

# Structural Optimization of Molecular Clusters With Density Functional Theory Combined With Basin Hopping

Hainam Do<sup>1</sup> and Nicholas A. Besley<sup>1, a)</sup>

*School of Chemistry, University of Nottingham, University Park, Nottingham, NG7 2RD, UK.*

(Dated: 14 September 2015)

Identifying the energy minima of molecular clusters is a challenging problem. Traditionally, search algorithms such as simulated annealing, genetic algorithms or basin hopping have been used in conjunction with empirical force fields. We have implemented a basin hopping search algorithm combined with density functional theory to enable the optimization of molecular clusters without the need for empirical force fields. This approach can be applied to systems where empirical potentials are not available or may not be sufficiently accurate. We illustrate the effectiveness of the method with studies on water, methanol and water + methanol clusters as well as protonated water and methanol clusters at the B3LYP+D/6-31+G\* level of theory. A new lowest energy structure for  $\text{H}^+(\text{H}_2\text{O})_7$  is predicted at the B3LYP+D/6-31+G\* level. In all of the protonated mixed water and methanol clusters, We find that  $\text{H}^+$  prefers to combine with methanol rather than water in the lowest-energy structures.

---

<sup>a)</sup>nick.besley@nottingham.ac.uk

## I. INTRODUCTION

Finding the global minimum on a complex potential energy surface (PES) is a fundamental problem in many disciplines, including physics, chemistry and biology. For example, finding the crystalline ground state structure of a solid or identifying the native state structure of a protein require searching for the lowest energy structure. Characterizing these structures via experiment can be difficult and simulation often provides the most efficient route to determine such structures. The potential energy landscapes of many chemical and biological systems are very complicated with many minima, with similar energies, separated by high-energy barriers. For such systems, an exhaustive search of configuration space is not viable and manual search procedures not reliable. This creates the demand for efficient searching algorithms, and in the last few decades, several algorithms have emerged for global minimization. Genetic algorithms (GAs)<sup>1-4</sup> are based on evolution theory, and have become successful tools in optimization problems in many fields. GAs belong to the class of stochastic optimization methods, which also comprise many other techniques. Amongst these, one of the most versatile techniques is Monte Carlo (MC) simulated annealing.<sup>5</sup> This method starts by exploring the configurational space at high temperature where the free energy landscape is relatively simple and gradually approaches the free energy global minimum by decreasing the temperature to 0 K, where the global minimum of the PES and the free energy coincide.

Despite its popularity, simulated annealing has several drawbacks. Finding the correct global minimum is only guaranteed for infinitely slow cooling, and this necessitates extensive sampling of conformational space. Furthermore, if the global minimum of free energy changes at low temperatures where dynamical relaxation is slow, the system can get stuck in the structure corresponding to the high temperature free energy global minimum, which corresponds to a local minimum on the PES. Hypersurface deformation methods represent another class of global optimization techniques, in which a transformation to smooth the PES and reduce the number of local minima is performed.<sup>6,7</sup> The global minimum of the deformed PES is mapped back to the original surface under the assumption that it corresponds to the global minimum of the original PES. However, difficulties in locating global minima of small Lennard-Jones clusters using this method were reported.<sup>8,9</sup> Other techniques which

attempt to reduce the effects of barriers on the PES have also been proposed.<sup>10–12</sup> However, in general these techniques are computationally demanding and have not yet been tested for large clusters. In the search for an efficient global optimization algorithm for Lennard-Jones clusters, Wales and Doye proposed the basin hopping (BH) algorithm<sup>13</sup>, which shares the same principle as the MC minimization algorithm of Li and Scheraga<sup>14</sup>. This method has proved to be an effective stochastic global search algorithm. Recently, other variants of the BH method have also been proposed,<sup>15–17</sup> which further strengthen the technique.

In addition to a suitable search algorithm, an accurate force field is critical to characterize the global minimum energy structure successfully. The extensive sampling required by all search algorithms means that calculations predominantly use a combination of empirical force field and mixing rules, and this approach has proved successful in a wide range of applications.<sup>18,19</sup> However, there are some problems associated with the reliance on empirical force fields. For example, a suitable force field may not be available for the system of interest, necessitating the development of an appropriate force field before any calculations to determine the global minimum can commence. Furthermore, the transfer of a force field from the framework in which it was parameterized to the system being studied is an approximation, and strictly force field parameters should be computed for all interactions in the system. This has been achieved based on quantum mechanical calculations.<sup>20–22</sup> However, the development of such force fields is demanding and their availability is limited. Another approach is to refine structures determined using empirical force fields by posteriori geometry optimizations using quantum chemical methods.<sup>23</sup>

In recent years, the advent of density functional theory (DFT) and the growth in available computational resources has reduced dramatically the time required for accurate quantum chemical calculations. These developments have made it possible to interface search algorithms directly to quantum chemical methods, namely DFT, without reference to any empirical potential. This will be substantially more computationally expensive, but it is possible to conceive of many systems where such an approach is an attractive alternative, for example where an empirical potential is not available, or the complex electronic structure and changes in electronic structure cannot be captured by a simple potential. Some authors have already begun to pursue such approaches, predominantly using DFT in con-

junction with GAs with great promise.<sup>23–28</sup> The Gradient Embedded Genetic Algorithm (GEGA) technique<sup>25</sup> was successfully used for identifying the lowest-energy structures of small lithium clusters  $\text{Li}_n^{0/+1/-1}$  ( $n = 5 - 7$ ). The method was extended to search for the global minima on the potential energy surface of mixed clusters formed by a radical H atom solvated by one to four water molecules.<sup>26</sup> Recently, Johnston and co-workers have used a genetic algorithm coupled with DFT calculations to perform global optimisations for all compositions of 8-atom Au-Ag bimetallic clusters.<sup>28</sup>

In this paper, we describe our implementation that combines the basin hopping search algorithm with DFT, and show that it provides an effective approach for global optimization. DFT is well suited to a basin hopping approach, since efficient geometry optimization routines to find local minima are well established,<sup>29</sup> and in some cases the BH method has outperformed GAs.<sup>30</sup> The structure of atomic and molecular clusters has been studied extensively with computational approaches, and we utilize our method to search for the global minima of  $(\text{H}_2\text{O})_n$  (where  $4 \leq n \leq 11$ ),  $(\text{CH}_3\text{OH})_m$  (where  $4 \leq m \leq 7$ ) and  $(\text{H}_2\text{O})_n(\text{CH}_3\text{OH})_m$  (where  $4 \leq n + m \leq 7$ ). In addition, we also explore the protonated  $(\text{H}_2\text{O})_n$  clusters (where  $3 \leq n \leq 9$ ) and  $(\text{H}_2\text{O})_n(\text{CH}_3\text{OH})_m$  clusters (where  $n = 1$  and  $2 \leq m \leq 5$ ). Water is the most important liquid, and there has been many investigations of water cluster structures using different optimization techniques and potential functions.<sup>31</sup> The study of water clusters can aid understanding of the formation of ice, clouds, solution chemistry, and many biochemical processes.<sup>32</sup> Methanol is a simple organic molecule that has both hydrogen bonding and hydrophobic interactions, and small methanol clusters have been studied recently.<sup>33</sup> Methanol is miscible with water at all proportions by forming multiple hydrogen bonds, and the nature of how these multiple hydrogen bond networks are formed has been a subject of interest.<sup>34–39</sup> Both water and methanol clusters have been well characterized using either empirical or semi-empirical force fields or quantum chemistry combined with a manual search procedures, which provide good test systems for the method used here. Subsequently, some stable structures of the binary mixture of water and methanol, and some protonated structures are reported for the first time.

## II. COMPUTATIONAL METHODS

We have implemented the BH method within the Q-Chem software package,<sup>40</sup> to provide a powerful tool for global energy optimization. In principle, BH can be used in conjunction with any quantum chemical method for which energy and gradients are available. However, in practice DFT provides a good balance between computational cost and accuracy. Dispersion forces have a significant effect on the structure of water clusters.<sup>43</sup> For the results presented here, DFT augmented with an empirical dispersion correction (DFT-D) is used. More specifically, the B3LYP exchange-correlation functional<sup>44,45</sup> with the +D dispersion correction of Grimme<sup>46</sup> in combination with the 6-31+G\* basis set is used. The BH algorithm is essentially a combination of the Metropolis sampling technique<sup>47</sup> and a gradient-based local search method. This has the effect of sampling the energy basins, where an energy basin is a certain part of the configuration space around a minimum on the PES that contains all the configurations that will relax into this minimum using downhill relaxations, instead of sampling the configuration space. The PES is effectively mapped onto a set of interpenetrating staircases with plateaux corresponding to the energy basins. Figure 1 shows a schematic view of the staircase topography that results from this transformation.

To enhance the efficiency of the method, BH with occasional jumping<sup>15</sup> is used, which incorporates a jumping move in addition to the standard MC moves. Jumping is a MC move without local minimization at infinite temperature and, consequently, is always accepted. When the usual MC moves are rejected a number of times, the system is judged to be trapped at the local minimum. The temperature is raised to  $T = \infty$ , and the MC jumping moves are executed several times to allow the system to escape from the local minimum (Figure 1). If the number of the jumping moves is too large, the technique constitutes a random multi-start strategy. Therefore, it is important to keep a partial memory of the previous state by employing a small number of jumping moves. This provides an efficient way to escape from a local minimum and to explore the next basin of the valley when it is separated by high barriers.

