

Chapter 2

How Can One Locate the Global Energy Minimum for Hydrogen-Bonded Clusters?

Sergey Kazachenko and Ajit J. Thakkar

Abstract An important problem in many areas of chemistry and physics is finding the global energy minimum on a potential energy surface. The difficulty stems from the exponential increase in the number of local minima with the size of the system. An efficient algorithm to find the global minima of water clusters is described and tested. It works well for clusters containing up to about 55 water molecules. A generalization to other hydrogen-bonded clusters is outlined. Applications of this algorithm to water clusters and methanol clusters have already been reported in the literature.

2.1 Introduction

Material particles consisting of a few to a few thousand atoms are called clusters. Cluster properties can have dramatic size and shape dependence. Clusters can serve as building blocks for new materials and electronic devices. Hence clusters of metals, semiconductors, ionic solids, rare gases, and small molecules have been studied using both theoretical and experimental methods. Atomic and molecular clusters [1] are held together by hydrogen bonds [2] or by relatively weak intermolecular forces [3, 4].

In particular, small clusters of hydrogen-bonded water molecules have received a lot of attention; see, for example, spectroscopic work [5–16], and density functional theory and *ab initio* investigations [17–31]. Water is not a simple substance and has anomalous physical and chemical properties. More than a century of work has been devoted to modeling and understanding these properties. Nevertheless, many aspects of water remain unsolved puzzles and the development of water models continues

S. Kazachenko · A.J. Thakkar (✉)
Department of Chemistry, University of New Brunswick,
Fredericton, NB E3B 5A3, Canada
e-mail: ajit@unb.ca

S. Kazachenko
Present Address: Department of Chemistry, Queen's University,
Kingston, ON K7L 3N6, Canada

to attract much interest. Water models can use water clusters, $(\text{H}_2\text{O})_n$, as building blocks [32–36].

An important part of cluster research is a characterization of the global minimum of the potential energy surface—the structure of the most stable cluster. A scholarly and comprehensive survey of the difficult problem of locating the global minimum on a potential energy surface is available [37]. Computational approaches to the determination of global minima are very attractive because the experimental determination of the structure of the ground state is extremely difficult for complex systems. Unfortunately, the computational solution is rather expensive because the number of local minima on the surface grows exponentially with the dimensionality of the surface or, in other words, with the number of atoms in the system. An algorithm searching for a global minimum can easily get trapped in one of these local minima because the traversal of many intermediate minima and the crossing of high barriers separating them may be required to find a lower local minimum. In high-dimensional cases, which are the norm rather than the exception in physically important clusters, a global minimum can only be discovered and verified after a sufficiently large number of low energy minima have been located and compared.

We have been studying water clusters for some time now. Hence we have been gradually developing and refining our own algorithm for locating a global energy minimum for water clusters. We previously described successive improvements to our algorithm in an incremental manner in a series of papers [38–41] and a thesis [42]. The purpose of this chapter is to try and describe the current state of our algorithm in a relatively self-contained and cohesive manner without the reader having to consult our previous work.

All algorithm tests on water clusters discussed in this chapter are done with the empirical TIP4P model [43], a reparametrization of the venerable Bernal-Fowler model [44]. The interactions are considered to be pair-wise additive and the water monomers are held rigid so that they do not vibrate. The interaction between a pair of water molecules is given by a Lennard-Jones (12,6) interaction between the oxygen atoms and electrostatic interactions between three point charges on each water molecule. There are positive charges on the hydrogen atoms and a balancing negative charge between the hydrogen atoms along the C_2 symmetry axis. The TIP4P model gives a reasonable thermodynamic description of liquid water. Many studies have been devoted to finding the global minima of TIP4P water clusters [38–40, 45–51]. Generally good agreement has been found [48] between its predictions for small clusters with $n \leq 12$ and both *ab initio* and experimental results. The TIP4P global minima for cluster sizes up to 47 are now firmly established [41] except for $n = 39$ and $n = 45$. Putative TIP4P global minima have been reported [41] for larger clusters with $n \leq 55$. Low-energy TIP4P structures for much larger clusters with selected sizes have been located [7, 49]. Pure water clusters based on the TIP4P model are now a benchmark for methods for global optimization of hydrogen-bonded clusters.

Some of the basic terminology commonly used in the global optimization literature is summarized in Sect. 2.2 which also contains a brief description of basin and minima hopping. Next, the long Sect. 2.3 details the various ways by which we

optimize the topology of water clusters with a fixed oxygen skeleton. The overall algorithm is then described in Sect. 2.4 and its generalization to other hydrogen-bonded clusters is outlined in Sect. 2.5. A few concluding remarks are made in Sect. 2.6.

2.2 Basin Hopping and Minima Hopping

A discussion of search algorithms is facilitated by the introduction of some terminology that is widely used in the literature. A *basin* is a region of geometrical configuration space around a minimum on the potential energy surface. The basin contains all structures, or configurations, from which a search can reach this minimum using only small steps and downhill moves. A *super-basin* is the union of several neighboring basins. A *funnel* $\mathcal{F}$ is a super-basin with the property that starting at any point in $\mathcal{F}$ one can reach the lowest minimum in $\mathcal{F}$ without crossing barriers that are very high relative to the average energy difference between local minima in $\mathcal{F}$. Figure 2.1 is a schematic illustration of two super-basins, one of which is a funnel and the other one is not.

Use of the Boltzmann factor, $\exp(-\Delta E/kT)$, to control all the steps used to leave a basin makes the crossing of high barriers a rare event. Hence, global minimization methods based exclusively on thermodynamic principles can be extremely slow because they may find it hard to exit a funnel and may repeatedly visit neighboring configurations that are close in energy. Many standard algorithms such as simulated annealing [53–55] and basin hopping [46, 56–59] are based on thermodynamic principles. However, genetic algorithms [60–62] and minima hopping [52] are not.

The fundamental idea in basin hopping is that the potential energy surface is effectively transformed to a stepped surface [46, 56, 63]. This is done by ending each search step with a local optimization so that one effectively searches over local minima. How one starts the next search step distinguishes various algorithms based on this seminal idea. We chose minima hopping [52] as the global optimization

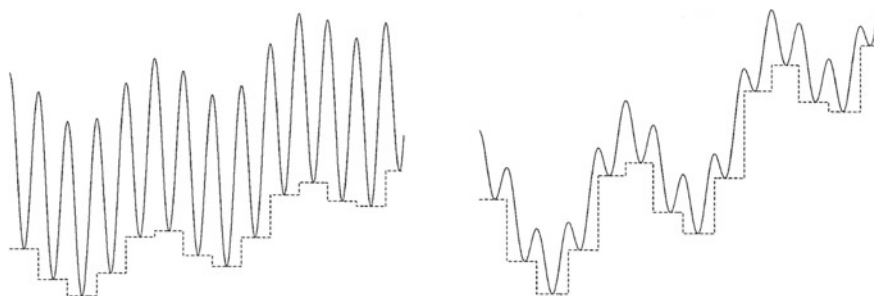


Fig. 2.1 Two super-basins: the one on the right is a funnel but the one on the left is not. Figure adapted from Goedecker [52]

algorithm in our preliminary work. Minima hopping can be thought of as a version of basin-hopping [46, 56, 63]. Its goal is to avoid revisiting previously located local minima without forbidding repeated passages through any transition basins that may separate many funnels. This is achieved in the minima-hopping method by maintaining a list of visited minima and using adaptive thresholds to leave a basin more and more vigorously each time the algorithm revisits it. Molecular dynamics (MD) steps are used as an efficient mechanism for crossing energy barriers.

The minima-hopping algorithm has an inner part for jumping to a new local minimum and an outer part for accepting or rejecting the new local minimum. In the inner part, one tries to escape the current minimum $\mathbf{M}_c$ by a short MD simulation in which the atoms have a Boltzmann velocity distribution such that their kinetic energy is fixed at E_{kin} . The simulation is stopped as soon as a minimum is encountered along the trajectory or the maximum number of MD steps is exceeded. Then one optimizes to the closest local minimum $\mathbf{M}$ using a suitable local optimization method. If this minimum has been visited previously, then multiply E_{kin} by $\beta > 1$ to make the next escape attempt more vigorous and repeat the inner part. If this is a newly found local minimum, divide E_{kin} by β to make the next escape attempt less vigorous and go to the outer part. The latter accepts or rejects the local minimum $E(\mathbf{M})$ as follows. If $E(\mathbf{M}) - E(\mathbf{M}_c) < E_{\text{diff}}$, the minimum is accepted and E_{diff} is divided by $\alpha > 1$ to make the next acceptance more difficult. Otherwise it is rejected and E_{diff} is multiplied by α to make the next acceptance easier. We tried various values of α and β but were unable to improve upon the values of $\alpha = 1.02$ and $\beta = 1.05$ recommended by Goedecker [52]. The minima-hopping algorithm stops when E_{kin} reaches or exceeds a maximum value $E_{\text{kin}}^{\text{max}}$ or the number of inner steps exceeds a preset limit N_{iter} .

It is important to do local optimizations as efficiently as possible because they are a time-consuming part of minima hopping. In our implementation, we perform local minimization in two steps. The first optimization uses the limited memory L-BFGS method [64, 65] with a loose convergence threshold. The optimization is refined in the second step which uses Davidon's optimally conditioned variable metric method [66] and a more stringent convergence criterion.

It soon became apparent to us that minima hopping did not always succeed in finding the optimum hydrogen bond topology for water clusters [39]. Nevertheless, minima hopping served us as the basic algorithm upon which improved and specialized methods for water clusters were built. Special topology refining algorithms, described in the next section, are used both within the global search algorithm and independently for a refining step on the list of saved low energy minima. Although many features described here were designed to work only with water, the core algorithm is general and has been applied successfully to other hydrogen-bonded clusters including pure methanol clusters [67] and pure clusters of ethanol, n-propanol, and iso-propanol.

