Global Minima of Water Clusters $(H_2O)_N$, $N \leq 25$, Described by Three Empirical Potentials

H. Kabrede and R. Hentschke*

FB Physik and Institut für Materialwissenschaften, Bergische Universität, D-42097 Wuppertal, Germany

Received: December 20, 2002; In Final Form: February 19, 2003

We compare apparent global minimum structures of three widely used empirical point-charge models for water. These are SPC/E (Simple Point Charge/Extended), TIP3P (Transferable Intermolecular Potential 3 Point), and TIP4P. We obtain results for $(H_2O)_N$ with $N \leq 25$. We compute various quantities characterizing the clusters such as the average energy per molecule, the average oxygen-to-oxygen distance, the average hydrogen bond length and number, and the quantity μ^2/V, where V is an effective cluster volume and μ is the cluster's dipole moment. The quantity μ^2/V is interesting because it is approximately proportional to the free energy of solvation of a cluster in water. We also compare our genetic algorithm to the basin-hopping Monte Carlo method.

1. Introduction

Water clusters are considered important building blocks in bulk models aimed at explaining the "anomalous" properties of water.[1,2] In addition, isolated clusters are an important testing ground for global minimization techniques.[3] The structure of small isolated water clusters can be determined by combining spectroscopy[4] and theoretical models.[5] The predictions based on measurements require both accurate water models as well as theoretical techniques suitable for this type of multiple minimum problem.

In this work we apply a genetic global minimization algorithm, which we have developed recently,[6] to water clusters with up to 25 molecules. We compare three rigid point charge water models, i.e., SPC/E,[8] TIP3P,[9] and TIP4P,[9] which are widely used in molecular modeling of aqueous systems. Because of this we believe it is interesting to compare the three models directly in a single calculation. Wales and Hodges[10] have calculated the global minima of water clusters with up to 21 molecules based on the TIP4P potential. The introduction to their paper compiles the computational state of the art until 1998, which has not changed significantly up to now. Experimentally, cluster structures with $N \leq 6$ molecules are well established. Cluster structures in the range $6 < N \leq 10$ have also been obtained but with less precision.[4] Most water models, simple or complex and including the above-mentioned, agree up to $N = 5$ among each other and with the experimental structures. Beyond $N = 5$ differences occur, because of the different molecular geometries and interactions used by the various models. In addition, the multiple minimum problem becomes harder to solve with increasing N. Ab initio calculations loose first as a result of the high computational effort involved. Empirical force fields, on the other hand, allow the use of Monte Carlo and Molecular Dynamics techniques in combination with certain cooling procedures.[11] Other approaches to the global minimum search attempt to simplify the potential energy hypersurface via smoothing, e.g., the diffusion equation method,[12] or by varying the interaction range.[13] A third strategy copies principles of natural evolution. The resulting algorithms are called genetic algorithms,[14] which only in recent years have been applied to molecular systems such as water.[3]

Besides presenting our results for the apparent global minimum structures for $N \leq 25$ for the three model potentials, we also compare our genetic algorithm to the basin-hopping Monte Carlo algorithm used in Ref 10. We compute various quantities characterizing the clusters such as the average energy per molecule E/N, the average oxygen-to-oxygen distance, the average hydrogen bond length and number, and the quantity μ^2/V, where V is an effective cluster volume. The quantity μ^2/V is interesting because it is approximately proportional to the free energy of solvation of a cluster in water.[7] It therefore provides a measure for the relative stability of a given cluster in bulk water, which in turn is important if one intends to construct a hierarchical model involving structures built upon clusters. Here we find that certain cluster sizes are distinctly preferred.

2. Method

Two of the water models used in this work, SPC/E[8] and TIP3P,[9] are flat three-center models; the other model, TIP4P,[9] is a flat 4-center model. The potential energy of a water cluster in such a model is the sum of Coulomb- and Lennard-Jones terms:

$$\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\left\{\frac{A}{r^{12}_{O_iO_j}}-\frac{C}{r^{6}_{O_iO_j}}+\sum_{k_i=1}^{3}\sum_{k_j=1}^{3}\frac{q_{k_i}q_{k_j}}{r_{k_ik_j}}\right\} \qquad (1)$$

Here i, j denote the molecules. The Lenard-Jones interaction centers are the oxygen atoms O_i, O_j in all three model potentials. A and C are the Lennard-Jones parameters of the overlap plus nuclear repulsion and the dispersion attraction, respectively. Coulomb interactions are described via point charges q_{ki}, q_{kj} located at the atom sites for SPC/E and TIP3P, whereas in TIP4P the oxygen charge is shifted to a fourth site located closer but still equidistant from the two hydrogen atoms. The respective parameters are compiled in Table 1.

The genetic algorithm used in this work is based on a genetic algorithm developed previously to compute the global minimum structures of neutral sodium chloride clusters.[6] Avoiding a lengthy discussion of the present algorithm we concentrate on the differences from the original algorithm explained in detail

* Corresponding author.