The transformed energy landscape is explored using a canonical MC simulation at a constant temperature of 300 K. Initially, the molecules are placed randomly in a cubic box

with a density equal 1.5 times lower than that of the liquid phase under ambient conditions. The precise details of the container have little effect on the results, and it is only required to prevent the molecules from evaporating off. At each MC step, all molecules are translated and rotated twice on average. The nature of the transformed surface allows for relatively large step sizes, and the maximum displacement in the translation move is set to  $2.5 \text{ \AA}$  and the maximum rotation set to  $\pi/2$ . The atoms are not allowed to get closer than  $1.5 \text{ \AA}$  to avoid overlapping which can result in the self-consistent field (SCF) calculation failing to converge. Thus, any move that brings at least one atom pair within  $1.5 \text{ \AA}$  is discarded, and when a molecule walks out of the box, it will be placed back in the opposite side of the box. If the MC moves are rejected ten times, a jumping procedure consisting of seven random walks in the configurational space will take place.

For each cluster, ten separate runs, each consisting of 1000 MC steps starting from different randomly generated configurations of molecules, are performed. To speed up the calculations, the integration grid is reduced from the default SG1 grid to a 25 point Euler-Maclaurin radial grid and 50 point Lebdev angular grid, and the SCF energy convergence criteria is loosened to  $10^{-5}$  a.u., the root mean squared gradient to fall below  $2 \times 10^{-3}$  a.u. and the convergence criterion for displacement is set to  $2 \times 10^{-3}$  a.u. This provides a good balance between the quality of the evaluated energy and efficiency. Once the global minimum is established the structure is refined with optimization with default optimization parameters.

### III. RESULTS AND DISCUSSION

#### A. Water clusters

The stable structures of  $(\text{H}_2\text{O})_n$  clusters (where  $4 \leq n \leq 11$ ) predicted using B3LYP+D/6-31+G\* are shown in Figure 2, and the corresponding binding energies are given in Table 1. Additional binding energies computed with the larger 6-311+G\*\* basis set evaluated using the same structures are also shown in Table 1. An additional set of calculations without the +D dispersion correction has also been performed in order to explore the effect of dispersion on the predicted structures. With the exception of  $(\text{H}_2\text{O})_6$ , the dispersion

correction does not alter the overall shape of the lowest energy structures, although there are small differences in the precise intermolecular distances.

Cyclic rings are found for  $(\text{H}_2\text{O})_4$  and  $(\text{H}_2\text{O})_5$ , which agree with previous global optimization of small water clusters using the diffusion-equation method combined with an empirical potential.<sup>42</sup> In fact, these cyclic rings are widely accepted as the lowest-energy structures in the literature.<sup>43</sup> There is more ambiguity regarding the structure of  $(\text{H}_2\text{O})_6$ . Without the dispersion correction, a book-like structure is preferred, while with the inclusion of dispersion a trigonal prism configuration is found. The trigonal pyramid structure agrees with previous studies using MP2, and DFT coupled with dispersion-enhanced functionals (M05-2X, M06-2X, M06-L)<sup>43</sup>. In fact, the water hexamer is rather an interesting case, as controversies also arise from the experimental determination of its stable structure. For example, a cage-like structure is favorable for  $(\text{H}_2\text{O})_6$  in Liu’s experiment,<sup>48</sup> while a cyclic configuration is observed in Nauta and Miller’s experiment.<sup>49</sup> For  $(\text{H}_2\text{O})_6$ , there are clearly several low-energy structures and this is reflected in the calculations, where for some of the calculations with different starting geometries, alternative structures, such as the cage-like and ring-like structures, are found. This illustrates the need to use different runs to ensure the global minimum is found in difficult cases.

For  $n = 7$ , an edge-capped trigonal prism (cubelike structure with a corner missing), is the most stable. This structure has ten hydrogen bonds, which is the greatest number of hydrogen bonds that  $(\text{H}_2\text{O})_7$  can possibly form. Other structures which have eight and or nine hydrogen bonds are also found. However, these structures all have higher energy than the edge-capped trigonal prism one. The most stable geometry of  $(\text{H}_2\text{O})_8$  that is found is a cubic shape with  $D_{2d}$  symmetry and 12 hydrogen bonds. A slightly higher energy structure with  $S_4$  symmetry is also observed, which is the structure that was found using the TIP4P force field.<sup>31b</sup> However, the cubic  $D_{2d}$  structure is in agreement with experiment.<sup>50</sup> For  $(\text{H}_2\text{O})_9$ , the lowest energy structure has one water molecule hydrogen bonded to the cubic octamer. In order to form this structure, a hydrogen bond within the cube is broken and the two water molecules at the edges of the cube form an additional two hydrogen bonds with the ninth water molecule. This structure has 13 hydrogen bonds and is consistent with the prediction from the experimental studies of Buck et al.<sup>50b</sup> For  $n = 10$ , a

fused pentameric structure has the lowest energy. This structure has two water molecules hydrogen bonded to the corner water molecule of a cube. This structure agrees with all theoretical studies in the literature,<sup>43</sup> but experiments<sup>50b</sup> suggest a butterfly structure with 14 hydrogen bonds. However, the butterfly structure has been confirmed to be energetically less stable than the fused pentameric structure in a theoretical study by Maheshwary et al.<sup>31d</sup> The most stable structure for  $(\text{H}_2\text{O})_{11}$  contains 17 hydrogen bonds. For this cluster size, there are no available experimental data and theoretical studies have predicted different low energy structures. The structure found by Bulusu et al.<sup>51</sup> is consistent with ours, but alternative structures have been suggested by Maheshwary<sup>31d</sup> and Wales and Hodges using the TIP4P force fields.<sup>31b</sup> However, separate calculations confirm that these structures are energetically higher than the structure shown at the B3LYP+D/6-31+G\* level of theory.

The binding energies for the water clusters are given in Table 1. There is a small variation between the results computed with the two different basis sets considered, with the binding energies for the smaller basis set being larger. However, if the structures were optimized with the larger basis set, the magnitude of the binding energy for this basis set may increase. The O-H bonds that participate in the hydrogen bonds are longer than those which are free. This detail of the structure will not be captured by rigid empirical force fields, revealing an advantage of using methods such as DFT. The binding energies computed here are greater than the previously reported values obtained using Hartree-Fock theory.<sup>31d</sup> The difference between the DFT values reported here and the Hartree-Fock values increases with the size of the cluster and is most likely associated with the dispersion interaction, which is included in our calculations but will be totally absent at the Hartree-Fock level.

## B. Methanol clusters

Hydrogen bonding is also important in methanol clusters, and the most stable structures maximise the number of hydrogen bonds whilst minimizing steric effects. The most stable structures found by the BH search are shown in Figure 3. For the  $(\text{CH}_3\text{OH})_4$  cluster, a cyclic arrangement with  $S_4$  symmetry with the methyl groups alternating between above and below the plane of the ring is the most stable. Other conformers including a cyclic tetramer with  $C_i$  symmetry and a cyclic trimer + 1 molecule are also found. However, these isomers



are higher in energy than the  $S_4$  symmetry one. A ring structure is also predicted for the pentamer. Since there are an odd number of methanol molecules having the methyl groups, alternating above and below the ring is not possible, and this results in the ring being less planar and a methyl group almost in the plane of the ring. Other local minima with higher energies including a cyclic trimer + 2 and a cyclic tetramer + 1 are also observed. For the larger  $(\text{CH}_3\text{OH})_6$  and  $(\text{CH}_3\text{OH})_7$  clusters, the most stable structures are still cyclic but they deviate significantly from planarity. It is expected that for larger clusters ( $m \geq 6$ ), there will be attempts to minimize the volume of the systems. The flat ring structures lead to an excluded volume, and it is more favorable for these clusters to form a folded ring structure and fill these excluded volume with a dielectric.<sup>33d</sup>

It is of interest to explore whether the dispersion energy plays an important role in determining the lowest energy structure of the methanol clusters. While hydrogen bonding is a dominant interaction, in the case of methanol it may be expected that the dispersion interaction be an important factor in the position of the methyl groups with respect to each other. The lowest energy structures predicted using DFT B3LYP/6-31+G\* (without dispersion correction) are shown in Figure 4. There is little qualitative change in the structure for the  $m = 4$  cluster. However, noticeable differences for structures of  $m \geq 5$  are observed. For these structures, the cyclic ring is more planar than the corresponding structures when dispersion is included. For  $(\text{CH}_3\text{OH})_6$  and  $(\text{CH}_3\text{OH})_7$  the tendency for methyl groups to orientate towards the center of the ring is also reduced. The binding energies of the methanol clusters are given in Table 1. The size of the binding energies for the methanol clusters is greater than for the corresponding water cluster, presumably because of the additional dispersion interaction arising from the methyl groups.

### C. Water + Methanol clusters

One of the attractive features of performing the calculations entirely using DFT is that mixed component systems present no additional difficulties or the need for additional potentials. We have studied mixed water and methanol clusters of the form  $(\text{H}_2\text{O})_n + (\text{CH}_3\text{OH})_m$ , and the resulting lowest energy structures are shown in Figures 5-7 and the binding energies given in Table 2. Four different types of  $(\text{H}_2\text{O})_n + (\text{CH}_3\text{OH})_m$  (where  $4 \leq n + m \leq 7$ )

cluster have been studied. For  $n + m = 4$ , there are three combinations namely  $(\text{H}_2\text{O})_1 + (\text{CH}_3\text{OH})_3$ ,  $(\text{H}_2\text{O})_2 + (\text{CH}_3\text{OH})_2$  and  $(\text{H}_2\text{O})_3 + (\text{CH}_3\text{OH})_1$ . For the pure clusters  $(\text{H}_2\text{O})_4$  and  $(\text{CH}_3\text{OH})_4$ , the lowest energy structure is the symmetric cyclic  $S_4$  cluster, and it would be expected that the mixed clusters adopt a similar structure. Figure 5 shows that the predicted global minima, do indeed adopt a cyclic arrangement. When there is more than one methanol molecule present, they arrange themselves to be adjacent to each other. This is likely to be a result of the favorable dispersion interaction from this arrangement. For the  $(\text{H}_2\text{O})_2 + (\text{CH}_3\text{OH})_2$  combination, there are two different low-lying isomers: one where the two methanol molecules are arranged adjacent to one another with the two methyl groups oriented above and below the plane of the ring. In the other isomer, the two methyl groups lie across the ring and are oriented in the same direction with reference to the ring plane. The first of these isomers has the lower energy, and this agrees with a recent study by Mandal et al.<sup>39</sup> The lowest energy structures of the pentamer mixtures ( $n + m = 5$ ) are also shown in Figure 5. For all of the combinations studied, the cyclic ring geometry is to be the most energetically favorable structure with the methanol molecules adjacent to each other. Other structures with higher energies including cyclic trimer + 2 and cyclic tetramer + 1 are also observed.

