PCCP

Accepted Manuscript

This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the **Information for Authors**.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard **Terms & Conditions** and the **Ethical guidelines** still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

Journal Name

ARTICLE TYPE

Global Optimization of Clusters of Rigid Molecules by the Artificial Bee Colony Algorithm[†]

Jun Zhang[*a] and Michael Dolg[*a]

The global optimization of molecular clusters is an important topic encountered in many fields of chemistry. In our previous work (*Phys. Chem. Chem. Phys.* 2015, 17, 24173), we successfully applied the recently introduced artificial bee colony (ABC) algorithm to the global optimization of atomic clusters and introduced the corresponding software "ABCluster". In the present work, ABCluster was extended to the optimization of clusters of rigid molecules. Here "rigid" means that all internal degrees of freedom of the constituent molecules are frozen. The algorithm was benchmarked by TIP4P water clusters $(H_2O)_N$ ($N \leq 20$), for which all global minima were successfully located. It was further applied to various clusters of different chemical nature: 10 microhydration clusters, 4 methanol microsolvation clusters, 4 nonpolar clusters and 2 ion-aromatic clusters. In all the cases we obtained results consistent with previous experimental or theoretical works.

1 Introduction

Molecular clusters are aggregates containing several molecules. They are gaining more and more attention among researchers due to their chemical importance. For experimental studies, clusters often exhibit an interesting size-dependence of their properties when going from a couple of molecules to the bulk substance[1]. Clusters can also reveal some local structural information of liquids[2]. Many metal clusters exhibit special catalytic and optical properties[3]. In theoretical studies, modelling solvation processes, e.g. calculating the hydration energy, requires one to introduce explicit water molecules in the inner hydration spheres as well as an implicit solvation model to guarantee a sufficient accuracy[4], i.e. such studies involve the treatment of a solute-solvent cluster[5,6].

For theoretical studies of a molecular cluster, the first step is often to find the global minimum (GM) on its potential energy surface (PES), since this corresponds usually to its most stable structure at low temperature. However, for larger systems this is a difficult task. One can identify a local minimum (LM) on the PES by the zero gradient condition, but a robust condition for identifying a GM does not exist. Therefore, a deterministic search of the GM is usually impossible, and for such global optimization problems, nondeterministic algorithms are more popular. These algorithms can find the true GM beyond a significant probability after a sufficient number of iterations. For cluster optimiza-

tion, there are two kinds of algorithms: the biased and unbiased ones. The former class is designed for specific clusters, since it uses the known information of the GMs of small clusters as much as possible to search those of the larger ones. The latter class can be applied for general clusters, and makes no assumptions how their GMs should look like. The biased algorithms are often more efficient for specific clusters than the unbiased ones, e.g. basin-hopping[7,8], an unbiased algorithm, can find reliable GMs for $(H_2O)_N$ when $N \leq 21$[9], but a biased algorithm designed specifically for water clusters can work well up to $N = 30$[10]. Obviously the biased algorithms are not robust and transferable. The unbiased algorithms can be classified as individual-based (starting the global optimization from a single cluster, e.g. simulated annealing[11], Monte Carlo minimization[12] and basin hopping[7,8]) and population-based (starting the global optimization from a set of clusters, e.g. differential evolution[13] and particle swarm optimization[14]) ones. For a comprehensive discussion of these methods please refer to related reviews[15–17].

In our previous work[18], we introduce a recently proposed population-based algorithm, i.e. the "artificial bee colony" (ABC) algorithm to the field of global optimization of clusters. The ABC algorithm requires only three parameters, thus it is very easy to learn and apply by non-experts in global optimization. It has been wrapped in a black-box way in the software ABCluster (ABC for clusters). We have proved that ABCluster is very efficient in searching the GMs for ionic and metal atomic clusters. In this work, we extend the ability of ABCluster to clusters formed by rigid molecules and show its excellent performance for these cases. For similar results as well as references for other optimization schemes and applications for

[a] Theoretical Chemistry, University of Cologne, Greinstr. 4, 50939 Cologne, Germany. Fax: ++49 (0)221 470 6896; Tel: ++49 (0)221 470 6893; E-mail: zhangjun-qcc@gmail.com, m.dolg@uni-koeln.de.

atomic clusters, the reader is refered to Ref. 18. Now, ABCluster is available on our group site (http://www.uni-koeln.de/math-nat-fak/tcchem/mitarbeiter/zhang/zhang/software-abcluster-introduction.html).

2 Theory

2.1 The Potential Energy Function

Searching the GM of a molecular cluster is mathematically an unconstrained global optimization problem. The variables to be optimized are the atomic coordinates $\mathbf{X} \equiv \{x_1, y_1, z_1, \cdots, x_M, y_M, z_M\}$ where M is the number of atoms in the cluster, and the objective function is the potential energy function $U(\mathbf{X})$ which can be an empirical or first-principle one. In this work, we only consider clusters of *rigid* molecules, where "rigid" means that all internal degrees of freedom (DOFs) of a molecule (bond lengths, bond angles and dihedral angles) are kept unchanged during the optimization. This approximation is not suitable for all cases, e.g. it is not suitable for molecules with soft DOFs like those with long, rotatable side chains. However it works quite well for most small and medium-sized molecules and can significantly reduce the number of DOFs of the cluster (i.e. the dimension of the global optimization problem). Within this approximation, each molecule can be described by a six-component external DOF $\mathbf{q}$. In ABCluster we use the following $\mathbf{q}$: the coordinates of its geometrical center $\mathbf{R} \equiv \{X, Y, Z\}$ and three Euler angles $\Omega \equiv \{\alpha, \beta, \gamma\}$ relative to its pre-defined body-fixed coordinate system (see Figure 1). Other choices of coordinates like angle-axis representation[19] are possible but their performance in the global optimization shows no significant difference[20]. Thus, for a cluster containing N rigid molecules, the total number of DOFs is $6N$, i.e.

$$\begin{aligned} \mathbf{Q} &\equiv \{\mathbf{q}_1, \cdots, \mathbf{q}_N\} \equiv \{\mathbf{R}_1, \Omega_1, \cdots, \mathbf{R}_N, \Omega_N\} \\ &\equiv \{X_1, Y_1, Z_1, \alpha_1, \beta_1, \gamma_1, \cdots, X_N, Y_N, Z_N, \alpha_N, \beta_N, \gamma_N\} \end{aligned} \quad (1)$$