10.1021/jp027783q CCC: $25.00
Published on Web 04/02/2003

TABLE 1: Parameters of the Empirical Water Models Used in This Work[a]

model	r_{OH}/Å	Φ/deg	A/kJ^{12}mol^{-1}	C/kJ^{6}mol^{-1}	q_O/e	r_{OM}/Å	q_M/e
SPC/E	1.0	109.47	2633.24	2616.91	−0.8476		
TIP3P	0.9572	104.52	2436.72	2491.15	−0.834		
TIP4P	0.9572	104.52	2512.08	2553.95		0.15	−1.04

[a] Bond length: r_{OH}; valence angle: Φ; Lennard-Jones parameters: A, C; point charges: q_O, q_M. Note that q_M is located at a distance r_{OM} measured from the oxygen atom in the direction of the hydrogen atoms on the line bisecting Φ.

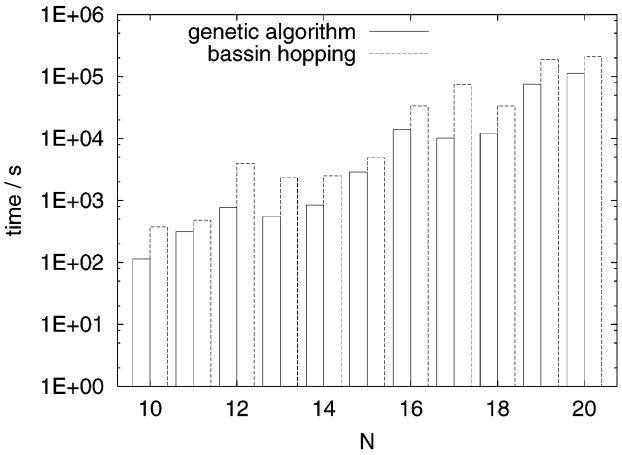

Figure 1. Comparison of our genetic algorithm (left axis and solid columns) and Monte Carlo basin-hopping (right axis and dashed columns). Shown is the average time it takes the respective algorithm to find the apparent global minimum. This time is calculated as (total number of trials) × (average time per trial)/(successful trials).

in ref 6 (cf. the flowchart in Figure 1 of this reference). Again, we start with a randomly generated initial or starting generation of typically $N_{pop} = 10$ clusters. As before, the next step is to define each cluster's fitness based on its potential energy. Mutation is now completely inherent in the genetic operators, therefore this step is omitted here. Subsequently, genetic operators are selected to construct the next generation. In this algorithm we use five different genetic operators (for an operator description, cf. Table 2). As in the original algorithm, each operator initially has the same probability of being applied. This probability changes dynamically on the basis of an operator's success. In the original algorithm an operator was successful if it created a low-lying minimum with $E \leq 0.85E_{opt}$ and $E_{new} \neq E_{old}$. E_{opt} is the best energy found so far, E_{new} is the energy of the new cluster(s), and E_{old} is the energy of the parent cluster(s). Here we use the slightly stricter criterion that an operator is successful if it creates a cluster with $E_{new} < E_{old}$. An operator's probability of being applied is raised by 1% per success. As before, an operator's probability of being applied may also be reduced. The minimal probability is 5% (for more details, cf. Operator Selection in ref 6). In the next step, N_{pop} parent clusters are chosen on the basis of their fitness. Then the new clusters are constructed. All new clusters with a potential energy $E_{new} \leq 0.4E_{opt}$ are relaxed to their closest local minimum by a conjugate gradient algorithm. The next generation is built up differently than in the original algorithm. Here we copy the new clusters to a sorted list containing the 100 best clusters found so far. We build a "combined" population of the N_{pop} new clusters and the N_{pop} best clusters found so far and calculate each cluster's fitness in this combined population. The next generation is selected from this combined population. The algorithm is repeated, beginning with the calculation of the cluster fitness, as long as a new local minimum is added to the sorted list, of the 100 best local minima during a fixed number (of typically 100) generations. If the sorted list remains unchanged for this number of generations the algorithm terminates, and the first element of the list is what we call the apparent global minimum.

TABLE 2: The Five Genetic Operators Used in Our Algorithm[a]

genetic operators			
1	crossoverI:	2→2	In cluster one, a molecule is chosen randomly. In cluster two, a molecule with similar distance to the origin is chosen. Both clusters are rotated so that the chosen molecule's center of mass is on the z-axis. The two linear arrays containing the molecular coordinates of each cluster are sorted according to the z-coordinates. Subsequently both arrays are fused crosswise at a randomly chosen molecular position along the arrays.
2	crossoverII:	2→2	The same as crossoverI, but after rotating the clusters the linear arrays containing the molecular coordinates are sorted according to the molecular distances to the origin. Subsequently, both arrays are fused crosswise as before.
3	crossoverIII:	2→2	The linear arrays containing the molecular coordinates are sorted according to the molecular z-coordinates. Again, a molecular position along the arrays is chosen at random. But now only the valence angle is exchanged between the molecules in the two clusters.
4	reflection:	1→2	The molecules are sorted according to their z-coordinates. For one cluster a random number of molecules is reflected at the $y-z$ plane and for the second cluster at the $x-z$ plane.
5	arithmetic mean:	2→2	The two new clusters' oxygen atoms coordinates are the arithmetic mean of the two parent oxygen atoms coordinates. The coordinates of the hydrogen atoms relative to the oxygen atoms remain unchanged.

[a] Note that Operator 3 acts on the angles only, while Operator 5 acts on the oxygen positions only. Inherent mutation is present in Operators 1-4. [b] For every operation each cluster's center of mass is the origin.