The lowest energy structures for the hexamer mixtures are shown in Figure 6. All of the structures for the various different combinations resemble those low-lying minimum energy structures of water hexamer cluster. For  $(\text{H}_2\text{O})_1 + (\text{CH}_3\text{OH})_5$  and  $(\text{H}_2\text{O})_3 + (\text{CH}_3\text{OH})_3$ , a book-like structure is the global minimum, while a prism structure is found for  $(\text{H}_2\text{O})_2 + (\text{CH}_3\text{OH})_4$  and  $(\text{H}_2\text{O})_5 + (\text{CH}_3\text{OH})_1$ . For  $n = 4$  and  $m = 2$ , a cage-like structure is the most stable structure. The lowest-energy structures for the  $n + m = 7$  mixtures are shown in Figure 7. The most stable structure for all of the various combinations is a prism shape, similar to the cluster of seven water molecules, where the number of hydrogen bonds is maximized.

#### D. Protonated systems

There have been many previous studies of the global minimum structures of protonated water clusters using empirical force fields.<sup>31,52,53</sup> In this section, we present results for the

global minima of  $\text{H}^+(\text{H}_2\text{O})_n$  (where  $n \leq 9$ ) and  $\text{H}^+(\text{H}_2\text{O})_1(\text{CH}_3\text{OH})_m$  (where  $m \leq 5$ ) clusters from B3LYP+D/6-31+G\* in conjunction with BH. As before, ten independent BH runs with different starting structures, consisting of 1000 MC steps each, were performed. However, the lowest-energy minimum was located in under 1000 steps in all cases. The lowest energy structures for the protonated water clusters are shown in Figure 8 with the binding energies given in Table 3. For  $n \leq 4$ , the chain-like structure, including a branched chain for  $n = 4$  are preferred. For  $n \geq 5$ , the structures are dominated by four- and five-membered rings, and only for  $n = 5$  is a water molecule found outside a closed ring. All of the global minima of the protonated water clusters found in this study are qualitatively the same as those reported by Hodges and Wales using an empirical potential<sup>31c</sup>, apart from the global minimum of  $n = 7$  cluster. For this cluster, we predict a different lowest energy structure, which has a binding energy that is 0.515 kcal/mol lower than that of the structure found by Hodges and Wales at the B3LYP+D/6-31+G\* level.

The protonated mixed water + methanol clusters raise the interesting question of whether the  $\text{H}^+$  is associated with a water or methanol molecule. We have studied the  $\text{H}^+(\text{H}_2\text{O})_1(\text{CH}_3\text{OH})_m$  clusters where  $m \leq 5$ , and the lowest energy structures are shown in Figure 9 with the binding energies given in Table 3. For all of the clusters studied,  $\text{H}^+$  preferentially associated with  $\text{CH}_3\text{OH}$  rather than  $\text{H}_2\text{O}$ . For  $m = 2$ , a chain-like structure is preferred, while for  $m = 3$  and  $m = 4$  ring-like structures are observed. The chain-like structure is also observed for cluster with  $m = 3$  and  $m = 4$ . However, these structures have higher energy than the ring-like structure. For  $m = 5$ , a book-like structure similar to the water hexamer cluster is preferred which maximizes the number of hydrogen bonds that can be formed by this mixture cluster. Other low-lying minima including the chain-like structure and pentamer ring ( $\text{H}_2\text{O}$  stays in the ring) + 1 are also found for this cluster.

#### IV. CONCLUSION

The basin hopping method has been implemented within the Q-Chem software package to allow the lowest energy structures of complex system to be characterized purely using quantum chemical methods without the need for empirical potential functions. While the computational expense of this approach is considerably greater than using empirical energy

functions, it may be useful for systems where empirical potentials are not available or where the complex electronic structure or changes in electronic structure make the system unsuitable to be described by a simple potential energy function. This method has been used to study water, methanol, water + methanol, protonated water, and protonated water + methanol clusters at the B3LYP+D/6-31+G\* level of theory. Overall, the predicted structures of the water clusters using our approach agree with those reported in the literature. Although these systems are largely governed by electrostatic interactions and hydrogen bonding, the inclusion of dispersion does have a significant effect. For the water hexamer cluster, DFT without dispersion correction predicts a book-like structure to be the global minimum, while with the inclusion of dispersion a preference for a trigonal prism configuration is found. Dispersion plays a greater role in the structure of the methanol clusters. For example, without the dispersion correction, a cyclic ring structure rather than folded structures are observed for  $(\text{CH}_3\text{OH})_5$  and  $(\text{CH}_3\text{OH})_6$ . Furthermore, dispersion also has a large effect on the position of the methyl groups in the methanol clusters.

For water and methanol mixed clusters, a cyclic ring geometry is the most energetically favorable structure for all combinations of the tetramer and pentamer clusters. This is consistent with the corresponding pure  $\text{H}_2\text{O}$  and  $\text{CH}_3\text{OH}$  clusters which also have a cyclic ring structure. For the larger mixed clusters, the lowest energy structures resemble the relevant structures for the pure component clusters. Since this approach does not require potentials, it is straightforward to study protonated clusters. For protonated water, a different lowest energy structure is predicted for  $\text{H}^+(\text{H}_2\text{O})_7$ , which has a binding energy of 0.515 kcal/mol lower than that of the structure found by Hodges and Wales<sup>31c</sup> at the B3LYP+D/6-31+G\* level. For the protonated mixed water and methanol clusters, we find that the  $\text{H}^+$  preferentially combines with methanol rather than water in the lowest-energy structures.

The combination of the basin hopping approach with DFT provides a useful tool for identifying minima on complex PESs, which can be readily applied to a wide variety of systems. As the availability of computational resources continues to increase, the range and scope of systems that can be studied using this approach will continue to grow.

## V. ACKNOWLEDGMENTS

We thank the University of Nottingham High Performance Computing facility for providing computer resources and the Engineering and Physical Sciences Research Council (EPSRC) for funding (Grant No. EP/I012303). We are grateful to prof. Jonathan Hirst for useful comments on the manuscript.

## REFERENCES

- <sup>1</sup>D. M. Deaven and K. M. Ho *Phys. Rev. Lett.* **75**, 288 (1995).
- <sup>2</sup>J. A. Niesse and H. R. Mayne *J. Chem. Phys.* **105**, 4700 (1996).
- <sup>3</sup>S. K. Gregurick, M. H. Alexander, and B Hartke. *J. Chem. Phys.* **104**, 2684 (1996).
- <sup>4</sup>J. Mesters and G. E. Scuseria *J. Comput. Chem.* **16**, 729 (1995).
- <sup>5</sup>S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi *Science* **220**, 671 (1983).
- <sup>6</sup>F. H. Stillinger and T. A. Weber *J. Stat. Phys.* **52**, 1429 (1988).
- <sup>7</sup>L. Piela, K. A. Olszewski, and J. Pillardy *J. Mol. Struct. (THEOCHEM)* **308**, 229 (1994).
- <sup>8</sup>J. Kostrowicki, L. Piela, B. J. Cherayil, and H. A. Scheraga *J. Phys. Chem.* **95**, 4113 (1991).
- <sup>9</sup>H. A. Scheraga *Int. J. Quant. Chem.* **42**, 1529 (1992).
- <sup>10</sup>A. B. Finnila, M. A. Gomez, C. Sebenink; C. Stenson, and J. D. Doll *Chem. Phys. Lett.* **219**, 343 (1996).
- <sup>11</sup>C. D. Maranas and C. A. Floudas *J. Chem. Phys.* **97**, 7667 (1992).
- <sup>12</sup>C. D. Maranas and C. A. Floudas *J. Chem. Phys.* **100**, 1247 (1994).
- <sup>13</sup>D. J. Wales and J. P. K. Doye *J. Phys. Chem. A* **101**, 5111 (1997).
- <sup>14</sup>Z. Li and H. A. Scheraga *Proc. Natl. Acad. Sci. U.S.A.* **84**, 6611 (1987).
- <sup>15</sup>M. Iwamatsu and Y. Okabe *Chem. Phys. Lett.* **399**, 396 (2004).
- <sup>16</sup>L. Zhan, J. Z. Y. Chen, and W. Liu *Phys. Rev. E* **73**, 15701 (2006).
- <sup>17</sup>S. Shanker and P. Bandyopadhyay *J. Phys. Chem. A* **115**, 11866 (2011).
- <sup>18</sup>W. L. Jorgensen and J. Tirado-Rives *J. Am. Chem. Soc.* **110**, 1657 (1988).
- <sup>19</sup>W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives *J. Am. Chem. Soc.* **118**, 11225 (1996).