As the molecules are rigid, $U(\mathbf{X})$ is a function of $\mathbf{Q}$. The elimination of the translation and rotation DOFs reduces $\mathbf{Q}$ to $6N - 6$ coordinates. In principle, an "exact" solution to this global optimization problem requires an ergodic sampling over the $\mathbf{Q}$ space. This is impossible for large clusters. The sampling difficulty has been discussed for atomic clusters in our previous work[18] where we pointed out that the number of LMs of a cluster of size N increases exponentially[15,21] leading to a rugged PES. For molecular clusters the difficulty manifests itself in an additional aspect. A molecule can have several directional interaction sites, e.g. H_2O has four (two H atoms and two lone electron pairs on O), CH_3OH has three (one H atom and two lone electron pairs on O), and C_6H_6 has two (the π electron system on each side of the molecular plane). For large N, the number of the possible interaction network topologies in a cluster can be extremely large and depends strongly on the nature of its components. Thus, a convergent global optimization may require a very long computation time. For biased algorithms, one could use some graph theory approaches to accelerate the optimization for specific clusters (e.g. Ref. 10). Since the algorithm in ABCluster is an unbiased one and is designed for general rigid molecular clusters, we do not apply

these approaches. To alleviate the ruggedness of the PES, we use a smoothed PES function $\tilde{U}$ rather than the original one[18]:

$$\tilde{U}(\mathbf{Q}) = \min : \{U(\mathbf{Q})\} \quad (2)$$

where "min" stands for performing a local minimization of U starting from $\mathbf{Q}$. The advantage of $\tilde{U}$ over U is that the former one removes the energetic barriers along the downhill movement towards a funnel, leading to a more efficient optimization[18,22]. It has been used in the pioneering work of the global optimization of proteins[23]. However, (2) cannot remove the barriers between the funnels. These barriers make sampling clusters of quite different interaction network topologies require a long time, being the bottleneck of the global optimization problem.

The potential energy function U is essential in the description of a molecular cluster. In this work we only consider the two-body empirical potential function of the following form:

$$\begin{aligned} U(\mathbf{Q}) = \sum_{I=1}^{N} \sum_{I<J}^{N} \sum_{i_I \in I} \sum_{j_J \in J} &\left[\frac{e^2}{4\pi\varepsilon_0} \frac{q_{i_I} q_{j_J}}{r_{i_I j_J}} \right. \\ &\left. + 4\varepsilon_{i_I j_J} \left(\left(\frac{\sigma_{i_I j_J}}{r_{i_I j_J}} \right)^{12} - \left(\frac{\sigma_{i_I j_J}}{r_{i_I j_J}} \right)^6 \right) \right] \end{aligned} \quad (3)$$

Here I and J are the indices of the molecules, i_I and j_J are the indices of the atoms in molecule I and J, respectively. $r_{i_I j_J}$ is the distance between atom i_I and j_J. Obviously (3) only considers the intermolecular Coulomb and Lennard-Jones interactions. Although this form is simple, it is used in many modern force fields like CHARMM[24], OPLS[25] and AMBER[26] and has been tested in numerous studies, confirming its reliability. Using a more sophisticated U is more expensive. Also, since the GM is very sensitive to the form and parameters of U (e.g. the GM of $(H_2O)_6$ with the TIP4P and TIP5P force field is the cage and ring isomer, respectively[9,20]), using different forms of potentials will cause confusion. Therefore, we decide to use only the simplest form (3) in this work. In practice one can first obtain a set of LMs with (3) and then study them further with, e.g., quantum chemical methods.

For the technical details of computing (2) and (3) by using the external DOFs $\mathbf{q}$ please refer to the Appendix.

2.2 The Artificial Bee Colony Algorithm

The artificial bee colony (ABC) algorithm was proposed in 2005 by Karaboga[27]. It is a swarm intelligence based algorithm, modelling the foraging behavior of honey bee colonies. The bees want to find the best nectar as food source and have developed an efficient methodology to accomplish this mission. In terms of the global optimization problem, a rigid molecular cluster with external DOFs $\mathbf{Q}$ is a nectar, its energy $U(\mathbf{Q})$ is the quality of the nectar. A lower energy implies a higher quality or a good solution. The ABC algorithm simulates bees' methodology by introducing three kinds of bees: employed, onlooker and scout bees. In each search cycle, first employed bees perform a coarse exploration of the $\mathbf{Q}$ space, obtaining some trial solutions; then onlooker bees do the search in the neighborhood of some "good" solutions; finally
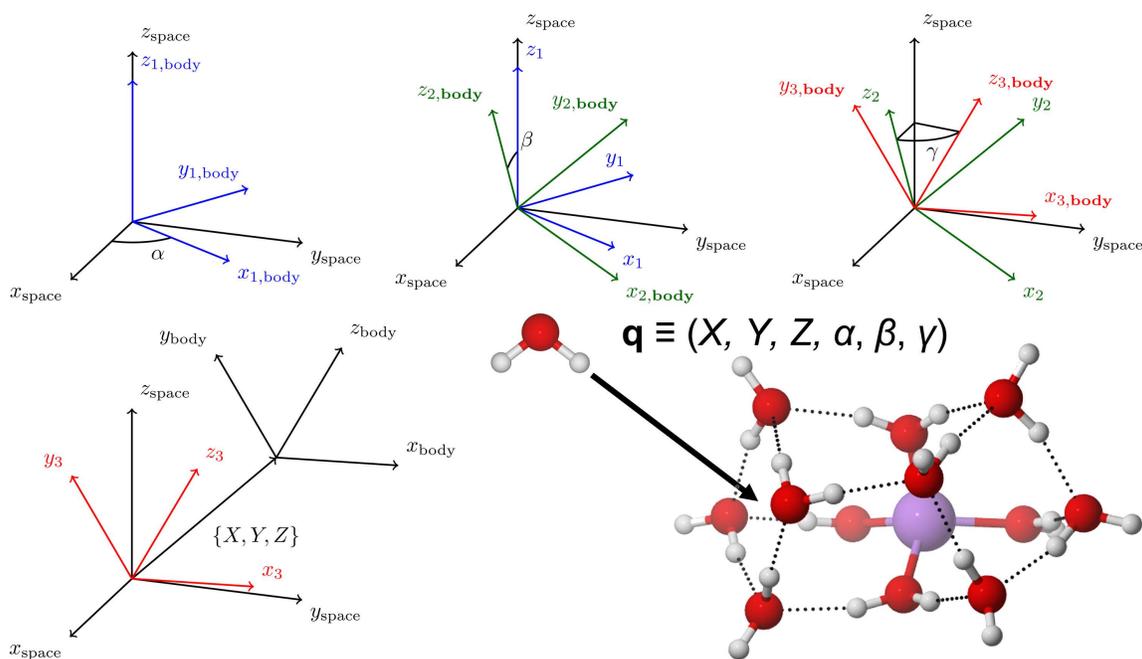
Fig. 1 The external DOFs **q** of a rigid molecule used in ABCluster. A molecule will first be rotated by α, β and γ as shown in the Figure and then translated by $\{X,Y,Z\}$ to its final position in a cluster. For explicit expressions please refer to the Appendix.

scout bees examine the obtained solutions and discard the ones that had little contribution to the improvement of the solutions during the past several cycles and replace them by new random ones. The search cycles until some stopping criteria are satisfied, and the best solution obtained so far is assumed to correspond to the GM. The mechanism and performance of the ABC algorithms have been discussed in several papers[28–30]. Especially, its specific implementation in ABCluster has been discussed in details in our previous work[18]. For the global optimization of rigid molecular clusters, the ABC algorithm is very similar to the one for atomic clusters[18]. Therefore we will only briefly describe it here.

