Deriving Intermolecular Potential Functions for the Water Dimer from ab Initio Calculations[1]

William L. Jorgensen

Contribution from the Department of Chemistry, Purdue University, West Lafayette, Indiana 47907.
Received September 21, 1978

Abstract: An intermolecular potential function for the water dimer is developed from ab initio molecular orbital calculations with a minimal basis set. This allows the thorough study of the basis set dependence of Monte Carlo simulations of liquid water. General techniques for generating intermolecular potential functions are presented. A procedure for increasing the occurrence of dimer configurations with low energy when they are selected in a random fashion is discussed and utilized. A careful analysis of the forms of potential functions for the water dimer reveals several interesting observations. First, it is found that the quality of fits using 12-6-3-1 potentials is the same as for functions with exponentials ($\alpha e^{-\beta r}$) for the short-range interactions. However, the 12-6-3-1 functions reduce the computation times for condensed-phase simulations by 11%. It is further established that models for a water monomer with four point charges are preferable to alternatives with three charges. Distinct improvement in the predicted geometry and dimerization energy for the linear water dimer are obtained in this way. Testing procedures including searches for unwanted minima in potential functions are applied. The final four point charge 12-6-3-1 potential is found to provide an excellent representation of the potential surface for the water dimer from STO-3G calculations. In all, 291 points on the surface were considered. The standard deviations for the fit of the 12-6-3-1 function are ca. 0.3 kcal/mol.

In view of the chemical and biological importance of water, it has been the focus of theoretical studies of molecular liquids.[3] A key component in the research is the use of intermolecular potential functions for the water dimer. Numerous functions have been developed either from empirical or quantum mechanical origins. In the empirical approach, sums of standard functions, such as Coulomb and Lennard-Jones 6:12, are parametrized to reproduce experimental properties of water, e.g., multipole moments, geometries, hydrogen bond energies, lattice energy, etc. Some famous examples are the Rowlinson,[4a] Ben-Naim and Stillinger (BNS),[4b] and ST2 potentials.[4c] A new series of "central force" potentials has appeared which also permits vibrational motion.[5]

Quantum mechanical potentials are derived by using ab initio molecular orbital methods to compute dimerization energies for many orientations of the dimer. The results are then fit to functional forms similar to those in the empirical potentials. An obvious advantage of the quantum mechanical approach is that systems can be treated for which there is little or no experimental data. A complication is that the potentials depend on the basis set quality and methods for correlation energy corrections in the ab initio calculations. Consequently, Clementi and co-workers have developed potentials for the water dimer at several levels of theory.[6] Their first effort involved a large polarization basis set which yielded a potential function representative of near Hartree-Fock (HF) level calculations.[6a] This function was subsequently modified by a variety of empirical corrections for the correlation energy.[6b] However, Monte Carlo simulations with these potentials predicted too little structure for liquid water beyond the first solvation shell. The problem was remedied by a potential function derived from ab initio calculations including explicit corrections for the correlation energy via configuration interaction (CI).[6c]

Several important issues remain to be resolved with regard to generating the quantum mechanical potentials. First, if it is essential to perform calculations at the CI level to obtain reasonable results, the approach will be restricted to small systems with limited study of their potential surfaces owing to the costliness of the computations. The HF results for liquid water were discouraging; however, there is no reason to expect a smooth enhancement of the results with increasing sophistication in the ab initio calculations. In fact, the problems with the HF potential could largely be attributable to the function's high estimate of the dimerization energy for the water dimer (−4.55 kcal/mol) in comparison to the CI value (−5.84 kcal/mol). This made it particularly intriguing to note that minimal basis set ab initio calculations without CI also yield dimerization energies of ca. −6 kcal/mol.[7] Consequently, it seemed worthwhile to generate a potential function from minimal basis set calculations and use it in simulations of liquid water. At least a more complete picture of the basis set dependence of the simulations could then be obtained. The results of such studies are presented here and in the accompanying paper. The findings are encouraging because the structural and thermodynamic results are in far better agreement with experiment than from the HF potential and are in fact comparable in quality to the data from the best empirical functions. The CI potential still provides the closest agreement with experiment for the structure of liquid water. Nevertheless, it is established that reasonable intermolecular potential functions for hydrogen-bonded molecules containing second-row atoms can be obtained from minimal basis set calculations. Extension to larger systems and the development of a preliminary trimer potential for water at reasonable cost are possible directions for future research.

During the course of this work, numerous issues of general interest arose concerning the formulation of quantum mechanical potentials. Topics that receive attention below include the selection and number of dimer geometries, proper and efficient forms for the potentials, choice of point-charge models, and unwanted minima.

