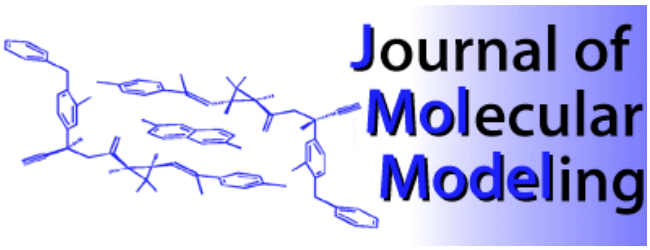
J. Mol. Model. **1999**, *5*, 281 – 286



FULL PAPER

Global Potential Energy Minima of SPC/E Water Clusters without and with Induced Polarization Using a Genetic Algorithm

Jiang Qian, Elmar Stöckelmann, and Reinhard Hentschke

Max-Planck-Institut für Polymerforschung, Postfach 3148, 55021 Mainz, Germany

Received: 26 August 1999/ Accepted: 19 November 1999/ Published: 10 December 1999

Abstract The global minimum potential energy structures of water clusters, $(H_2O)_n$, n = 2-14, have been calculated for the SPC/E (Simple Point Charge/Extended) model and a recent fluctuating charge version of the latter using a simple genetic algorithm. The SPC/E cluster geometries are in good agreement with previous TIP3P (Transferable Intermolecular Potential-3 Point) and TIP4P (Transferable Intermolecular Potential-4 Point) calculations as well as the interpretation of experimental measurements. In contrast to this, the polarizable version of the SPC/E model, which is based on the fluctuating charge approach, deviates rather strongly for n = 6 with few exceptions. However, comparing the polarizable model to *ab initio* results for identical cluster geometries we find reasonable agreement for the magnitude of the average molecular dipole moment, the corresponding energy per molecule, and the average oxygen-oxygen distance as functions of n.

Keywords Global energy minimum, Genetic algorithm, Water cluster, Polarization

Introduction

Determining the global energy minimum of a molecular system is a challenging problem [1]. The methods employed in this context can be categorized as either deterministic or stochastic. The conceptionally interesting diffusion equation method [2-4] is an example of a deterministic global minimization technique. This method searches for the global minimum on a smoothed potential energy hypersurface. If the diffusion-like smoothing procedure is successful, then a single minimum corresponding to the original global minimum remains and can be found by any local minimization method. Subsequently the deformation of the original energy hypersurface is gradually reversed, and the minimum is traced back to the original global minimum. Unfortunately, in general the method is not as straightforward as it sounds. A representative of stochastic global minimization is simulated annealing [5]. Here the originally high temperature of a system is reduced gradually to zero, following various cooling schedules. Eventually the system will be frozen at the global minimum. However, this only happens when the cooling rate is infinitely slow, which is not feasible in practice.

Genetic algorithms [6-8] are also stochastic global minimization techniques which are inspired by Darwinian natural evolution, and they appear to be rather robust. Through operations like crossover, which are conceptionally rather

Correspondence to: R. Hentschke
Bergische Universität-Gesamthochschule Wuppertal, FB 8 Physik, Gauss-Str. 20, D-42097 Wuppertal, Germany. Tel: +49-202-439 2628; Fax: +49-202-439 2811. E-mail: hentschk@uni-wuppertal.de

simple, the "fit" members of a population can pass their characteristics, in the present context this means a low potential energy, to their descendents. On the other hand, genetic operations like mutation create descendents whose characteristics may be very different from their parents. This prevents the method from getting trapped in local minima, a problem which other methods may not solve efficiently. The application of genetic algorithms to molecular optimization problems is relatively recent (cf. references 20-33 in reference [9]). In this paper we apply a genetic algorithm to study small clusters of SPC/E water molecules, which is similar to a previous genetic algorithm application to water clusters based on the TIP3P model [9].