In the ABC algorithm, three parameters are needed: the size of the population of trial solutions SN, the scout limit g_{limit} and the maximum cycle number g_{max}. The cluster is described by its external DOFs **Q**, size N, an estimated length L, and the potential parameters. The global optimization then begins:

1. Initialize the population: $\mathbf{Q}_1^1, \cdots, \mathbf{Q}_{SN}^1$. One can use random initial guesses, i.e. each component of **R** and Ω are randomly taken from the range $[0,L]$ and $[0,2\pi)$, respectively. Next all the clusters are locally optimized by the limited-memory-Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) algorithm[31]. The LM (or GM) property of the obtained structures is here tacitly assumed.

2. Modelling employed bees: In cycle g, for each $\mathbf{Q}_i^g$ $(i = 1,\cdots,SN)$, a new trial solution $\mathbf{V}_i$ is generated by the trigonometric mutation operator[32]:

$$\mathbf{V}_i = \frac{1}{3}\left(\mathbf{Q}_{k_1}^g + \mathbf{Q}_{k_2}^g + \mathbf{Q}_{k_3}^g\right) + (p_2 - p_1)\left(\mathbf{Q}_{k_1}^g - \mathbf{Q}_{k_2}^g\right)$$
$$+ (p_3 - p_2)\left(\mathbf{Q}_{k_2}^g - \mathbf{Q}_{k_3}^g\right) + (p_1 - p_3)\left(\mathbf{Q}_{k_3}^g - \mathbf{Q}_{k_1}^g\right) \quad (4)$$

where k_1, k_2 and k_3 are random integers in $\{1,\cdots,SN\}$ and $k_1 \neq k_2 \neq k_3 \neq i$, and

$$p_{k_m} = \frac{\left|\tilde{U}\left(\mathbf{Q}_{k_m}\right)\right|}{\left|\tilde{U}\left(\mathbf{Q}_{k_1}\right)\right| + \left|\tilde{U}\left(\mathbf{Q}_{k_2}\right)\right| + \left|\tilde{U}\left(\mathbf{Q}_{k_3}\right)\right|} \quad (m = 1,2,3) \quad (5)$$

$\tilde{U}$ is the smoothed potential energy function (2). This solution is updated with a greedy selection scheme (6):

$$\mathbf{Q}_i^{g+1} = \begin{cases} \mathbf{V}_i & \text{if } \tilde{U}\left(\mathbf{V}_i\right) < \tilde{U}\left(\mathbf{Q}_i^g\right) \\ \mathbf{Q}_i^g & \text{otherwise} \end{cases} \quad (6)$$

3. Modelling onlooker bees: For SN times, a "good" solution $\mathbf{Q}_k^g$ is selected by the tournament scheme[16] and a new trial solution $\mathbf{V}_k$ is generated by the "ABC/current/2+ABC/best/2" strategy (7):

$$\mathbf{V}_k = \begin{cases} \mathbf{Q}_k^g + F\left(\mathbf{Q}_{k_1}^g + \mathbf{Q}_{k_2}^g - \mathbf{Q}_{k_3}^g - \mathbf{Q}_{k_4}^g\right) & \text{if } \eta < 0.5 \\ \mathbf{Q}_{\text{best}}^g + F\left(\mathbf{Q}_{k_1}^g + \mathbf{Q}_{k_2}^g - \mathbf{Q}_{k_3}^g - \mathbf{Q}_{k_4}^g\right) & \text{otherwise} \end{cases} \quad (7)$$

where k_1, k_2, k_3 and k_4 are random integers in $\{1,\cdots,SN\}$ and $k_1 \neq k_2 \neq k_3 \neq k_4 \neq k$. F and η are random numbers in $[0,1)$. $\mathbf{Q}_k^g$ is again updated with the greedy selection scheme (6).

4. Modelling scout bees: Now each $\mathbf{Q}_i^g$ $(i = 1,\cdots,SN)$ is examined. A $\mathbf{Q}_i^g$ which does not change in the last g_{limit} cycles will be replaced by a random trial solution $\mathbf{Q}_i^{g+1}$ regardless of whether it is better than $\mathbf{Q}_i^g$ or not.

5. If $g \geq g_{\text{max}}$, the algorithm is finished, otherwise go to step 2.

3 Applications

In the section, we examine the performance of the ABC algorithm in the global optimization of rigid molecular clusters. All the potential parameters in (3) were taken from the CHARMM36 force field[33] and the details can be found in the Supplementary Information. All optimizations were performed by ABCluster[18]. The graphs of the clusters were rendered by CYLView[34].

3.1 Water Clusters as Benchmark

Water clusters are of fundamental importance in chemistry[35] and thus their GMs have attracted much attention from the scientific community. Water molecules can form complex hydrogen bond networks. For $(H_2O)_N$ in ice-I_h structure, Pauling pointed out that the number of possible networks scales as $(3/2)^N$ [36]. This is a big challenge for a global optimization. The global minima of small water clusters are well documented in the literature (see Ref. 10, 37 and references therein). Therefore we take this system as a first benchmark of our algorithm. Here the water molecule is described by the TIP4P model[38]. The algorithm parameters and optimization results are given in Table 1. Some GMs are shown in Figure 2.