Selection of Dimer Geometries

The genesis of a quantum mechanical potential (QMP) requires four main elements: selecting the geometries for the dimers, performing the ab initio calculations, fitting the computed dimerization energies to a functional form, and testing the function. The process is often iterative. Since a series of such undertakings is planned in this laboratory, careful consideration has gone into establishing general procedures. The application of the techniques is described here for each step in the development of an intermolecular potential for the water dimer.

Until now there have been two principal ways to obtain the dimer geometries. A grid search can be performed in which the range of each geometric variable is broken into increments and each combination of values for the variables then yields one geometry. The problems here are that the number of resultant geometries increases exponentially with the number of vari-

0002-7863/79/1501-2011$01.00/0

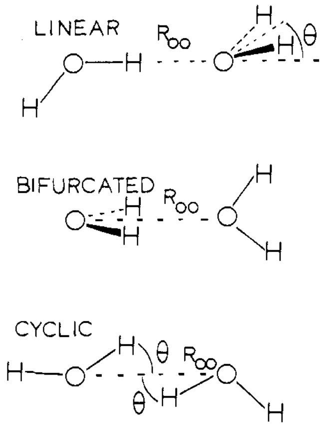


Figure 1. Definition of geometric variables for configurations of the water dimer.

ables and most combinations of the variables yield geometries that would rarely be sampled in a condensed-phase simulation. As a consequence of the first problem, Yarkony et al. used an angular variation of 45° in their grid for the HF dimer.[8] Clearly this implies that strongly hydrogen bonded configurations could not be well represented: only 8 of the 294 points computed by Yarkony et al. were bound by more than 3 kcal/mol. Clementi et al. have used nonuniform grids to help focus on more interesting regions of the water dimer potential surface.[6]

A clever alternative was proposed by Beveridge et al.[9] Briefly, the method is a statistically based procedure which begins by fitting the potential function to an initial set of *randomly* generated dimer configurations. The potential's predictive ability is then checked with a new sample of computed configurations. A new fit is made to the expanded data base, additional random geometries are generated, and the procedure is repeated until the predictive ability of the potential converges. Beveridge obtained potentials for water–formaldehyde and water–methane in this manner.[9] The author also used the procedure for the HF dimer.[2a] However, it was found to be desirable in the latter work to augment the 200 randomly generated points with a set of 50 dimer geometries that were anticipated to have low energies. This was necessary to guarantee that linear dimers were more bound than cyclic forms. The difficulty stems from the fact that randomly generating the dimer geometries also does not produce many low-energy points. A further illustration is provided by the following observation. A program that randomly produces dimer geometries was used in conjunction with Clementi's CI potential. The range of the OO separation was restricted from 2.38 to 4.50 Å and all angular variations were allowed. Only 20 in 11 700 or 1 in 585 random geometries yielded energies in the range −5.8 to −4.6 kcal/mol; 1 in 65 were between −4.6 and −3.4 kcal/mol.

To generate the water dimer geometries for the present work, a straightforward modification of Beveridge's procedure has been made. Instead of producing the points completely randomly, the CI potential has been used to provide 200 geometries with energies evenly distributed in the range −5.8 to 6.2 kcal/mol. The range was broken into 10 bins and 20 geometries with energies in each bin were randomly selected. As indicated, 11 700 points had to be considered to obtain the 200 "energy-distributed" dimer geometries. In addition, the 66 geometries used by Clementi to generate the CI potential were added to the initial data base. These points had been selected carefully; furthermore, this would permit an analysis of whether or not the 66 points by themselves could be expected to generate a comprehensive potential (vide infra). In all aspects of this work, the water monomers have been held rigid in their experimental geometries: $r(\mathrm{OH}) = 0.9572$ Å, $\angle\mathrm{HOH} = 104.52°$.[10]

Table I. Optimized Geometries and Dimerization Energies for $(H_2O)_2$[a]

	optimized STO-3G[b]	CI potential	fixed STO-3G[c]
		Linear	
r(OO)	2.73[d]	2.88	2.70
θ	58.7	30	52
$-\Delta E$	6.0	5.84	6.41
		Bifurcated	
r(OO)		2.96	2.78
$-\Delta E$		3.49	3.46
		Cyclic	
r(OO)		2.90	2.78
θ		46	53
$-\Delta E$		3.75	2.06

[a] Distances in angstroms; angles in degrees; energies in kcal/mol. [b] Data from ref 7. [c] The water monomers have been held fixed in their experimental geometries. [d] Experimental values for the linear dimer are 2.98, 60, and 5.2 ± 1.5, respectively: ref 15 and 16.

The procedure for obtaining energy distributed random geometries can often be adopted. Crude potential functions could be constructed for many intermolecular interactions. For example, the water–amino acid functions developed by Clementi could be modified appropriately.[11] Or the potentials for force-field calculations may be adequate.[12] Clearly, the closer the potential being used to generate the points is to the potential being constructed, the better the geometries will be distributed in energy. Therefore, iterative procedures could be envisioned in which an initial set of geometries is used to construct a crude potential which is then used to generate the next set of geometries, etc.