- <sup>20</sup>M. T. Oakley and R. J. Wheatley *J. Chem. Phys.* **130**, 34110 (2009).
- <sup>21</sup>M. T. Oakley, H. Do, and R. J. Wheatley *Fluid Phase Equilib.* **290**, 48 (2009).
- <sup>22</sup>M. T. Oakley; H. Do, J. D. Hirst, and R. J. Wheatley *J. Chem. Phys.* **134**, 114518 (2011).
- <sup>23</sup>D. T. Tran and R. L. Johnston *Proc. R. Soc. A* **467**, 2004 (2010).
- <sup>24</sup>G. Johannesson, T. Bligaar, H. Ruban, H. Skriver, K. Jacobsen, and J. Norskov *Phys. Rev. Lett.* **88**, 255506 (2002).
- <sup>25</sup>A. N. Alexandrova and A. I. Boldyrev *J. Chem. Theory Comput.* **1**, 566 (2005).
- <sup>26</sup>A. N. Alexandrova *J. Phys. Chem. A* **114**, 12591 (2010).
- <sup>27</sup>R. Ferrando, J. Jellinek, and R. L. Johnston *Chemical Reviews* **108**, 846 (2006).
- <sup>28</sup>S. Heiles, A. J. Logsdail, R. Schafer, and R. L. Johnston *Nanoscale* **4**, 1109 (2012).
- <sup>29</sup>B. Paizs, G. Fogarasi, and P. Pulay *J. Chem. Phys.* **109**, 6571 (1998).
- <sup>30</sup>H. G. Kim, S. K. Choi, and H. M. Lee *J. Chem. Phys.* **128**, 144702 (2008).
- <sup>31</sup>(a) C. J. Tsai and K. D. Jordan *J. Chem. Phys.* **95**, 3850 (1991). (b) D. J. Wales and M. P. Hodges *Chem. Phys. Lett.* **286**, 65 (1998). (c) M. P. Hodges and D. J. Wales *Chem. Phys. Lett.* **324**, 279 (2000). (d) S. Maheshwary, N. Pate, N. Sathyamurthy, A. D. Kulkarni, and S. R. Gadre *J. Phys. Chem. A* **105**, 10525 (2001). (e) T. James, D. J. Wales, and J. Hernandez-Rojas *Chem. Phys. Lett.* **415**, 302 (2005). (f) J. Cui, H. Liu, and K. D. Jordan *J. Chem. Phys.* **110**, 18872 (2006). (g) P. T. Kiss and A. Baranyai *J. Chem. Phys.* **131**, 204310 (2009). (h) G. L. Holden and D. L. Freeman *J. Phys. Chem. B* **115**, 4725 (2011). (i) S. Shanker and P. Bandyopadhyay *J. Phys. Chem. A* **115**, 11866 (2011).
- <sup>32</sup>R. Ludwig *Angew. Chem. Int. Ed.* **40**, 1808 (2001).
- <sup>33</sup>(a) U. Buck and J. Siebers *J. Chem. Phys.* **108**, 20 (1997). (b) F. C. Hagemeister, C. J. Gruenloh, and T. S. Zwier *J. Phys. Chem. A* **102**, 82 (1998). (c) S. L. Boyd and R. J. Boyd *J. Chem. Theory Comput.* **3**, 54 (2007). (d) M. M. Pires and V. F. DeTuri *J. Chem. Theory Comput.* **3**, 1073 (2007). (e) J. David, D. Guerra, and A. Restrepo *J. Phys. Chem. A* **113**, 10167 (2009).
- <sup>34</sup>(a) B. S. Jursic *J. Mol. Struct. (THEOCHEM)* **466**, 203 (1999). (b) E. Eudes and F. S. Canuto *Int. J. Quantum Chem.* **102**, 554 (2005). (c) E. Ruckenstein, I. Shulgin, and J. L. Tilson *J. Phys. Chem. A* **109**, 807 (2005). (d) E. Eudes and F. S. Canuto *Int. J. Quantum Chem* **104**, 808 (2005).
- <sup>35</sup>(a) S. J. Suresh, A. I. Prabhu, and A. Arora *J. Chem. Phys.* **126**, 134502 (2007). (b) M. M. Pires and V. F. DeTuri *J. Chem. Theory Comput.* **3**, 1073 (2007). (c) A. Krishtal,

- P. Senet, and C. V. Alsenoy *J. Chem. Theory Comput.* **4**, 426 (2008). (d) J. David, D. Guerra, and A. Restrepo *J. Phys. Chem. A* **113**, 10167 (2009).
- <sup>36</sup>(a) N. Nishi, K. Koga, C. Ohshima, K. Yamamoto, U. Nagashima, and K. Nagami *J. Am. Chem. Soc.* **110**, 5246 (1988). (b) N. Nishi, K. Yamamoto *J. Am. Chem. Soc.* **109**, 7353 (1987). (c) S. Dixit, J. Crain, W. Poon, J. Finney, and A. Soper *Nature* **416**, 829 (2002).
- <sup>37</sup>L. Gonzalez, O. Mo, and M. Yanez *J. Chem. Phys.* **109**, 139 (1998).
- <sup>38</sup>S. S. Xantheas *J. Chem. Phys.* **102**, 4504 (1995).
- <sup>39</sup>A. Mandal, M. Prakash, R. M. Kumar, R. Parthasarathi, and V. Subramanian *J. Phys. Chem. A* **113**, 2250 (2010).
- <sup>40</sup>Y. Shao., L. F. Molnar, Y. Jung, J. Kussmann, C. Ochsenfeld, S. T. Brown, A. T. B. Gilbert, L. V. Slipchenko, S. V. Levchenko, D. P. O'Neill; R. A. DiStasio Jr, R. C. Lochan, T. Wang, G. J. O. Beran; N. A. Besley, J. M. Herbert, C. Y. Lin, T. V. Voorhis, S.-H. Chien, A. Sodt, R. P. Steele, V. A. Rassolov, P. E. Maslen, P. P. Korambath, R. D. Adamson, B. Austin, J. Baker, E. F. C. Byrd, H. Dachsel, R. J. Doerksen, A. Dreuw, B. D. Dunietz, A. D. Dutoi, T. R. Furlani, S. R. Gwaltney; A. Heyden, S. Hirata, C.-P. Hsu, G. Kedziora, R. Z. Khalliulin, P. Klunzinger, A. M. Lee, M. S. Lee, W. Liang, I. Lotan, N. Nair, B. Peters, E. I. Proynov, P. A. Pieniazek, Y. M. Rhee, J. Ritchie, E. Rosta, C. D. Sherrill, A. C. Simmonett, J. E. Subotnik, H. L. Woodcock III, W. Zhang, A. T. Bell, A. K. Chakraborty, D. M. Chipman; F. J. Keil, A. Warshel, W. J. Hehre, H. F. Schaefer III, J. Kong, A. I. Krylov, P. M. W. Gill, and M. Head-Gordon *Phys. Chem. Chem. Phys.* **2006**, 8, 3172-3191.
- <sup>41</sup>P. Qian, W. Song, L. Lu, and Z. Yang *Int. J. Quantum Chem* **110**, 1923 (2010).
- <sup>42</sup>R. J. Wawak, M. M. Wimmer, and H. A. Scheraga *J. Phys. Chem.* **96**, 5138 (1992).
- <sup>43</sup>F. Li, L. Wang, J. Zhao, J. R. Xie, K. E. Riley, and Z. Chen *Theor. Chem. Acc.* **130**, 341 (2011).
- <sup>44</sup>A. D. Becke *J. Chem. Phys.* **98**, 1372 (1993).
- <sup>45</sup>A. D. Becke *J. Chem. Phys.* **98**, 5648 (1993).
- <sup>46</sup>S. Grimme, J. Antony, S. Ehrlich, and H. Krieg *J. Chem. Phys.* **132**, 154104 (2010).
- <sup>47</sup>N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. N. Teller, and E. Teller *J. Chem. Phys.* **21**, 1087 (1953).
- <sup>48</sup>K. Liu, M. G. Brown, C. Carter, R. J. Saykally, J. K. Gregory, and D. C. Clary *Nature* **381**, 501 (1996).

- <sup>49</sup>K. Nauta and R. E. Miller *Science* **287**, 293 (2000).
- <sup>50</sup>(a) Ch. Janzen, D. Spangenberg, W. Roth, and K. Kleinermanns *J. Chem. Phys.* **110**, 9898 (1999). (b) U. Buck, I. Ettischer, M. Melzer, V. Buch, and J. Sadlej *Phys. Rev. Lett.* **80**, 2578 (1998). (c) C. J. Gruenloh, J. R. Carney, C. A. Arrington, T. S. Zwier, S. Y. Fredericks, and K. D. Jordan *Science* **276**, 1678 (1997).
- <sup>51</sup>S. Bulusu, S. Yoo, E. Apra, S. Xantheas, and X. C. Zeng *J. Phys. Chem. A Lett.* **110**, 11781 (2006).
- <sup>52</sup>R. E. Kozack and P. C. Jordan *J. Chem. Phys.* **96**, 3131 (1992).
- <sup>53</sup>T. James and D. J. Wales *J. Chem. Phys.* **122**, 134306 (2005).



