

A combination of Monte Carlo Temperature Basin Paving and Graph theory: Water cluster low energy structures and completeness of search

RAJAN SHRIVASTAVA, AVIJIT RAKSHIT, SUDHANSHU SHANKER, LOVEKESH VIG and PRADIPTA BANDYOPADHYAY*

School of Computational and Integrative Sciences, Jawaharlal Nehru University, New Delhi 110 067, India
e-mail: lovekesh@jnu.ac.in; praban07@gmail.com

MS received 26 March 2016; revised 29 June 2016; accepted 18 July 2016

Abstract. The knowledge of degree of completeness of energy landscape search by stochastic algorithms is often lacking. A graph theory based method is used to investigate the completeness of search performed by Monte Carlo Temperature Basin Paving (MCTBP) algorithm for $(\text{H}_2\text{O})_n$, ($n=6, 7$, and 20). In the second part of the work, a combination of MCTBP and graph theory was used to devise a new algorithm for finding low energy structures of $(\text{H}_2\text{O})_n$, ($n=21-25$), where input structures for $(\text{H}_2\text{O})_n$ comes from the graphs of $(\text{H}_2\text{O})_{n-1}$. The new algorithm can be a complementary tool to the MCTBP method.

Keywords. Monte Carlo sampling; water cluster; graph theory.

1. Introduction

Exploration of the energy landscape (EL) of flexible molecules and molecular clusters has been generating considerable attention in the scientific community, since the EL of a molecule reveals complete information about its structure and dynamics. The knowledge of EL can also help in understanding the progression of properties from gas phase to bulk phase.¹ However, exploration of EL is a non-trivial problem, since the complexity of the landscape increases sharply with the size of molecules. Even if the search is restricted to finding only the minima on the landscape, the problem remains intractable as the number of minima increases rapidly with the size of the system. Since, deterministic optimization algorithms usually find the nearest minimum, many of the algorithms developed for searching the energy landscape are based on heuristic stochastic and evolutionary optimization methods.²⁻¹⁰ One formal issue with these algorithms is that it is difficult to determine if the search of the EL is complete. In other words, is it possible to quantify the degree of completeness of the search? This is the first question we would like to address in the current work.

The formal definition of completeness of search is the following: In order to claim a search algorithm is “complete”, one must show that if there is an answer, the algorithm will find that in a finite time. This means that it must be shown that if there is an answer, there cannot

exist, with any probability a non-finishing (or finishing with wrong answer) path. So, one needs to show that a solution will be found with probability one (exactly, not approximately!), to show a stochastic algorithm is “complete”. There is also the notion of probabilistic approximate completeness, which requires that the probability of discovering a solution tends towards one as the running time approaches infinity, and this can be used to compare completeness of two algorithms if one can compare how the probabilities of finding a solution increases with time.¹¹

For large molecules, generally speaking, information about the full search space is not available. Hence, the formal definition as given above cannot be used directly. However, there have been attempts to understand the completeness of the EL for molecular systems. For instance, for chain molecules, a distribution of potential minima is given in terms of rotational isomerism approximation.^{12,13} Sullivan *et al.*, devised a method to assess the accessible conformation space explored by an ensemble of protein structures.¹⁴ Nonetheless, this problem has not got enough attention in the literature. In the current work, we have taken the Monte Carlo Temperature Basin Paving (MCTBP) method^{15,16} developed in our group as the sampling algorithm to be tested and the quality of search is assessed by a complete enumeration of water cluster structures represented as graphs¹⁷⁻²⁶ for water clusters of size 6 and 7. The difficulty of the method used in this work for larger water clusters such as $(\text{H}_2\text{O})_{20}$ has also been illustrated. As the stationary points are the most important features of the

*For correspondence

energy landscape, we have restricted our investigation only to those points on the EL. In the second part of this work, a combination of graph theory and MCTBP was used to find low energy structures of $(\text{H}_2\text{O})_n$, $n=21-25$ with a novel incremental algorithm, where the input for $(\text{H}_2\text{O})_n$ comes from the graphs of low energy structures of $(\text{H}_2\text{O})_{n-1}$.²⁷

It was found that for $(\text{H}_2\text{O})_6$ and $(\text{H}_2\text{O})_7$, we can approximately assess the completeness of the search by the MCTBP method. However, this method cannot be used directly for $(\text{H}_2\text{O})_{20}$, since full enumeration of all the graphs was not possible to generate for this cluster. Low energy structures obtained from the new algorithm, combining graph theory and MCTBP method, were found to be close to the structures obtained from the previous studies and gave more diverse structures than obtained by the MCTBP method alone.

2. Computational

2.1 Monte Carlo Temperature Basin Paving (MCTBP) Method

MCTBP method is similar to the Basin Hopping (BH) method, where the original energy surface is converted into a surface with minima only.^{28,29} This is done by minimizing each trial structure generated by a Monte Carlo move. The MCTBP method uses a non-Boltzmann weight factor based on the history of sampling as opposed to the Boltzmann weight factor used in the BH method. At the beginning of the sampling, an energy range to be sampled is defined and the energy range is divided into bins. An effective temperature of each bin can be defined as the following:

$$T(E, t) = T_{\text{initial}} + c' T_{\text{initial}} H(E, t), \quad (1)$$

where $T(E, t)$ is the effective temperature for energy bin E at MC step t , and T_{initial} is the initial temperature. $H(E, t)$ is the histogram of energy at MC step t . c' is a normalization constant. Along with this linear dependence of $T(E, t)$ on $H(E, t)$, an exponential dependence was also used.¹⁶ For details of acceptance/rejection in this scheme, readers are referred to the original literature.^{15,16} The MCTBP method has been used successfully for water clusters of size 20, 25, 30.^{15,16,30-32} In the current work, the number of water molecules moved in each step was between 1 and 4 chosen randomly. Both translational and rotational moves were given in the MC move. The maximum step length for translational and rotational moves were taken as 0.4 Å and 80 degrees based on our previous work.¹⁵ The bin size was taken as 0.5 kcal/mol.