Ab Initio Calculations

As mentioned above, it was decided to obtain a QMP for the water dimer from minimal basis set calculations. Consequently, the GAUSSIAN/74 program was utilized with the STO-3G basis of Pople et al.[13] The computations were executed on the dual CDC/6500 system at Purdue. Although the dimerization energy seemed fine, concern could be expressed for the geometry of the dimer from STO-3G results. The predictions from STO-3G calculations and the CI potential are summarized in Table I. The geometric parameters for the linear, cyclic, and bifurcated forms of the water dimer are reviewed in Figure 1. As shown, the OO distance from fully optimized STO-3G data is 0.15 Å shorter than the result from the CI potential.[14] The experimental value is an additional 0.1 Å longer (2.98 ± 0.04 Å); however, this result corresponds to ambient temperature as compared to 0 K for the theoretical estimates.[15] Since the CI potential yields the correct position for the first peak in the OO radial distribution function for liquid water,[6c] an STO-3G based potential may be expected to err by about 0.15 Å. This difference seemed tolerable; in addition, θ for the linear dimer is in much closer agreement with experiment (60°)[15] from STO-3G results (59°) than the CI potential (30°).

Since the water monomers in the calculations for the STO-3G QMP have been fixed in their experimental geometries, the proper point of comparison for the potential is with optimized STO-3G calculations constrained in the same fashion. Thus, the three illustrated dimer geometries were optimized using STO-3G calculations with the monomers rigid. The results are given in the last column of Table I. Surprisingly, the dimerization energy dropped to −6.41 kcal/mol.

Table II. Standard Deviations for Fitting the STO-3G Dimerization Energies to Potential Functions[a]

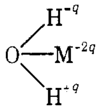
	potential function			
	exp	12-6-3-1	12-6-3-1	12-6-3-1
no. points	266	266	266	291
no. point charges	3	3	4	4
α^b	20	20	20	100
σ	0.35	0.35	0.37	0.39
$\sigma_{<0}$	0.34	0.34	0.32	0.32
$\sigma_{<-2}$	0.40	0.39	0.36	0.35
$\sigma_{<-4}$	0.45	0.51	0.47	0.34

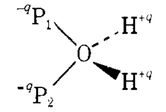
[a] Standard deviations in kcal/mol. [b] Coefficient for weighting function.

Apparently, the monomers gain more from complete optimization than the dimer. This worsens the agreement with the CI value. Since in Monte Carlo simulations the CI potential yields an energy for liquid water that is too positive, the discrepancy could be desirable. However, it might create too much structure in the radial distribution functions for the liquid.

Potential Functions

After the 266 STO-3G calculations were performed, the geometries and energies were stored for use by a nonlinear least-square program to obtain fits to trial potentials. A variety of functional forms were considered. Point charge models for water commonly involve three or four charges. The former was used by Clementi in the HF and CI potentials and features a center of negative charge whose position along the HOH bisector is optimized. The four-point alternative has two pseudo-lone pairs that are oriented tetrahedrally (∠P1OP2 = 109.47°) as in the ST2 potential.[4c] The distance between oxygen and the pseudo-lone pairs may be optimized. The point charges establish the Coulombic part of the potential for the dimer. In both cases





$$\Delta E_{\text{coul}} = \sum_{i<j}^{N} \frac{q_i q_j}{r_{ij}}$$

where the sum is over all pairs of point charges between the two monomers and N is the number of charges.

The other obvious component is terms for short-range interactions. The Lennard-Jones 6:12 form[4] or exponential functions ($\alpha e^{-\beta r}$)[5,6] are usually employed. In the earlier study of $(HF)_2$ and in this work, 12-6-3-1 potentials have proven to be most useful.[2a] The power series suggests that Coulomb, dipole, and short-range interactions are considered. The

$$\Delta E(\text{12-6-3-1}) = \sum_{i<j}^{N} \frac{q_i q_j}{r_{ij}} + \sum_{\mu<\nu}^{M} \left(\frac{b_{mn}}{r_{\mu\nu}{}^3} + \frac{c_{mn}}{r_{\mu\nu}{}^6} + \frac{d_{mn}}{r_{\mu\nu}{}^{12}} \right)$$

Coulombic term is the same as before; however, the second double sum is over pairs of atoms (μ,ν) with types m and n in the different monomers. The adjustable parameters are then the charge, q, and b, c, and d for each type of atomic interaction (OO, OH, and HH) and the distance of the negative point charges from the oxygen. As discussed previously,[2a] it is also necessary to constrain the $1/r^3$ terms such that they cancel at large r. This minimizes dependence on the sample size in condensed-phase simulations using the potential. The requirement for the water dimer is $b_{OO} + 4b_{OH} + 4b_{HH} = 0$ or $b_{OH} = -(b_{OO} + 4b_{HH})/4$. Consequently, the number of adjustable parameters for the 12-6-3-1 functions is ten, which is exactly the same as for the CI potential.