Clusters	$N_{HB}$	B3LYP+D/6-31+G*	B3LYP+D/6-311+G**	Others' work
(H <sub>2</sub> O) <sub>4</sub>	4	-34.905	-33.236	-26.06 <sup>41</sup>
(H <sub>2</sub> O) <sub>5</sub>	5	-44.950	-43.303	-38.47 <sup>41</sup>
(H <sub>2</sub> O) <sub>6</sub>	7	-59.142	-56.055	-50.53 <sup>41</sup>
(H <sub>2</sub> O) <sub>7</sub>	10	-72.868	-69.468	-61.68 <sup>41</sup>
(H <sub>2</sub> O) <sub>8</sub>	12	-92.833	-88.365	-77.48 <sup>41</sup>
(H <sub>2</sub> O) <sub>9</sub>	13	-103.515	-98.847	-86.67 <sup>41</sup>
(H <sub>2</sub> O) <sub>10</sub>	15	-117.672	-112.187	-98.57 <sup>41</sup>
(H <sub>2</sub> O) <sub>11</sub>	17	-129.767	-124.470	-107.82 <sup>41</sup>
(CH <sub>3</sub> OH) <sub>4</sub>	4	-36.905	-35.850	-36.35 <sup>33e</sup>
(CH <sub>3</sub> OH) <sub>5</sub>	5	-47.977	-47.247	-
(CH <sub>3</sub> OH) <sub>6</sub>	6	-59.855	-58.795	-
(CH <sub>3</sub> OH) <sub>7</sub>	7	-70.720	-69.901	-

<sup>a</sup> Ref.<sup>41</sup> is at the HF/6-31G(d) level.

<sup>b</sup> Ref.<sup>33e</sup> is at the B3LYP/6-31+G\* level.

TABLE I. Calculated binding energies in kcal/mol and number of hydrogen bonds for the H<sub>2</sub>O and CH<sub>3</sub>OH clusters.

Clusters	$N_{HB}$	B3LYP+D/6-31+G*	B3LYP+D/6-311+G**	Others' work
(H <sub>2</sub> O) <sub>1</sub> (CH <sub>3</sub> OH) <sub>3</sub>	4	-36.381	-35.150	-28.180 <sup>39</sup>
(H <sub>2</sub> O) <sub>2</sub> (CH <sub>3</sub> OH) <sub>2</sub>	4	-35.857	-34.471	-27.850 <sup>39</sup>
(H <sub>2</sub> O) <sub>3</sub> (CH <sub>3</sub> OH) <sub>1</sub>	4	-35.382	-33.839	-27.540 <sup>39</sup>
(H <sub>2</sub> O) <sub>1</sub> (CH <sub>3</sub> OH) <sub>4</sub>	5	-47.739	-46.481	-
(H <sub>2</sub> O) <sub>3</sub> (CH <sub>3</sub> OH) <sub>2</sub>	5	-46.342	-44.837	-
(H <sub>2</sub> O) <sub>4</sub> (CH <sub>3</sub> OH) <sub>1</sub>	5	-45.763	-44.224	-
(H <sub>2</sub> O) <sub>1</sub> (CH <sub>3</sub> OH) <sub>5</sub>	7	-61.070	-59.305	-
(H <sub>2</sub> O) <sub>2</sub> (CH <sub>3</sub> OH) <sub>4</sub>	9	-61.983	-59.148	-
(H <sub>2</sub> O) <sub>3</sub> (CH <sub>3</sub> OH) <sub>3</sub>	7	-60.076	-58.038	-
(H <sub>2</sub> O) <sub>4</sub> (CH <sub>3</sub> OH) <sub>2</sub>	8	-60.167	-57.511	-
(H <sub>2</sub> O) <sub>5</sub> (CH <sub>3</sub> OH) <sub>1</sub>	9	-59.655	-56.908	-
(H <sub>2</sub> O) <sub>2</sub> (CH <sub>3</sub> OH) <sub>5</sub>	10	-74.142	-71.143	-
(H <sub>2</sub> O) <sub>3</sub> (CH <sub>3</sub> OH) <sub>4</sub>	10	-75.365	-72.840	-
(H <sub>2</sub> O) <sub>4</sub> (CH <sub>3</sub> OH) <sub>3</sub>	10	-74.659	-71.808	-
(H <sub>2</sub> O) <sub>5</sub> (CH <sub>3</sub> OH) <sub>2</sub>	10	-74.111	-71.107	-

<sup>a</sup> Ref.<sup>39</sup> is at the MP2/6-31G(d) level.

TABLE II. Calculated binding energies in kcal/mol and number of hydrogen bonds for the H<sub>2</sub>O plus CH<sub>3</sub>OH clusters.

Clusters	$N_{HB}$	B3LYP+D/6-31+G*	B3LYP+D/6-311+G**	Others' work
$H^+(H_2O)_3$	2	-65.004	-66.449	-52.786 <sup>31c</sup>
$H^+(H_2O)_4$	3	-83.849	-84.641	-75.247 <sup>31c</sup>
$H^+(H_2O)_5$	5	-101.384	-101.116	-88.901 <sup>31c</sup>
$H^+(H_2O)_6$	8	-118.309	-116.630	-101.326 <sup>31c</sup>
$H^+(H_2O)_7$	9	-134.711	-134.012	-114.700 <sup>31c</sup>
$H^+(H_2O)_8$	11	-149.854	-147.766	-126.779 <sup>31c</sup>
$H^+(H_2O)_9$	13	-169.126	-165.658	-140.360 <sup>31c</sup>
$H^+(H_2O)_1(CH_3OH)_2$	2	-59.996	-61.433	-
$H^+(H_2O)_1(CH_3OH)_3$	4	-81.446	-81.603	-
$H^+(H_2O)_1(CH_3OH)_4$	5	-101.844	-102.182	-
$H^+(H_2O)_1(CH_3OH)_5$	7	-116.199	-116.182	-

<sup>a</sup> Ref.<sup>31c</sup> uses KJ( $H_3O^+$ ) empirical force fields.

TABLE III. Calculated binding energies in kcal/mol and number of hydrogen bonds for the protonated  $H_2O$  and  $(H_2O)_1(CH_3OH)_m$  clusters.

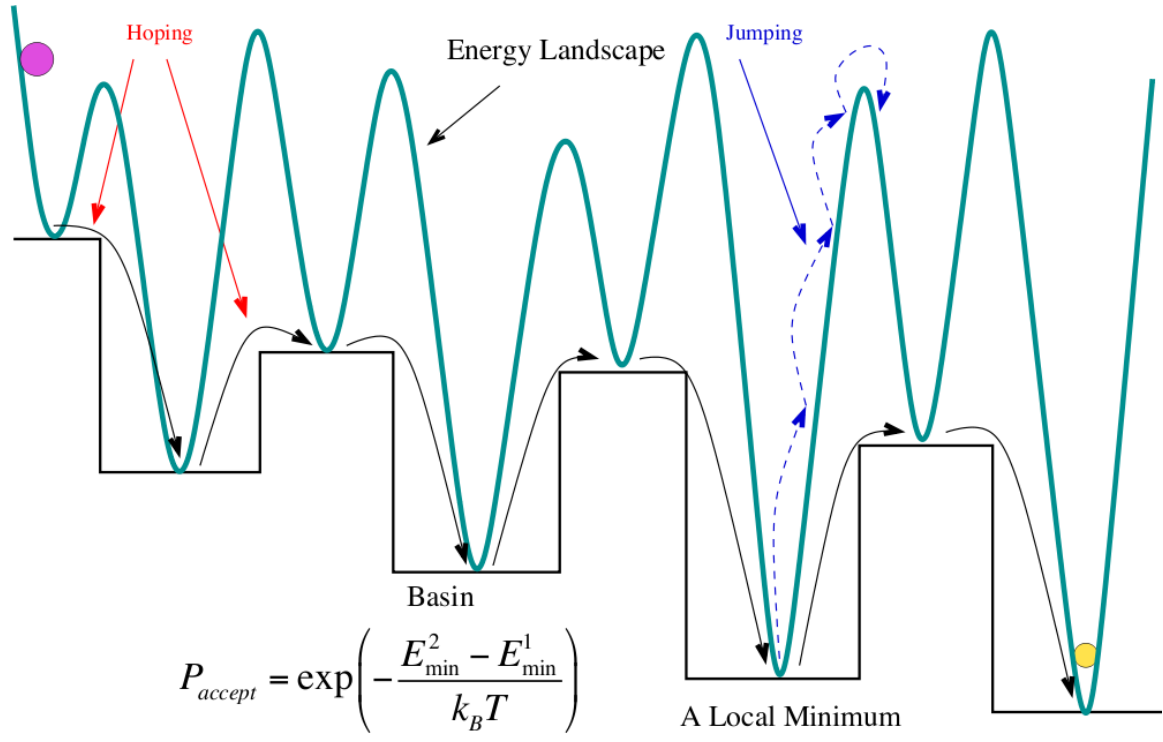


FIG. 1. A schematic diagram illustrating the Basin Hopping algorithm for a one-dimensional energy landscape