2.3 Finding the Optimal Hydrogen Bond Topology

2.3.1 Representation of H-Bonded Clusters

An algorithm for global optimization of hydrogen-bonded clusters requires a convenient description of a cluster and its features. At the most basic level, the cluster is defined by the Cartesian coordinates and types of the atoms it consists of. We use that information to derive some important properties and find a convenient way to represent them.

A basic property of any H-bonded cluster is the presence or absence of a hydrogen bond between a pair of given molecules. We use simple geometric criteria to decide whether there is an O–H...O hydrogen bond: the H...O distance should be less than 2.5 Å and the O–H...O angle should be greater than 90°. It is convenient to describe the H-bonds in a cluster of n molecules by a graph. The latter is represented as an $n \times n$ adjacency matrix $\mathbf{A}$ with elements obeying the usual rules: $A_{ij} = 1$ if there is a bond between molecules i and j and $A_{ij} = 0$ otherwise. By definition, $A_{ii} = 0$ because molecules are not connected to themselves and $A_{ij} = A_{ji}$ because the bond direction is not taken into consideration. The adjacency matrix allows one to calculate, for example, the total number of hydrogen bonds in a cluster as $N_{\text{bond}} = \frac{1}{2} \sum_{ij} A_{ij}$. One can also calculate the number of rings of a given size formed by connected molecules. We implemented the counting of rings using a backtracking algorithm based on the work of Franzblau [68]. There are several possible definitions of a ring in a graph. We count all rings in which each monomer is connected to exactly two other monomers belonging to the same ring.

The H-bond directionality is important in topology optimization. Keeping track of both the existence and direction of the H-bonds in a cluster can be accomplished by using a digraph. The latter can be represented by a directed adjacency matrix $\mathbf{D}$. As with the adjacency matrix, the absence of a bond between monomers i and j is indicated by $D_{ij} = 0$ and so all diagonal elements vanish, $D_{ii} = 0$. If there is a bond in the $i \rightarrow j$ direction then $D_{ij} = 1$ and $D_{ji} = 0$. The $\mathbf{D}$ matrix contains, for example, information about the number of donor (don) and acceptor (acc) H-bonds for each molecule: $N_{\text{don}}^i = \sum_j D_{ji}$, $N_{\text{acc}}^i = \sum_j D_{ij}$. Note that the summation index depends on the definition of the H-bond direction. In a computer program, the sparsity of $\mathbf{A}$ and $\mathbf{D}$ can be exploited by storing them as linked lists of non-zero elements.

2.3.2 Why Is Topology Optimization Needed?

A simple characterization of a cluster of n water molecules is its skeleton or graph by which we mean the connectivity of the monomers described by an adjacency matrix $\mathbf{A}$. The positions of the oxygen atoms and the skeleton define the shape of a cluster as in the left panel of Fig. 2.2. However, virtually all hydrogen bonds between water molecules can be assigned a direction, say from donor to acceptor. Hence,

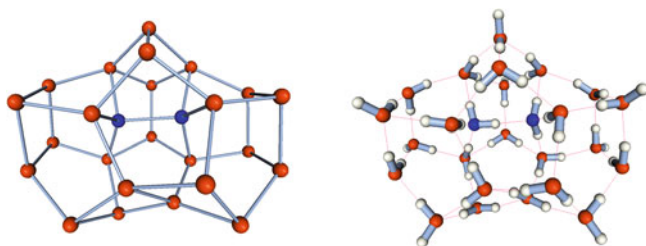


Fig. 2.2 An example of a framework and a complete cluster with one of the possible H-bond topologies for a $(\text{H}_2\text{O})_{26}$ cluster. Two internal molecules are shown in *blue*

there can be many water cluster geometries with the same skeleton but different hydrogen bond topologies; that is to say, cluster structures with the same skeleton but different directions in one or more of the hydrogen bonds. In the terminology of Sect. 2.3.1, any given adjacency matrix $\mathbf{A}$ can correspond to many directed adjacency matrices $\mathbf{D}_k$. The oxygen framework and the H-bond topology together define a complete cluster as in the right panel of Fig. 2.2. The stability of a given framework can vary significantly depending on the hydrogen bond distribution [69–75].

Since each cluster framework maybe paired with a large number of H-bond topologies, the resulting cluster structures can have significant energy differences. Therefore, it is important that we are able to locate the one with the lowest energy. It might also be of interest to know how many topologies exist for a given framework and possibly separate them into categories. Locating the minimum energy topology, or topology optimization, can be done either as a separate procedure on selected water clusters or as a part of a global optimization.

The problem of finding the best hydrogen bond topology for water clusters has been studied extensively; see, for example, Refs. [49, 51, 69–74, 76–83]. Polyhedral and cubic water clusters were described in several publications using graph theory [69, 74, 78, 79]. The effects of H-bond topology on the stability and spectroscopic properties of water octamers were studied by Francisco et al. [73, 80]. The relation between the topology and interaction energy was studied for polyhedral clusters [72, 81] and for some other shapes and sizes [71, 81]. As a result, several formulas were developed to predict the relative energy of a water cluster based on its monomer connectivity. A different approach is to use a proton transfer to change the direction of hydrogen bonds and so sample a number of topologies using general optimization methods [49, 51]. There is also the brute-force approach of examining all possible topologies for an arbitrary cluster shape [77, 82, 83]. The last two approaches will be discussed later in more detail.

Unfortunately, most of the suggested methods were designed only for a particular oxygen framework or turned out to be inefficient. We created several algorithms suitable to our goal of reliably finding the lowest energy topology for a water cluster of an arbitrary shape. The algorithms were efficient at locating lowest-energy topologies when applied to clusters with no more than 55 molecules.

2.3.3 *Comparing and Storing Water Cluster Minima*

In our approach for H-bond topology optimization, a water cluster is defined as a structure with a particular oxygen framework. A framework is assumed to have a unique adjacency matrix $\mathbf{A}$. No two frameworks will have the same matrix and, therefore, clusters can be compared by comparing their $\mathbf{A}$ matrices. All possible topologies for a framework are considered to be variations of the same cluster. Therefore, each cluster defined by its $\mathbf{A}$ matrix has a set of local minima corresponding to all possible H-bond topologies. Such separation of the framework and topology helps to reduce the number of local minima that are stored during a global minimum search. Only the version of a cluster with the best topology is kept.

How can we determine if two sets with an equal number of water molecules form clusters with an identical shape? $\mathbf{A}$ matrices can be compared directly if the position of the molecules has changed only slightly, i.e. after a small distortion. However, in general we want the comparison to be independent of the order of molecules in a coordinate list. In our case this is done in three steps which helps to reduce the number of time-consuming operations. First, the number of H-bonds must be the same, which is easy to calculate and to check. Next, we use the idea that the way molecules are connected has an effect on the number of rings formed by H-bonded monomers in a cluster. Moreover, the number of rings is independent of the order of molecules in a list. Therefore, as the second step, it is required that the number of rings of each size from 3 to 10 molecules must be the same. For those cases where the ring rule is also satisfied, an alignment of $\mathbf{A}$ matrices is used to perform the final check which must allow for the possibility that the order of the elements (monomers) is different. A backtracking procedure is used to examine the connectivity of each monomer and to try and match pairs of molecules from the clusters being compared. This procedure allows us to determine whether two matrices correspond to the same framework even when the order of the elements is not the same.

When the list of visited minima is large enough, it is no longer feasible to keep either the Cartesian coordinates or the adjacency matrix for each local minimum. In that case, the third step that includes alignment of $\mathbf{A}$ matrices can be omitted. Nine integers, the number of H-bonds and the numbers of three- through ten-membered rings, are saved for each local minimum and provide a robust and reliable way to compare cluster frameworks.

2.3.4 *Filters to Screen Topologies*

Finding the lowest-energy topology of a large cluster framework would be impossible if it required the local optimization of all or even most of its topologies. Fortunately, the number of expensive local optimizations required can be reduced by using simple filters or criteria based solely on geometrical considerations to weed out the topologies that are likely to have a high energy. Note that although the shape of a

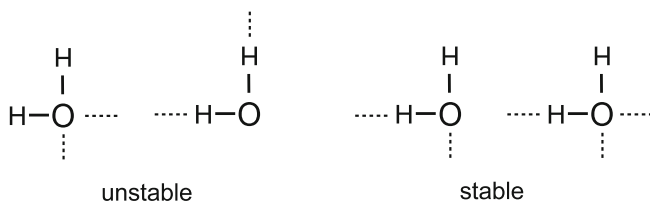


Fig. 2.3 Examples of less stable and more stable hydrogen bonding of a monomer

water cluster depends somewhat on its topology, a stable (i.e. low-energy) framework does not break or change when its topology is changed.

The first thing to do is to count the number of H-bonds and their types for each molecule. We define a donor bond as a bond created by a hydrogen atom of a given molecule and an acceptor bond as the one created by an oxygen atom of the molecule. We use a modification of the venerable Bernal-Fowler “ice rules” [44] and require that

$$1 \leq \text{number of donor bonds} \leq 2,$$

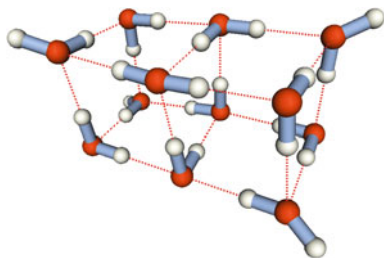
and

$$1 \leq \text{number of acceptor bonds} \leq 2 \text{ or } 3.$$

In words, each water molecule in a cluster must have at least one donor and one acceptor bond. Moreover, there can be at most two donor bonds (one for each hydrogen) and at most two acceptor bonds. These rules allow one to avoid unstable connectivity of molecules; see Fig. 2.3. In rare cases a water molecule in a cluster accepts three hydrogen bonds and it is necessary to account for such a possibility. However, allowing all molecules to accept three H-bonds would significantly increase the number of topology combinations. Thus, the algorithm detects penta-coordinated molecules in the input geometry and only those can have up to three acceptor bonds, while the rest are restricted to have no more than two acceptor bonds.