Table III. Geometric Results for Potential Functions[a]

	potential function			
	exp	12-6-3-1	12-6-3-1	12-6-3-1
no. points	266	266	266	291
no. point charges	3	3	4	4
α	20	20	20	100
	Linear			
r(OO)	2.65	2.63	2.64	2.66
θ	19	20	49	51
$-\Delta E$	5.99	6.10	6.42	6.46
	Bifurcated			
r(OO)	2.71	2.75	2.71	2.71
$-\Delta E$	3.51	3.38	4.06	4.19
	Cyclic			
r(OO)	2.74	2.76	2.78	2.76
θ	47	47	50	48
$-\Delta E$	3.15	3.07	2.53	2.75

[a] Bond lengths in angstroms; angles in degrees; energies in kcal/mol.

Results for Potential Functions

In order to enhance the fit for low-energy points, a weighting function was used in the least-squares procedure. As in the earlier studies,[2a,9] its form is $w_i = 1 + \alpha \exp(-(\Delta E_i - \Delta E_0)/kT)$. For the present purposes, ΔE_0 and kT were set at −7.0 and 0.6 kcal/mol, respectively. α was varied depending on the potential as described below.

A number of interesting questions were examined. To begin, potential functions were fit to the data for the 266 dimer configurations using the three point charge model. A function identical with the form of Clementi's CI potential[6c] was tried and termed the "exp" potential. A 12-6-3-1 function was also fit. In both cases α was taken as 20. Again, the number of parameters is 10 in all cases, so these potentials are directly comparable. The standard deviations for the two functions are given in the second and third columns of Table II. $\sigma_{<\chi}$ refers to all points with STO-3G dimerization energies less than χ kcal/mol; e.g., $\sigma_{<-2}$ is the standard deviation for points with dimerization energies below −2 kcal/mol. The quality of the fits is very good and nearly identical for both potentials. Thus, there is no statistical advantage for using exponential functions rather than the 12-6-3 power series for short-range interactions. However, a serious computational issue in simulations of condensed phases is the speed of evaluating the potentials. To examine this, Monte Carlo programs were written for water using the exp and 12-6-3-1 functions. Timing tests for carefully optimized FORTRAN/IV code revealed that simulations with the 12-6-3-1 potential are 11% faster than for the exp. Although the difference is not dramatic, it is still significant in view of the lengthy computations times for Monte Carlo and molecular dynamics calculations; e.g., a Monte Carlo run for 125 water molecules and 500 000 configurations requires over 50 h of central processor time on the CDC/6500 or about 3–4 h on a CDC/7600.

Energy minimizations were carried out using the potential functions for the linear, cyclic, and bifurcated forms of the water dimer. The results are presented in Table III and may be compared with the data in Table I. Both functions underestimate the OO distances for all three forms of the dimer as compared to the fixed STO-3G optimizations. Another difference is that the linear dimer is found to be too flat by 30–40° when contrasted with the STO-3G and experimental results for θ. The CI potential has the same flaw (Table I) so it seems inherent in the three point charge model. The other troublesome discrepancy is that both potentials predict the dimeri-

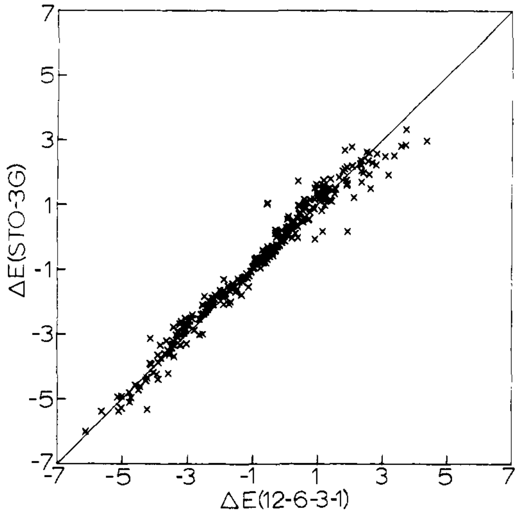


Figure 2. Comparison of the dimerization energies from the STO-3G calculations and the 12-6-3-1, four point charge potential for the 291 dimer geometries.