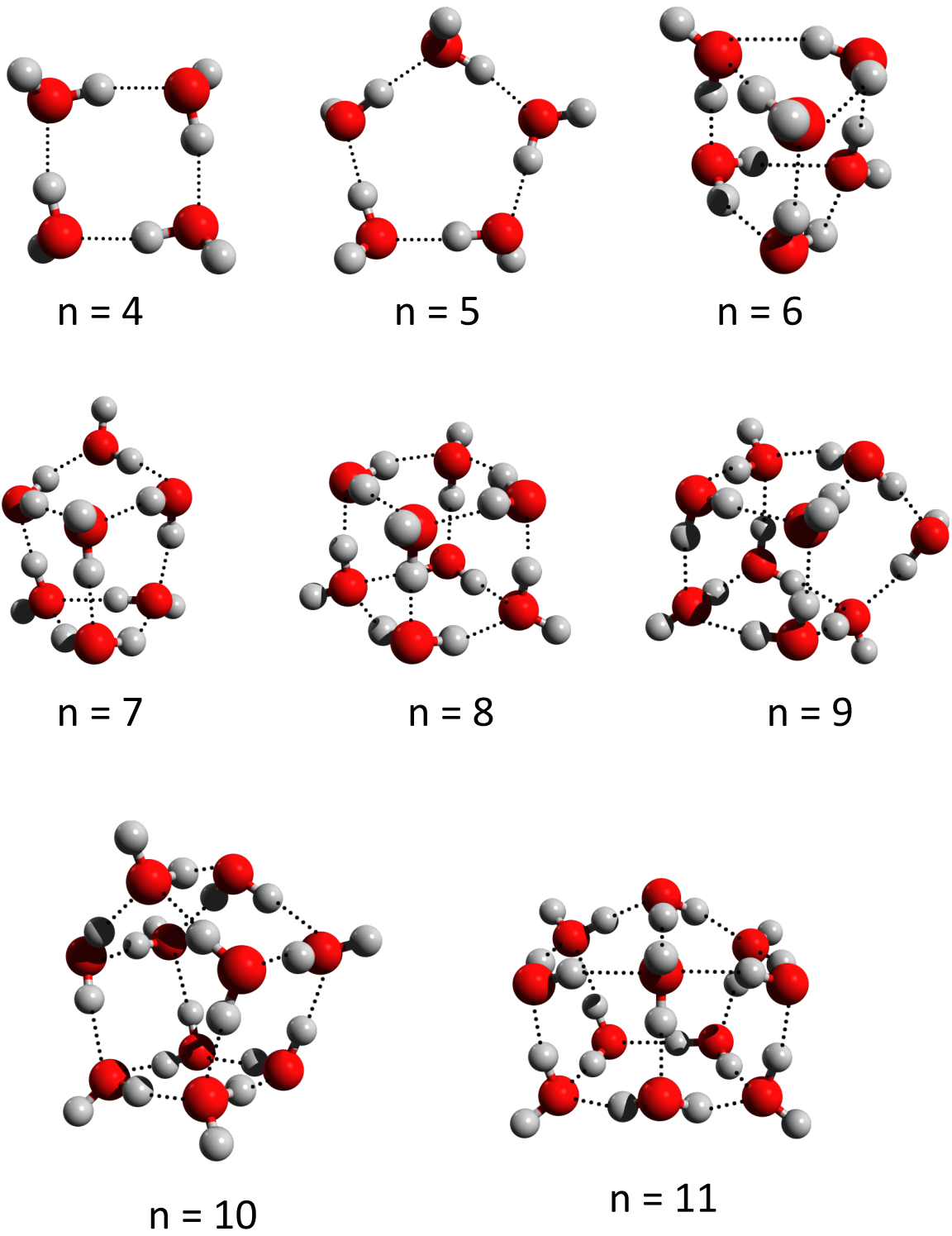
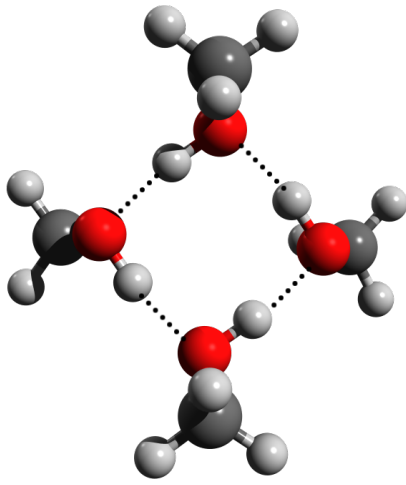
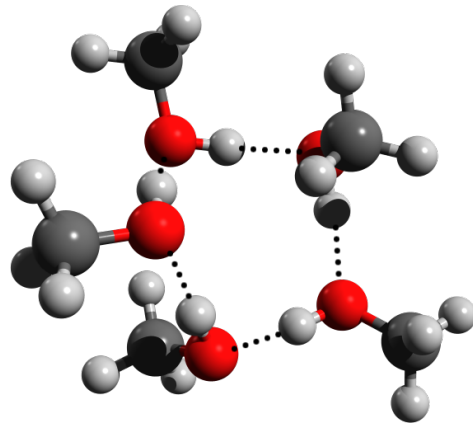


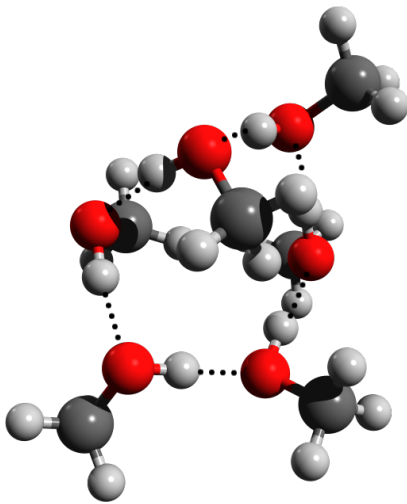
FIG. 2. Predicted lowest-energy structures for the  $(\text{H}_2\text{O})_n$  clusters ( $4 \leq n \leq 11$ ). White and red balls denote hydrogen and oxygen atoms, respectively. The dashed lines represent hydrogen bonds.



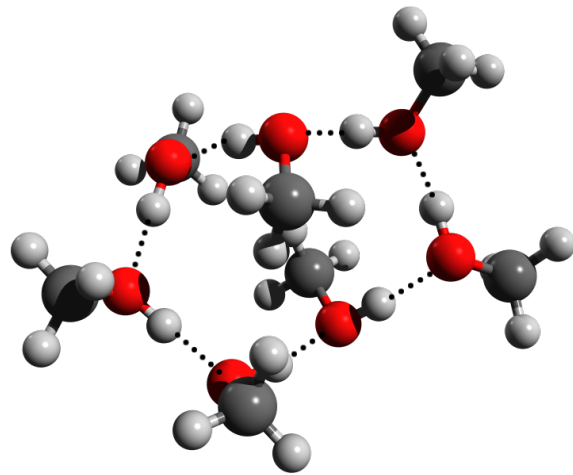
$m = 4$



$m = 5$



$m = 6$



$m = 7$

FIG. 3. Predicted lowest-energy structures for the  $(\text{CH}_3\text{OH})_m$  clusters ( $4 \leq m \leq 7$ ). Grey, white and red balls denote carbon, hydrogen and oxygen atoms, respectively. The dashed lines represent hydrogen bonds.

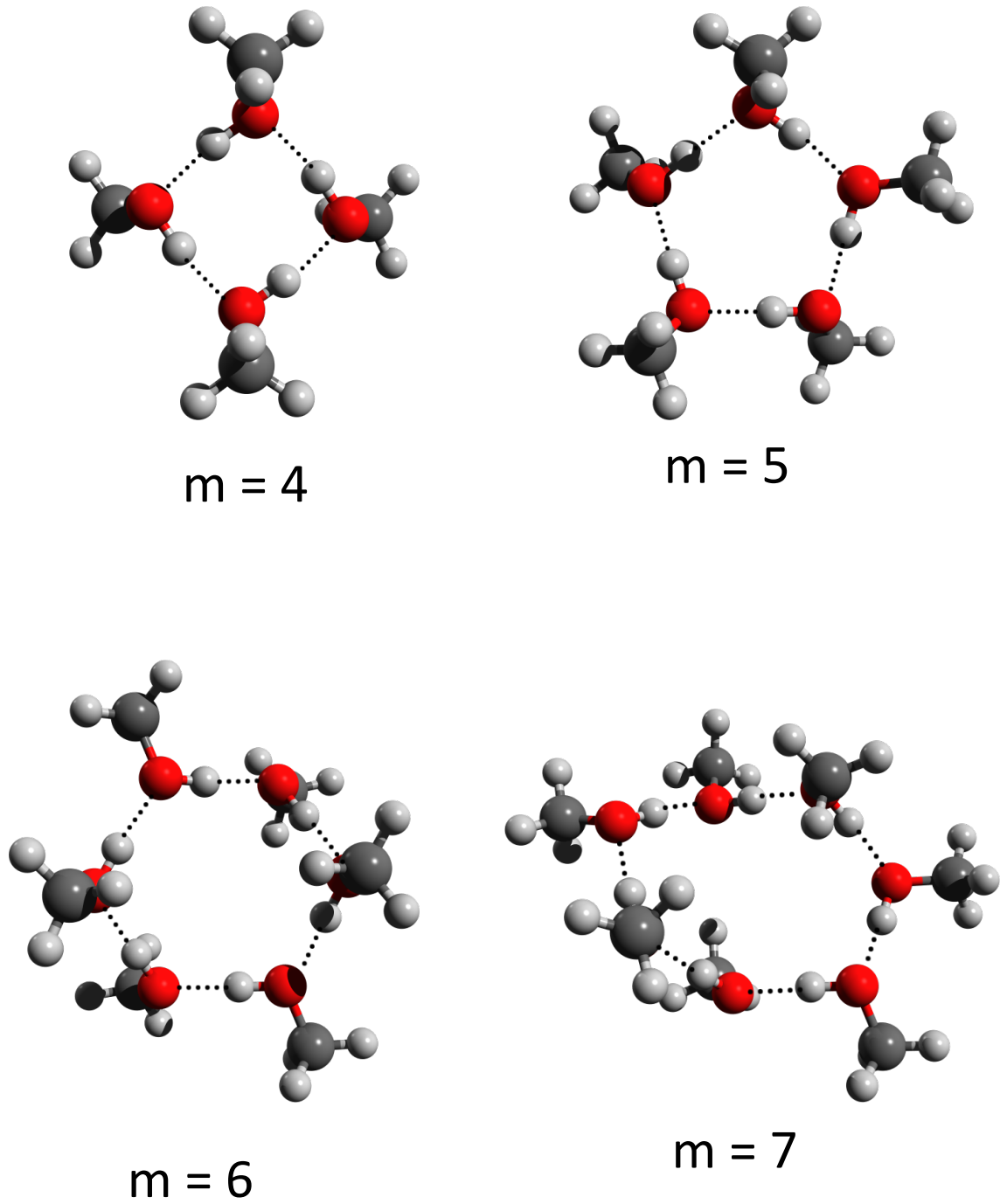


FIG. 4. Predicted lowest-energy structures for the  $(\text{CH}_3\text{OH})_m$  clusters ( $4 \leq m \leq 7$ ) without dispersion correction. Grey, white and red balls denote carbon, hydrogen and oxygen atoms, respectively. The dashed lines represent hydrogen bonds.