We found that in some cases changing the cluster topology leads to unrealistic monomer angles; see Fig. 2.4. Note that the molecules in a cluster shift somewhat during a local optimization of a new topology. Thus we must allow for a range of angles to be accepted. Test calculations suggest that a topology can reliably be

Fig. 2.4 An example of an unrealistic monomer angle in one of the H-bond topologies for an $(\text{H}_2\text{O})_{12}$ cluster



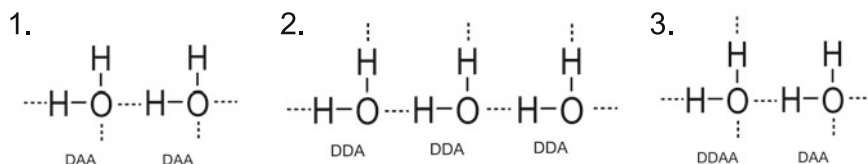


Fig. 2.5 Adjacent monomers with a connectivity that leads to higher energy (D—donor, A—acceptor)

considered unfavorable if it leads to a monomer angle larger than 150° or smaller than 65° . These values span a large enough range to ensure that no important topologies are missed because of the angle filter.

It is well-established that adjacent dangling hydrogen atoms lead to a higher cluster energy [69, 71–73, 79]. In addition, Anick found two other patterns leading to an increased energy [81]. These three motifs of connectivity are shown in Fig. 2.5. The types of H-bonds are labeled by D for donor and A for acceptor. The number of motifs 1 and 2 that occur in a cluster must be kept as low as possible. This means that one can safely discard all topologies with more such motifs than the lowest number found at any given stage. The third rule is not as strong. We found that a topology can be discarded safely only when the number of such motifs exceeds the lowest number found *plus two*.

2.3.5 Topology Optimization by Enumeration

A straightforward way of finding the best topology is to generate all possible H-bond distributions, do local optimizations on each one, and compare their energies. Miyake and Aida used adjacency and directed adjacency matrices (graphs and digraphs) to describe the framework and the topology of a cluster, respectively [82]. Knowing the digraph and the coordinates of the oxygen atoms, one can recreate the topology and then perform a local optimization to relax the cluster geometry. The use of adjacency matrices was a promising idea; however, in their algorithm all possible **D** matrices were generated first and only later were they checked to see if they were useful. This led to a large wasted computational effort and the largest cluster size they were able to study was limited to just eight water molecules. Vukićević et al. improved the performance of the method by eliminating the need to generate unrealistic matrices [77, 83] and analyzed the topology for clusters with up to 12 water molecules.

The ideas of using graphs and digraphs and eliminating undesired combinations on the fly were used in the creation of our method, called topology enumeration (NT). Consider the cage hexamer shown in Fig. 2.6a as an example. We start from the given framework and its adjacency matrix **A**. Each row and column of **A** with the same index corresponds to a molecule. Defining all connections of a molecule to the

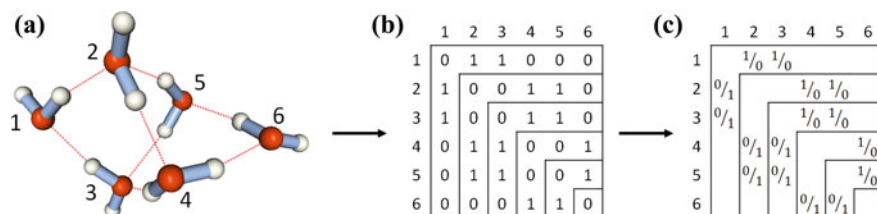


Fig. 2.6 Adjacency matrices for the enumeration method. **a** sample geometry; **b** adjacency matrix; **c** possible directed adjacency matrices

other molecules automatically defines the connections of all molecules to the current one. Therefore, it is sufficient to consider each row and column starting from the corresponding diagonal element; see Fig. 2.6b. The connectivity of the last molecule (number 6), as expected, is completely defined by the connectivity of the previous molecules.

The positions with ones in each row (or column) of A tell us which molecules are connected to a given molecule. Let us choose rows as a reference. Then, by going through all combinations of 1 and 0 for positions marked with “1/0” in Fig. 2.6c we can generate all possible H-bond directions for the given skeleton. This can be achieved by taking a bit representation of an integer value that ranges from 0 to $2^n - 1$. From the definition of D , the values in a column are the opposite of the values in a row (positions with “0/1” in Fig. 2.6c). H-bond directions are generated for each molecule in this manner using a backtracking loop. The backtracking allows one to apply geometry filters for each molecule right away without generating a complete matrix. The steps constituting our backtracking method are shown in Listing 2.1.

Listing 2.1 Backtracking algorithm for topology enumeration.

```

-----
Mark all bit combinations of all molecules as not used
Select molecule 1 as the current molecule 'M'
WHILE ('M' > 0)
  IF (all bit combinations for 'M' have been used) THEN
    Mark all bit combinations for 'M' as not used
    Set 'M' = 'M' - 1
  ELSE
    Choose next bit combination for 'M'
    Mark this bit combination as used
    IF (topology filters are satisfied) THEN
      Set 'M' = 'M' + 1
    END IF
    IF ('M' > number of molecules) THEN
      Create and save a cluster geometry for local optimization
      Set 'M' = 'M' - 1
    END IF
  END IF
END WHILE
-----

```

In the beginning it is important to sort molecules by their distance from the first one to allow the algorithm to backtrack from bad combinations sooner rather than later. The bicoordinated molecules are a special case because they can generally have two stable directions for the dangling hydrogen. This is accounted for by generating all combinations of the directions for bicoordinated molecules for each topology. A symmetrical framework is not considered as a special case because it is a rare case for water clusters. During creation of the geometry, hydrogen atoms are placed on the line between two corresponding oxygen atoms. In the case of rigid monomers (as in the TIP4P potential), and to save optimization time in the case of non-rigid ones, the monomer bond lengths and angles are adjusted to the equilibrium values of the given potential using a simple geometrical transformation.

A large number of trial topologies is usually generated. It is not possible to perform a tight local optimization for all topologies even with an inexpensive potential. Therefore, a local optimization is done in several steps of increasing precision. A single-point energy is calculated for the initial geometry of each topology. The geometries of the structures with the 5000 lowest single-point energies are then optimized until a 1.0 kcal/mol/Å gradient threshold is reached. Then the 500 best resulting geometries are further optimized with a 0.1 kcal/mol/Å threshold. Next, the 50 lowest-energy geometries that result are optimized using a 0.01 kcal/mol/Å threshold. Finally, a very tight optimization is performed at the best geometry.

The number of topologies generated for three cluster sizes and the effects of topology filters are shown in Table 2.1. Note that in these test cases the best topology found is the same no matter what filters are used; the difference is in the number of geometries that need to be locally optimized. The use of the modified ice rules still leads to a huge number of topologies left for energy calculations. Checking for monomer angle and adjacent dangling (non-H-bonded) hydrogen atoms further reduces the number of accepted topologies by three orders of magnitude. The next filter is not as effective but reduces the number of topologies in all cases. The final filter makes a difference only for the larger cluster with $n = 31$. Even with all the filters included the number of resulting topologies shows an exponential growth. Thus the method becomes increasingly inefficient for clusters with more than about 35 water molecules.

The backtracking loop for generating **D** matrices was also used in the routine that produces a random topology. Such a topology should be random but still satisfy

Table 2.1 The number of topologies to be locally optimized for a sample oxygen framework in $(\text{H}_2\text{O})_n$

Filters used	$n = 23$	$n = 28$	$n = 31$
Modified ice rules	$\sim 10^7$	$\sim 10^8$	$\sim 10^9$
+ HOH angle	$\sim 10^6$	$\sim 10^7$	$\sim 10^8$
+ 2 adjacent DAA	11556	46922	403465
+ 3 adjacent ADD	6892	34971	337986
+ AADD – AAD	6892	34971	318728

The plus sign means that a filter is added to all previous ones

topology filters. To achieve this, random bond directions are selected for each line of the **D** matrix. Then the backtracking search is performed starting from this random matrix until all topology filters are satisfied.

2.3.6 Short Topology-Altering Optimization

Kazimirski and Buch [49] foresaw and Takeuchi [51] refined a different approximate but fast H-bond topology optimization which can be used as part of a global minimum search. We now describe our version of this method which we call the “short topology-altering optimization”.

Tri-coordinate water molecules in a cluster are either double-acceptor, single-donor (AAD) molecules or single-acceptor, double-donor (ADD) molecules. The basic H-bond topology-altering operation is the reversal of the direction of the H-bonds in a contiguous sequence of water molecules. It can also be viewed as a proton transfer. It is possible to perform such a reversal without altering the connectivity in rings and in those chain substructures which have an AAD water at one end and an ADD water at the other end. Examples of such a H-bond topology-altering operation are shown in Fig. 2.7.

Sometimes this operation has the effect of creating an extra pair of non-H-bonded or “dangling” hydrogen atoms on adjacent water molecules. However, as discussed earlier in Sect. 2.3.4, structures with adjacent AAD water molecules are energetically unfavorable [69, 71–73, 79]. To reduce the number of such structures generated, we introduce a split-chain H-bond topology-alteration. This new operation consists of splitting a chain substructure into two shorter chains between a pair of adjacent ADD and AAD waters, and performing the H-bond reversal in both subchains without altering the H-bond connecting the subchains. An example is shown in Fig. 2.8. Adding this type of operation helps to distribute “dangling” hydrogen atoms on the surface of a cluster.