Table IV. Parameters for the 12-6-3-1 Potential Function for $(H_2O)_2$ Fit to 291 STO-3G Dimerization Energies[a]

	parameter		
	OO	OH	HH
b	1.641 846	(−0.861 296 4)[b]	0.450 834 9
c	−47.352 50	4.827 449	−3.656 219
d	1 074 064	2148.026	3808.762
	$q^2 = 0.011\ 512\ 42$	$r(\mathrm{OP}) = 1.945\ 861$	

[a] All values use energies and distances in atomic units. [b] Determined from b_{OO} and b_{HH}; see text.

zation energy for the linear form to be too positive by about 0.4 kcal/mol in view of the fixed STO-3G value. Since the region near this minimum would be heavily sampled in Monte Carlo simulations, it is important to have it well described in the QMP. Otherwise, the potential is not truly representative of the ab initio potential surface.

Four Point Charge Potential

In an attempt to to solve these difficulties, the four point charge model was tried with a 12-6-3-1 potential. The results for $\alpha = 20$ are shown in the fourth columns of Tables II and III. The fit for low-energy points was somewhat enhanced (Table II). However, the major gain was the near-perfect agreement between the potential and the fixed STO-3G results for θ and the dimerization energy in the linear dimer (Tables I and III). The pseudo-lone pairs clearly make the hydrogen bonding more tetrahedral. Unfortunately, the OO distances were not improved, so the value for the linear dimer is 0.24 Å less than from the CI potential (2.88 Å).

In a final attempt to obtain better OO separations, 25 points were added to the data base near the minima for the linear, cyclic and bifurcated dimers. Great improvement could not be expected because the difference between the OO distance for the linear form from the potentials and the STO-3G results was only ca. 0.05 Å. At this stage, α was also optimized to enhance the fit for the low-energy points. The results of increasing α are striking. As shown in the last column of Table II, $\sigma_{<-4}$ is reduced to 0.34 kcal/mol when α is set at 100 with the expanded data base containing 291 points. Increasing α beyond this point provides only modest enhancement of the fit for low-energy points and substantial deterioration in the fit for points with $\Delta E > 0$. The OO distance for the linear dimer was slightly increased to 2.66 Å with this final potential. The parameters for the function are given in Table IV.

The quality of the fit of the potential to the STO-3G dimerization energies is illustrated in Figure 2. The figure also reveals a much more even distribution of energies than has been obtained previously; cf. Figures 3 and 1 of ref 2a and 9a, respectively. This can be largely attributed to the procedure for obtaining energy distributed random geometries. Sixty-five of the 200 geometries yielded STO-3G dimerization energies below −1 kcal/mol and 23 were below −3 kcal/mol.

An additional important test of the potential was made. In the same spirit as in Beveridge's and our earlier study,[2a,9] 40 more dimer geometries were generated with distributed energies. The predictive ability of the 291-point potential was then tested against the STO-3G energies for the 40 points. The standard deviations for the predictions were σ (0.31), $\sigma_{<0}$ (0.30), $\sigma_{<-2}$ (0.35), and $\sigma_{<-4}$ (0.30). This confirmed that the function has good predictive value and provides a reasonable representation of the water dimer energy surface.

In view of the large number of points used for the STO-3G QMP, it seemed interesting to question the validity of only using 66 points as for the CI potential. Could such a small number of points produce a function representative of the entire energy surface? A fit was made to the STO-3G energies for Clementi's original 66 points using a 12-6-3-1 potential and the three point charge model with $\alpha = 20$. Naturally, an excellent fit was obtained: σ (0.27), $\sigma_{<0}$ (0.25), $\sigma_{<-2}$ (0.32). The predictive ability of this 66-point function was then tested against the STO-3G energies for the first 266 points, i.e., Clementi's 66 plus the 200 energy distributed random geometries. The standard deviations for the predictions were σ (1.08), $\sigma_{<0}$ (0.33), $\sigma_{<-2}$ (0.33). Amazingly, the predictions for bound dimers were excellent; however, the potential's representation of the repulsive part of the energy surface was disastrous. The potential is too soft at short range. Similar shortcomings may be anticipated for the CI potential. Experience here and elsewhere[2a,9] indicates that 200–300 points are needed to obtain truly representative QMPs for systems the size of the water dimer.

The relative softness of the CI potential is illustrated in Figure 3 in comparison with the 266-point, three point charge and 291-point, four point charge 12-6-3-1 potentials. The figure shows the variation of dimerization energy with OO distance for the linear water dimer. In each case θ has been fixed at its optimum value for the minimum. The figure also reveals that the STO-3G QMPs have narrower wells than the CI potential. The difference is probably due to a combination of effects stemming from the basis set truncation and lack of correlation energy corrections in the STO-3G calculations. It will be interesting to ascertain any consequences in Monte Carlo simulations of the liquid.

Figure 4 illustrates the effect on the dimerization energy as θ is varied for the linear dimer. Results are given for the four point charge 12-6-3-1 and CI potentials and for STO-3G calculations. In each case the OO separation was fixed at its optimum value for the linear dimer (Tables I and III). The shape of the curve for the 12-6-3-1 potential is similar to that for the ST2 potential including the double minima, while the flatness of the CI results is reminiscent of the earlier Rowlinson potential (cf. ref 4c, Figure 3). The STO-3G calculations reveal one minimum and an inflection point, so they appear intermediate in shape between the curves for the two potential functions in Figure 4.