Table 1 confirms that ABCluster successfully located the GMs for all $(H_2O)_N$ ($N = 5 - 20$) clusters. The number of steps required for convergence increases rapidly for $N \geq 10$, reflecting the exponential scaling of the number of their LMs. It is observed that this quantity increases as $\exp(0.60N)$. Interestingly, Takeuchi found a similar dependency of the number of local optimizations performed during a search for the GM of $(H_2O)_N$ clusters using another method, i.e. $\exp(0.63N)$ [10]. Therefore, a reliable search for the GMs of $(H_2O)_{21}$ would already require more than 10^5 steps. The optimization becomes more difficult for unbiased methods like basin hopping[20] and even for biased algorithms[10]. Nevertheless, this benchmark confirms the reliability of our algorithm. In principle, any GM can be found with sufficiently large g_{max}. In the remainder we will apply the ABC algorithm to more complex systems.

Table 1 Benchmark for the water clusters (Energy unit: kJ mol^{-1})$^{a,\,b}$

ABC algorithm parameters: $SN = 60$, $g_{limit} = 4$, $g_{max} = 30000$ Initial guess: random					
$(H_2O)_N$	Step^c	Energy	$(H_2O)_N$	Step^c	Energy
5	1	−152.1371	13	304	−533.0679
6	1	−197.8168	14	437	−583.0969
7	1	−243.6168	15	1485	−628.4856
8	1	−305.5747	16	783	−681.3157
9	6	−344.4982	17	1940	−723.9389
10	16	−391.0943	18	2221	−773.3718
11	169	−431.5672	19	4285	−821.1843
12	28	−492.9979	20	28054	−873.1465

^{a} The reference GM energies ($N = 5 - 20$) are from Ref. 9. Note that our energy is slightly larger in magnitude than theirs (e.g. for $(H_2O)_{10}$ −391.0943 vs −391.0227). This is probably due to the slightly different accuracy of "$\frac{e^2}{4\pi\varepsilon_0}$" in (3). thus they are in fact identical (For this we use $\frac{e^2}{4\pi\varepsilon_0} = 1389.506$ Å kJ mol^{-1}).
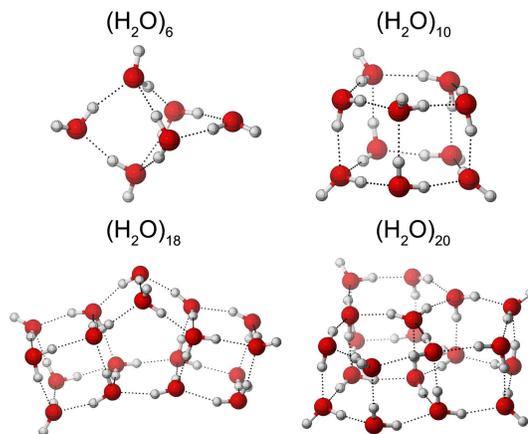^{c} "Step" is the step at which the final energy is obtained.

Fig. 2 Some GMs of water clusters obtained by ABCluster.

3.2 Microhydration Clusters

Next we want to examine the performance of ABCluster for some microhydration clusters, i.e. $X(H_2O)_N$. We chose $N = 20$ in most cases in order to model the behavior of solute X in a sufficient amount of water to form at least a complete first hydration sphere. Each of the systems discussed in the following could be part of an independent project but here they are merely used as examples to prove the accuracy and robustness of our methodology. The optimization results are given in Table 2 and Figure 3.

First we consider the three alkali cations Na$^+$, K$^+$ and Cs$^+$. The subtle difference in hydration properties of Na$^+$ and K$^+$ makes them play important but completely different roles in biological processes. An essential factor is their charge density. The water molecules interact stronger and thus tend to be closer to X with higher charge density such as Na$^+$. This leads to an increased repulsion between the directly coordinating water molecules and to a preference of a smaller water coordination number (CN). Our optimization results (see Figure 3) confirmed this: in the GM of Na$^+(H_2O)_{20}$, Na$^+$ takes an off-center position with CN = 6, while in the GM of K$^+(H_2O)_{20}$ and Cs$^+(H_2O)_{20}$ the cations are found in the center and exhibit larger CNs, being of clathrate-like structure. These observations are in agreement with previous studies on Li$^+$ to Cs$^+$ and Ca^{2+} [39–41].

As an example for a molecular cation we considered guanidinium (Gmd$^+$), which is characterized as "most weakly" hydrated since *no* recognizable hydration shell is observed by neutron diffraction[42]! Figure 3 reveals that in the GM of Gmd$^+(H_2O)_{20}$, the cation is pushed to the periphery of the cluster formed by the water molecules. The unfavorable interaction of Gmd$^+$ with water is compatible with the experimental observations mentioned above. This phenomenon is critical to understanding the fact that Gmd$^+$ is a strong protein denaturant[42].

Next we look at the two anions Cl$^-$ and SO$_4^{2-}$. Figure 3 suggests that both anions locate in the center of their GMs. Since an anion can accept more O−H bonds than a water molecule, its microhydration cluster contains much less dangling O−H bonds than that of a cation. Indeed, the GM of Cl$^-(H_2O)_{20}$ has only two and the GM of SO$_4^{2-}(H_2O)_{20}$ even has none! In the infrared pho-

todissociation (IRPD) spectroscopy of small $SO_4^{2-}(H_2O)_N$ clusters ($N \leq 43$), no bands corresponding to dangling O−H bonds were observed[2]. This supports the result of our as well as of a recent work[43].

Further we examined two cation-anion microhydration clusters, i.e. $K^+Cl^-(H_2O)_{20}$ and $Mg^{2+}SO_4^{2-}(H_2O)_{20}$. Since cations and anions lock different DOFs of a water molecule, their effects on surrounding water can be strongly interdependent and nonadditive, reaching sometimes even beyond the first hydration shells of the ions[44]. In Figure 3, their GMs are both a water cage containing an ion pair. With more water molecules the ion pair may separate to lower the cluster energy, since small cages can lead denser water hydrogen bond networks to compensate the energy loss of the ion pair separation.

Finally we consider the microhydration clusters of nonpolar molecules such as $CH_4(H_2O)_{20}$ and coronene-$(H_2O)_{10}$, where the latter might be considered as a model for graphite-$(H_2O)_{10}$. We observe in Figure 3 that their GMs look like CH_4 or coronene are loosely binding a relaxed GM of $(H_2O)_N$. For $C_{60}(H_2O)_N$ similar phenomena were observed[45]. However, when the water cluster has very low-lying LMs, the nonpolar molecule may change their energy order, e.g. in the GM of $C_{60}(H_2O)_6$ and coronene-$(H_2O)_6$, the $(H_2O)_6$ part is of book rather than of cage conformation as it is in the unbound state[45,46] (see Figure 2). We note here in passing that the clathrate-like conformation (i.e. CH_4 in the center of a dodecahedral $(H_2O)_{20}$ cage) is not the GM of $CH_4(H_2O)_{20}$, even when pressure is taken into account[47].