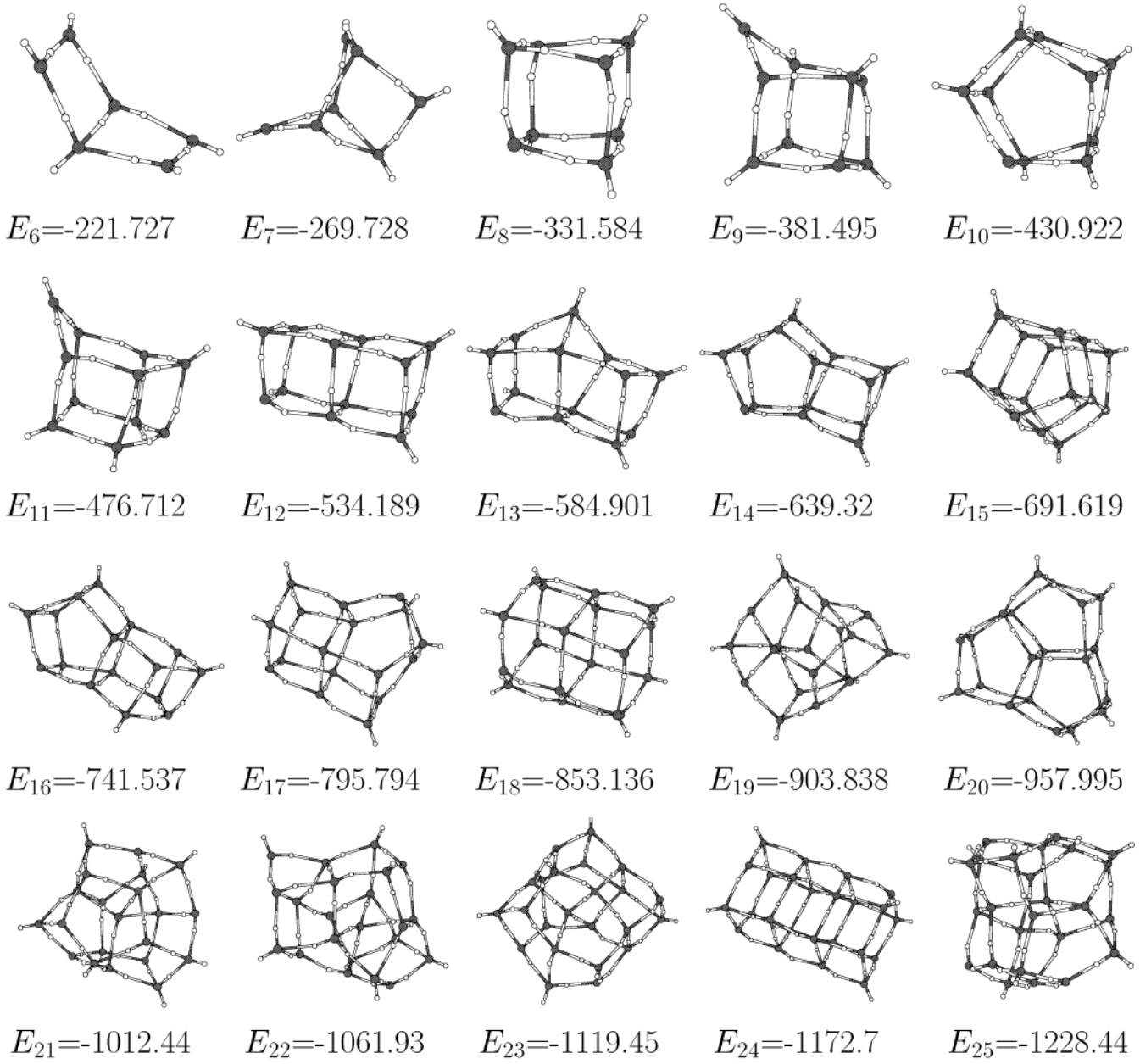

Figure 2. Apparent global minimum structures of water clusters consisting of between 6 and 25 molecules. The potential energy is calculated using the SPC/E potential. Energies are given in kJ/mol.

To test the validity of the apparent global minima we found with this method we also implemented Monte Carlo basin-hopping. This method was used by Wales and Doye[10] to compute the local minima of TIP4P water clusters with up to 21 molecules. This is the most extensive global minima study for water clusters so far. Our genetic algorithm outperforms Monte Carlo basin-hopping in the range $10 \leq N \leq 20$ by a factor of more than two, cf. Figure 1. However, one should keep in mind that these search algorithms are very sensitive to details of the implementation. Most important is the algorithm's ability to solve the problem of global optimization, which both algorithms do well for all cluster sizes and all three model potentials discussed here.

3. Discussion

The apparent global minimum structures we have calculated for the three model potentials are presented in Figures 2, 3, and 4. To the best of our knowledge, other studies have presented results for the SPC/E model potential for up to 14 molecules[15] and additionally for $N = 16$, and 20;[16] TIP3P clusters were studied up to $N = 13$;[3,10] TIP4P clusters were studied up to $N = 22$.[10,17] For SPC/E, we obtain the same results obtained by Quian et al. in ref 15. The apparent global minimum structures for $N = 12$ and $N = 20$ in ref 16 (SPC/E) are the same as ours. The potential energy of their $N = 16$-structure, however, exceeds ours by about 0.8%. The TIP3P results obtained by Wales and Hodges[10] agree with our results as do their TIP4P results. Differences in energy are solely due to the parameter values, but the structures provided by them via the Internet are in exact agreement with ours. The TIP4P $N = 22$-cluster found by Hartke[17] is higher in energy than ours by 0.5%. The small clusters with 2 to 5 molecules agree for all three model potentials. These cluster structures are in accord with ab initio calculations[18] and experimental findings.[4] While the hexamer structure predicted by ab initio calculations[19] and verified by experiments is not found to be the global minimum