2.2 Graph theoretical generation of structures

Since hydrogen bonds exhibit directionality, a water cluster with hydrogen bonds between water molecules can be modeled by a digraph where the vertices represent water molecules and the edges represent the direction of the hydrogen bond between them (depending on which vertex is the hydrogen-donor and which one is the hydrogen acceptor).¹⁷⁻²⁶ Apart from this, there are several criteria that a digraph must satisfy in order to accurately represent a water cluster. A water molecule can accept up to two hydrogens from other molecules and has the capacity to donate two hydrogens to other molecules. Therefore, any digraph representing a water molecule can have at most two incoming and outgoing edges for any vertex. Although it is possible that a water molecule can accept more than two hydrogens, this is not considered in the present work. This implies that any row or column of the corresponding graph's adjacency matrix cannot have more than two 1's. Also every molecule in a water cluster must be connected by a hydrogen bond either as an acceptor or as a donor. This implies that the digraph representing the cluster must be weakly connected *i.e.*, every node in the graph must have at least one incoming or outgoing edge (indegree/outdegree at least 1). An example of a digraph of $(\text{H}_2\text{O})_3$ is shown in Figure 1.

2.3 Enumerating Relevant Water Clusters

Enumeration of water clusters in terms of graph has been done by several previous workers.¹⁷⁻²⁶ In order to generate the possible graphs for the water clusters, we initially enumerated the N possible rows of the desired adjacency matrix. Since each vertex can have at most two outgoing edges, only rows with a maximum of two 1's were allowed. All such possible rows were generated and all possible combinations of such rows were recursively explored. The candidate graphs are then checked for weak connectivity and for possible isomorphic matches with previously generated cluster graphs in order to avoid duplicates. For $(\text{H}_2\text{O})_{20}$, we used the following two criteria to reduce the number of graphs, (1) all nodes of the digraph must be at least doubly connected, and (2) digraphs should contain cycles of length five or six, although smaller rings can be part of the larger rings.

2.4 Conversion of 2D graph to 3D structures

The conversion from 2D graph to 3D structures is, strictly speaking, an unsolvable problem, since there could be multiple 3D structures arising from the same

graph. However, although it is not possible to find all the solutions, some solutions to this problem can be found. For this purpose, an optimization algorithm is developed to generate 3D structures of water clusters from a digraph. The algorithm works in the following way:

- (1) In the first step, an approximate model 3D structure is generated from a graph. An energy function which treats the connected and non-connected nodes with different weights is introduced. The energy function utilized is shown below:

$$E = -K \sum_{i,j} d_{ij}^2 + nk \sum_{i'j'} \frac{1}{d_{i'j'}^3} \quad (2)$$

where the first and second term represent the connected and non-connected nodes, respectively. d_{ij} is the distance between connected nodes i and j . The dash on the indices i and j of the second term denotes that the sum is over all pairs of non-connected nodes. K and nk are the force constants for connected and non-connected nodes. Obviously, this is one of the many ways to choose an energy function. This energy function is minimized with respect to d_{ij} to get an approximate 3D structure for this energy function starting from an initial approximate 3D model.

- (2) The structure obtained from the first step was refined by reoptimizing with effective fragment potential (EFP) using the GAMESS package after adding hydrogens.^{33,34} It is to be noted that all energies are calculated with EFP potential in this work. For comparison with previous works, the structures reported by previous workers have been reoptimized in EFP and then compared. To find similarity (or dissimilarity) among different structures, dissimilarity index (DSI) introduced by the group of Buch was used.³⁵ For a pair of water

cluster structures, all sorted O-O distances are taken as a vector and Euclidian distance between the two vectors is the DSI.

2.5 Incremental approach for large size cluster

Finding low energy structure for a large size water cluster (>20) using graph theory is a computationally complex problem, as the number of possible digraphs increases exponentially with the size of the cluster. For example, for water cluster of size 20, total number of possible digraphs will be of the order of $({}^{20}C_1 + {}^{20}C_2 + {}^{20}C_0)$.²⁰ Since, in water clusters, only one and two connections per row/column is allowed, ${}^{20}C_2$ and ${}^{20}C_1$ represent the number of ways the connections can be arranged for 2 connections and 1 connection, respectively.

In order to find low energy structure of large size (21–25) water clusters, we have developed a new

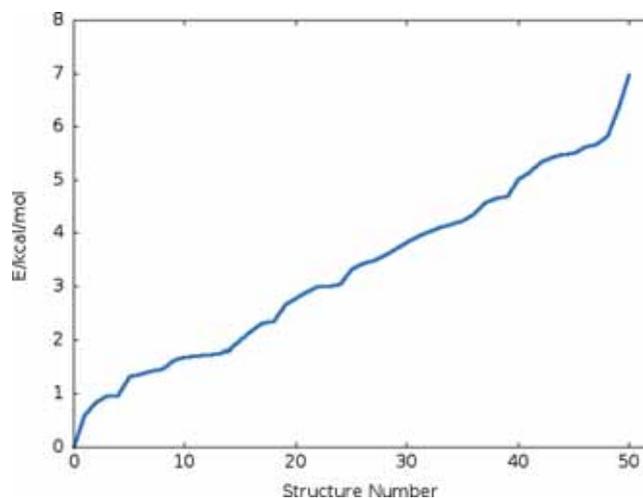


Figure 2. Distribution of structures of $(\text{H}_2\text{O})_6$ as a function of Energy. Zero energy indicates the energy of the most stable prism structure.