A short topology-altering search consists of applying the operations described above to all rings containing no more than 10 monomers and all chain and split-chain substructures with no more than 5 monomers. Note that Takeuchi used only rings and size 2 chain substructures while Buch and Kazimirski used only chains and rings of size 4. New topologies are created by changing bond directions of one substructure at a time. If a new topology has a lower energy, it is accepted as a new

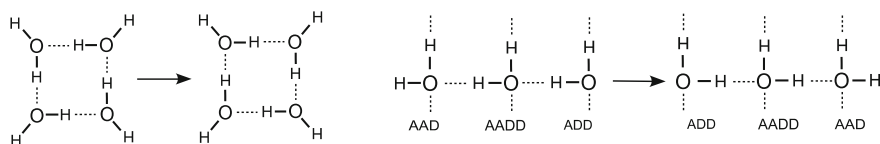


Fig. 2.7 Examples of a H-bond topology-altering operation applied to (left) a ring and (right) to a chain of water molecules

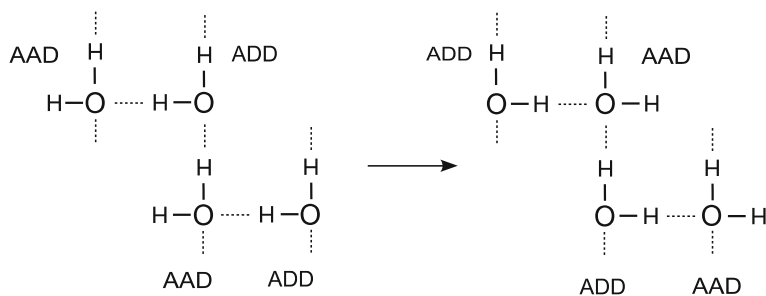


Fig. 2.8 A split-chain, H-bond topology-altering operation. In this example, the split is between the second and third waters

reference point and optimization returns to the smallest substructure. The search is allowed to move only in a downhill direction of the potential energy surface to limit the number of topologies considered. The framework of the initial cluster is allowed to change during optimization. In fact, a framework change does happen often when the input structure has a relatively high energy. The search is complete when no substructure alteration leads to energy lowering.

Since the algorithm goes through all possible substructure changes, it is likely that the same topology would be generated and locally optimized more than once. Saving optimized geometries in a list removes that problem at the cost of adding complexity to the algorithm. A list entry consists of an energy, Cartesian coordinates, and a **D** matrix. New structures are saved in the list. If a structure with a given **D** matrix is already in the list, the geometry and energy from the list are used instead of performing a local optimization. The list is created for a given framework. It can be limited to a single optimization call or can be global in case the parent process uses short topology-altering optimization as part of a larger algorithm as in Sect. 2.3.8. If the framework changes during the optimization, the list is either reset to be used with the new framework, or, in the case of a global list, the search is terminated. The algorithm can only go downhill in energy. Listing 2.2 describes the general flow of the ST algorithm.

Listing 2.2 Short topology-altering optimization (ST).

```

-----
Set substructure size M = 2
WHILE (M <= 10)
  FOR each substructure of type 'T'
    Find substructures 'T(M)' that allow altering H-bond direction
    FOR each substructure 'T(M)'
      Alter H-bond direction of the substructure
      IF (topology filters are satisfied) THEN
        Perform local optimization
        Keep the best topology generated for the current M
      END IF
    END FOR
  END FOR
END FOR

```

```

IF (lower energy geometry has been found) then
  M = 2
else
  M = M + 1
end if
END WHILE
-----

```

The effects of using topology filters and the minima list have been examined for three large cluster sizes: $n = 34, 43$, and 52 . For each size, 1000 low energy clusters with randomized topologies were generated as a test set. Short topology-altering (ST) optimization was applied to each cluster, first with no filters or lists as a reference point, and then adding one filter at a time. The average speed up of run times was calculated for each case; see Table 2.2.

The direction of a search often changes when filters are added. In some cases, optimization finishes at a higher energy minimum than the one reached without the use of the additional filter. However, with all of the improvements, the energy lowering is the same on average while the run time is reduced by about 40 %. Note that the AADD–AAD pair motif and the three adjacent ADD motif were not used as filters because the improvement in speed was not significant.

This algorithm has also been implemented to work with mixed clusters containing many water molecules and a few other molecules. The latter are treated as a special case and only the H-bond network between water molecules is allowed to change. Of course, for such an optimization to make sense, there should be enough water molecules to form a topological network. The algorithm can be used to optimize an initial distribution of a large number of water molecules around a solvated molecule, for example a protein, before starting a molecular dynamics simulation.

2.3.7 Extended Topology-Altering Optimization

The short topology-altering optimization allows only downhill moves. However, it is desirable to extend it to cross energy barriers and explore larger areas of the potential energy surface. Using a Monte Carlo search in the manner of Buch and Kazimirski [7] is one option. We wanted, however, to keep the algorithm more

Table 2.2 Effect of topology filters shown as an average percentage improvement in calculation time relative to the trial with no filters

Features used	$(\text{H}_2\text{O})_n$		
	$n = 34$	$n = 43$	$n = 52$
+ HOH angle	11.68	7.64	6.39
+ 2 adjacent DAA	41.82	37.18	31.31
+ list of minima	45.30	40.16	33.99

deterministic. Our algorithm, called extended topology-altering optimization $ET(m)$, uses a population of m topologies that act as different solutions and are refined independently. This way some topologies with an energy higher than the best are saved as part of a population thus allowing the algorithm to search in several directions.

The algorithm is shown in Listing 2.3. A population member is described by Cartesian coordinates, energy, **A** and **D** matrices, and its substructure level—the size of the largest substructure for which the altering of H-bond directions has been performed. At the beginning, the population contains only one member, an initial structure with its substructure level set to 1. A single substructure level is considered during a search iteration. An H-bond alternation search is performed with the ST algorithm of Sect. 2.3.6 for all members whose substructure level is $M - 1$ in which M is the current level. In the case a new topology is accepted, it replaces the highest energy one in the population if its energy is lower and the population does not already have such a topology. The new member's substructure level is set to 1, and M is reset to 2. Thus, only the part of the population with smaller substructure levels is considered until all members have the same level again. The search stops when substructures of size 10 have been examined for all members. The lowest-energy member of the population is the result of the optimization.

Listing 2.3 Extended topology (ET) optimization.

```

-----
Set current population size to 1
Use initial structure as the first member
Set the substructure level of the first member to 1
Set current substructure size M = 2
WHILE (M <= 10)
  Copy current population into parent population
  FOR all current population members with substructure level < M
    Increase substructure level by 1
  END FOR
  FOR all parent population members
    IF (the substructure level of the member < M) THEN
      FOR each substructure of type 'T'
        Find all substructures 'T(M)' that allow
          altering H-bond direction
        FOR each substructure 'T(M)'
          Alter H-bond direction of the substructure
          IF (topology filters are satisfied) then
            Local optimization
            IF (new topology is better than the worst in
              the current population) THEN
              Add structure if population is not full
              or replace the worst
              Set the substructure level of a new member to 1
            END IF
          END IF
        END FOR
      END FOR
    END FOR
  END IF
END WHILE

```

```

END FOR
IF (current population has a new member) then
  Set M = 2
else
  Set M = M + 1
end if
END WHILE

```

We can control the extent of the search by changing the size of the population. Using a larger population allows us to explore a larger part of the potential energy surface (PES) but with the computational effort being significantly larger. An ET optimization with a population of one is equivalent to an ST optimization. Since each population member has its own adjacency matrix, the framework is allowed to change during the optimization. Therefore, if the input structure is not sufficiently stable, it is possible that the population will contain several different frameworks. We have not implemented a list of visited geometries for this algorithm.

2.3.8 Genetic Topology Optimization

The enumeration algorithm is limited to clusters with up to 35 water molecules while the ST and ET algorithms that alter the H-bond directions of the substructures are not designed to perform an exhaustive search. We need a different approach to reliably find the best topology for larger clusters.

If we consider the positions of ones in an $\mathbf{A}$ matrix and the matrix dimensions, we can create a transformation rule that converts a corresponding $\mathbf{D}$ matrix into a string containing only information on the direction of existing bonds. In other words, all zeroes that indicate the absence of a H-bond and do not change when the topology is altered are removed. It suffices to use only half of the $\mathbf{D}$ matrix because D_{ji} can be determined from D_{ij} . Since the bond direction is encoded in a binary manner by 0 and 1, we can create a bit string that unambiguously represents a topology for a given oxygen framework; see Fig. 2.9.

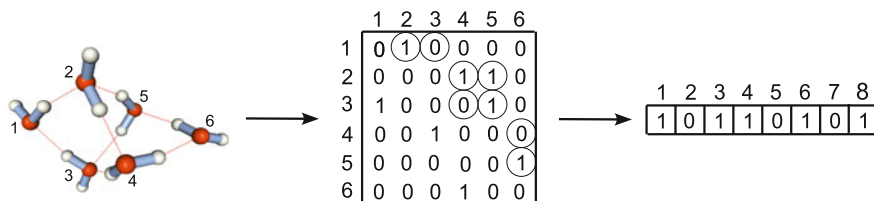
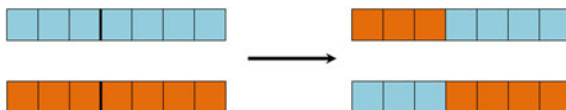


Fig. 2.9 Representation of a hydrogen bond topology using a bit string

Fig. 2.10 Crossover operation



In the jargon of genetic algorithms, a new bit string can be created by a crossover operation on a pair of parent strings as in Fig. 2.10. The definition of a suitable crossover operation makes it possible to search over a large number of topologies and find the best one with genetic optimization methods.

