosen indicates that fitting the energies only is insufficient.

The correlation between the energies calculated with the effective potential and the cooperative potential (calculated after the MC run on the selected configurations) is consistently lower than the corresponding correlations on the set of configurations where the fitting was done (calculated during the fitting). This indicates that the effective potential allows the simulation to sample regions of the configuration space that differ significantly from the regions sampled in the previous iteration, decreasing the confidence in the result. This is true even for the LJQ fits where the cooperatively calculated energies are remarkably close to the effective energies used to generate the configurations.

The apparent tendency of the simulation with a newly fitted effective potential based on the cooperative energies alone to move into newer regions of the configuration space may hold the clue to the strong dependence of the simulated structure on the functional form used since the energies of the "newer" configurations are obtained essentially by extrapolation, which is clearly more sensitive to the functional form. Inclusion of the virial sum into the fitting thus can be expected to help "localize" the iteration process since the virial sum, eq 8, depends on the derivatives of the energy.

Results of Fitting the Cooperative Energies and Virial Sums. The fitting of the energy and virial sum combination again started from the original three sets of configurations. The weight factors w_{E} and w_{V} were selected empirically in a two-step process. At first

$$w_{\mathrm{E}} = w'_{\mathrm{E}}/\langle -E^{\mathrm{eff}}\rangle \tag{23}$$

$$w_{\mathrm{V}} = w'_{\mathrm{V}}/\max\,(|V^{\mathrm{eff}}|,\ V^{\mathrm{eff}}_{\max} - V^{\mathrm{eff}}_{\min}) \tag{24}$$

was used with both w'_{E} and w'_{V} set to 1.0 in an attempt to factor out the difference in magnitude between the energy and virial sum. This consistently yielded much better correlations for the virial sum than for the energies. It was found that raising w'_{E} to 5.0 yielded much better correlations for the energies without significantly deteriorating the virial fit, and these w values were kept for the subsequent fits. The fit indicators showed little change once w'_{E} was set above 5. For the combined dataset MT24B this gave $w_{\mathrm{E}} = 0.9465$ and $w_{\mathrm{V}} = 0.0535$. The near singularity mentioned in the previous section showed up less frequently (only for fits using near-zero R_{OQ}).

Following the pattern established at the calculations using the energy fit only, two iterations were performed by using each of the three functional forms. The resulting sets of configurations were labeled LJQV, RRQV, and MQV for the first iteration and LJQV1, RRQV1, and MQV1 in the second. In each iteration, the optimal R_{OQ} value was established first for the LJQ and RRQ forms (they consistently came out to be the same in any given

---

[34] Soper, A. K.; Philips, M. G. *Chem. Phys.* **1986**, *107*, 47.

TABLE VI: Results of Fitting Potentials of the Form LJQ to the Energies and Virial Sums[a]

data	R_{OQ}	C_{12}^{OO}	C_6^{OO}	q_H	Δ	Δ_E	c_E	Δ_V	c_V
MT24B	0.21	827343	1914.73	0.49429	0.312	0.147	0.92	1.173	0.98*
LJQV	−0.10	850960	2380.63	0.22474	0.515	0.167	0.59	2.572	0.56*
RRQV	−0.10	731224	1471.03	0.36536	0.291	0.118	0.83	1.905	0.76
LJQV1	−0.11	584777	1598.28	0.28754	0.704	0.328	0.50	10.91	0.09*
RRQV1	−0.25	552379	873.017	0.34063	0.830	0.163	0.79	1.576	0.76

[a] C_{12}^{OO} and C_6^{OO} are in kcal/(mol·Å^{12}) and kcal/(mol·Å^6), respectively; the hydrogen charge q_H is in electrons; Δ, Δ_E, and Δ_V, defined by eqs 13–15, are in kcal/mol; c_E and c_V are the correlations defined by eqs 16 and 17; and R_{OQ} is in Å. An asterisk marks the fits chosen for new simulations. Data labels refer to the potential function (see Results) that generated the dataset used in the fit.

TABLE VII: Results of Fitting Potentials of the Form RRQ to the Energies and Virial Sums[a]

data	E_{OQ}	C_{12}^{OO}	C_{12}^{OH}	q_H	Δ	Δ_E	c_E	Δ_V	c_V
MT24A	0.21	75266.0	1469.01	0.64751	0.364	0.208	0.85	0.208	0.98*
LJQV	0.05	47975.3	1974.50	0.53227	0.582	0.157	0.68	2.956	0.53
RRQV	−0.10	197498	1244.89	0.42733	0.218	0.096	0.89	1.406	0.86*
LJQV1	−0.10	−281127	6777.09	0.43401	0.687	0.306	0.57	10.76	0.23
RRQV1	−0.10	127690	1239.79	0.42167	0.647	0.132	0.88	1.227	0.85*
RRQV2	−0.10	149200	1073.48	0.42090	0.645	0.115	0.78	1.532	0.76
RRQV2	0.00	135994	1103.67	0.47855	0.642	0.115	0.78	1.524	0.76
RRQV3	−0.10	138800	1118.65	0.41964	0.625	0.110	0.82	1.504	0.82
RRQV3	0.06	119351	1153.37	0.52001	0.613	0.127	0.75	1.478	0.82

[a] C_{12}^{OO} and c_{12}^{OH} are in kcal/(mol·Å^{12}); the hydrogen charge q_H is in electrons; Δ, Δ_E, and Δ_V, defined by eqs 13–15, are in kcal/mol; c_E and c_V are the correlations defined by eqs 16 and 17; and R_{OQ} is in Å. An asterisk marks the fits chosen for new simulations. Data labels refer to the potential function (see Results) that generated the dataset used in the fit.