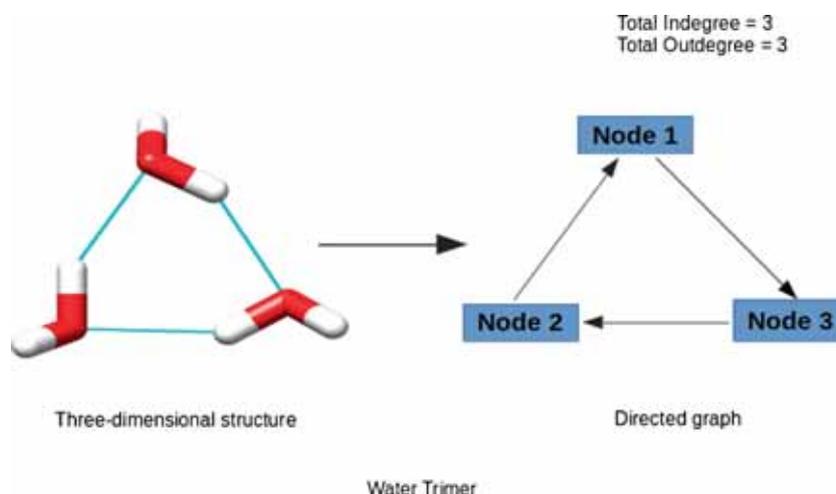


Figure 1. Representation of $(\text{H}_2\text{O})_3$ as a digraph.

incremental digraph approach, where we incrementally find low energy structures of water cluster of size $(n + 1)$, starting with known low energy structures of cluster size n . Our new approach works in the following way:

(1) At first, reported or calculated low energy structures of water cluster of size n were converted to their digraph representation; (2) Then one water molecule (one node) is added to the digraphs; (3) The digraphs were then filtered based on the same criteria used for $(\text{H}_2\text{O})_{20}$ and then used to generate low energy three dimensional structures of water clusters of size $n+1$; (4) From each digraph, a number of three dimensional structures, in general, are generated; (5) We used an algorithm, explained later, to find the most diverse set of structures for the total set of 3D structures obtained in step 4; and (6) Finally, a short (~ 1000 steps) MCTBP calculation is done for each structure obtained in step 5.

3. Results and Discussion

3.1 Completeness of search for water clusters

Figure 2 shows the plot of energy versus number of structures generated by the MCTBP method for $(\text{H}_2\text{O})_6$ for the energy range approximately up to 7 kcal/mol higher from the lowest energy prism structure. 51 unique structures (using DSI threshold 0.3) were found from the MCTBP run. The known low energy structures of $(\text{H}_2\text{O})_6$ such as cage, book, prism, bag, cyclic, boat³⁶ have been found by the MCTBP method and shown in Figure 3. However, even for this small water cluster, there is no easy way to know if all low energy stationary points have been obtained by the MCTBP method; *i.e.*, if the search is complete at least for the energy minima located within 7 kcal/mol from the lowest energy prism

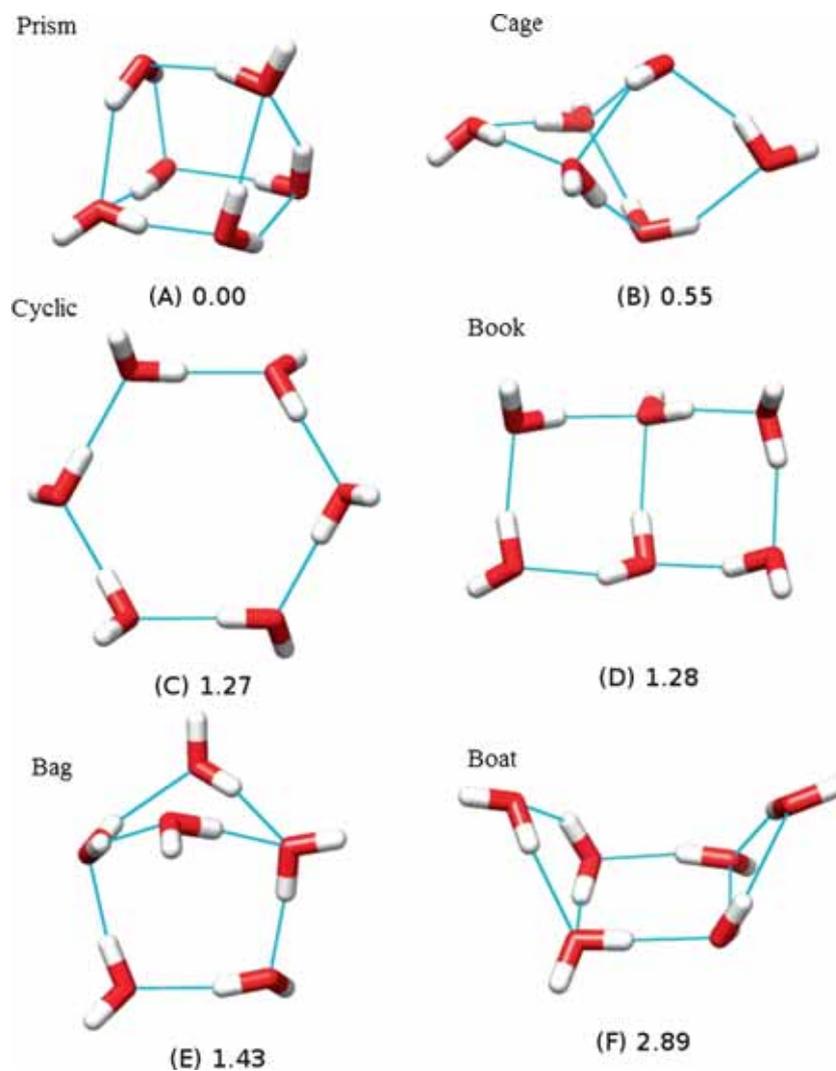


Figure 3. Different known structures of $(\text{H}_2\text{O})_6$ obtained from the MCTBP simulation. Numbers below the structures represent the relative energy in kcal/mol from the most stable structure.