Finally, the long-range attraction is seen to be greater in Figure 3 for the CI than the 12-6-3-1 potentials. This is consistent with the dipole moments for the monomers that can be calculated from the Coulombic parts of the potentials. The nonuniqueness of the parameters and the coupling of terms in the potentials make such calculations far from rigorous.

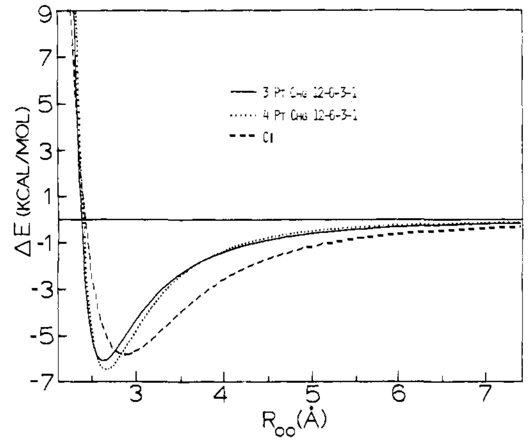


Figure 3. Comparison of dimerization energies from the CI and three and four point charge, 12-6-3-1 potentials for the linear water dimer.

Nevertheless, the dipole moments are 1.68 and 1.22 D for the three and four point charge 12-6-3-1 potentials and 2.19 D for the CI potential. The low value for the four point charge potential is compensated by the negative coefficient for the $1/r^3$ OH term (Table IV). The experimental dipole moment is 1.85 D.[17] Some enhancement of a monomer's polarization upon dimer formation is apparent in ab initio calculations.[2b,16] A similar effect is observed in ice.[19]

Extraneous Minima

A hazard in empirical or quantum mechanical intermolecular potential functions is the occurrence of unwanted minima. This is an understandable consequence of the multidimensional, multiparameter nature of the functions. Clearly, careful checking of the potentials is required. Also, the risk is diminished in the quantum mechanical approach by using a data base that is large enough to span the possible dimer geometries so well as to leave little room for flaws.

Besides the tests and optimizations described above, the following points have been found helpful. For the 12-6-3-1 potentials, the least-squares procedure occasionally yields a negative coefficient for a $1/r^{12}$ term with a perfectly reasonable statistical fit. Such a feature in a potential would be undesirable, even if at moderate distances it is offset by a $1/r^6$ term, because at some small internuclear separations it would probably dominate and yield an extraneous minimum. Generally, the difficulty is easily overcome by either altering the starting guessses for the parameters or by increasing the size of the data base.

A brute-force method for searching for false minima is to couple the potential with a program that randomly generates dimer geometries. For example, such a procedure helped find that the CI potential has a minimum for a cyclic dimer with $r(\text{OO}) = 1.122$ Å, $\theta = -12°$, and a dimerization energy of $-115\,028$ kcal/mol. It is apparent that this minimum is operationally irrelevant for Monte Carlo calculations assuming that the simulations are begun with reasonable dimer geometries. Otherwise, the Monte Carlo results of Clementi,[6c] Scheraga,[18a] and Beveridge[18b] would have been absurd. A scan of the potential surface for planar geometries showed that the extraneous minimum is isolated by a barrier of nearly 250 kcal/mol. This also points out that normal gradient procedures would have difficulty locating the minimum. Incidentally, this was the only extraneous minimum found for the CI potential. A thorough search involving hundreds of thousands of dimer geometries for the four point charge, 12-6-3-1 potential failed to produce any points with energy lower than for the linear minimum. To the best of our knowledge, the function is well behaved and contains no extraneous minima.

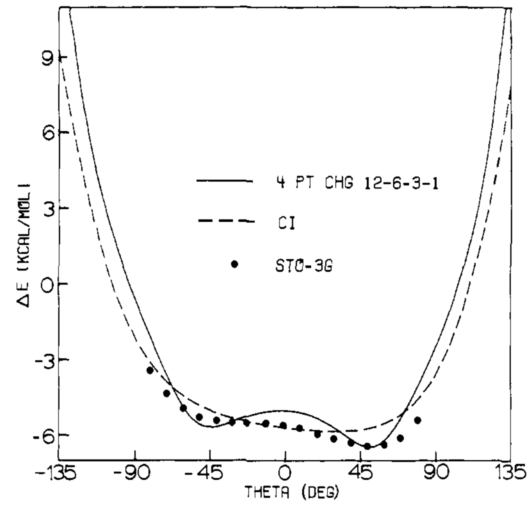


Figure 4. Variation of the dimerization energy with θ for the linear water dimer.