Table 2 The global optimization of some microhydration clusters (Energy unit: kJ mol^{-1})

ABC algorithm parameters: $SN = 60$, $g_{limit} = 4$, $g_{max} = 35000$ Initial guess: random		
System	Step^a	Energy
$Na^+(H_2O)_{20}$	2627	−1167.6197
$K^+(H_2O)_{20}$	1349	−1099.8384
$Cs^+(H_2O)_{20}$	103	−1050.8635
$Gmd^+(H_2O)_{20}$	2201	−1016.3934
$Cl^-(H_2O)_{20}$	3011	−1143.7146
$SO_4^{2-}(H_2O)_{20}$	2798	−1659.7028
$K^+Cl^-(H_2O)_{20}$	14617	−1585.2550
$Mg^{2+}SO_4^{2-}(H_2O)_{20}$	45	−3937.4321
$CH_4(H_2O)_{20}$	30921	−879.6630
Coronene-$(H_2O)_{10}$	18	−430.6415

^{a} "Step" is the step at which the final energy is obtained.

3.3 Methanol Microsolvation Clusters

Organic solvents are extremely important for chemical reactions. They can change the reaction mechanism[48] or even the spin state of the solute[49]. Here we will focus on methanol, which is a unique solvent since it contains both a hydrophilic (OH) and hydrophobic CH_3 group. The opposite character of these two possible interactions of methanol with solutes causes the solvation properties to be somewhat different from those of water. We searched the GMs of some methanol microsolvation systems, the results of which shown in Table 3 and Figure 4.

For $(CH_3OH)_{13}$, the GM is a chain linked by hydrogen bonds, which is folded to a helix by the methyl packing (quasi-one-

dimensional structure denoted by "q1D"). Indeed, many experimental[50,51] and theoretical[52] studies implied that unlike water, which tends to form a complex hydrogen bond network, a large number of methanol molecules prefer quasi-one-dimensional structures. However, for this species, the B3LYP-DCP method predicted that the GM is a six-membered and a seven-membered ring in a face-to-face pose (denoted by "6+7")[?]. This structure is also detected by ABCluster, however, only as a LM lying 2.3 kJ mol^{-1} above the GM. We extracted the structure of "q1D" and "6+7" and optimized them at B3LYP-D3/def2-tzvp level[53–55] using ORCA[56]. It turned out that "6+7" is more stable than "q1D" by 7.5 kJ mol^{-1}. Thus this inconsistency is attributed to a defect of the force field parametrization of methanol. In fact, CHARMM as well as some other force fields are known to tend to overestimate the stability of helix structures[57].

In the GM of $Na^+(CH_3OH)_{12}$, Na^+ still shows a CN of 6 as in water, the remaining methanol molecules forming a ring. This is also observed for $Li^+(CH_3OH)_N$[58]. For $Cl^-(CH_3OH)_{12}$, Cl^- also shows a CN of 6 and it is "interior" solvated, in agreement with the fact[59] that the vibrational spectrum of $Cl^-(CH_3OH)_{12}$ exhibits a broad peak between 3500 to 3100 cm^{-1}. For $C_6H_6(CH_3OH)_{12}$, the GM is a "surface" solvated structure: the benzene molecule is weakly bound by two methyl groups of the $(CH_3OH)_{12}$ cluster. Similar structures were observed by quantum chemical methods for $C_6H_6(CH_3OH)_N$ ($N \leq 6$)[60].

Table 3 The global optimization of some methanol microsolvation clusters (Energy unit: kJ mol^{-1})

ABC algorithm parameters: $SN = 60$, $g_{limit} = 4$, $g_{max} = 5000$ Initial guess: random		
System	Step^a	Energy
$(CH_3OH)_{13}$	1012 (GM, "q1D")	−464.1312
$(CH_3OH)_{13}$	LM, "6+7"	−461.8823
$Na^+(CH_3OH)_{12}$	36	−795.9935
$Cl^-(CH_3OH)_{12}$	1750	−603.9984
$C_6H_6(CH_3OH)_{12}$	460	−449.0779

^{a} "Step" is the step at which the final energy is obtained.

3.4 Nonpolar Molecular Clusters

We also consider four nonpolar molecular clusters: $(CO_2)_{13}$, $(CH_4)_{13}$, $(C_6H_6)_{13}$ and $(C_6H_{12})_{13}$. All these clusters are bound by weak dispersion interactions, thus their PES are very flat. Their GMs are of S_6, C_3, C_3 and C_1 symmetry, respectively, having nutshell structures ("1+12", see Table 4 and Figure 5). Our results for $(CO_2)_{13}$ and $(C_6H_6)_{13}$ are very similar to those of previous studies[61,62]. The orientation of these molecules in large clusters shows a strong correlation with the one in bulk substance[61].

As an illustration, we also "solvated" a sodium and a chloride ion in 12 benzene molecules, respectively. The GMs (see Table 4 and Figure 6) imply that Na^+ can form a special solvation shell in a similar way as it does in water. In contrast, the Cl^- is pushed out of the cluster of benzene molecules. This is explained by the fact that the anion repulses the π electron cloud of benzene. Thus our algorithm can be used in the study of cation- and anion-aromatic interactions, especially the latter which is currently much less developed.
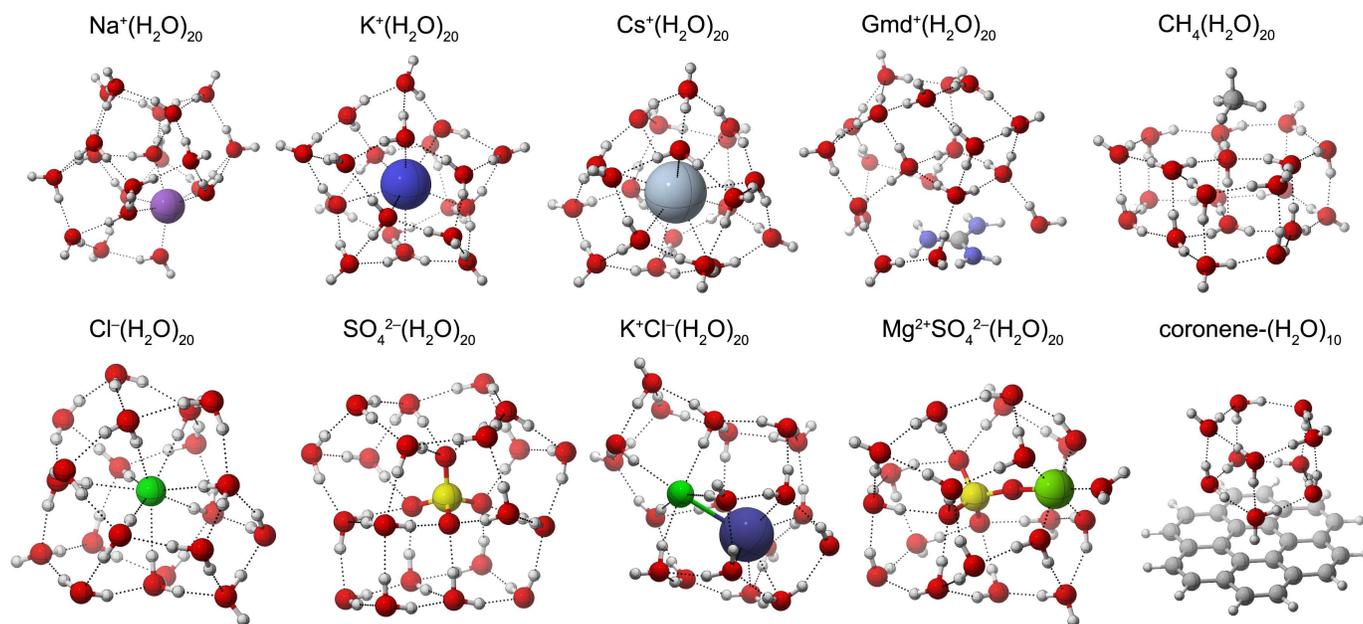
Na$^+$(H$_2$O)$_{20}$ K$^+$(H$_2$O)$_{20}$ Cs$^+$(H$_2$O)$_{20}$ Gmd$^+$(H$_2$O)$_{20}$ CH$_4$(H$_2$O)$_{20}$