structure. In the next step, we generated all possible 1406 digraphs of $(\text{H}_2\text{O})_6$.¹⁸⁻²⁰

Before we converted all the digraphs to 3D structure, we tested the efficiency of our 2D to 3D conversion algorithm by thoroughly investigating 10 graphs taken randomly. We varied starting coordinates, value of K (see eq. 2, K was varied from 2 to 20 with an increment of 2 and n_k was fixed at -2) and the distance between the non-connected nodes. For each parameter, 10 different calculations were done, so in total 1000 optimizations were done for each graph. We have found that although all together 10000 optimizations were done, only 17 unique 3D structures were obtained. This shows that our simple algorithm of finding 3D structures from 2D graph is robust with respect to the variation of input parameters.

By converting all 2D digraphs to 3D structures, 176 unique structures were generated out of which 125 were

found as new (with DSI 0.3). This means that these structures are not part of 51 unique structures generated by MCTBP method on the basis of DSI value of 0.3. On the other hand, we have found that graph theory and 2D to 3D conversion found all the unique structures found by MCTBP.

Figure 4 shows six low energy structures, which were not obtained by the Monte Carlo TBP method. It can be seen from the Figure 4 that structures with energy as low as 0.21 kcal/mol more than the prism structure was obtained. Of course, running more MC trajectories may find the new structures found by the graph theory. However, $(\text{H}_2\text{O})_6$ case is shown as an illustration.

3.1a $(\text{H}_2\text{O})_7$: Total number of digraphs possible for $(\text{H}_2\text{O})_7$ is 14241 and we generated all of those.¹⁸⁻²⁰ From those graphs, we have generated 141890 3D structures. With a DSI threshold of 0.3, the number of

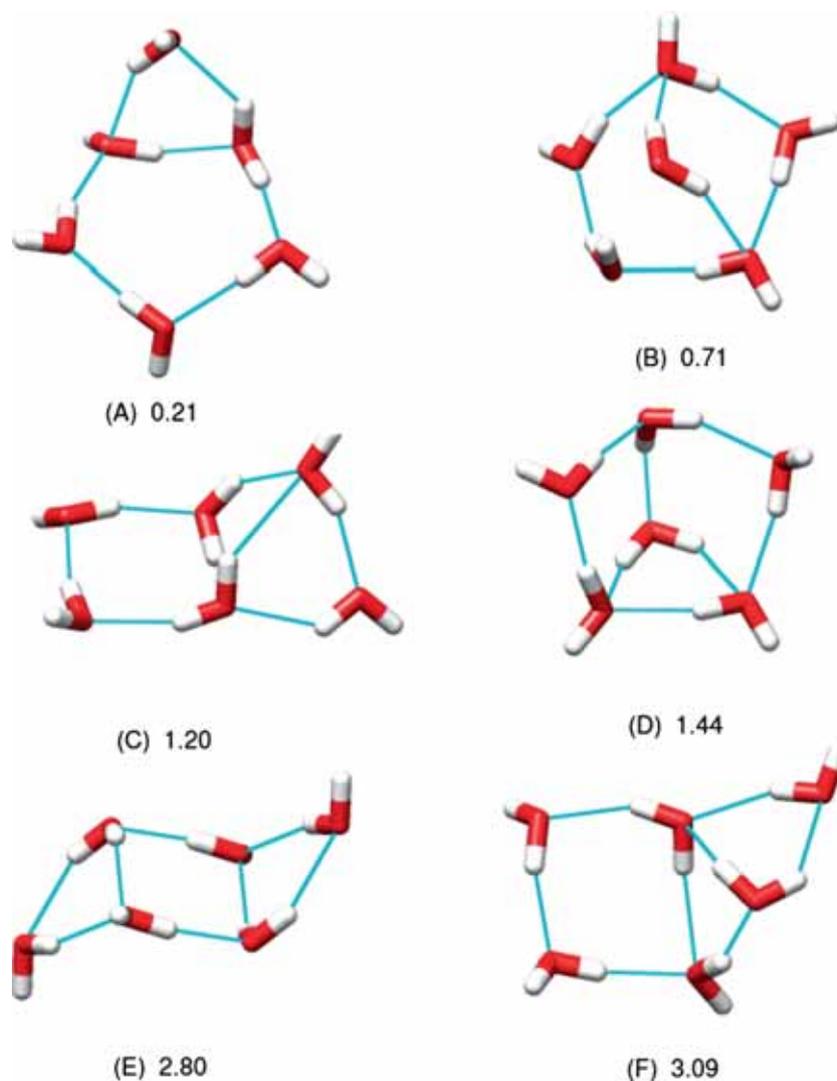


Figure 4. New structures of $(\text{H}_2\text{O})_6$ obtained from graph theory based approach. Numbers in kcal/mol below the structures represent the relative energy from the most stable structure.

structures reduced to 938 unique structures within about 7 kcal/mol from the lowest energy structure as given in references 37 and 38.