TABLE VIII: Results of Fitting Potentials of the Form MQ to the Energies and Virial Sums[a]

data	R_{OQ}	R_{eq}	A	D	q_H	Δ	Δ_E	c_E	Δ_V	c_V
MT24B	0.21	3.1	2.3	0.7408	0.5740	0.384	0.132	0.91	1.565	0.97
MT24B	0.21	3.1	2.1	1.0254	0.5436	0.342	0.113	0.93	1.401	0.98
MT24B	0.00	3.1	2.1	1.1099	0.3978	0.384	0.132	0.91	1.565	0.97
MT24B	0.00	3.1	2.0	1.3108	0.3794	0.304	0.111	0.94	1.227	0.98*
MQV	0.00	3.1	2.0	0.9048	0.3959	0.803	0.167	0.71	1.822	0.76
MQV	−0.10	3.1	2.0	0.8281	0.3487	0.780	0.147	0.73	1.768	0.78*
MQV1	0.00	3.0	1.8	1.4143	0.3419	0.814	0.193	0.68	2.139	0.68
MQV1	−0.55	3.1	1.8	1.4062	0.1933	0.796	0.190	0.68	2.092	0.70
MQV1	−0.55	3.1	1.7	1.7077	0.1722	0.792	0.201	0.63	2.073	0.69

[a] R_{OQ}, Re_q, and A are in Å; D is in kcal/mol; the hydrogen charge q_H is in electrons; Δ, Δ_E, and Δ_V, defined by eqs 13–15, are in kcal/mol; and c_E and c_V are the correlations defined by eqs 16 and 17. An asterisk marks the fits chosen for new simulations. Data labels refer to the potential function (see Results) that generated the dataset used in the fit.

TABLE IX: Results of Simulations with the Effective Potentials Fitted to the Energies and Virial Sums[a]

data	$-E^{eff}$	V^{eff}	p	K	r^g_{M1}	g_{M1}	r^g_{M2}	g_{M2}	$-E^c$	c_E	V^c	c_V
LJQV	13.43	8.23	−4883	7.2	2.80	4.16	5.60	1.42	12.46	0.64	6.83	
RRQV	15.45	4.16	−1811	4.4	2.65	5.12	4.45	1.33	12.53	0.48	−39.6	
MQV	14.53	7.18	−4106	7.3	2.80	3.96	5.50	1.41	14.10	0.61	6.27	
LJQV1	12.54	9.50	−5873	8.3	2.85	5.32	5.10	1.59	11.56	0.45	12.87	0.13
RRQV1	14.44	1.01	535	4.5	2.65	4.26	4.75	1.17	14.56	0.79	−7.85	0.83
MQV1	14.82	0.00	−3552	6.9	2.70	3.57	5.40	1.32	14.09	0.69	13.31	0.61
RRQV2	14.95	3.90	−1614	4.5	2.65	4.47	4.75	1.21	14.42	0.82	4.39	0.77
2SD	0.1		240	0.4		0.1		0.2				
RRQV3	14.86	3.41	−1243	4.6	2.65	4.47	4.75	1.17	14.30	0.78	14.59	0.83
2SD	0.1		533	0.04		0.05		0.01				

[a] E^{eff} and E^c are the calculated average internal energy and the calculated cooperative energy of 120 configurations generated by the new effective potential in kcal/mol, respectively; p is the liquid pressure in atm; V^{eff} and V^c are the calculated average virial sum and the calculated cooperative virial sum of the 120 configurations generated by the new effective potential in kcal/mol, respectively; K is the coordination number defined as the average number of neighbors within 3.3 Å; r^g_{M1} and r^g_{M2} are the locations of the first and second peaks in the water–water radial distribution function $g(r)$, respectively; c_E is the correlation between E^{eff} and E^c; c_V is the correlation between V^{eff} and V^c; and 2SD is the 95% confidence interval estimated from the Monte Carlo simulation block averages. Data labels refer to the potential function (see Results) that generated the dataset used in the fit.

iteration, although there is no guarantee for it in general). For the MQ form, the other nonlinear parameters were optimized with this R_{OQ} value, and R_{OQ} was reoptimized afterward. The salient characteristics of the fits are collected in Tables VI–VIII, and the simulation results are given in Table IX.