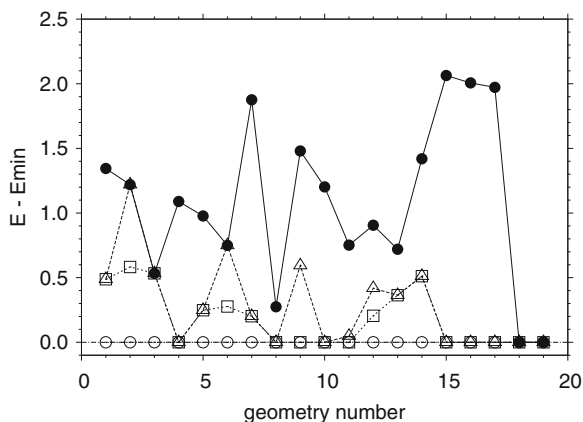
Genetic topology optimization (GT) proceeds in the following manner. A list of visited topologies is maintained in the form of bit strings. A new generation is created using steps standard in genetic algorithms: elitism, crossover, and mutation. Elitism copies 20 % of the previous population into the next generation to keep several of the lowest energy topologies in the pool. During the crossover phase, all possible pairs of the population members exchange sequences of bits. All possible crossover positions and sequence lengths are examined. Each child that has not been visited previously is checked for satisfaction of the topology filters. Each accepted child is added to the list of visited topologies and processed using a multiple step procedure described in the next paragraph. Sometimes there are not enough new topologies generated by the crossover phase to fill the population. In that case, random topologies are added to the population during a mutation phase. The search stops either when the best energy remains unchanged after 20 generations or when neither crossover nor mutation can generate an unvisited topology.

When a new topology is accepted, it is processed in several steps. First, a local optimization is performed in two stages of increasing precision. After the first stage, the topology is discarded if the energy of a child is higher than the highest energy of the parent population by 3 kcal/mol or more. Otherwise, a tighter optimization is performed. The adjacency matrices of the initial and current geometries are compared at the second step to ensure that the oxygen framework has not changed. Allowing the framework to change during the genetic topology search would add unnecessary complications. Next, a short topology-altering (ST) optimization is applied. A crossover operation generates diverse starting points, sampling different areas of the potential energy surface, while an ST optimization allows the algorithm to reach a low energy minimum at the bottom of a funnel. The framework distortion is checked again. If the energy was lowered during the ST step, the new topology is checked against the visited list.

The initial population is generated as follows. Random topologies corresponding to the framework of the input structure are created using the enumeration (NT) method described in Sect. 2.3.5 and processed as described in the previous paragraph until the desired population size is obtained.

The genetic topology algorithm is successful in optimizing topologies of large clusters. However, a significant number of random topologies is often added to the population during optimization. Perhaps the random topologies are more important starting points for the ST optimization than the results of a crossover operation.

Fig. 2.11 Comparison of four topology optimization methods. Energy differences $E - E_{\min}$ are in kcal/mol. Filled circles ST, triangles ST + ET(10), squares ST + ET(20), circles ST + ET(10) + GT



A careful assessment of this issue would help to improve algorithm performance. However, one should remember that test runs over a large variety of structures are required to obtain information on the details of the algorithm performance, since topology optimization is much more difficult for some oxygen frameworks than others.

2.3.9 Comparison of Topology Optimization Methods

The ability of the ST, ET, and GT methods described in Sects. 2.3.6–2.3.8 to locate the best topology was tested. To perform a comparison, 19 low-energy (H_2O)₅₂ clusters with randomized topologies were subject to topology optimization. During topology optimization, we start with a faster algorithm and then refine results employing more precise methods. Therefore, different combinations of methods were examined to see whether the inclusion of more complex algorithms leads to a better topology.

The results are shown in Fig. 2.11. $E_{\min}$ is the lowest energy that has been found for a cluster whereas E is the cluster energy after a sequence of topology optimizations. The initial random topologies have an energy difference, $E - E_{\min}$, ranging from 5 kcal/mol to 25 kcal/mol. The short topology-altering optimization lowers the energy significantly, leading to $E - E_{\min} < 2.5$ kcal/mol. The addition of a more complex method helps to lower the energy further in all but two cases. Comparison of ET(10) and ET(20) shows that increasing the population size of ET(m) is advantageous for 5 geometries out of 19. In nine cases GT is necessary to find the best topology.

2.4 Global Optimization of Water Clusters

An improved parallel evolutionary algorithm [84] and basin hopping combined with vibrational modes [50] are examples of successful approaches to the problem of global optimization of water clusters. However, these methods are not able to find the lowest energy geometry for systems with more than 30 water molecules. A detailed comparison of results for each cluster size can be found in our publication that describes an application of minima hopping to water clusters [41].

2.4.1 Improved Minima Hopping

Goedecker's minima-hopping algorithm was shown to be efficient for atomic clusters [52, 85]. However, our test calculations revealed that the original minima-hopping algorithm does not have a satisfactory performance on water clusters with more than 25 molecules. To improve the algorithm, we decided [38] to add geometry altering operations that are different from molecular dynamics or Monte Carlo steps. These were allowed to be specific to a water cluster system in some cases. Some examples of operations that take into account the current energy or geometry of a cluster are direct mutation [84], internal, surface, and rotational operators [51], and cluster surface smoothing [86].

Our improved algorithm uses several types of geometry transformation steps which we call operators. An operator acts as a black box. It takes a cluster geometry, performs some transformations independent of the other operators, and returns a new geometry. For example, the MD simulation step in standard minima hopping [52] is considered to be an operator. The main purpose of the MD operator is to "hop" over higher energy minima, thereby sampling different areas of the PES. The other operators are designed to lower the energy of a cluster starting from a minimum obtained by MD steps. Four types of geometry transformations were added. The topology operations, ST and ET, described in Sect. 2.3 are used to lower a cluster's energy through topology variation. A translational operator (TRN) moves a molecule to a more favorable position in a cluster [38]. Lastly, a distortion operator (DST) changes the position of several molecules in a selected area of a cluster.

A global optimization using the improved minima-hopping method consists of many cycles. A cycle takes search parameters and cluster coordinates as an input. Each cycle combines operator calls in an optimal sequence. During a cycle, the output list and the visited minima list are updated and the final search parameters and coordinates are returned. To account for the different computational expense of the different operators, a cycle is divided into an initial exploration phase and a subsequent energy lowering phase. The exploration phase attempts to move the search to a slightly different area of the PES by crossing energy barriers using the MD and ST operators. We found that the ST operator is cheap enough to be applied to each accepted minimum during the search, significantly improving the algorithm

performance. The other three operators are more expensive and hence they are used sparingly. In the second phase of the cycle, the energy of a structure is lowered using sequences of TRN, DST, and ET(4) operators. The aim of the energy lowering phase is to reach the bottom of a funnel. Each operator is followed by a tight local optimization and an acceptance or rejection of the resulting local minimum.

2.4.2 Accept or Reject?

The decision process closely follows the original minima-hopping [52] algorithm and its modification [85]. An essential item for deciding what to do with a new minimum is the list of visited minima which contains a description of each accepted minimum and a count of the number of times it has been visited. Our improved algorithm searches over unique cluster frameworks rather than unique topologies thus reducing the number of possible minima to be considered. To achieve this, the list contains not only the energy for each minimum but also its nine framework parameters, the number of hydrogen bonds and eight ring counts, as described in Sect. 2.3.3. The improved minima-hopping search is controlled by two adjustable energy criteria, E_{kinetic} and E_{diff} . Five parameters control the rate of change of these two thresholds: $\beta_1 = \beta_2 = \beta$, $\beta_3 = 1/\beta$, $\alpha_1 = \alpha$, $\alpha_2 = 1/\alpha$. Values of the parameters are $\alpha = 1.02$ and $\beta = 1.05$ as in Goedecker's work [52].

Detailed pseudo code for the processing of a newly found minimum is shown in Listing 2.4. The first section checks whether the minimum has not been changed by the operator or has been visited before. The rate of kinetic energy increase is modified by an *enhanced feedback* mechanism [85]; it depends on the number of visits to a minimum according to the formula $\beta'_2 = \beta_2(1 + c \log N)$, where N is the number of visits and $c = 0.1$ is a feedback coefficient. With this modification, repeated visits to a minimum lead to a greater repulsion from that minimum. A special case arises when the topology of a minimum is improved but the framework remains the same. In such a case, the energy in the visited list is updated and the minimum is accepted. The second section of the processing checks the energy of the new minimum. The MD operator uses E_{diff} to determine whether to accept or reject the minimum. For all other operators a new minimum is accepted only if the energy is lowered. Finally, the accepted minimum is written to the output and visited minima lists.