To investigate the efficacy of the MCTBP method, we have run 15 independent trajectories with 50,000 steps each. Figure 5 shows the cumulative percentage of unique structures (*i.e.*, the percentage of 938 structures obtained from graph theory with DSI value of 0.3 Å) recovered from MCTBP simulation as a function of number of trajectories. There are 106 structures obtained by the graph theory but not obtained by the MCTBP method; *i.e.*, MCTBP recovered about 88.66% of structures obtained by graph theory. This percentage can be taken as an approximate measure of completeness of search by the 15 MCTBP trajectories. There are also 9 structures obtained by MCTBP but not by graph theory. Hence, our graph theory method cannot be used for completeness of search. However, it can still give approximate idea about the completeness of search. Our MCTBP sampling found all the twelve structures reported by previous workers.^{37,38} Six of those structures are given in Figure 6.

From the investigations on water clusters of sizes 6 and 7, it was found that use of graph theory can, at least, partially answer the questions we posed in the introduction. It can indicate the extent of sampling by the MCTBP. Moreover, it can help in locating new structures not yet found by the MCTBP method.

3.1b $(H_2O)_{20}$: Although, ideally we would like to use our algorithm for large water clusters, at present, it turns out that use of this for $(H_2O)_{20}$ would require the generation of $({}^{20}C_1 + {}^{20}C_2 + {}^{20}C_0)^{20}$ no. of digraphs.

Although the actual number of graphs would be less because of graph isomorphism, it is impractical to generate all the graphs as we did for $(H_2O)_6$ and $(H_2O)_7$. Another approximate way is to use a sub-set of digraphs taken randomly. We tried the second approach by taking about 42000 digraphs and used the filtering as described in the method section.

We found the lowest energy structure is about 5.2 kcal/mol more than the lowest energy structure given in Cambridge Energy Landscape Data Base (CELDB).³⁹ Since the number of potential structures grows exponentially with the number of water molecules, proving exact completeness is challenging for large systems. However, Probabilistic Approximate Completeness (PAC) can be shown for larger systems as the space of graphs is finite for a given number of nodes, therefore the more graphs the algorithm is allowed to process, the higher is the probability of discovering the associated 3D molecular structures.

In the second part of this work, we have focused on finding low energy structures for large water clusters using the graph theory and MCTBP for size $(H_2O)_n$, $n > 20$ using the incremental approach described before.

3.2 Incremental approach to find low energy structures of $(H_2O)_n$, $n = 21 - 25$

Table 1 shows a comparison between MCTBP and incremental methods. In the first step, an MCTBP run was performed for the different water clusters. For $(H_2O)_n$ ($n=21-23$), the runs were stopped after getting 300 unique structures. For the larger two clusters, runs were continued beyond 20000 steps but stopped after

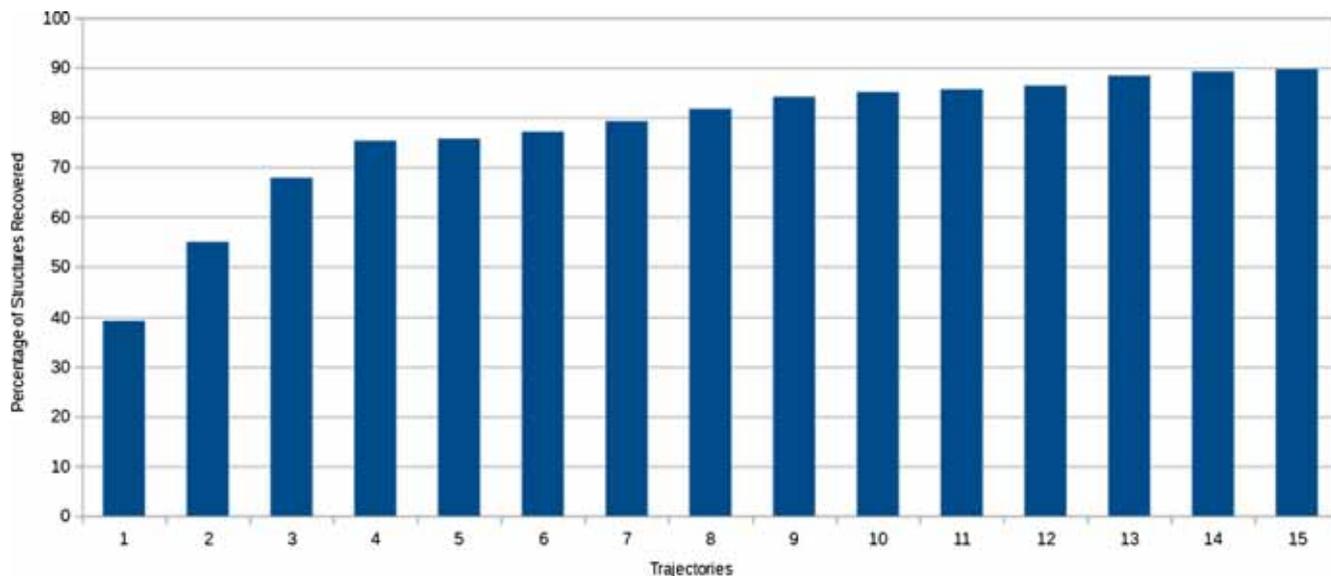


Figure 5. Cumulative Percentage of structures from graph theory obtained from MCTBP method as a function of number of trajectories of $(H_2O)_7$.