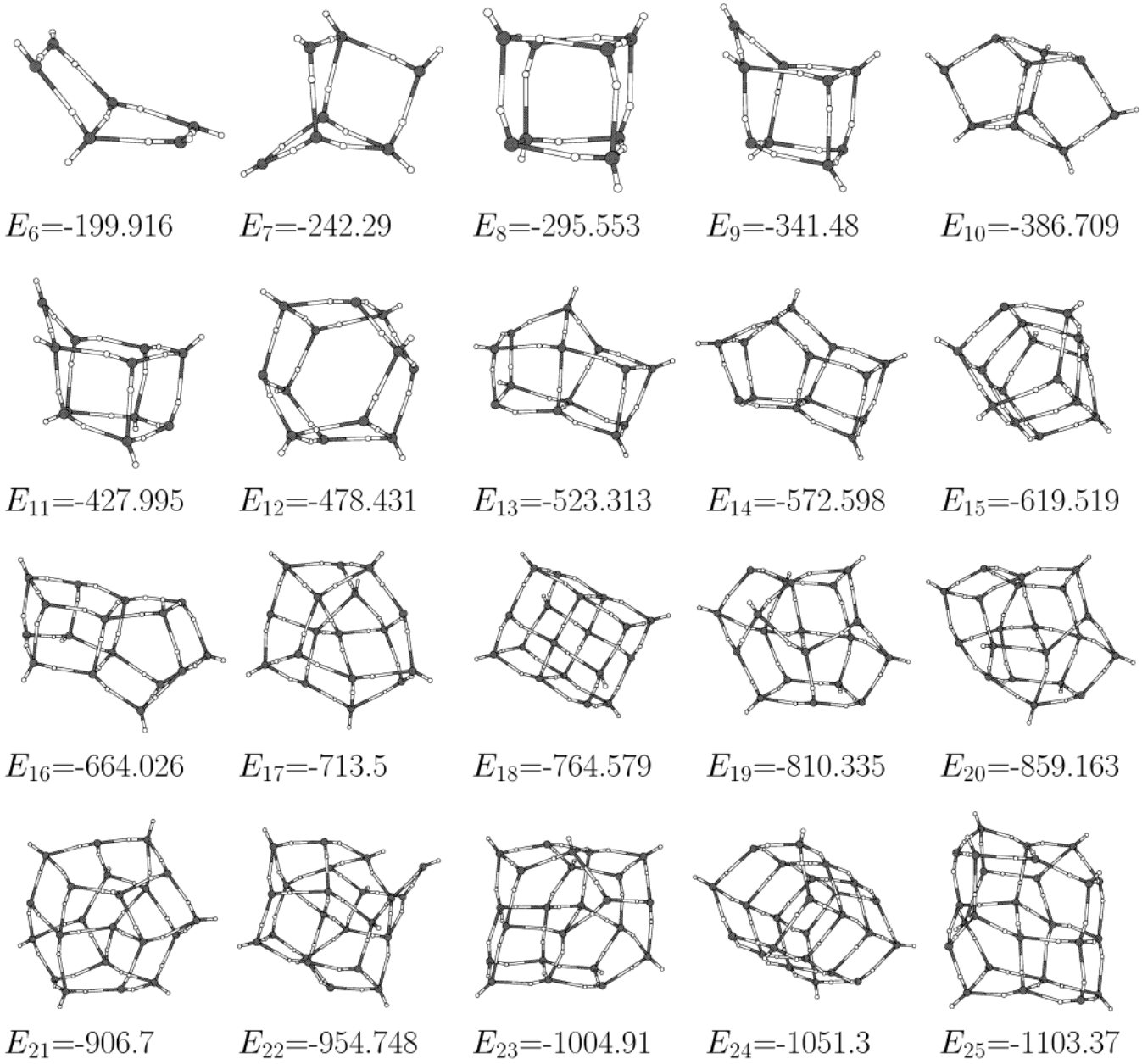

Figure 3. Apparent global minimum structures of water clusters consisting of between 6 and 25 molecules. The potential energy is calculated using the TIP3P potential. Energies are given in kJ per/mol.

structure by any of the three model potentials, they do perform better for the bigger clusters with up to twelve molecules. The SCP/E potential predicts structures similar to the ab initio results for $N = 8$, 9, 11, and 12,[20,21] the TIP3P potential for $N = 8-11$, and the TIP4P potential for $N = 7-9$, and 12.