Listing 2.4 Analysis of a new minimum. 'Mcurrent' is the current minimum and 'M' is a new minimum. 'M' is rejected unless said otherwise.

```
-----
---Checking for repeated visits---
if (energy('M') equals energy('Mcurrent')) then
  if (operator is MD) then
    Ekinetic=Ekinetic*beta1
  end if
GOTO CHECKEND
```

```

end if
Look for 'M' in a list of visited minima
if ('M' matches minimum 'Mlist' in the list) then
  if (energy('M') larger than energy('Mlist')) then
    if (operator is MD) then
      Increase number of visits 'N' to 'Mlist' by one
      Ekinetic=Ekinetic*beta2*(1+c*log('N'))
    end if
    GOTO CHECKEND
  else
    energy('Mlist') is substituted by energy('M')
  end if
end if
---Checking energy---
if (operator is MD) then
  Ekinetic=Ekinetic*beta3
  if (energy('M') - energy('Mcurrent') < Ediff) then
    Ediff=Ediff*alpha1
    Accept 'M'
  else
    Ediff=Ediff*alpha2
  end if
else
  if (energy('M') < energy('Mcurrent')) then
    Accept 'M'
  end if
end if
CHECKEND
-----

```

2.4.3 Operator Call Sequences

Next, we consider in greater detail the search during a cycle of our improved minima-hopping method. During the exploratory phase the MD operator is called repeatedly. The ST operator is used to relax the topology whenever MD leads to a new accepted minimum. To use the expensive energy lowering operators efficiently, the exploration phase must stop at an optimal minimum which is not too high in energy and preferably belongs to a different funnel of the PES. To achieve this, a change in geometry is first calculated for each new minimum. A shift of the molecules' centers of mass (CM) between the initial cluster geometry of a cycle and the current geometry is calculated as

$$S = \left[\frac{\sum_{i=1}^{N_{\text{mol}}} [\text{CM}_{\text{init}}(i) - \text{CM}_{\text{new}}(i)]^2}{N_{\text{mol}}} \right]^{1/2}.$$

The geometry is considered to be different when S is higher than an empirically selected threshold value of 0.95 Å. The exploration phase ends when the last three

accepted minima satisfy the geometry condition, and the energy of the second minimum in this sequence is the lowest of the three. The latter condition prevents the search from stopping when the energy of the minima is changing monotonically. The lowest-energy minimum of these minima is the result of the exploratory phase.

The energy lowering operators take a significant amount of time. Often, the result of an operator is the same minimum or a minimum much higher in energy than the best one. The following procedure is used to avoid wasting time on expensive calculations far from the lowest energy. The energy of a minimum obtained in the exploration phase is compared with the best energy found so far. The translation and distortion operators are designed to have two different sets of parameters: one for a more thorough search and the other for a less thorough search. The settings for a more thorough search are used if the difference in energy between the initial minimum for the energy-lowering phase and the best energy $\Delta E = E_{\text{new}} - E_{\text{best}} < E_{\text{high}}$, where $E_{\text{high}} = 5$ kcal/mol. If $E_{\text{high}} \leq \Delta E < E_{\text{low}}$, where E_{low} is an automatically adjusted threshold, then the less thorough settings are used and E_{low} is decreased. If $\Delta E \geq E_{\text{low}}$, then the energy lowering phase is not used at all, E_{low} is increased, and the cycle goes back to the beginning of the exploration phase. In the energy-lowering phase, operators are called in the sequence

$$\text{TRN} \Rightarrow \text{DST} \Rightarrow \text{ETOP} \Rightarrow \text{TRN} \Rightarrow \dots$$

An operator is called again if it is successful in lowering the energy; otherwise, the next one is called. The energy-lowering phase is completed when application of all three operators in a row fails to lower the energy.

2.4.4 *How the Operators Work*

An effective operator should take into account the properties of a system and employ transformations that make use of the way molecules are connected. An efficient operator should use as few local optimizations as possible. This can be achieved by using geometry filters to predict high energy minima without geometry optimization, avoiding acceptance of the same geometry repeatedly by using a list of visited minima, and by doing local optimization in stages of increasing precision so that it can be stopped when it becomes clear that it is not needed.

2.4.4.1 **The MD Operator**

The basic operator of our algorithm is a short MD simulation as in standard minima hopping. How short should the simulation be? How many MD steps should be taken before continuing with a geometry relaxation? Three options were considered: (1) stop after a fixed number of steps, (2) stop at the first energy local minimum which is not too shallow, and (3) stop after a certain number of energy minima. Goedecker's

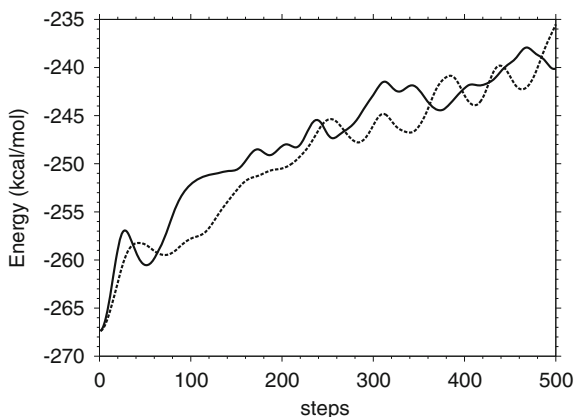
minima hopping uses the third criterion [52]; however, there were no clear rules presented on how to choose the number of minima. In our first implementation we used a fixed number of steps for simplicity but this was not optimal. If we examine the usual energy trend during the MD steps, as shown in Fig. 2.12, we can see that there is a minimum within about 100 steps while other minima could occur much later. Since the primary object of the MD simulation is to distort the geometry and reach a new local minimum, it suffices to stop the MD trajectory at the first energy minimum. Therefore, we settled on the second option. The trajectory continues until a minimum is reached provided that the energy decreases during at least the last five steps before the minimum is reached. The latter restriction prevents termination of the MD simulation at a very shallow local minimum. Since this prescription could lead to a long MD simulation, the trajectory is terminated if such energy lowering is not found within the first 100 steps. Our algorithm also checks for the situation when monomers move too far away from the cluster since such distortions are not useful.

In principle, following the directions of soft vibrational normal modes should result in easier barrier crossing. Such a technique was used by Kabrede in a vibrational mode basin hopping study of water clusters [50]. A method of choosing a low curvature direction for minima hopping was suggested by Goedecker et al. [85]. However, they tested it only on atomic clusters. They also noted that excessive use of low curvature directions could reduce the randomness of the PES exploration. We tried several approaches to make use of low curvature directions for the initial MD velocities; however, nothing gave any improvement over randomly directed velocities.

2.4.4.2 The Translation (TRN) Operator

Translation of one or more molecules to a different part of a cluster provides a good addition to the MD moves in which all molecules are shifted by a relatively

Fig. 2.12 Energies of two typical MD trajectories for a water cluster



small amount. The translation operator is general and can be applied with little modification to a cluster of any type. It was used in various forms in previous studies [38, 51, 84, 86, 87]. The purpose of a translational operator is to find a few molecules that contribute least to the cluster binding energy and put them in a different location that will lower the total energy. We use an algorithm that translates up to two molecules at once. While it can be easily extended to the simultaneous translation of three or more molecules, we came to the conclusion that going beyond two molecules increases the cost-efficiency ratio significantly. Further, only bi- and tri-coordinated water molecules are moved as they are the ones most likely to have higher energies.

Suppose that the cluster is composed of n monomers. First consider that only one monomer is to be moved. We begin by selecting a scaffold composed of $n - 1$ monomers around which the remaining monomer is moved. This scaffold is chosen by requiring that it minimizes disruption of the stability of the cluster. Candidates for the scaffold are created as follows. Each bi- and tri-coordinated molecule is removed in turn from the cluster and the residual structure with one less molecule is locally optimized. The candidate with the lowest energy is selected to be the scaffold geometry around which the molecule is to be moved. Next, a grid of points is set up around the cluster. The grid size depends on the minimal distance that needs to be maintained between the centers of mass of the monomers. The grid has a rectangular prism shape to allow for different lengths of the cluster in the three principal directions. The n th monomer is moved to a grid point only if it is not too close to its original position and if it is within a threshold distance from any three other molecules. The n th molecule's center of mass is placed at an accepted grid point and the position and orientation of the molecule are then adjusted. This simple adjustment might involve several random rotations followed by a single-point energy calculation or a loose local optimization. A more sophisticated adjustment mechanism was developed for water clusters and it will be described below. For each grid point, the energy of the obtained structure is compared with the initial structure's energy. If the energy is not lowered by moving a single molecule, two molecules are moved simultaneously in an analogous fashion. In this case, the $n - 2$ scaffold is created as follows. Each possible pair of bi- and tri-coordinated molecules is removed in turn, the resulting scaffold candidate is locally optimized, and the lowest-energy candidate is selected to be the scaffold geometry. As soon as the energy is improved, the operator stops and returns the structure without checking any unexplored grid points. This does not leave other grid points unexplored because an energy-lowering translation is always followed by another translation as the next operator in the minima-hopping cycle.

2.4.4.3 The Distortion (DST) Operator

The distortion operator attempts to improve a selected area of the cluster by moving several adjacent molecules a relatively short distance while the rest of the cluster remains intact. This operator is expected to be useful for sufficiently large clusters

for which distortion of one section of the surface will not significantly perturb the rest of the cluster. The DST operator works as follows. The molecules are sorted by their contribution to the cluster binding energy. The half of the molecules with the smaller contributions serve as the centers of distortion. The molecules closest to a distortion center are shifted and the resulting geometry is analyzed. The number of molecules being shifted varies in turn from three to five. If energy lowering is found, the operator stops immediately. If the final energy is higher than the starting one, but within 0.5 kcal/mol, the operator is repeated from that higher energy point.

In our first attempt, the shifting was performed by short MD simulations with only the selected molecules allowed to move. This did not work well enough and a more efficient and complex distortion mechanism was created as described below. A stable hydrogen bond network in a water cluster creates high energy barriers. Unlike the case of an atomic cluster, it is not enough to just move a molecule; one also has to change the existing H-bond network in the area where the molecule is placed and in the area from which the molecule was removed. Initially, we left the creation of new hydrogen bonds to chance using random rotations. To improve upon this, we created a H-bond adjustment mechanism. The general idea is to remove H-bonds that create energy barriers, change the positions of the pertinent water molecules, and then recreate hydrogen bonds in a meaningful way for a new distribution. When we remove hydrogen bonds, we remove the hydrogen atoms from the cluster and switch to a representation of each water monomer as a bead or point particle. To reproduce reasonable water molecule positions without the presence of the hydrogen bonds we employ Molinero and Moore's mW coarse-grained model of water [88] based on the Stillinger-Weber silicon potential [89]. The energy of a system in the mW model depends on distances between pairs of beads and angles between triplets of beads. The short-ranged potential sets to zero the forces between beads farther apart than 4.32 Å. The mW model successfully reproduces bulk properties of water. For us, the important feature of the model is that it maintains a tetrahedral distribution of beads and a correct distance between them. The angle term is a three-body term which leads to more elaborate code than is needed for pairwise potentials.