The optimal geometry of water clusters is an interesting subject for both experiment and simulation [10]. For $(H_2O)_n$-clusters with n=5 the configurations are not in doubt [11]. However, for n = 6 the configurations are very sensitive to the water model, i.e., the potential function. The motivation of this work is to explore the performance of the genetic approach applied to the SPC/E model, which is probably the simplest and, together with the TIP4P model, the most widely used general purpose empirical water model. In addition, we want to study the effect of polarization in the context of this model. The approach used here is based on charge equilibration as described by Goddard and co-workers [12]. Subsequently Berne and co-workers [13] have shown how this description can be included very efficiently in molecular dynamics algorithms enabling on the fly calculation of the partial charge distribution. The low extra cost of about 15% additional computer time in comparison to the fixed charge case makes such a water model very appealing, and thus it is interesting to study its performance not only in the pure liquid phase. Note for instance that the magnitude of the dipole moment per molecule increases from about 2.0 D for the dimer to values in the vicinity of 3 D in the bulk (cf. below). Fixed charge models account for the polarization contribution to the dipole moment in a mere mean field sense via a fixed increase of the vacuum dipole moment. In the light of this it is somewhat unexpected that the cluster geometries based on the simple SPC/E model are in good accord with the interpretation of experimental measurements, and, less surprisingly, with previous TIP3P and TIP4P calculations. In contrast to this, the polarizable version of the SPC/E model deviates rather strongly for n = 6 with few exceptions. However, comparing the polarizable model to *ab initio* results for identical cluster geometries we find reasonable agreement for the magnitude of the average molecular dipole moment, the corresponding energy per molecule, and the average oxygen-oxygen distance as functions of n, indicating that the model is basically sound.

Method

The Water Model

The geometry of the polarizable, rigid water model used here is that of the SPC/E model [14], i.e., the H-O-H angle is 109.47°, and the O-H bond length is 1 Å. The charges are located on the nuclei. The difference between this model and SPC/E is the treatment of polarization. The latter includes induced polarization in a mean field sense via an increased dipole moment, i.e., 2.35 D in comparison to the vacuum value of 1.85 D. In the following we use a fluctuating charge approach adjusting the atomic charges according to their local environment. The fluctuating charge model is that of Berne and co-workers [13] with some refinements explained in detail in reference [15]. The potential energy of a system containing N atoms is given by

$$U\left(\{\vec{r}\},\{q\}\right) = \sum_{\substack{i<j \\ \text{oxygens only}}}^{N} u_{ij}^{dr}\left(r_{ij}\right) + \sum_{i<j}^{N} J_{ij}\left(r_{ij}\right) q_i q_j + \sum_{i=1}^{N}\left[\tilde{\chi}_i^0 q_i + \frac{1}{2} J_{ii}(0) q_i^2\right] - U^{ref} \quad (1)$$

where $\{\vec{r}\}$ and $\{q\}$ are the positions and charges of all atoms. The indices i and j indicate atoms, and r_{ij} is their separation. The first term on the right describes Lennard-Jones interactions between the oxygen atoms $u_{ij}^{dr}(r_{ij}) = 4\varepsilon\left[\left(\sigma / r_{ij}\right)^{12} - \left(\sigma / r_{ij}\right)^{6}\right]$. As in the original SPC/E model the dispersion and overlap interactions between water molecules are modeled solely in terms of Lennard-Jones sites centered on the oxygen atoms. The second term describes the Coulomb interaction between two atomic charge distributions and differs from usual fixed charge models. The quantity $J_{ij}(r_{ij})$ is a Coulomb integral, i.e., $J_{ij}(r_{ij}) \approx r_{ij}^{-1}$ for large r_{ij}, whereas $J_{ij} \to const$ for $r_{ij} \to 0$. The third term is the energy of creating a partial charge, q_i, on the isolated atom i in the form of a Taylor series including the second order term. The last term is a reference energy, which is taken to be the total energy of the system in the gas phase at infinite dilution. The explicit expression for J_{ij} and U^{ref} including the parameters $\tilde{\chi}_i^0$ can be found in reference [15]. In the following $U\left(\{\vec{r}\},\{q\}\right)$ is the fitness function characterizing the members of the cluster population. Note that $U\left(\{\vec{r}\},\{q\}\right)$ is minimized with respect to both $\{\vec{r}\}$ and $\{q\}$.