The clusters consisting of 13 to 15 molecules basically possess the same global minimum structures for all three model potentials. The $N = 13$-cluster structure is based on the cage structure of the TIP4P and SPC/E $N = 12$ apparent global minimum structure. One molecule is added to the larger side of this cuboid to form a pentagonal ring replacing the square. The apparent global minimum structure of the $N = 14$-cluster is composed of a cage and two layered pentagonal rings. In all three models, the apparent global minimum structure of the water cluster consisting of 15 molecules is a "tube" composed of three layers of pentagonal rings. For the larger clusters, the three potentials produce a wide variety of apparent global minimum structures. The clusters consisting of 16, 18, 20, and 24 molecules in the SPC/E model, 16, 18, and 24 in the TIP3P model, and 16, 18, 20, and 24 in the TIP4P model have simple geometric structures based on cubic cages and pentagonal rings or, in the 18 molecules SPC/E and TIP3P apparent global minimum structures, hexagonal rings. The first cluster with an interior molecule is the $N = 17$ TIP3P apparent global minimum structure. In the other two models, the first cluster with an interior molecule appears at $N = 19$. This change in structure constitutes a second step toward the bulk phase. Clusters with up to six molecules are flat. Subsequently a transition to three-dimensional structures occurs, where all molecules are on the cluster surface. In the case of the three center models, SPC/E and TIP3P, there are only two clusters with no interior molecule after the first cluster with an interior molecule. In the case of the TIP4P model, we find four such clusters. The clusters with interior molecules do not show obvious symmetry and are best described as amorphous. For the cluster sizes we are able to study, there is no cluster with more than one interior molecule. The next step toward the bulk phase should take place when the interior molecules outnumber the surface molecules. On a

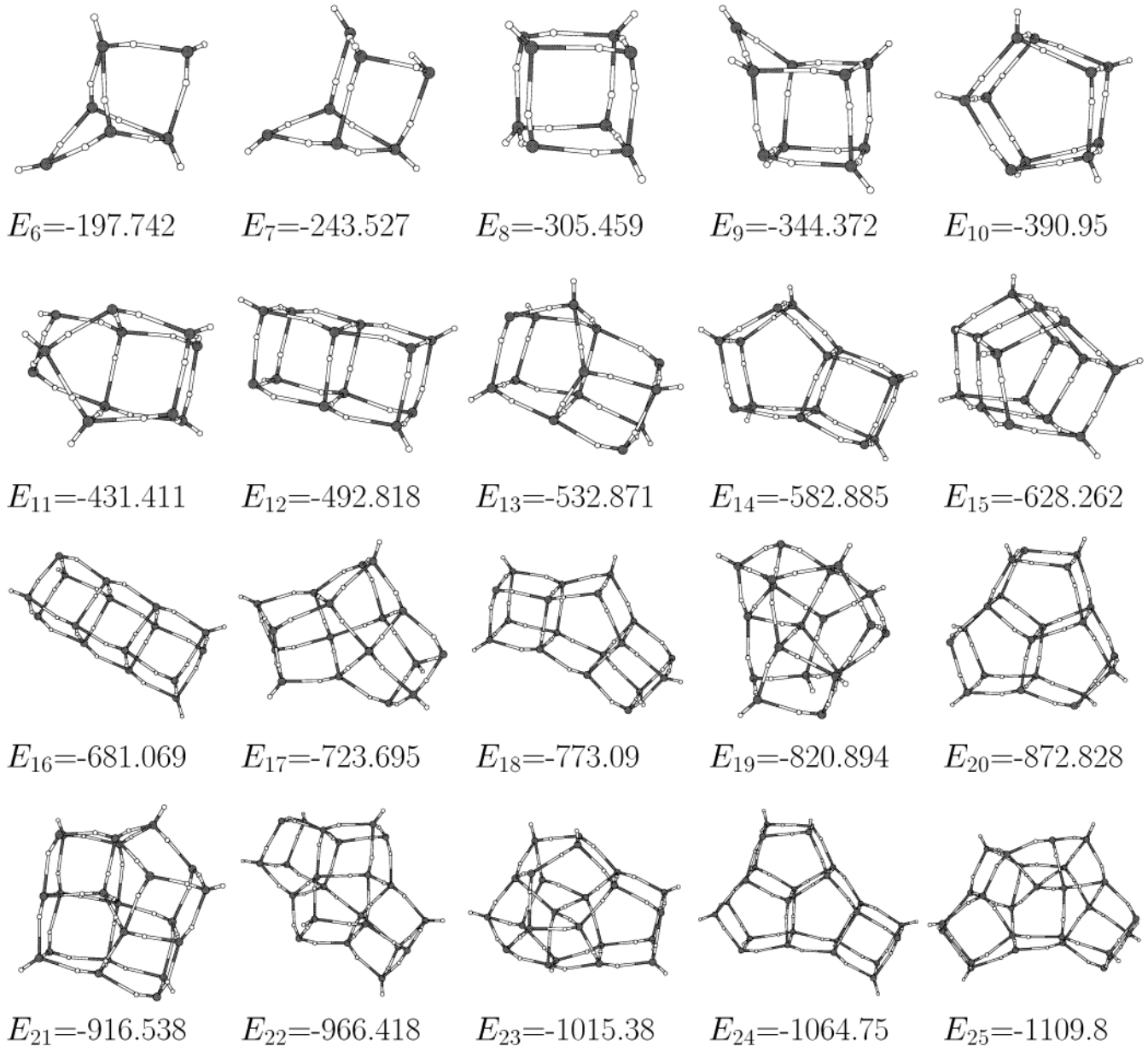

Figure 4. Apparent global minimum structures of water clusters consisting of between 6 and 25 molecules. The potential energy is calculated using the TIP4P potential. Energies are given in kJ/mol.

cubic lattice this happens for edge lengths between 9 and 10, corresponding to cluster sizes of about 500 to 1000 molecules.