The adjustment process consists of redistributing selected molecules and then recreating hydrogen bonds between them. The process starts by selecting a central point of distortion. For the DST operator, it is the geometrical center of the several molecules selected to be shifted. The cluster is converted into the bead representation and divided into two sections. The N_{free} molecules adjacent to the center of distortion are allowed to move during the local optimization and adjust their positions depending on a new distribution. There are $N_{\text{frozen}} = N_{\text{total}} - N_{\text{free}}$ molecules which have their coordinates frozen and their connectivity does not change. Next, the $N_{\text{move}} < N_{\text{free}}$ molecules closest to the center of distortion are selected to be relocated. A new distribution of beads is created by placing them at random positions very close to the central point of distortion and performing a local optimization using the mW model potential. During the optimization beads will move away from the central point and assume a tetrahedral distribution. N_{try} such attempts are made for each point of distortion. The mW energy is not reliable. Therefore, the N_{pos} lowest energy distributions with different connectivity are saved for analysis. Adjacency

matrices of the mW geometries are compared to avoid saving identical distributions. In addition, bi- and penta-coordinated beads are avoided to reduce the number of possible distributions considered. A process of hydrogen bond restoration is performed for the saved geometries.

The process of creating new H-bonds for free molecules makes use of the backtracking algorithm described in Sect. 2.3.5. At first, possible bonds are determined and a directed adjacency matrix is created for a cluster using three bond categories: existing bonds, no bonds, and possible bonds. Existing and absent bonds belong to the frozen molecules and possible bonds to the free ones. Hydrogen atoms are placed back into the cluster in such a way that water molecules would form hydrogen bonds according to the modified ice-rules discussed in Sect. 2.3.4. A backtracking algorithm is used to analyze all possible combinations of bonds in an efficient way. The following filters are used to avoid high energy combinations: bicoordinated monomers, non-bonded molecules that are too close, three-membered rings, adjacent dangling hydrogen atoms, and unrealistic monomer angles. A single point energy for each accepted structure is calculated and local optimization is performed for the low energy ones. An ST optimization is used to relax the topology for several of the energetically-best geometries.

An example of distortion using the adjustment mechanism is shown in Fig. 2.13 which highlights in blue the molecules selected to be moved (1). The molecules are stripped of hydrogen atoms (2) and then placed close to each other (3). The next image depicts the result of mW optimization, where highlighted molecules and

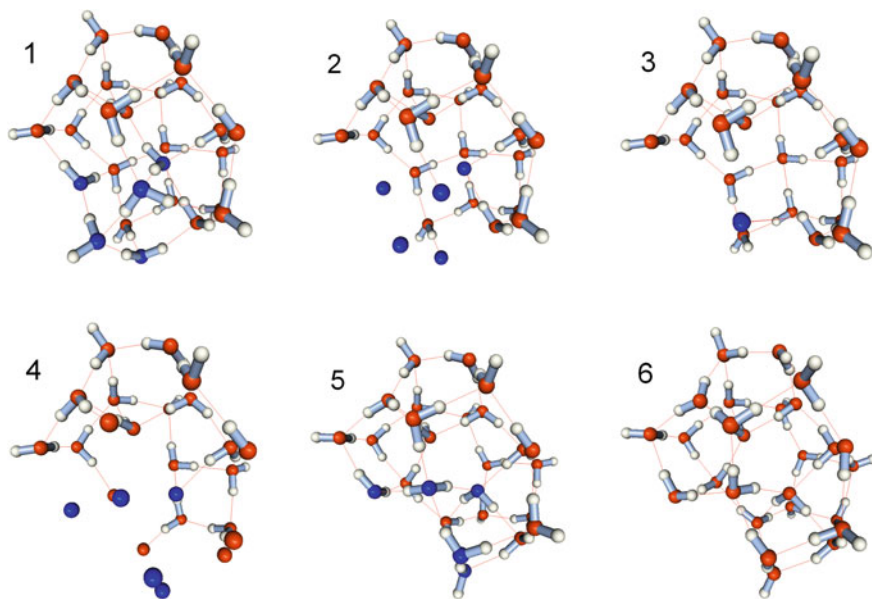


Fig. 2.13 An example of distortion using the adjustment mechanism

Table 2.3 Values of adjustment mechanism parameters for higher and lower accuracy operators

	TRN(high)	TRN(low)	DST(high)	DST(low)
N_{move}	2	2	3–5	4
N_{free}	12	8	12	8
N_{try}	4	2	10	4
N_{pos}	4	2	5	2

several adjacent ones were moved to new positions (4). Then hydrogen bonds are restored (5) and ST optimization is performed (6).

The translation and distortion operators become similar when the adjustment mechanism is used. In the translation, movement of a molecule is followed by an adjustment procedure. In the distortion, the adjustment is applied right away on a selected area of a cluster; however, more adjustment attempts are made in that case.

Different sets of parameters for the adjustment mechanism are used to create versions of operators with higher and lower accuracy. The values used are listed in Table 2.3.

2.4.5 Performance Tests

Test runs were made to check the effects of the various operators on the performance of our improved minima hopping. It is hard to judge the change in performance of the algorithm from a single run due to its random search nature. We need an average result from several global optimization runs on the same cluster size. The latter must be small enough to allow the algorithm to reach the global minimum in a reasonable time. Following these considerations, we performed 20 global optimization runs each on TIP4P (H_2O) $_n$ clusters with $n = 19, 23$, and 27 . A run was stopped if a global minimum was not found within two days. The average time used to locate the global minimum framework in each cluster size was calculated. The operators with a simple adjustment mechanism were used.

Table 2.4 shows the results of the test runs. In four cases, not all of the 20 runs were able to locate the global minimum within two days and the number of successful runs is indicated instead of an average time. We can conclude that all operators improve the performance and are necessary to obtain the global minimum in a reasonable time for large clusters. The inclusion of the mW potential and an advanced adjustment mechanism led to improvement in performance. The time taken to locate the global minimum was almost halved from 54.9×10^3 to 29.1×10^3 s for $n = 27$. The algorithm performance using a simple list of visited minima was also assessed. Treating every topology as a unique minimum leads, as expected, to an increase in the average search time to 35.5×10^3 s.

Table 2.4 Average computer time (seconds) used to locate the global minima for $(\text{H}_2\text{O})_n$ clusters using different sets of the search operators

n	MD	+ST	+ TRN	+ DST	+ ET(4)
19	1741	371	246	150	110
23	(6 of 20)	4199	2200	1336	1191
27	(0 of 20)	(6 of 20)	(11 of 20)	78775	54927

The operators with a simple adjustment mechanism were used. If the global minimum was not found in all 20 runs within two days, the number of successful runs is shown in parentheses instead of the time

Table 2.5 Operator success rate (%)

n	MD	ST	TRN	DST	ET(4)
44	25.63	72.73	36.67	17.94	31.33
45	25.63	72.51	36.93	17.90	31.59
46	25.64	72.31	36.40	17.46	31.87
47	25.64	72.02	37.22	17.52	31.86
48	25.64	71.89	37.43	18.26	31.70
49	25.65	71.85	38.00	17.38	32.15
50	25.65	71.56	38.01	17.48	32.38
51	25.64	71.52	36.47	16.76	32.42
52	25.66	71.18	36.30	16.94	32.63
53	25.66	71.11	35.76	16.07	32.67
54	25.64	70.88	36.07	16.31	32.70
55	25.64	70.88	35.17	15.94	32.89

Table 2.5 lists the percentage of calls that led to energy lowering during the search. Values for the TRN and DST operators combine both low and high precision calls. A 25 % rate for the MD operator follows from the minima-hopping logic with only 1/4 of generated minima being accepted. Although the other operators did not have any constraints, their success rate is consistent over a number of cluster sizes. The ST operator shows very high success rates showing its importance. The ET(4) success rate grows slightly as the cluster size increases while the ST success rate decreases. This is a result of the hydrogen bond topology becoming more complicated.

Despite obvious improvements, the average time to locate the global minimum grows rapidly. The algorithm can reliably locate the global minimum for $(\text{H}_2\text{O})_n$ clusters up to $n = 35$. Reliability means that it always finds the same lowest energy minimum within several days of searching on a single CPU. With a significant amount of calculation, it was possible to find the same lowest energy minima at least twice for sizes up to $n = 47$, except for $n = 39$ and 45 [41]. The best minima for $n > 47$ reported earlier [41] were found only once.

2.5 General Version of Improved Minima Hopping

Our improved minima-hopping algorithm can be used with any type of cluster. In that case, all operators specific to water are disabled. By design, the MD operator can be used with any monomers and does not require changes. However, the improved minima-hopping cycle parameters should be adjusted for each type of cluster to achieve the best performance. The list of visited minima is simplified by using only the energy and the number of visits for each minimum as in standard minima hopping.

A general version of the translational operator was created as well. The same ideas of scaffold and grid points are used. The changes from the water version include moving only one molecule at a time because advanced adjustment cannot be used. Instead the highest and the second highest energy molecules are used. The adjustment mechanism consists of several random rotations. A collision check is performed after each rotation and a single point energy is calculated. Low energy trial structures are accumulated in a list and the best one is selected after a local optimization. As before, higher and lower quality settings are available.