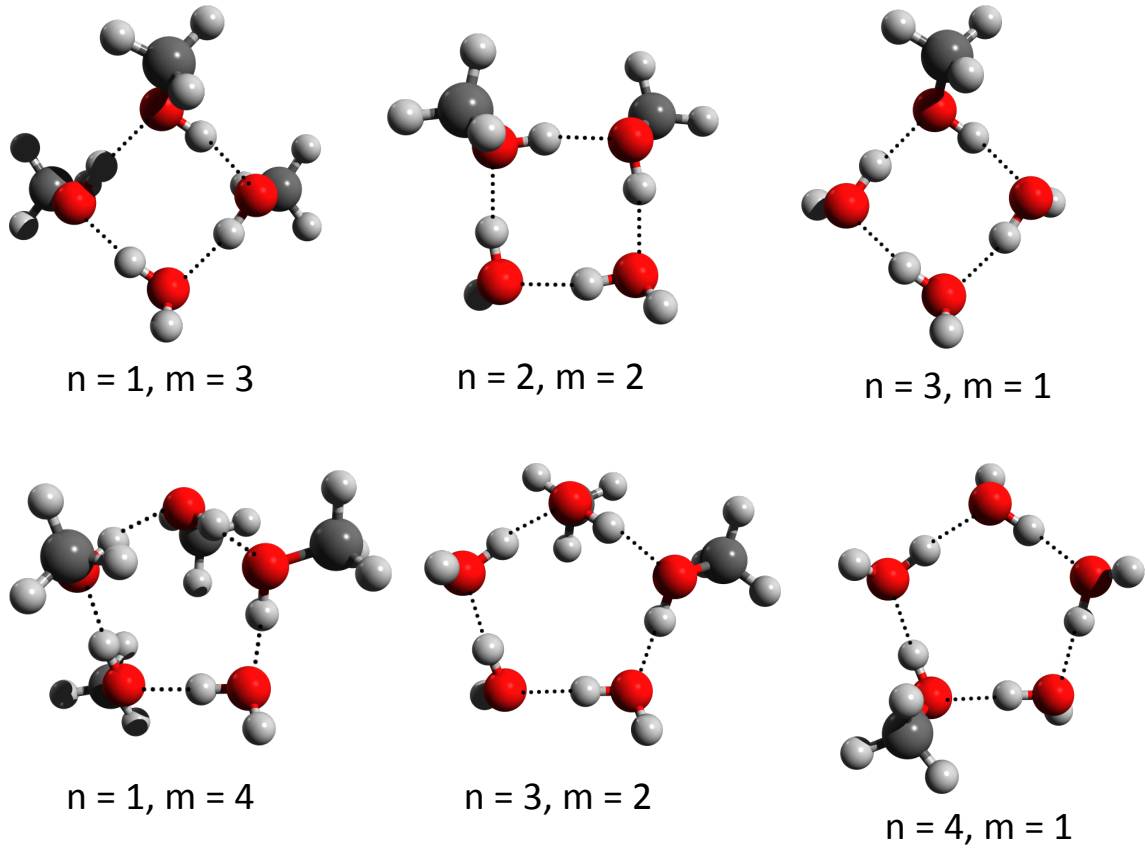


FIG. 5. Predicted lowest-energy structures for the  $(\text{H}_2\text{O})_n(\text{CH}_3\text{OH})_m$  clusters ( $n + m = 4$  and  $n + m = 5$ ). Grey, white and red balls denote carbon, hydrogen and oxygen atoms, respectively. The dashed lines represent hydrogen bonds.



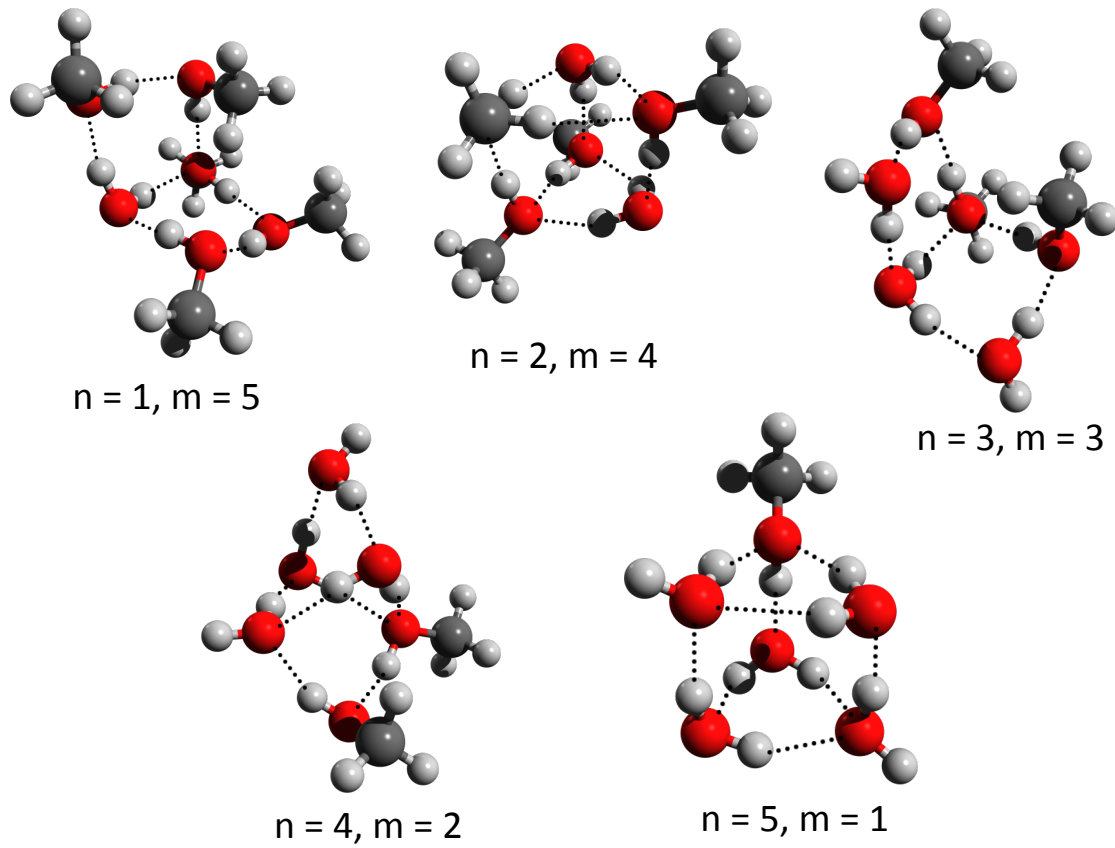
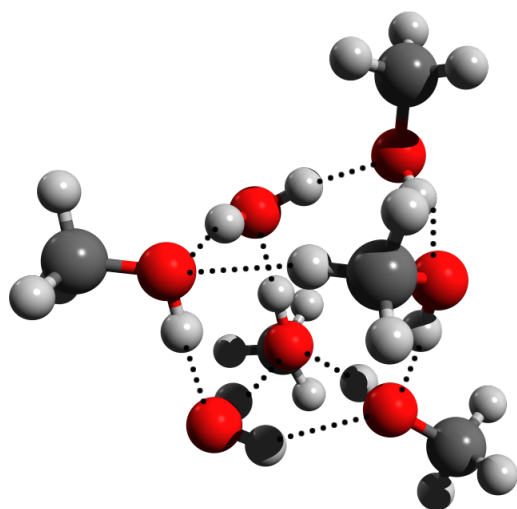
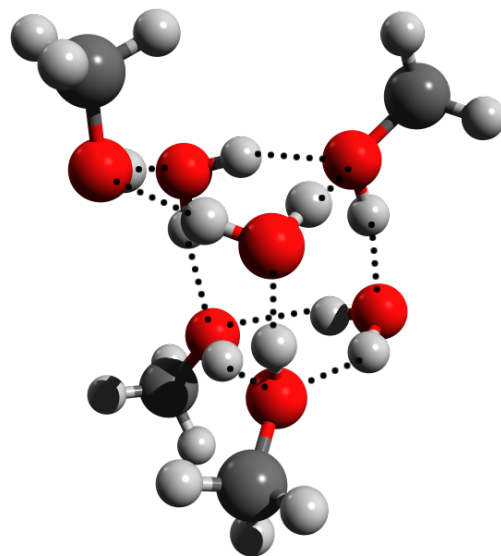


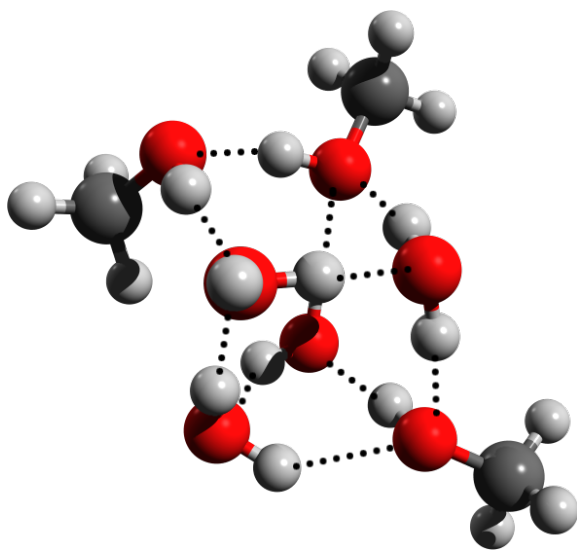
FIG. 6. Predicted lowest-energy structures for the  $(\text{H}_2\text{O})_n(\text{CH}_3\text{OH})_m$  clusters ( $n + m = 6$ ). Grey, white and red balls denote carbon, hydrogen and oxygen atoms, respectively. The dashed lines represent hydrogen bonds.