The following figures compare various geometric quantities and related energies for the aforementioned apparent global minimum structures. Figure 5 shows the average O−O distance in the various clusters. The error bars display the shortest and the longest O−O distance between neighboring molecules. The open circles represent the average number of neighboring molecules. Contrary to what one finds with polarizable water[15] we observe only little change in the average O−O distance. However, certain clusters have unusually large maximum neighbor distances, like the $N = 22$ SPC/E cluster or the 17, 19, and 20 molecule TIP3P clusters. The geometries of those apparent global minimum structures are clathrate-like. Figure 6 is similar to Figure 5 but it displays the average hydrogen-bond distance and the average number of hydrogen-bonds per molecule. The error bars again refer to the shortest and the longest hydrogen-bond in the respective cluster. Note that the number of hydrogen-bonds in a tetrahedral network is 2, and the number of hydrogen-bonds as a function of N is similar in all three models. Note also that the number of O−O neighbors is the same as the number of hydrogen-bonds except for three TIP3P clusters. These clusters are the three clusters consisting of 17, 19, and 20 molecules possessing an unusually large maximum O−O neighbor distance.

Assuming that the cohesive energy of a cluster can be written as the sum of a volume term plus a surface term, we write the total energy per molecule as

$$\frac{E}{N} = E_{\text{bulk}} - \frac{E_{\text{surf}}}{\sqrt[3]{N}} \tag{2}$$

Figure 7 shows the energy per molecule of the apparent global minimum clusters together with fits corresponding to eq 2. The difference between the energy per molecule and the fit is shown in the insets. The fits yield the "bulk" energies per molecule: $E_{\text{bulk}} = -66.46$ kJ/mol for the SPC/E model, -59.76 kJ/mol for the TIP3P model, and -59.67 kJ/mol for the TIP4P model. From the fits we see that the $N = 8$-cluster is especially stable in the SPC/E and TIP4P models but not in the TIP3P model,

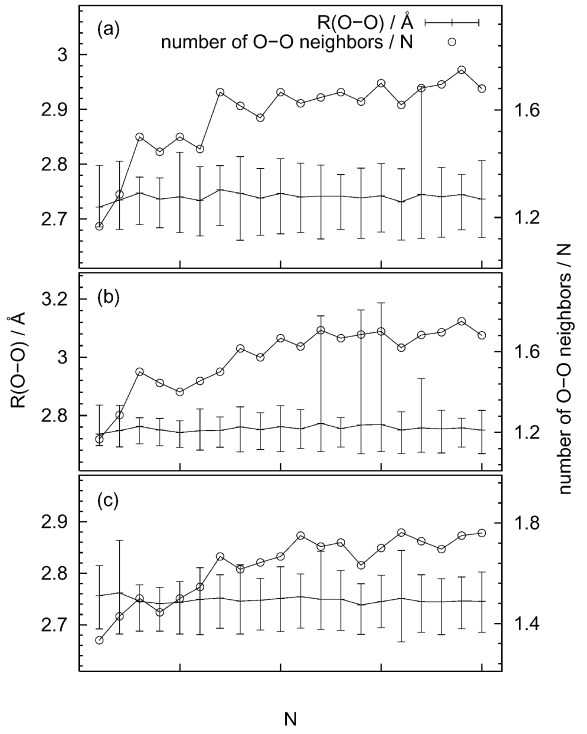

Figure 5. The average (O−O) distance and the number of neighbors for (a) SPC/E, (b) TIP3P, and (c) TIP4P. The error bars display the minimum and maximum distances.

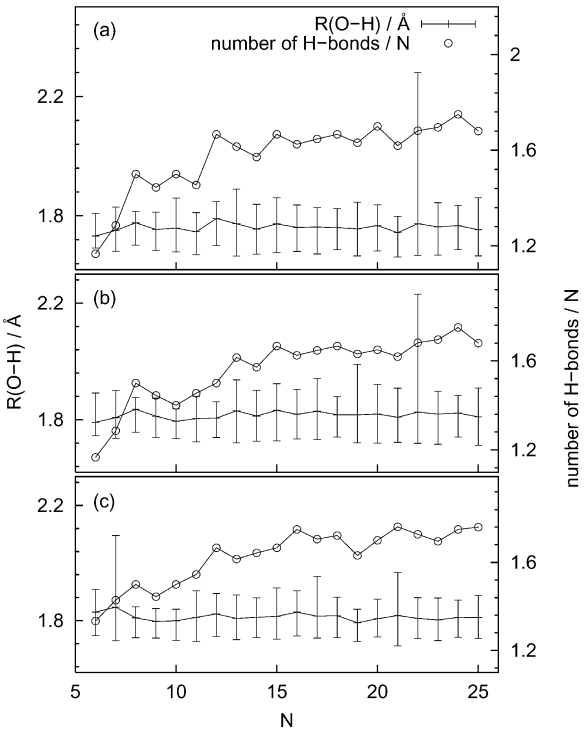

Figure 6. The average hydrogen-bond distance and the number of hydrogen-bonds for (a) SPC/E, (b) TIP3P, and (c) TIP4P. The error bars display the minimum and maximum distances.