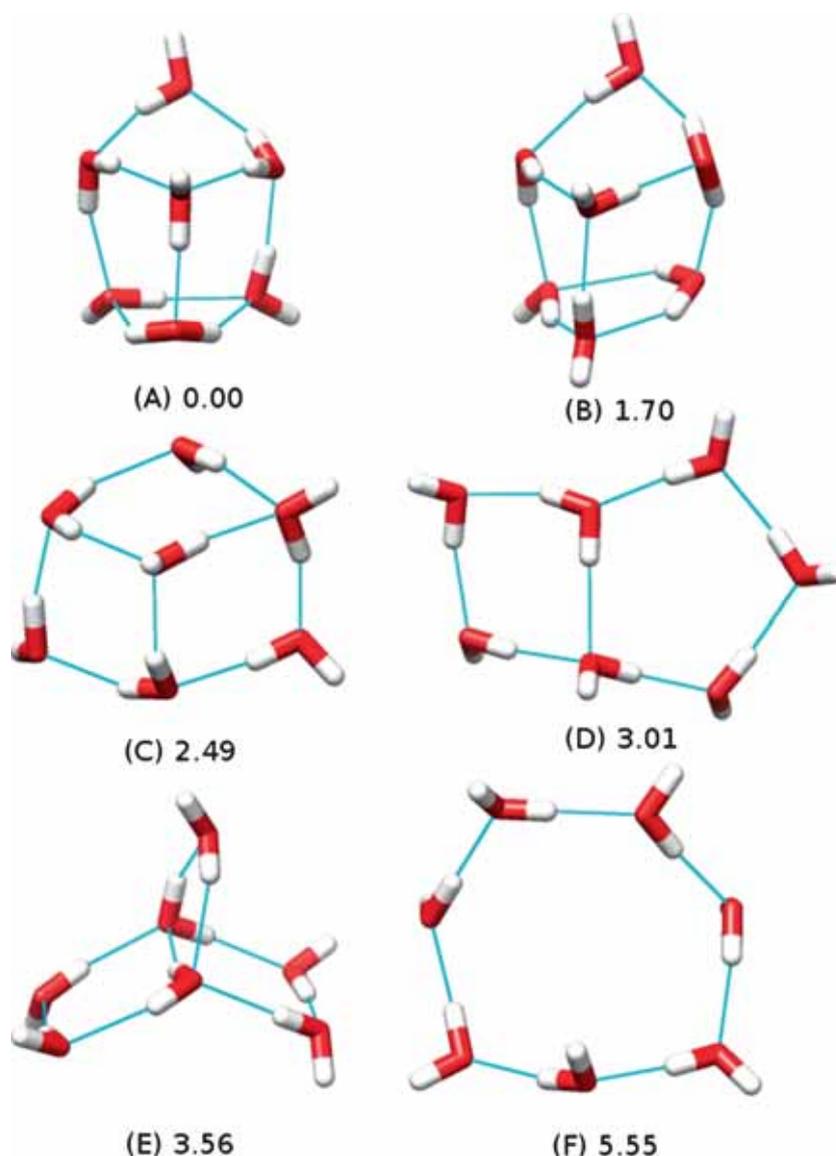


Figure 6. Six best structures of $(\text{H}_2\text{O})_7$ found by our MCTBP simulation. Numbers in kcal/mol below the structures represent the relative energy from the most stable structure.

getting 230 and 160 unique structures for $(\text{H}_2\text{O})_{24}$ and $(\text{H}_2\text{O})_{25}$ respectively.

To compare the pure MCTBP method mentioned above with the incremental approach, we ran the MCTBP part of the incremental method for the most diverse structures obtained from the 2D to 3D conversion outlined in section 2.4. The most diverse 50 structures were obtained in the following way: For each structure S , the neighboring structures within a DSI radius of 3 Å from S were enumerated. The idea is that structures with fewer neighbors are more likely to be unique. The 50 structures with the fewest neighbors from the enumerated list were chosen and shortened MCTBP runs (1000 steps) were performed for each of these 50 structures. This method of choosing diverse structures based on neighbors is an approximation of spectral

clustering⁴⁰ to the case where the dimensionality of the distance matrix is comparable to the number of desired clusters, and ensures that we choose a diverse set of structures from our original set.

The unique structures are determined with DSI threshold of 1.0 Å. It can be seen from Table 1 that to get essentially same number of unique structures, incremental method takes about twice (for $(\text{H}_2\text{O})_{21}$) to over thrice (for $(\text{H}_2\text{O})_{20}$) more time than that by the MCTBP method alone. The major cost of the incremental method is coming from the MCTBP runs in the incremental method, which is of 50,000 steps for all 50 structures. For comparison, we have also performed Monte Carlo Basin Hopping (MCBH) calculation for all the structures with 20,000 steps each using the same protocol used for MCTBP calculations. The number of

Table 1. Comparison of time for MCTBP and incremental methods.

Water Clusters (H ₂ O) _n	Total No. of MCTBP steps	MCTBP Time (in hours)	No. of unique Structures	Incremental method time (approx. in hours)
(H ₂ O) ₂₁	18183	6.12	300	19
(H ₂ O) ₂₂	12520	4.58	300	8.5
(H ₂ O) ₂₃	17862	6.47	300	12.5
(H ₂ O) ₂₄	27414	9.07	230	19
(H ₂ O) ₂₅	24614	8.38	160	19

Table 2. Comparison of performance between incremental and MCTBP method.