Each water molecule is fully characterized by eight coordinates $\left(x_i, y_i, z_i, \alpha_i, \beta_i, \gamma_i, q_i^O, q_i^H\right)$. The first three are the Cartesian coordinates of the oxygen atom. The following three describe the molecular orientation via Euler-type rotation angles. The last two coordinates are the partial charges of the oxygen and the hydrogen atoms. Because the molecules are neutral, the partial charge of the second hydrogen is not independent. Of course, for the original SPC/E molecule the first six coordinates are sufficient.

Genetic Algorithm

The genetic algorithm employed here consists of the operations crossover mating and mutation. Traditional genetic algorithms operate on chromosome-like linear information, where crossover points are selected at random. For three dimensional clusters the information is no longer linear. Here we perform the following operations upon two parent clusters: (a) The clusters are rotated with respect to the X-, Y-, and Z- axes of their respective center of mass coordinate systems. The rotation angles are chosen at random. Crossover mating then recombines two clusters by cutting along their X-Y plane. Cutting means that we select half of the molecules in each cluster according to the z-values of their molecular center of mass. Subsequently we cross-wise fuse the resulting half-clusters. Note that there is no cutting of individual water molecules. (b) At certain intervals during the evolution the positions (x_i, y_i, z_i) within a range of 1 Å, and the orientations $(\alpha_i, \beta_i, \gamma_i)$ of all molecules within a cluster are changed randomly. This is called mutation. In practice, if a random number chosen between 0 and 1 is less than 0.5, then every cluster in a generation is subject to a mutation. In the cases of mutation as well as crossover mating numerical problems due to overlapping atoms are avoided via a suitably chosen inner-cutoff (here: 1 Å).

The overall algorithm is as follows. First the positions, within certain limits, and the orientations of molecules in the initial clusters are generated randomly. For the polarizable model the initial partial charges are the same as in the SPC/E model. A local energy minimization is then applied to each cluster with $(x_i, y_i, z_i, \alpha_i, \beta_i, \gamma_i, q_i^O, q_i^H)$ as variables using a Powell-algorithm which requires no derivatives [16]. Subsequent generations are obtained by crossover mating (a) and mutation (b) as described above. Note that every crossover mating and mutation is followed by a local energy minimization applied to each cluster. Note also that $N(N\text{-}1)/2$ children are generated by crossover mating of N parents. N clusters with the lowest energy are selected from the N parents and $N(N\text{-}1)/2$ children. These N clusters form the next generation. In this work the population size $N = 4$ (cf. below). The program stops if no energy changes are detected during 500 generations. At this point the cluster with the lowest potential energy defines the global minimum.

Results

The upper panel in Figure 1 shows an example run for SPC/E clusters of size $n = 14$. The energy, E, is the lowest cluster potential energy $U(\{\vec{r}\},\{q\})$ within a population divided by the cluster size, n. E decreases rapidly as a function of the number of generations, N_gene, at the beginning of the run. The number of generations after which the lowest energy cluster has closely approached the predicted global minimum structure is indicated by an arrow. There still is a slight decrease of E beyond N_gene^{min}, but this has no significant effect on the cluster structure. The inset shows N_gene^{min} as function of the mutation rate calculated for 10 independent start configurations based on 8-clusters of SPC/E water. Here the optimal rate is around 50%, which is quite large. N_gene^{min} as function of cluster size, n, is shown in the lower panel of Figure 1. Note that N_gene^{min} increases exponentially with n for both the SPC/E and the fluctuating charge model. Each error bar is based on 10 runs starting from different configurations. Of course, the value of N_gene^{min} depends on the population size. A large population size yields smaller values for N_gene^{min}, but increases the computational effort within each generation. Note also that the computation time for each generation increases with cluster size as $O(n^2)$. As a test of our genetic algorithm we calculated the global minimum energy structure of Lennard-Jones clusters containing up to 29 particles. The results agree with those of reference [17]. Again N_gene^{min} increases exponentially with n, but with a smaller rate than that of the genetic algorithm in reference [9].