The general version of the distortion operator uses several molecular dynamics steps to create a local distortion, followed by geometry relaxation. The operator uses several molecular dynamics steps in which only certain molecules are allowed to move. Such distortions are applied to several areas of a cluster, each followed by a geometry relaxation.

We have reported [67] application of the general version of improved minima hopping to methanol clusters $(\text{CH}_3\text{OH})_n$ with $n \leq 15$ to generate a set of low-energy geometries for subsequent electronic structure calculations. In unpublished work, we have applied it to pure clusters of ethanol, n-propanol, and iso-propanol. In that work, a conformational rotation operator was added to speed up the optimization.

2.6 Concluding Remarks

Our improved minima-hopping algorithm has proved its worth in applications to water clusters using five different potential functions including two which had vibrating monomers [41]. Clusters sizes up to $n = 55$ were studied successfully although the minima for sizes with $n > 47$ must be considered putative since they were found only once. Our ambitious goal of global optimization of water clusters containing up to $n = 100$ monomers has not yet been reached and requires further development of the algorithm.

Acknowledgments The Natural Sciences and Engineering Research Council of Canada (NSERC) supported this work. SK is grateful to the University of New Brunswick for the award of a Frank J. and Norah Toole Graduate Scholarship.

References

1. Johnston RL (2002) Atomic and molecular clusters. CRC Press, Boca Raton
2. Scheiner S (1997) Hydrogen bonding: a theoretical perspective. Oxford, New York
3. Stone AJ (1996) The theory of intermolecular forces. Oxford, New York
4. Thakkar AJ (2001) In: Moore J, Spencer N (eds) Encyclopedia of chemical physics and physical chemistry, vol I. Fundamentals, chap. A1.5, pp 161–186. Institute of Physics Publishing, Bristol
5. Keutsch FN, Saykally RJ (2001) Proc Nat Acad Sci USA 98:10533
6. Keutsch FN, Cruzan JD, Saykally RJ (2003) Chem Rev 103:2533
7. Buch V, Bauerecker S, Devlin JP, Buck U, Kazimirski JK (2004) Int Rev Phys Chem 23:375
8. Prell JS, Williams ER (2009) J Am Chem Soc 131:4110
9. Hamashima T, Mizuse K, Fujii A (2011) J Phys Chem A 115:620
10. Sliter R, Gish M, Vilesov AF (2011) J Phys Chem A 115:9682
11. Nilsson A, Pettersson LGM (2011) Chem Phys 389:1
12. Ceponkus J, Uvdal P, Nelander B (2012) J Phys Chem A 116:4842
13. Pérez C, Muckle MT, Zaleski DP, Seifert NA, Temelso B, Shields GC, Kisiel Z, Pate BH (2012) Science 336:897
14. Pradzynski CC, Forck RM, Zeuch T, Slavíček P, Buck U (2012) Science 337:1529
15. Saykally RJ, Wales DJ (2012) Science 336:814
16. Richardson JO, Wales DJ, Althorpe SC, McLaughlin RP, Viant MR, Shih O, Saykally RJ (2013) J Phys Chem A 117:6960
17. Bulusu S, Yoo S, Aprà E, Xantheas S, Zeng XC (2006) J Phys Chem A 110:11781
18. Lenz A, Ojamäe L (2010) J Mol Struct (Theochem) 944:163
19. Loboda O, Goncharuk V (2010) Chem Phys Lett 484:144
20. Yoo S, Aprà E, Zeng XC, Xantheas SS (2010) J Phys Chem Lett 1:3122
21. Shanker S, Bandyopadhyay P (2011) J Phys Chem A 115:11866
22. Góra U, Podeszwa R, Cencek W, Szalewicz K (2011) J Chem Phys 135:224102
23. Temelso B, Archer KA, Shields GC (2011) J Phys Chem A 115:12034
24. Ramírez F, Hadad CZ, Guerra D, David J, Restrepo A (2011) Chem Phys Lett 507:229
25. Liu X, Lu WC, Wang CZ, Ho KM (2011) Chem Phys Lett 508:270
26. Bates DM, Tschumper GS (2011) J Chem Theory Comput 7:2753
27. Pruitt SR, Addicoat MA, Collins MA, Gordon MS (2012) Phys Chem Chem Phys 14:7752
28. Wang Y, Babin V, Bowman JM, Paesani F (2012) J Am Chem Soc 134:11116
29. Gillan MJ, Manby FR, Towler MD, Alfè D (2012) J Chem Phys 136:244105
30. Miliordos E, Apra E, Xantheas SS (2013) J Chem Phys 139:114302
31. Iwata S, Bandyopadhyay P, Xantheas SS (2013) J Phys Chem A 117:6641
32. Weinhold F (1998) J Chem Phys 109:373
33. Chaplin MF (2000) Biophys Chem 83:211
34. Ludwig RW (2007) ChemPhysChem 8:938
35. Bukowski R, Szalewicz K, Groenenboom GC, van der Avoird A (2007) Science 315:1249
36. Lenz A, Ojamäe L (2009) J Chem Phys 131:134302
37. Wales DJ (2003) Energy landscapes. Cambridge University Press, New York
38. Kazachenko S, Thakkar AJ (2009) AIP Conf Proc 1108:90
39. Kazachenko S, Thakkar AJ (2009) Chem Phys Lett 476:120
40. Kazachenko S, Thakkar AJ (2010) Mol Phys 108:2187
41. Kazachenko S, Thakkar AJ (2013) J Chem Phys 138:194302
42. Kazachenko S (2013) Global energy optimization of hydrogen-bonded clusters. PhD thesis, University of New Brunswick, Fredericton, Canada
43. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) J Chem Phys 79:926
44. Bernal JD, Fowler RH (1933) J Chem Phys 1:515
45. Tsai CJ, Jordan KD (1993) J Phys Chem 97:5208
46. Wales DJ, Hodges MP (1998) Chem Phys Lett 286:65

47. Hartke B (2000) *Z Phys Chem* 214:1251
48. Kabrede H, Hentschke R (2003) *J Phys Chem B* 107:3914
49. Kazimirski JK, Buch V (2003) *J Phys Chem A* 107:9762
50. Kabrede H (2006) *Chem Phys Lett* 430:336
51. Takeuchi H (2008) *J Chem Inf Model* 48:2226
52. Goedecker S (2004) *J Chem Phys* 120:9911
53. Kirkpatrick S, Gelatt CD Jr, Vecchi MP (1983) *Science* 220:671
54. Černý V (1985) *J Optim Theory Appl* 45:41
55. Salomon P, Sibani P, Frost R (2002) *Facts, conjectures and improvements for simulated annealing*. SIAM, Philadelphia
56. Li Z, Scheraga HA (1987) *Proc Natl Acad Sci USA* 84:6611
57. Wales DJ, Doye JPK (1997) *J Phys Chem A* 101:5111
58. Doye JPK, Wales DJ (1998) *Phys Rev Lett* 80:1357
59. Doye JPK, Wales DJ, Miller MA (1998) *J Chem Phys* 109:8143
60. Fraser AS (1959) *Biometrics* 15:158
61. Booker LB, Goldberg DE, Holland JH (1989) *Artif Intell* 40:235
62. Goldberg D (1989) *Genetic algorithms in search, optimization, and machine learning*. Addison-Wesley, New York
63. Wales DJ, Scheraga HA (1999) *Science* 285:1368
64. Nocedal J (1980) *Math Comp* 35:773
65. Liu DC, Nocedal J (1989) *Math Prog* 45:503
66. Davidon WC (1975) *Math Prog* 9:1
67. Kazachenko S, Bulusu S, Thakkar AJ (2013) *J Chem Phys* 138:224303
68. Franzblau DS (1991) *Phys Rev B* 44:4925
69. McDonald S, Ojamäe L, Singer SJ (1998) *J Phys Chem A* 102:2824
70. Kuo JL, Ciobanu CV, Ojamäe L, Shavitt I, Singer SJ (2003) *J Chem Phys* 118:3583
71. Lenz A, Ojamäe L (2005) *Phys Chem Chem Phys* 7:1905
72. Anick DJ (2003) *J Chem Phys* 119:12442
73. Belair SD, Francisco JS (2003) *Phys Rev A* 67:063206
74. Shi Q, Kais S, Francisco JS (2005) *J Phys Chem A* 109:12036
75. Khan A (2008) *J Mol Struct (Theochem)* 850:144
76. Tissandier MD, Singer SJ, Coe JV (2000) *J Phys Chem A* 104:752
77. Vukičević D, Grubeša T, Graovac A (2005) *Chem Phys Lett* 416:212
78. Anick DJ (2002) *J Mol Struct (Theochem)* 587:87
79. Anick DJ (2002) *J Mol Struct (Theochem)* 587:97
80. Suite BP, Belair SD, Francisco JS (2005) *Phys Rev A* 71:043204
81. Anick DJ (2010) *J Chem Phys* 132:164311
82. Miyake T, Aida M (2002) *Chem Phys Lett* 363:106
83. Vukičević D, Graovac A (2008) *Croat Chem Acta* 81:347
84. Bandow B, Hartke B (2006) *J Phys Chem A* 110:5809
85. Schönborn SE, Goedecker S, Roy S, Oganov AR (2009) *J Chem Phys* 130:144108
86. Cheng L, Feng Y, Yang J, Yang J (2009) *J Chem Phys* 130:214112
87. Cheng L, Cai W, Shao X (2005) *Chem Phys Lett* 404:182
88. Molinero V, Moore EB (2009) *J Phys Chem B* 113:4008
89. Stillinger FH, Weber TA (1985) *Phys Rev B* 31:5262

Practical Aspects of Computational Chemistry IV

Leszczynski, J.; Shukla, M.K. (Eds.)

2016, VIII, 398 p. 142 illus., 87 illus. in color., Hardcover

ISBN: 978-1-4899-7697-0