Water clusters (H ₂ O) _n	Lowest energy structures obtained*			DSI range (Angstrom)	
	MCTBP	incremental	MCBH	MCTBP	Incremental
(H ₂ O) ₂₁	-3.72(0.76)	-3.93(0.54)	-4.58(-0.09)	1-20.0	1-21.9
(H ₂ O) ₂₂	2.33(1.72)	0.20(-0.42)	1.15(0.53)	1-16.9	1-18.9
(H ₂ O) ₂₃	1.24(1.24)	1.14(1.14)	1.14(1.14)	1-11.2	1-13.5
(H ₂ O) ₂₄	2.37(2.37)	0.00(0.00)	2.01(2.01)	1-14.2	1-16.2
(H ₂ O) ₂₅	0.92(1.21)	1.41(1.71)	2.41(2.70)	1-9.4	1-16.5

*in kcal/mol relative to refs. ⁴¹ and ⁴². The value in parenthesis is relative energy with respect to the structures obtained in reference ⁴². Incremental means MCTBP+incremental.

unique structures (with 1 Å DSI) obtained were 89, 206, 110, 142 and 140 for (H₂O)₂₁, (H₂O)₂₂, (H₂O)₂₃, (H₂O)₂₄ and (H₂O)₂₅, respectively. Hence, MCBH appears to be less effective than either of MCTBP and MCTBP+incremental approach as far as finding unique structures are concerned. We have also checked the effect of choosing a different DSI threshold other than 1.0 Å. For a threshold of 1.3 Å, the number of unique structures obtained by both MCTBP and MCTBP+incremental methods decreases. However, the trend obtained from 1.0 Å DSI value remains same with the DSI value of 1.3 Å except for (H₂O)₂₁ (where the unique structures are 181 and 182 for MCTBP and MCTBP+incremental methods, respectively). To get a feeling of the same trend for other cases, we are giving the values for (H₂O)₂₂ and (H₂O)₂₅ (where the first and second value inside the parenthesis correspond to MCTBP and MCTBP+incremental methods, respectively); for (H₂O)₂₂ and (H₂O)₂₅ the values are (179, 368) and (98, 129), respectively. Hence, our conclusion is not dependent on the particular DSI value chosen.

Table 2 shows comparative performance, in terms of low energy structure and diversity of the structures, between the MCTBP and MCTBP+incremental approaches with respect to previous literature.⁴² For the low energy part we have also included the results obtained from the MCBH method. As seen from the table, as far as low energy structure determination is concerned, both MCTBP and MCTBP+incremental methods performed similarly. For (H₂O)₂₄, incremental

method gets structure more than 2 kcal/mol less than the MCTBP alone. However, for (H₂O)₂₂ the trend is reverse. Of course, one can extend an MCTBP run from the best structure obtained from the incremental approach with 1000 step MCTBP run. We tried this for (H₂O)₂₅ and got a structure, which is of essentially same energy as the best structure given by Kazachenko *et al.*,⁴¹ after running about 6000 additional MCTBP steps. This structure is shown in Figure 7. The performance of MCBH is comparable with the other two methods (for (H₂O)₂₁, this gives the lowest energy structure among the three methods) except for (H₂O)₂₄ and (H₂O)₂₅, where the lowest energy structure is about 2 and 1 kcal/mol more, respectively, in energy than that obtained from the MCTBP+incremental approach.

The most interesting part of Table 2 is the range of DSI values for the structures obtained using the two methods. It is clearly seen that the incremental method outperforms MCTBP for all the cluster sizes. Especially, for (H₂O)₂₅ the structures have DSI range 1-9.4 Å for MCTBP, while the incremental method gives 1-16.5 Å showing a much larger structural space is sampled.

To comment conclusively on the performance of the incremental method versus MCTBP, one needs to perform more extensive testing using different set of molecules. Nonetheless, the following points emerge: (1) the current implementation of the incremental method is slower than the MCTBP method as far as the determination of number of low energy structures are concerned; (2) The low energy structures obtained by both methods are similar in terms of energy; (3)

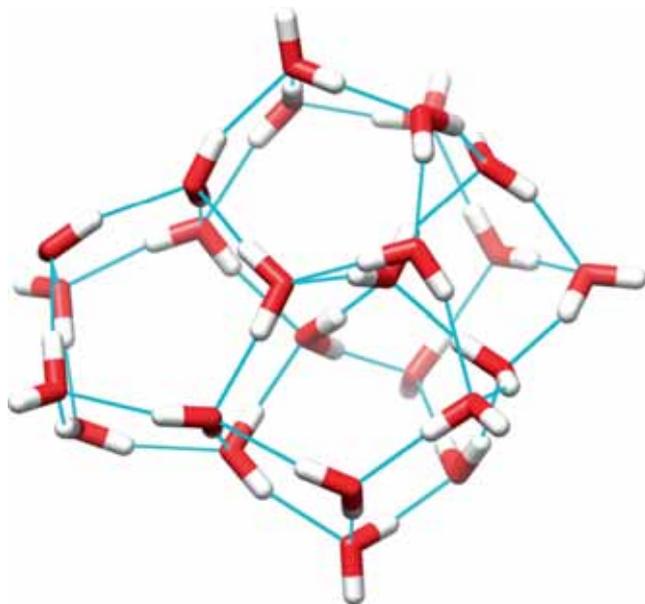


Figure 7. The lowest energy structure of $(\text{H}_2\text{O})_{25}$ obtained from graph theory based incremental approach.

The diversity of structures from the incremental method is clearly more than that obtained from the MCTBP method. The computational cost of the incremental method can be reduced further by choosing the most diverse graphs before performing the 2D to 3D conversion. To summarize, the new incremental method can be a complementary tool to the MCTBP-like method.