Concluding Remarks

In the present work, an intermolecular potential function representative of minimal basis set ab initio calculations has been constructed for the water dimer. This now permits thorough study of the basis set dependence of computer simulation results for liquid water. General procedures and considerations for the construction of quantum mechanical potentials were also delineated. The selection of dimer geometries, the form of potential functions, and testing the functions received special attention.

Acknowledgments. The author is grateful to Michael E. Cournoyer for programming assistance. Acknowledgment is made to the donors of the Petroleum Research Fund, administered by the American Chemical Society, for support of this work. Gratitude is also expressed to the Dreyfus Foundation, the Purdue Research Foundation, Purdue University Computing Center, and Research Corporation for aid.

References and Notes

(1) Quantum and Statistical Mechanical Studies of Liquids. 3. Part 2: ref 2b.
(2) (a) W. L. Jorgensen and M. E. Cournoyer, *J. Am. Chem. Soc.*, **100,** 4942 (1978); (b) W. L. Jorgensen, *ibid.*, **100,** 7824 (1978).
(3) For a historical review, see R. O. Watts and I. J. McGee, "Liquid State Chemical Physics", Wiley, New York, 1976, Chapter 9.
(4) (a) J. S. Rowlinson, *Trans. Faraday Soc.*, **47,** 120 (1951); (b) A. Ben-Naim and F. H. Stillinger in "Structure and Transport Processes in Water and Aqueous Solutions", R. A. Horne, Ed., Wiley-Interscience, New York, 1972, p 295; (c) F. H. Stillinger and A. Rahman, *J. Chem. Phys.*, **60,** 1545 (1974).
(5) (a) A. Rahman, F. H. Stillinger, and H. L. Lemberg, *J. Chem. Phys.*, **63,** 5223 (1975); (b) F. H. Stillinger and A. Rahman, *ibid.*, **68,** 666 (1978); (c) R. O. Watts, *Chem. Phys.*, **26,** 367 (1977).
(6) (a) H. Popkie, H. Kistenmacher, and E. Clementi, *J. Chem. Phys.*, **59,** 1325 (1973); (b) H. Kistenmacher, H. Popkie, E. Clementi, and R. O. Watts, *ibid.*, **60,** 4455 (1974); (c) O. Matsuoka, E. Clementi, and M. Yoshimine, *ibid.*, **64,** 1351 (1976); G. C. Lie, E. Clementi, and M. Yoshimine, *ibid.*, **64,** 2314 (1976).
(7) J. D. Dill, L. C. Allen, W. C. Topp, and J. A. Pople, *J. Am. Chem. Soc.*, **97,** 7220 (1975).
(8) D. R. Yarkony, S. V. O'Neil, H. F. Schaefer III, C. P. Baskin, and C. F. Bender, *J. Chem. Phys.*, **60,** 855 (1974).
(9) (a) S. Swaminathan, R. J. Whitehead, E. Guth, and D. L. Beveridge, *J. Am. Chem. Soc.*, **99,** 7817 (1977); (b) S. Swaminathan, S. W. Harrison, and D. L. Beveridge, *ibid.*, **100,** 5705 (1978).
(10) W. S. Benedict, N. Gailar, and E. K. Plyler, *J. Chem. Phys.*, **24,** 1139 (1956).
(11) E. Clementi, F. Cavallone, and R. Scordamaglia, *J. Am. Chem. Soc.*, **99,** 5531 (1977).
(12) E. L. Eliel, N. L. Allinger, S. J. Angyal, and G. A. Morrison, "Conformational Analysis", Wiley, New York, 1965, Chapter 7.
(13) GAUSSIAN/74 is a CDC version of GAUSSIAN/70: W. J. Hehre, W. A. Lathan, R. Ditchfield, M. D. Newton, and J. A. Pople, QCPE Program No. 236.

(14) The three geometries for the water dimer were not fully optimized with the CI calculations,[6c] so only the optimized results from the CI potential function can be reported here.
(15) T. R. Dyke and J. S. Muenter, *J. Chem. Phys.*, **60**, 2929 (1974).
(16) For a review, see P. A. Kollman in "Applications of Electronic Structure Theory", H. F. Schaefer III, Ed., Plenum Press, New York, 1977, p 109.
(17) R. D. Nelson, D. R. Lide, and A. A. Maryott, *Natl. Stand. Ref. Data Ser., Natl. Bur. Stand.*, **No. 10** (1967).
(18) (a) J. C. Owicki and H. A. Scheraga, *J. Am. Chem. Soc.*, **99**, 7403 (1977); (b) S. Swaminathan and D. L. Beveridge, *ibid.*, **99**, 8392 (1977).
(19) D. Eisenberg and W. Kauzmann, "The Structure and Properties of Water", Oxford University Press, London, 1969, p 106.