Figure 2 shows the global minimum energy geometries for SPC/E water clusters, $(H_2O)_n$, with n = 2-14. Compared to corresponding clusters of a recent calculation based on the TIP4P model [18] the main differences are the following. For $n = 6$ SPC/E has an "open book" structure, whereas TIP4P favors a "cage" structure. For $n = 7$ the SPC/E structure differs by one missing hydrogen bond. We note that this cluster is also obtained if we chose the $n = 7$-cluster of reference [18] as the starting structure. For $n = 11$ both SPC/E and TIP4P form cages, but these exhibit different structures. The visual comparison with the TIP3P clusters of reference [9] shows discernible differences only for $n = 10$ and 13. For $n = 10$ a "butterfly" structure is obtained (four pentagonal faces fused into a ring), and for $n = 13$ the only pentagonal face has switched position with an adjacent 4-ring. These TIP3P results were revised in [18] for $n = 11$-13, i.e., the $n = 13$-cluster is now very similar to the above SPC/E- and TIP4P-results, whereas the $n = 11$- and $n = 12$-cluster are cage structures deviating from the above SPC/E- and TIP4P-results.

Experimentally the transition from the cyclic structures obtained for $n = 3$-5 towards three dimensional clusters for larger n is not in doubt [11], and the main challenge is the structure evolution for $n = 7$ (even though the structure of the hexamer still appears somewhat unclear [18]). In reference [11] the cases $n = 8$-10 are studied in detail. The structures for $n = 8$ (see also [19]) and 9 agree with our SPC/E results. Actually, for the octamer SPC/E predicts a D_{2d} symmetry, whereas the experimentally studied OH stretch spectra are contributed by two energetically close isomers of D_{2d} and S_4 symmetry. For the decamer the experimental study analyses both of the above mentioned structures, the fused pentagon and the "butterfly" structure, and finds that the calculated vibrational OH stretch spectra are somewhat better reproduced by the "butterfly".

Figure 3 shows the global minimum energy structures obtained with the fluctuating charge model. For $n = 5$ the SPC/E and the fluctuating charge model yield nearly identical geometries, which mainly differ in the distances between