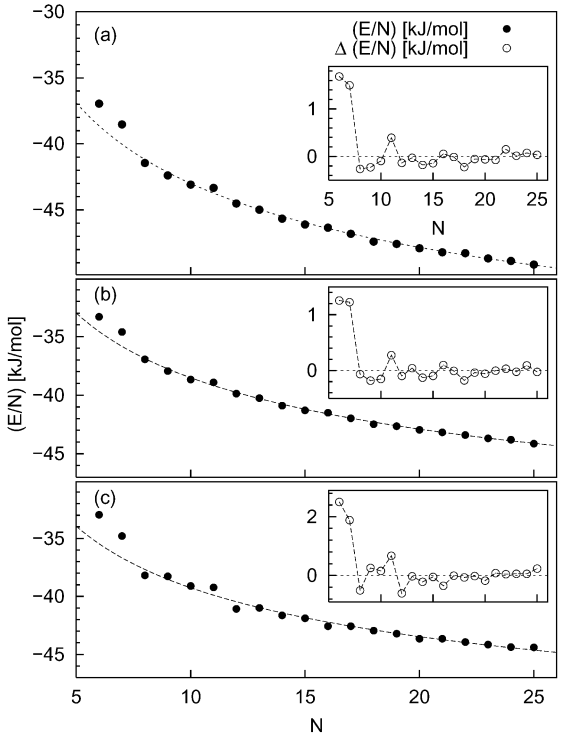

Figure 7. The apparent global minimum energy per molecule, E/N, vs N for (a) SPC/E, (b) TIP3P, and (c) TIP4P. In each panel the dashed line is a fit using eq 2 in the range $10 \leq N \leq 25$. The inset shows the difference between E/N and the fitted curve.

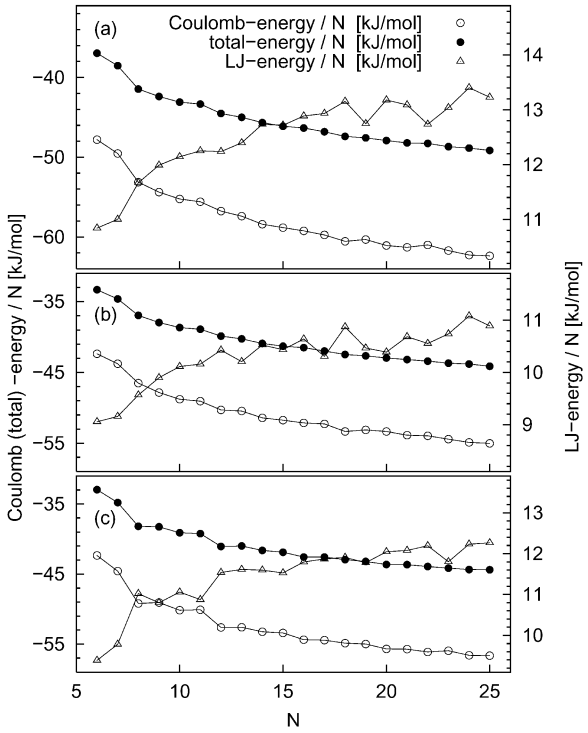

Figure 8. E/N together with its Lennard-Jones and Coulomb parts vs N for (a) SPC/E, (b) TIP3P, and (c) TIP4P.

where clusters consisting of 9 and 10 molecules appear more stable. Furthermore we see that clusters consisting of 11 molecules are relatively less stable in all three models. Note that in the TIP3P model, where clusters most often have interior molecules, the 24-molecule cluster is energetically even disfavored though it is highly symmetric due to the fact that all of its molecules are on the surface. This situation is different for the $N = 18$-cluster in the same model, which has a similar geometry, for it is especially stable. Thus we can assume that it is unlikely that global minimum structures with $N \geq 25$ possess only surface molecules at least for the TIP3P potential. Overall the insets in Figure 7a and Figure 7b are very similar. This indicates similarity of both models even though the

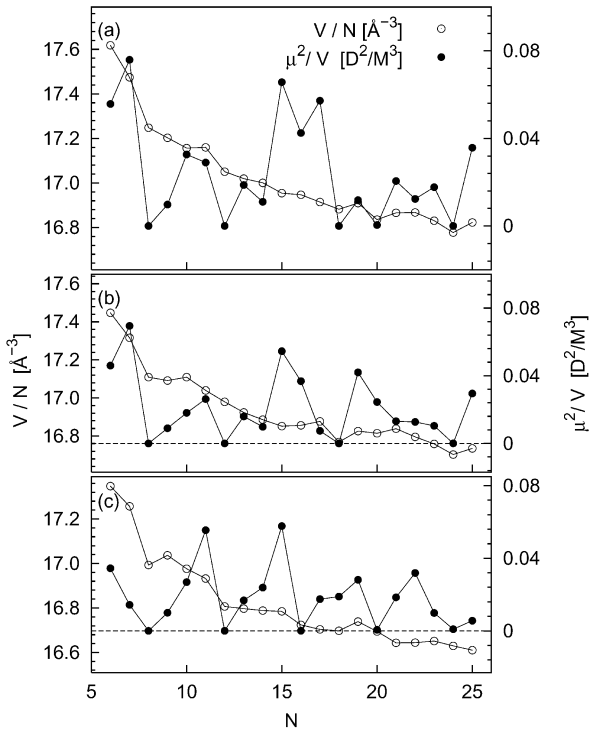

Figure 9. The cluster volume per molecule, V/N, and the quantity μ^2/V vs cluster size N for (a) SPC/E, (b) TIP3P, and (c) TIP4P.

apparent global minimum structures as well as E_{bulk} are quite different.