At the end of the second iteration, there were significant differences in the performance of the three functional forms. The best fit to the LJQV1 dataset showed a marked deterioration from the previous iterations and was much worse than the other two. Similarly, the MQ fit worsened after the second iteration and gave significantly larger Δ values than the RRQ fit. On these grounds, the LJQ and MQ forms were eliminated from further considerations, and a third iteration was performed by using the RRQ form (to establish the level of convergence). This simulation was twice as long as the previous ones, and the resulting set of configuration was analyzed in two halves, labeled RRQV2 and RRQV3. These results are also shown in Table VII.

Comparison of the RRQV2 and RRQV3 results and their calculated errors show the adequateness of the 1800K long runs. Comparison of the RRQV2 and RRQV3 results with the previous iterations shows that the coordination number, the location of the first and second peak of $g(r)$, the height of the second peak of the $g(r)$, and the energy and virial correlations between the cooperative and effective values converged. While the fluctuations from iteration to iteration in the energy, pressure, and first peak height of the $g(r)$ are significantly larger then their error, the RRQV2 and RRQV3 values are always between the first and second iteration values. Furthermore, unlike the fits using the energy only, the correlations between effective and cooperative energies and virials calculated during the fit were close to the corresponding correlations calculated on the newly generated configuration set.

Comparison of the calculated liquid properties (based on the final fit) with experiment shows that the basic water characteristics are present: the coordination number is around 4, and the second peak of the O–O radial distribution function is closer than the double of the first peak distance (although not as close as it should be) and is very small. However, the first peak of the O–O radial distribution function is about 0.15 Å closer than the experimental value and is too high, the calculated energy is more negative than the experimental (by 5 kcal/mol if quantum corrections are neglected), and the pressure is somewhat off (by ca. 0.5 kcal/mol in terms of pV). Altogether, these results are not too bad for a water model that was developed essentially completely ab initio.

Conclusions

The calculations described demonstrate the difficulty of providing a robust pairwise approximation to a cooperative model for water. Unless both the energies and the virial sums are fitted, different functional forms yielding equally good fits can result in simulated liquids that are qualitatively different. However, when the virial sum was also included into the fitting procedure, the functional forms under study showed significant differences, allowing a meaningful choice of best fit. Also, the fit parameters and the calculated liquid properties showed reasonable convergence.

The study described here focused on the fitting aspect of the problem. The fitting procedure was independent of the special features of the cooperative potential used. Therefore, it is expected that similar procedure would be applicable to different cooperative models, including semiempirical or, ultimately, ab initio calculations.

The fit can also be analyzed for the significance of the difference between the pairwise and cooperative energies. On the 120 configurations of the RRQV1 calculation the root mean square between the cooperative and fitted energies was 0.132 kcal/mol, and the largest deviation was 0.38 kcal/mol. As there were 125 molecules in the system, these values translate into a 10^{12} and 10^{34} correction factor, respectively, to the Boltzmann factor of the configuration involved. Thus it is safe to conclude that the configurations sampled by the two different potentials would be different. To quantify this difference, the energy changes resulted from random molecular displacements were calculated with center of mass displacement limited to 0.1 Å and angular displacment limited to 20°. It was found that with these displacements (averaging 0.05 Å and 10°) the average energy change is 1 kcal/mol. Thus it is reasonable to conclude that the cooperative energies can be reproduced by small perturbations of the configuration (well under 0.1-Å translational and 10° rotational displacement for each molecule).

Acknowledgment. This work was supported under RCMI Grant SRC5G12RR0307 from NIH to Hunter College and a CUNY/PSC grant. Computing resources were provided by the City University of New York, University Computing Center. The idea of this paper occurred to the author during a conversation with Dr. Richard Friedman. The manuscript significantly benefited from discussions with Prof. Edwin S. Campbell. The skillful programming of the least-squares fit by Mr. Boris Pekhman is also acknowledged.

Appendix. Efficient Calculation of the Pressure in the Canonical Ensemble Monte Carlo for Inverse Power Central Force Models

The pressure in the canonical ensemble is obtained as a function of the virial sum V:[31]

$$P = kT(N - V/3kT)/v \tag{A1}$$

where k is the Boltzmann factor, T is the absolute temperature, v is the volume, and V is the virial sum given by eq 8. In general, calculation of the virial sum requires the calculation of the forces on the particles, a nonnegligible amount of extra work (unless the force-biased displacement scheme[33] is used where the forces are also needed anyway), and is thus rarely done. However, if e_{ij} depends only on $|r_{ij}|$ (i.e., there is only one interaction center per particle) and the interaction follows an inverse power law (or is a sum of inverse power terms), then the contribution of particles i and j to the virial sum can be obtained with negligible extra work, since in this case eq 9 reduces to

$$(\mathbf{r}\cdot\nabla|\mathbf{r}|^{-x}) = -x|\mathbf{r}|^{-x} \tag{A2}$$

Thus the calculation of the virial sum in this case requires only the separate accumulation of the contributions to the total energy from the various distance powers during the simulation and their multiplication with the corresponding exponent x after the simulation.

Registry No. Water, 7732-18-5.