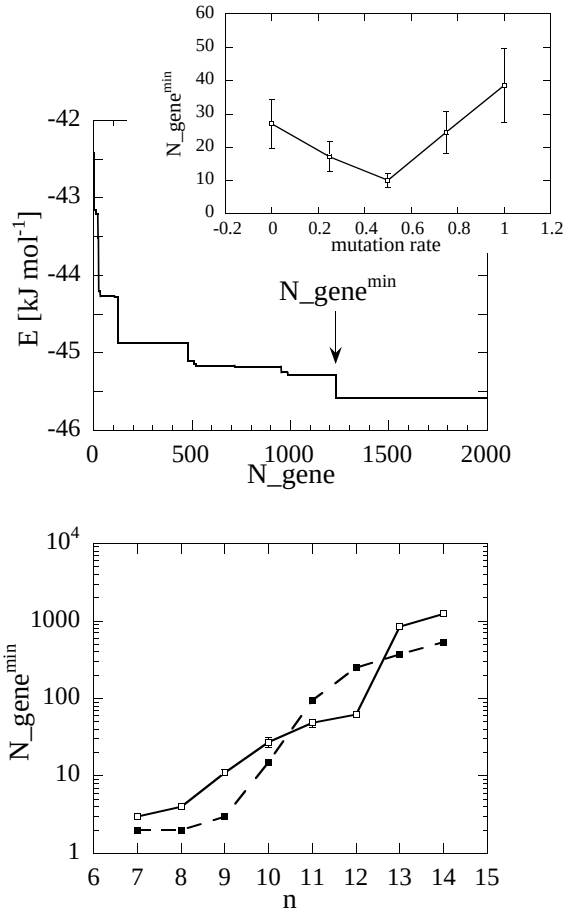


Figure 1 *Upper panel: Potential energy per water molecule,* E, *vs. the number of generations,* N_gene, *for one run. Here the water cluster consists of 14 SPC/E molecules. Note that* E *refers to the lowest cluster energy within each generation. Lower panel:* N_gene^{min}, *the number of generations before the global minimum is reached vs. the cluster size,* n. *Open squares: SPC/E; solid squares: polarizable model. The error bars are explained in the text. Inset:* N_gene^{min} *vs. mutation rate for 8-clusters of SPC/E water.*

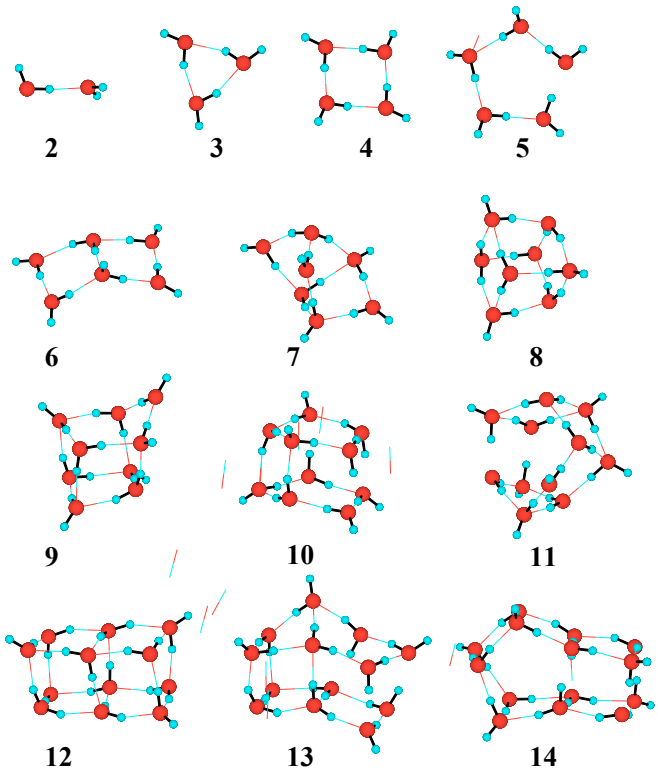


Figure 2 *Global minimum energy structures for SPC/E water clusters* $(H_2O)_n$ (n = *2-14)*

the molecules (cf. Figure 4). With the exception of the nonamer, however, the SPC/E structures differ from those obtained with the fluctuating charge model. Particularly troublesome is the prediction of planar structures for n = 6-8. The geometries of the larger clusters, however, are not necessarily to be dismissed based on this discrepancy alone, because they share numerous local structural similarities with the clusters predicted by the other models. The inclusion of polarization appears to favor cage-like structures composed of 5- or 6-ring faces over the simpler combinations of D_{2d}, S_4, and S_6 geometries, mainly consisting of fused cubes, observed for the fixed charge models. This is also seen in a study of $(H_2O)_n$ for n = 12, 16, and 20, where the POL1 model, which models polarization via iteratively computed induced dipole moments, is compared to SPC/E results [20]. The polarizable model prefers cage-like structures for all cluster sizes, whereas the nonpolarizable model predicts minima of fused cubic structures for n = 12 (same as above) and 16 but makes a transition to a cage-like minimum at n = 20. However, the molecular dynamics-quenching method used in [20] cannot reach the global minimum for the SPC/E model, and the minimum energy structures are constructed graphically instead. Returning to the octamer we remark that Stillinger and David [21] also predict a (different) two-ring structure using a polarizable model. Using both the cubic and Stillinger's structure as initial configurations of a local energy minimization based on Equation (1), however, yields a higher energy than the two-ring structure predicted by the genetic algorithm, i.e., -31.9kJ/mol and -31.3kJ/mol per molecule compared to -32.7kJ/mol.