The total energy of the clusters in these simple point charge models consists of a Coulomb part and a Lennard-Jones part. In Figure 8 these two parts are shown separately together with the total energy. As expected, the Lennard-Jones energy is positive and prevents the clusters from collapsing. For the larger clusters the Coulomb energy varies smoothly but the relative changes in the Lennard-Jones energy are more pronounced.

Finally in Figure 9, we show the cluster volume calculated via a Monte Carlo integration. Here we use the van der Waals radii 1.52 Å for oxygen and 1.2 Å for hydrogen as a measure of the atom size and obtain a molecular volume of 19.7 Å^3 for SPC/E and 19.3 Å^3 for TIP3P as well as for TIP4P. The molecular volume as a function of N decreases. From this quantity we can deduce how compact a cluster is. A surprising result is that the clusters consisting of 24 molecules in the SPC/E and TIP3P models are very compact, but have only surface molecules.

A more useful quantity than V itself is μ^2/V, which is a measure for the clusters solvation energy.[7] In this simple picture a point dipole is placed in a spherical cavity of radius a inside an infinite polarizable medium (dielectric constant ϵ). The quantity $\Delta E = (\epsilon - 1)\mu^2/(2\epsilon - 1)a^3$ is the attendant change in the energy of the total system. In this sense it is a crude measure for the energy (or even free energy) of solvation of a molecule. Guided by this idea we take μ^2/V to be a measure for the relative stability of clusters in an aqueous medium as a function of N. Note the strong effect of the dipole moment because it enters squared and changes drastically from cluster size to cluster size. The regular clusters consisting of 8, 12, 18, 20, 24 (SPC/E); 8, 12, 18, 24 (TIP3P); and 8, 12, 16, 20, 24 (TIP4P) molecules have a vanishing dipole moment. Thus, their solvation energy should be small. On the other hand certain small clusters and the clathrate-like structures have a non vanishing dipole moment, so they should be quite stable in solution. Another measure of cluster stability certainly is the number of "dangling" hydrogen bonds, which may be read off Figure 6. But since this number varies less strongly with N, we do expect that μ^2/V contributes the major effect when comparing clusters of not too different sizes.

4. Conclusion

With a robust and quite simple genetic algorithm we succeeded in finding apparent global minimum structures of water clusters consisting of up to 25 molecules. It seems that we have taken Monte Carlo basin-hopping as well as our version of the genetic algorithm to a limit, because we had to invest extensive computational power on a parallel machine to compute the large cluster structures. We have presented four new apparent global minima for the well-known TIP4P model potential and about 10 new apparent global minimum structures for the two other model potentials, SPC/E and TIP3P. All three model potentials yield the same result for the cluster sizes $N = 13$ to 15, and we expect these cluster structures to be verified experimentally in the future.

References and Notes

(1) Chaplin, M. F. *Biophys. Chem.* **2000**, *83*, 211.
(2) Ludwig, R. *Angew. Chem., Int. Ed.* **2001**, *40*, 1809.
(3) Niesse, J. A.; Mayne, H. R. *J. Comput. Chem.* **1997**, *18*, 1233.
(4) Keutsch, F. N.; Saykally, R. J. *PNAS* **2001**, *98*, 10533.
(5) Hartke, B. *Angew. Chem., Int. Ed.* **2002**, *41*, 1468.
(6) Kabrede, H.; Hentschke, R. *J. Phys. Chem. B* **2002**, *106*, 10089.
(7) Burkert, U.; Allinger, N. L. *Molecular Mechanics*; American Chemical Society: Washington, DC, 1982.
(8) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91*, 6269.
(9) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D. *J. Chem. Phys.* **1983**, *79*, 926.
(10) Wales, D. J.; Hodges, M. P. *Chem. Phys. Lett.* **1998**, *286*, 65.
(11) Leach, A. R. In *Reviews in Computational Chemistry*; Liplowitz, K. B.; Bogl, D, Eds.; VCH−Publishers: Amsterdam, 1991; Vol. 2, pp 1−55.
(12) Wawak, R. J.; Wimmer, M. M.; Scheraga, H. A. *J. Phys. Chem.* **1992**, *96*, 5138.
(13) Berry, R. S. *Chem. Rev.* **1993**, *93*, 2379.
(14) Goldberg, D. E. *Genetic Algorithms in Search Optimization and Machine Learning*; Addison-Wesley: Reading, 1989.
(15) Quian, J.; Stöckelmann, E.; Hentschke, R. *J. Mol. Model.* **1999**, *5*, 281.
(16) Sremaniak, L. S.; Perera, L.; Berkowitz, M. L. *J. Chem. Phys.* **1996**, *105*, 3715.
(17) Hartke, B. *Z. Phys. Chem.* **2000**, *214*, 1251.
(18) Laasonen, K.; Parrinello, M.; Car, R.; Lee. C.; Vanderbilt, D. *Chem. Phys. Lett.* **1993**, *207*, 208.
(19) Mhin, B. J.; Kim, J.; Lee, S.; Kim, K. S. *J. Chem. Phys.* **1994**, *100*, 4484.
(20) Sadlej, J. *Chem. Phys. Lett.* **2001**, *333*, 485.
(21) Lee, H. M.; Suh, S. B.; Kim, K. S. *J. Chem. Phys.* **2001**, *114*, 10749.