Cl$^-$(H$_2$O)$_{20}$ SO$_4$$^{2-}$(H$_2$O)$_{20}$ K$^+$Cl$^-$(H$_2$O)$_{20}$ Mg^{2+}SO$_4$$^{2-}$(H$_2$O)$_{20}$ coronene-(H$_2$O)$_{10}$

Fig. 3 The GMs of the microhydration clusters considered in this work.

(CH$_3$OH)$_{13}$(q1D) (CH$_3$OH)$_{13}$(6+7) Na$^+$(CH$_3$OH)$_{12}$ Cl$^-$(CH$_3$OH)$_{12}$ C$_6$H$_6$(CH$_3$OH)$_{12}$
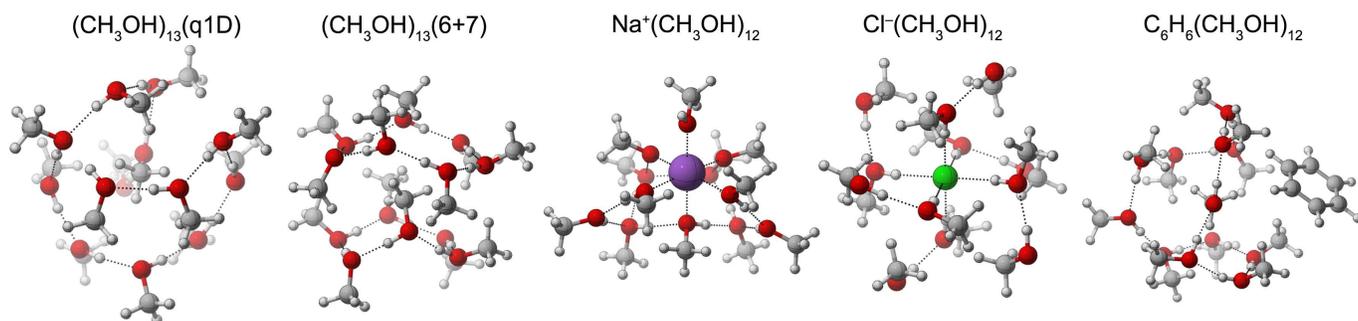
Fig. 4 The GMs of the methanol microsolvation clusters considered in this work.

(CO$_2$)$_{13}$ (CH$_4$)$_{13}$ (C$_6$H$_6$)$_{13}$ (C$_6$H$_{12}$)$_{13}$
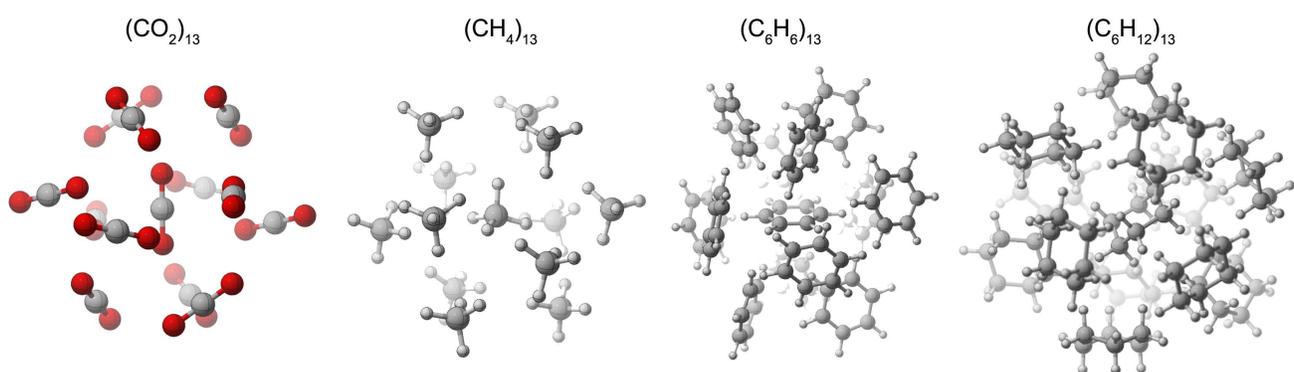
Fig. 5 The GMs of the nonpolar molecular clusters considered in this work.

Table 4 The global optimization of some nonpolar molecular clusters (Energy unit: kJ mol^{-1})

ABC algorithm parameters: $SN = 60$, $g_{limit} = 4$, $g_{max} = 5000$		
Initial guess: random		
System	Step^a	Energy
$(CO_2)_{13}$	1	-160.9737
$(CH_4)_{13}$	1	-66.1720
$(C_6H_6)_{13}$	23	-312.0627
$(C_6H_{12})_{13}$	43	-358.7142
$Na^+(C_6H_{12})_{12}$	32	-452.9565
$Cl^-(C_6H_6)_{12}$	31	-314.4801

^{a} "Step" is the step at which the final energy is obtained.