Regarding the ring structure predicted for n = 6 we add that this structure is also predicted to be the global minimum by Hartree-Fock (HF) calculations [22,23], second-order many body perturbation theory (MP2) calculations using HF-optimized geometries [23], and density functional calculations [24,25]. The comparison of experimental measurements and MP2 calculations of harmonic OH stretch frequencies of benzene-water clusters on the other hand indicates that n = 6 is not a cyclic structure, and the most likely candidate is a cage structure [26,27]. However, the presence of benzene may

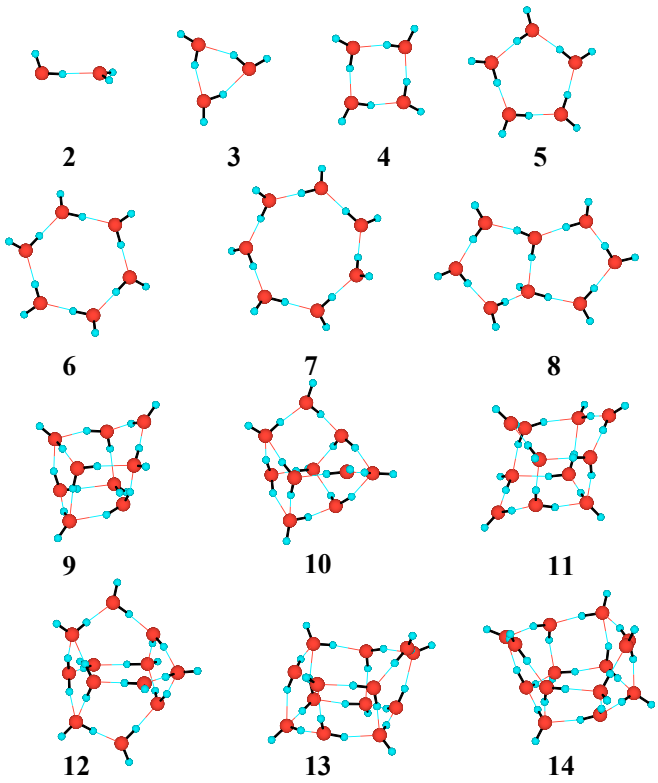


Figure 3 *Global minimum energy structures for polarizable water clusters* $(H_2O)_n$ (n = *2-14)*