4. Conclusions

Although there are a very large number of optimization algorithms for large molecules, degree of completeness of search space is usually not known. In the current work, full enumeration of graphs representing $(\text{H}_2\text{O})_6$ and $(\text{H}_2\text{O})_7$ are used to approximately check the completeness of sampling with MCTBP method. It was found that full enumeration of graphs and subsequent conversion of the graphs to 3D structure can be used as a measure to assess the completeness of search for water clusters. Graph theory and MCTBP method were combined to get an efficient optimization algorithm for water clusters which matches MCTBP in performance as far as finding the low energy structures is concerned.

Acknowledgements

The authors would like to thank Prof. S. R. Gadre (IIT Kanpur) for illuminating discussions.

References

1. Wales D J 2003 In *Energy Landscapes* (London: Cambridge University Press)
2. Lee J, Scheraga H A and Rackovsky S 1997 *J. Comp. Chem.* **18** 1222
3. Kabrede H and Hentschke R 2002 *J. Phys. Chem. B* **106** 10089
4. Goedecker S 2004 *J. Chem. Phys.* **120** 9911
5. Lee J, Lee I and Lee J 2003 *Phys. Rev. Lett.* **91** 080201
6. Kabrede H and Hentschke R 2003 *J. Phys. Chem. B* **107** 3914
7. Roy S, Goedecker S, Field M J and Penev E 2009 *J. Phys. Chem. B* **113** 7315
8. Bandyopadhyay P and Kharerin H 2011 *Chem. Phys. Lett.* **502** 130
9. Neogi S G and Chaudhuri P 2012 *J. Comp. Chem.* **33** 629
10. Neogi S G and Chaudhuri P 2014 *J. Comp. Chem.* **35** 51
11. Holger H H and Stützle T 2004 In *Stochastic Local Search: Foundations and Applications* (San Francisco, CA: Morgan Kaufmann/ Elsevier)
12. Perez J J, Villar H O and Arteca G A 1994 *J. Phys. Chem.* **98** 2318
13. Centeno N B and Perez J J 1995 *Chem. Phys. Lett.* **232** 374
14. Sullivan D C and Kuntz I D 2001 *Proteins: Struct. Funct. Genet.* **42** 495
15. Shanker S and Bandyopadhyay P 2011 *J. Phys. Chem. A* **115** 11866
16. Rakshit A and Bandyopadhyay P 2013 *Comp. Theo. Chem.* **1021** 206
17. Brinkmann G 2009 *J. Math. Chem.* **46** 1112
18. Miyake T and Aida M 2002 *Chem. Phys. Lett.* **363** 106
19. Miyake T and Aida M 2006 *Chem. Phys. Lett.* **427** 215
20. Vukičević D, Grubeša T and Graovac A 2005 *Chem. Phys. Lett.* **416** 212
21. Tissandier M D, Singer S J and Coe J V 2000 *J. Phys. Chem. A* **104** 752
22. Singer S J and Knight C 2012 *Adv. Chem. Phys.* **147** 1 (Eds.) S A Rice and A R Dinner (Hoboken, NJ: John Wiley)
23. Akase D and Aida M 2014 *J. Phys. Chem. A* **118** 7911
24. Radhakrishnan T P and Herndon W C 1991 *J. Phys. Chem.* **95** 10609
25. McDonald S, Ojamäe L and Singer S J 1998 *J. Phys. Chem. A* **102** 2824
26. Kirov M V, Fanourgakis G S and Xantheas S S 2008 *Chem. Phys. Lett.* **461** 180
27. Yeole S D and Gadre S R 2011 *J. Chem. Phys.* **134** 084111
28. Wales D J and Hodges M P 1998 *Chem. Phys. Lett.* **286** 65
29. Li J and Scheraga H A 1987 *Proc. Nacl. Acad. Sci. USA* **84** 6611
30. Iwata S, Bandyopadhyay P and Xantheas S S 2013 *J. Phys. Chem. A* **117** 6641
31. Furtado J P, Rahalkar A P, Shanker S, Bandyopadhyay P and Gadre S R 2012 *J. Phys. Chem. Lett.* **3** 2253
32. Sahu N, Gadre S R, Rakshit A, Bandyopadhyay P, Miliordos E and Xantheas S S 2014 *J. Chem. Phys.* **141** 164304

33. Gordon M S, Freitag M, Bandyopadhyay P, Jensen J H, Kairys V and Stevens W J 2001 *J. Phys. Chem. A* **105** 293
34. Schmidt M W, Baldrige K K, Boatz J A, Elbert S T, Gordon M S, Jensen J H, Koseki S, Matsunaga N, Nguyen K A, Su S J and Windus T L 1993 *J. Comp. Chem.* **14** 1347
35. Kazimirski J K and Buch V 2003 *J. Phys. Chem. A* **107** 9762
36. Day P N, Pachter R, Gordon M S and Merrill G N 2000 *J. Chem. Phys.* **112** 2063
37. Jensen J O, Krishnan P N and Burke L A 1995 *Chem. Phys. Lett.* **241** 253
38. Kim J, Majumdar D, Lee H M and Kim K S 1999 *J. Chem. Phys.* **110** 9128
39. Wales D J, Doye J P K, Dullweber A, Hodges M P, Naumkin F Y, Calvo F, Hernández-Rojas J and Middleton T F *The Cambridge Cluster Database* URL <http://www-wales.ch.cam.ac.uk/CCD.html> (Accessed on 10.05.2016)
40. Dhillon I S, Guan Y and Kulis B 2004 *Kernel k-means, Spectral Clustering and Normalized Cuts* In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)* (New York: ACM) pp. 551–556
41. Kazachenko S and Thakkar A 2009 *Chem. Phys. Lett.* **476** 120
42. Bandow B and Hartke B 2006 *J. Phys. Chem. A* **110** 5809