$Na^+(C_6H_6)_{12}$ $Cl^-(C_6H_6)_{12}$
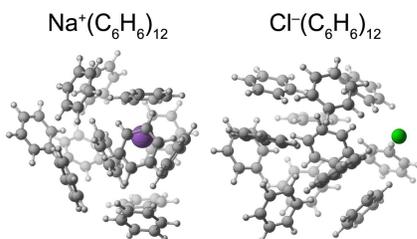
Fig. 6 The GMs of $Na^+(C_6H_{12})_{12}$ and $Cl^-(C_6H_{12})_{12}$.

4 Conclusion

In this work, we have successfully extended the artificial bee colony algorithm from the global optimization of atomic clusters to the one of clusters formed by rigid molecules. Of course, some sophisticated force fields U like AMOEBA[63] or even semiempirical quantum chemical methods can be combined with ABCluster for a finer search. When this is too time-consuming, a better strategy is to use ABCluster and a simple U to get a set of LMs and then use higher-level methods to compute their energies and other properties to select one or more most useful clusters. Through the illustrative applications described above, and due to the black-box nature and efficiency of ABCluster, we believe that it will be a useful tool in the study of inorganic, organic and biological molecular clusters, microsolvation of ions and molecules in polar or nonpolar solvents, and other systems where the molecular internal DOFs are not important. At the present stage there are two possible directions of improvement. One is to introduce ideas from biased optimization algorithms to increase the search efficiency especially for large clusters. Another one is to modify the current implementation to enable also the treatment of the internal DOFs of the molecular constituents, i.e. ABCluster could be applied to, e.g., polymer and large biomolecular clusters. The latest version of ABCluster can be obtained from our group site and will be in continuous development.

5 Appendix

5.1 Explicit Expressions of the Coordinates

In ABCluster, each rigid molecule has a set of atomic coordinates in a pre-defined body-fixed coordinate system. Their final positions in a cluster are defined in a space-fixed coordinate system. For an atom, the two sets of coordinates $\{x_{body}, y_{body}, z_{body}\}$ and $\{x_{space}, y_{space}, z_{space}\}$ are related by the external DOFs $\mathbf{q} \equiv$

$\{X, Y, Z, \alpha, \beta, \gamma\}$ as

$$\begin{pmatrix} x_{space} \\ y_{space} \\ z_{space} \end{pmatrix} = \mathbf{R}_z(\gamma)\mathbf{R}_y(\beta)\mathbf{R}_z(\alpha) \begin{pmatrix} x_{body} \\ y_{body} \\ z_{body} \end{pmatrix} + \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \quad (8)$$

where

$$\mathbf{R}_z(\theta) \equiv \begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (9)$$

$$\mathbf{R}_y(\theta) \equiv \begin{pmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{pmatrix} \quad (10)$$

This is shown in Figure 1.

5.2 The Computation of the Potential Energy Function

To compute the potential energy (3), one can first translate the coordinates from the body-fixed coordinate system to the space-fixed one through (8), then the computation becomes straightforward. The parameters σ and ε are obtained by the Lorentz–Berthelot rules:

$$\sigma_{ij} = \frac{\sigma_{ii} + \sigma_{jj}}{2} \quad (11)$$

$$\varepsilon_{ij} = \sqrt{\varepsilon_{ii}\varepsilon_{jj}} \quad (12)$$

To perform a local minimization (2), one needs to compute its derivative with respect to $\mathbf{q}$'s. Realizing that $r_{i_I j_J}$ in (3) is a function of the external DOFs $\mathbf{q}_I$ and $\mathbf{q}_J$, one can use the chain rule, e.g.

$$\frac{\partial U}{\partial \alpha_I} = \sum_{J=1, J \neq I}^{N} \sum_{i_I \in I} \sum_{j_J \in J} \frac{\partial u}{\partial r_{i_I j_J}} \left(\frac{\partial r_{i_I j_J}}{\partial x_{i_I}} \frac{\partial x_{i_I}}{\partial \alpha_I} + \frac{\partial r_{i_I j_J}}{\partial y_{i_I}} \frac{\partial y_{i_I}}{\partial \alpha_I} + \frac{\partial r_{i_I j_J}}{\partial z_{i_I}} \frac{\partial z_{i_I}}{\partial \alpha_I} \right) \quad (13)$$

where

$$u(r_{i_I j_J}) = \frac{e^2}{4\pi\varepsilon_0} \frac{q_{i_I} q_{j_J}}{r_{i_I j_J}} + 4\varepsilon_{i_I j_J} \left(\left(\frac{\sigma_{i_I j_J}}{r_{i_I j_J}} \right)^{12} - \left(\frac{\sigma_{i_I j_J}}{r_{i_I j_J}} \right)^6 \right) \quad (14)$$

For X, Y, Z, β and γ the expressions are similar.

References

1 *Molecular Clusters: A Bridge to Solid-State Chemistry*, ed. T. Fehlner, J.-F. Halet and J.-Y. Saillard, Cambridge University Press, Cambridge, UK, 2007, p. 257.

2 J. T. O'Brien, J. S. Prell, M. F. Bush and E. R. Williams, *J. Am. Chem. Soc.*, 2010, **132**, 8248–8249.

3 T. Shibata, B. A. Bunker, Z. Zhang, D. Meisel, C. F. Vardeman, and J. D. Gezelter, *J. Am. Chem. Soc.*, 2002, **124**, 11989–11996.

4 V. S. Bryantsev, M. S. Diallo and W. A. Goddard, *J. Phys. Chem. B*, 2008, **112**, 9709–9719.

5 J. Zhang, N. Heinz and M. Dolg, *Inorg. Chem.*, 2014, **53**, 7700–7708.

6 N. Heinz, J. Zhang and M. Dolg, *J. Chem. Theory Comput.*, 2014, **10**, 5593–5598.

7 D. J. Wales and J. P. K. Doye, *J. Phys. Chem. A*, 1997, **101**,

5111–5116.

8   M. T. Oakley, R. L. Johnston and D. J. Wales, *Phys. Chem. Chem. Phys.*, 2013, **15**, 3965–3976.

9   D. J. Wales and M. P. Hodges, *Chem. Phys. Lett.*, 1998, **286**, 65–72.

10 H. Takeuchi, *J. Chem. Infor. Model.*, 2008, **48**, 2226–2233.

11 L. Wille, *Chem. Phys. Lett.*, 1987, **133**, 405–410.

12 Z. Li and H. A. Scheraga, *J. Mol. Struct. (Theochem)*, 1988, **179**, 333–352.

13 R. Storn and K. Price, *Differential Evolution – A Simple and Efficient Adaptive Scheme for Global Optimization over Continuous Spaces*, Technical report, International Computer Science Institute, Berkley, 1995.