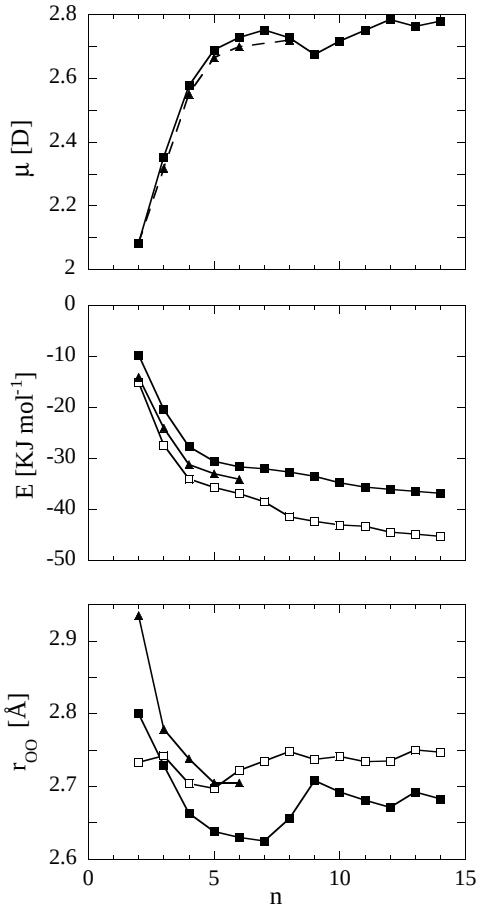


Figure 4 *Average magnitude of the water dipole moment,* μ*, average energy per water molecule,* E*, and average oxygen-oxygen distance,* r_{OO}*, vs. cluster size,* n*. Open squares: SPC/E; solid squares: polarizable model; solid triangles:* ab initio *results.*

alter the energy ordering of the water clusters. Tetrahertz laser vibration-rotation tunneling spectroscopy in conjunction with quantum Monte-Carlo simulations also indicates that the global minimum structure is a cage structure [28].

In Figure 4 we compare the *n*-dependence of the magnitude of the average molecular dipole moment, μ, of *E*, and of the average nearest neighbor oxygen-oxygen distance, r_{OO}, for the SPC/E and the fluctuating charge model. The values of μ and r_{OO} are calculated based on the lowest energy cluster. The dipole moment increases up to $n = 7$, and the values are in good agreement with the *ab initio* results of reference [10] obtained for the same cluster geometries. At the crossover from predominantly two-dimensional to three-dimensional clusters, i.e., between $n = 8$ to 10, μ exhibits a slight depression. The final plateau value is close to 2.8 D. We note that recently Silvestrelli and Parrinello [29] using *ab initio* molecular dynamics studied the behavior of the dipole moment of water molecules in the gas and in the liquid phase. In the liquid phase their dipole moment has an average value of about 3 D. Another recent calculation on the *ab initio* level gives an average dipole moment of 3.09 D per water molecule in ice Ih [30]. These values are significantly larger than a previous and extensively used value of 2.6 D, which was obtained by an approximate induction model [31]. In the middle panel of Figure 4 *E* is plotted for the SPC/E and the polarizable model in comparison to *ab initio* results [32]. The two phenomenological models bracket the quantum result, and the clusters of the polarizable model are the least strongly bound. With respect to r_{OO} vs. *n* the polarizable and the quantum models are quantitatively very similar. Compared to the SPC/E model they show a significant decrease of r_{OO} with increasing *n* (up to $n = 8$). Overall the result for the polarizable model mirrors the behavior of μ as function of *n*. Not shown are experimental data for r_{OO} [33], which fall roughly 0.15 Å above the values for the polarizable model. However, zero-point motion effects, which are not included here, significantly reduce this gap [32].

In conclusion, the genetic algorithm proposed here yields SPC/E clusters, which are in good accord with the currently accepted candidates for the respective global minimum structures. The only exception is the hexamer. The fluctuating charge version of the SPC/E model is less successful. Currently, we do not know why. Nevertheless, matters are not decided conclusively beyond $n = 9$. In favor of the model

speaks the reasonable agreement with the *ab initio* cluster calculations discussed above.

Supplementary Material Available The coordinates of 26 cluster structures in PDB format are available as supplementary material.