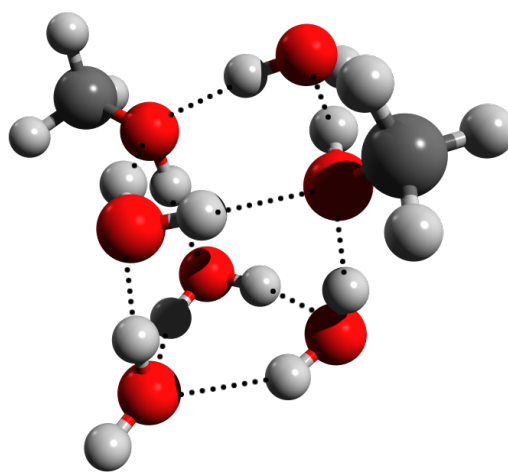
$$n = 2, m = 5$$



$$n = 3, m = 4$$



$$n = 4, m = 3$$



$$n = 5, m = 2$$

FIG. 7. Predicted lowest-energy structures for the  $(\text{H}_2\text{O})_n(\text{CH}_3\text{OH})_m$  clusters ( $n + m = 7$ ). Grey, white and red balls denote carbon, hydrogen and oxygen atoms, respectively. The dashed lines represent hydrogen bonds.

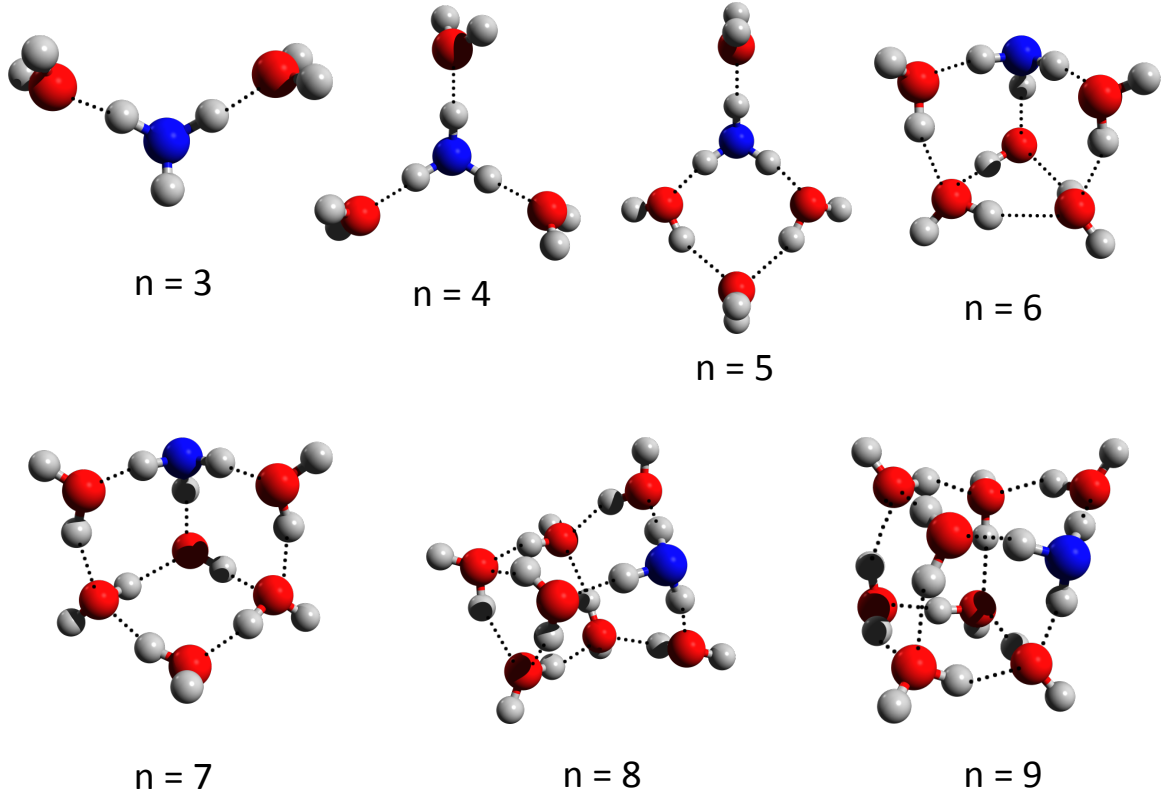
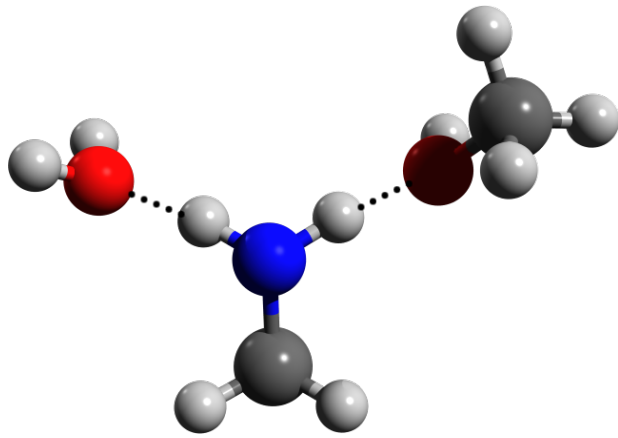
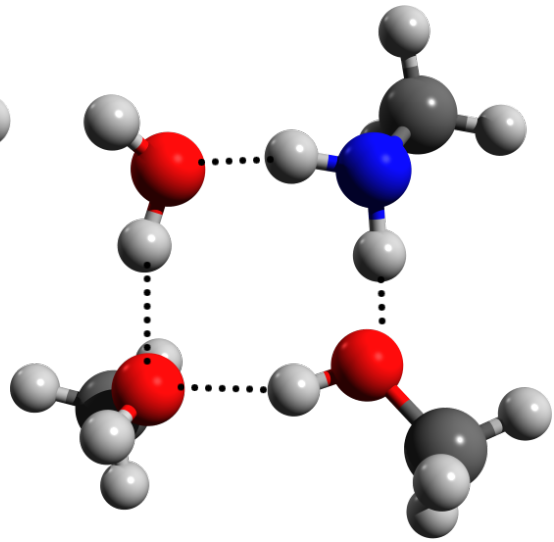


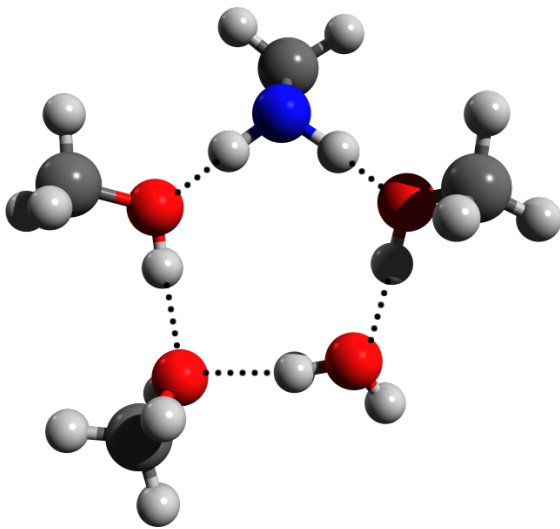
FIG. 8. The lowest-energy structures for protonated  $(\text{H}_2\text{O})_n$  clusters (where  $3 \leq n \leq 9$ ). Black, white and red balls denote carbon, hydrogen and oxygen atoms, respectively. Blue ball denotes protonated oxygen. The dashed lines represent hydrogen bonds.



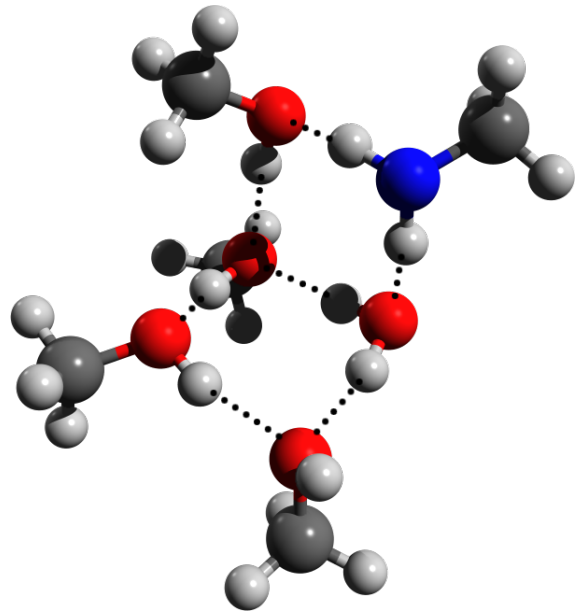
$n = 1, m = 2$



$n = 1, m = 3$



$n = 1, m = 4$



$n = 1, m = 5$

FIG. 9. The lowest-energy structures for  $\text{H}^+(\text{H}_2\text{O})_1(\text{CH}_3\text{OH})_m$  clusters (where  $2 \leq m \leq 5$ ). Black, white and red balls denote carbon, hydrogen and oxygen atoms, respectively. Blue ball denotes protonated oxygen. The dashed lines represent hydrogen bonds.