Minimal Basis Set Description of the Structure and Properties of Liquid Water[1]

William L. Jorgensen

Contribution from the Department of Chemistry, Purdue University, West Lafayette, Indiana 47907. *Received September 21, 1978*

Abstract: A Monte Carlo simulation of liquid water at 25 °C has been carried out using an intermolecular potential function derived from ab initio molecular orbital calculations with a minimal basis set (STO-3G). Comparisons are made with results for structural and thermodynamic properties from simulations based on potential functions representative of Hartree-Fock level (HF) and configuration interaction (CI) calculations. The quality of the results for the liquid is not enhanced smoothly as the ab initio calculations become more sophisticated. The STO-3G and CI potentials yield similar descriptions of liquid water, while the HF potential predicts too little structure beyond the first solvation shell and too high an energy. The first peak in the oxygen-oxygen radial distribution function is somewhat too high and displaced to shorter distances with the STO-3G potential than from the experimental or CI results. Nevertheless, the coordination number and energy distribution functions are similar in the simulations with the STO-3G and CI potentials. It is established that reasonable intermolecular potential functions for hydrogen-bonded dimers can be obtained from minimal basis set calculations.

I. Background and Rationale

It has been 10 years since the first computer simulation of liquid water.[3] In view of the great theoretical and experimental interest in molecular liquids, it is surprising that there has not been a mounting surge of simulations for progressively more complex fluids. In fact, the only systems that have been treated other than water and homonuclear diatomics[4] are carbon monoxide,[4,5] ammonia,[6] benzene,[7] *n*-butane,[8] and recently hydrogen fluoride.[9] Of these, the work on the hydrocarbons must be considered preliminary. The slow progress can largely be attributed to the difficulty in obtaining good intermolecular potential functions for the constituent dimers. The functions are obtained in two ways, which may be called empirical and quantum mechanical. The empirical approach involves the use of sums of standard functions, such as Coulomb and Lennard-Jones, parametrized to reproduce *experimental* properties of the monomers and dimers, e.g., dipole moment, vibrational frequencies, geometry, and dimerization energy. Rahman and Stillinger have formulated and used such functions with great success in their molecular dynamics simulations of water.[10] In the quantum mechanical approach, ab initio molecular orbital calculations are performed for many configurations of the dimer. The resultant dimerization energies are then fit to algebraic forms similar to those in the empirical potentials.

The problem with obtaining intermolecular potential functions empirically for systems more complex than water is the availability and reliability of the necessary experimental data. Water is unique in the extent to which it has been studied experimentally. Furthermore, reliable data on the geometries and dimerization energies for dimers in the gas phase are particularly critical and notoriously difficult to obtain. For example, gas-phase experiments did not confirm a linear structure for the water dimer until 1974, and the error bars for the dimerization energy are still ±30%.[12]

A clear advantage to the quantum mechanical approach is that, in principle, any system can be treated. The complications arise from the dependence of the results on the choice of basis set and correlation energy correlations. There are basically four levels of sophistication: (1) minimal basis set, (2) double ζ basis, (3) double ζ plus polarization (DZ + P) or Hartree-Fock limit (HF), and (4) DZ + P plus correlation corrections via configuration interaction (CI) or perturbation theory methods. Clementi and co-workers obtained potentials for the water dimer at levels 3 and 4.[13] The author's original calculations for the hydrogen fluoride dimer were at level 2.[14] It would seem that the quality of the potential functions should increase with the level of ab initio theory. This point has received almost no attention except in Clementi's work[15] and in the study of Swaminathan and Beveridge which used Clementi's potentials.[16] The issue is clouded by what is used to measure "quality". For the present purposes, the criteria can be radial distribution functions and thermodynamic properties, particularly the energy, which may be calculated for the liquid in Monte Carlo (MC) or molecular dynamics simulations. There is distinct improvement in both areas when the CI potential (level 4) is used for water rather than the Hartree-Fock (HF, level 3).[15,16] Thus, the energies of liquid water at 25 °C from MC calculations using the HF and CI potentials are −4.9 and −6.6 kcal/mol, while the experimental value is −8.1 kcal/mol. In addition, the OO radial distribution function from the HF potential reveals essentially no structure beyond the first solvation sphere, whereas the CI results agree with experiment in finding well-defined second and third solvation shells. If these results are interpreted to imply that acceptable intermolecular potential functions can only be obtained at the CI level, then the quantum mechanical approach could only be applied to small systems with limited survey of their potential surfaces in view of the expense of such calculations.

However, it is likely that the discrepancies between the HF and CI results are mainly due to the lower dimerization energy predicted by the CI potential (−5.84 kcal/mol) than the HF (−4.55 kcal/mol). Consequently, it is doubtful that a smooth

0002-7863/79/1501-2016$01.00/0