References

1. Leach, A. R. In *Reviews in Computational Chemistry*; Lipkowitz, K. B. and Boyd, D. B., Eds.; VCH Publishers: New York, 1991; Vol. 2; 1-55.
2. Piela, L.; Kostrowicki, J.; Scheraga, H. A. *Phys. Chem.* **1989**, *93*, 3339.
3. Kostrowicki, J.; Piela, L.; Cherayil, B. J.; Scheraga, H. A. *Phys. Chem.* **1991**, *95*, 4113.
4. Kostrowicki, J.; Scheraga, H. A. *Phys. Chem.* **1992**, *96*, 7442.
5. Kirkpatrick, S.; Gelatt, C. D.; Vecchi, M. P. *Science* **1983**, *220*, 671.
6. Goldberg, D. E. *Genetic Algorithms in Search, Optimization, and Machine Learning*; Addison-Wesley: Reading, 1989.
7. Davis, L. *Handbook of Genetic Algorithms*; Van Nostrand, Reinhold: New York, 1991.
8. *Genetic Algorithms in Molecular Modeling*; Devillers, J., Ed.; Academic Press: New York, 1996.
9. Niesse, J. A.; Mayne, H. R. *J. Comput.Chem.* **1997**, *18*, 1233.
10. Gregory, J. K.; Clary, D. C.; Liu, K.; Brown, M.; Saykally, R. J. *Science* **1997**, *275*, 814.
11. Buck, U.; Ettischer, I.; Melzer, M.; Buch, V.; Sadlej, J. *Phys. Rev. Lett.* **1998**, *80*, 2578.
12. Rappé, A. K.; Goddard, W. A. *J. Phys. Chem.* **1991**, *95*, 3358.
13. Rick, S. W.; Stuart, S. J.; Berne, B. J. *J. Chem. Phys.* **1994**, *101*, 6141.
14. Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91*, 6269.
15. Stöckelmann, E.; Hentschke, R. *J. Chem. Phys*. **1999**, *110*, 12097.
16. Press, W.; Flannery, B. P.; Teukolsky, S.; Vetterling, W. *Numerical Recipes in C: The Art of Scientific Computing*; Cambrodge University Press: Cambridge, 1988.
17. Wales, D. J.; Doye, J. P. K. *J. Phys. Chem.* **1997**, *101*, 5111.
18. Wales, D. J.; Hodges, M. P. *Chem. Phys. Lett.* **1998**, *286*, 65.
19. Gruenloh, C. J.; Carney, J. R.; Arrington, C. A.; Zwier, T. S.; Fredericks, S. Y.; Jordan, K. D. *Science* **1997**, *276*, 1678.
20. Sremaniak, L. S.; Perera, L.; Berkowitz, M. L. *Chem. Phys.* **1996**, *105*, 3715.
21. Stillinger, F. H.; David, C. W. *Chem. Phys*. **1980**, *73*, 3384.
22. Mhin, B. J.; Kim, H. S.; Yoon, C. W.; Kim, K. *Chem. Phys. Lett.* **1991**, *176*, 41.
23. Mhin, B. J.; Kim, J.; Lee, S.; Lee, J. Y.; Kim, K. *J. Chem. Phys.* **1994**, *100*, 4484.
24. Laasonen, L.; Parrinello, M.; Car, R.; Lee, C.; Vanderbilt, D. *Chem. Phsy. Lett.* **1993**, *207*, 208.
25. Lee, C.; Chen, H.; Fitzgerald, G. J. *J. Chem. Phys.* **1994**, *101*, 4472.
26. Pribble, R. N.; Zwier, T. S. *Science* **1994**, *265*, 75.
27. Kim, K.; Jordan, K. D.; Zwier, T. S. *J. Am. Chem. Soc*. **1994**, *116*, 11568.
28. Liu, K.; Brown, M. G.; Carter, C.; Saykally, R. J.; Gregory, J. K.; Clary, D. C. *Nature* **1996**, *381*, 501.
29. Silvestrelli, P. L.; Parrinello, M. *Phys. Rev. Lett.* **1999**, *82*, 3308.
30. Batista, E. R.; Xantheas, S. S.; Jonsson, H. *J. Chem. Phys.* **1998**, *109*, 4546.
31. Coulson, C.; Eisenberg, D. *Proc. R. Soc. London, Ser.A* **1966**, *291*, 445.
32. Gregory, J. K.; Clary, D. C. *J. Phys. Chem.* **1996**, *100*, 18014.
33. Liu, K.; Cruzan, J. D.; Saykally, R. J. *Science* **1996**, *271*, 929.