14 J. Kennedy and R. Eberhart, *Proc. IEEE Int. Conf.*, 1995, **4**, 1942–1948.

15 D. J. Wales and H. A. Scheraga, *Science*, 1999, **285**, 1368–1372.

16 R. L. Johnston, *Dalton Trans.*, 2003, 4193–4207.

17 S. Heiles and R. L. Johnston, *Int. J. Quantum Chem.*, 2013, **113**, 2091–2109.

18 J. Zhang and M. Dolg, *Phys. Chem. Chem. Phys.*, 2015, **17**, 24173–24181.

19 D. J. Wales, *Phil. Trans. R. Soc. A*, 2005, **363**, 357–377.

20 T. James, D. J. Wales and J. Hernández-Rojas, *Chem. Phys. Lett.*, 2005, **415**, 302 – 307.

21 T. Erber and G. M. Hockney, *Phys. Rev. Lett.*, 1995, **74**, 1482–1482.

22 J. P. K. Doye and D. J. Wales, *Phys. Rev. Lett.*, 1998, **80**, 1357–1360.

23 Z. Li and H. Scheraga, *Proc. Natl. Acad. Sci. USA*, 1987, **84**, 6611–6615.

24 J. A. D. MacKerell, D. Bashford, M. Bellott, J. R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin and M. Karplus, *J. Phys. Chem. B*, 1998, **102**, 3586–3616.

25 W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, *J. Am. Chem. Soc.*, 1996, **118**, 11225–11236.

26 W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, *J. Am. Chem. Soc.*, 1995, **117**, 5179–5197.

27 D. Karaboga, *An Idea Based on Honey Bee Swarm for Numerical Optimization*, Technical Report TR06, Erciyes University, 2005.

28 D. Karaboga and B. Akay, *Appl. Math. Comput.*, 2009, **214**, 108 – 132.

29 B. Verma and D. Kumar, *Int. J. Eng. Tech.*, 2013, **2**, 175–186.

30 D. Karaboga, B. Gorkemli, C. Ozturk and N. Karaboga, *Artif. Intell. Rev.*, 2014, **42**, 21–57.

31 D. Liu and J. Nocedal, *Math. Program.*, 1989, **45**, 503–528.

32 H.-Y. Fan and J. Lampinen, *J. Global Optim.*, 2003, **27**, 105–129.

33 *http://mackerell.umaryland.edu/charmm_ff.shtml.   Access time: Jun. 26, 2015*.

34 C. Y. Legault, *CYLview, 1.0b*, 2009, see http://www.cylview.org.

35 R. Ludwig, *Angew. Chem. Int. Ed.*, 2001, **40**, 1808–1827.

36 L. Pauling, *J. Am. Chem. Soc.*, 1935, **57**, 2680–2684.

37 S. Kazachenko, S. Bulusu and A. J. Thakkar, *J. Chem. Phys.*, 2013, **138**, 224303.

38 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926–935.

39 F. Schulz and B. Hartke, *ChemPhysChem*, 2002, **3**, 98–106.

40 B. S. González, J. Hernández-Rojas and D. J. Wales, *Chem. Phys. Lett.*, 2005, **412**, 23 – 28.

41 C. N. Rowley and B. Roux, *Journal of Chemical Theory and Computation*, 2012, **8**, 3526–3535.

42 P. E. Mason, G. W. Neilson, C. E. Dempsey, A. C. Barnes and J. M. Cruickshank, *Proc. Natl. Acad. Sci. U.S.A.*, 2003, **100**, 4557–4561.

43 L. C. Smeeton, J. D. Farrell, M. T. Oakley, D. J. Wales and R. L. Johnston, *J. Theory Comput. Chem.*, 2015, **11**, 2377–2384.

44 K. J. Tielrooij, N. Garcia-Araez, M. Bonn and H. J. Bakker, *Science*, 2010, **328**, 1006–1009.

45 J. Hernández-Rojas, J. Bretón, J. M. G. Llorente and D. J. Wales, *J. Phys. Chem. B*, 2006, **110**, 13357–13362.

46 B. S. González, J. Hernández-Rojas, J. Bretón and J. M. G. Llorente, *J. Phys. Chem. C*, 2007, **111**, 14862–14869.

47 B. Hartke, *J. Chem. Phys.*, 2009, **130**, 024905.

48 K. Doi, E. Togano, S. S. Xantheas, R. Nakanishi, T. Nagata, T. Ebata and Y. Inokuchi, *Angew. Chem. Int. Ed.*, 2013, **52**, 4380–4383.

49 P. Costa and W. Sander, *Angew. Chem. Int. Ed.*, 2014, **53**, 5122–5125.

50 K. J. Tauer and W. N. Lipscomb, *Acta Crystallogr.*, 1952, **5**, 606–612.

51 B. Torrie, S.-X. Weng and B. Powell, *Mol. Phys.*, 1989, **67**, 575–581.

52 I. Bakò, A. Bencsura, K. Hermannson, S. Bálint, T. Grósz, V. Chihaia and J. Oláh, *Phys. Chem. Chem. Phys.*, 2013, **15**, 15163–15171.

53 P. J. Stephens, F. J. Devlin, C. F. Chabalowski and M. J. Frisch, *J. Phys. Chem.*, 1994, **98**, 11623–11627.

54 F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3297–3305.

55 S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.

56 F. Neese, *Wiley Interdisc. Rev.: Comput. Mol. Sci.*, 2012, **2**, 73–78.

57 R. B. Best, N.-V. Buchete and G. Hummer, *Biophys. J.*, 2008, **95**, L07–L09.

58 J. Llanio-Trujillo, J. Marques and F. Pereira, *Comput. Theor. Chem.*, 2013, **1021**, 124 – 134.

59 O. M. Cabarcos, C. J. Weinheimer, T. J. Martínez and J. M. Lisy, *J. Chem. Phys.*, 1999, **110**, 9516–9526.

60 G. Matisz, A.-M. Kelterer, W. M. F. Fabian and S. Kunsági-Máté, *J. Phys. Chem. A*, 2011, **115**, 10556–10564.

61 H. Takeuchi, *J. Phys. Chem. A*, 2008, **112**, 7492–7497.

62 H. Takeuchi, *J. Phys. Chem. A*, 2012, **116**, 10172–10181.

63 P. Ren and J. W. Ponder, *J. Phys. Chem. B*, 2003, **107**, 5933–5947.