

# Monte Carlo-Simulated Annealing and Machine Learning-Based Funneled Approach for Finding the Global Minimum Structure of Molecular Clusters

Michal Roth, Yoni Toker,\* and Dan T. Major\*



Cite This: *ACS Omega* 2024, 9, 1298–1309



Read Online

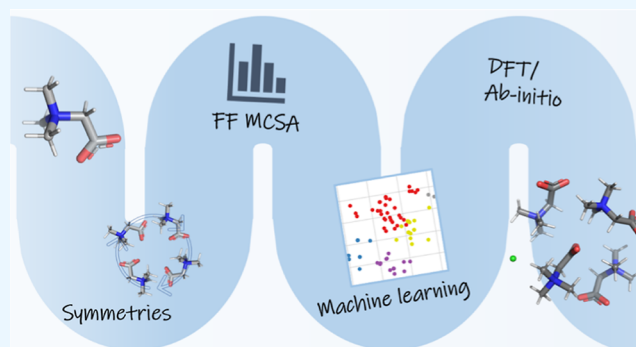
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

**ABSTRACT:** Understanding the physical underpinnings and geometry of molecular clusters is of great importance in many fields, ranging from studying the beginning of the universe to the formation of atmospheric particles. To this end, several approaches have been suggested, yet identifying the most stable cluster geometry (i.e., global potential energy minimum) remains a challenge, especially for highly symmetric clusters. Here, we suggest a new funneled Monte Carlo-based simulated annealing (SA) approach, which includes two key steps: generation of symmetrical clusters and classification of the clusters according to their geometry using machine learning (MCSA-ML). We demonstrate the merits of the MCSA-ML method in comparison to other approaches on several Lennard-Jones (LJ) clusters and four molecular clusters— $\text{Ser}_8(\text{Cl}^-)_2$ ,  $\text{H}^+(\text{H}_2\text{O})_6$ ,  $\text{Ag}^+(\text{CO}_2)_8$ , and  $\text{Bet}_4\text{Cl}^-$ . For the latter of these clusters, the correct structure is unknown, and hence, we compare the experimental and simulated fragmentation patterns, and the fragmentation of the proposed global minimum matches experiments closely. Additionally, based on the fragmentation of the predicted betaine cluster, we were able to identify hitherto unknown neutral fragmentation channels. In comparison to results obtained with other methods, we demonstrated a superior ability of MCSA-ML to predict clusters with high symmetry and similar abilities to predict clusters with asymmetrical structures.



## INTRODUCTION

Molecular clusters in the gas phase are a fascinating medium for investigating basic scientific ideas. For example, in action spectroscopy, molecular clusters are used as a means for applying perturbations to chromophores. An important example is the Zwitterion tag action spectroscopy (ZITA) technique which uses complexation with betaine, a molecule which has an extremely large dipole moment,<sup>1</sup> as a means of studying the shift in the photoabsorption spectrum resulting from an external electric field.<sup>2,3</sup> Another example is amino acid clusters, which have been extensively studied for analytical purposes, as well as in the context of origin of life theories.<sup>4–6</sup> The arguably most studied amino acid clusters are those of Serine, which have a prominent magic number at size  $N = 8$ ,<sup>7</sup> whose geometry has only recently been determined.<sup>5,6</sup> Theoretical studies of molecular clusters begin with identifying their geometry. In many cases, this is a difficult task due to the large number of configurations, i.e., local minima on the cluster's potential energy surface (PES). Moreover, in the gas phase, in contrast to the solid state, the clusters are not limited by any known symmetry groups or sizes. Many different algorithms have been used to tackle the problem of clusters in the gas phase.<sup>8–17</sup> A well-known algorithm for periodic systems is the ab initio

random structure searching (AIRSS).<sup>16</sup> The algorithm generates random “sensible” structures and then relaxes them into the closest minimum point using periodic DFT, possibly using symmetries to help to narrow down the search space. To the best of our knowledge, AIRSS has been employed mainly for molecular and ionic crystals and atomic clusters.<sup>16,18–22</sup> We note that the concept of employing symmetry could be helpful<sup>23</sup> for complex molecular clusters in the gas phase and we will return to this idea below. Other examples of algorithms for gas-phase clusters are the ABCluster<sup>24</sup> and NWPEsSe<sup>25</sup> which use the artificial bee colony method for searching stable clusters (i.e., global and local minima on the PES), OGOLEM,<sup>26</sup> AUTOMATON,<sup>27</sup> and GAFit<sup>28</sup> which use evolutionary algorithms to search for stable clusters and GMIN<sup>29</sup> that uses the basin-hopping algorithm<sup>30</sup> (among others) and employs symmetry.<sup>31</sup>

**Received:** October 1, 2023

**Revised:** November 9, 2023

**Accepted:** November 15, 2023

**Published:** December 19, 2023



Most gas-phase cluster algorithms usually contain two main steps: the first involves identifying local minima on the PES using low-cost energy calculations. There are a variety of methods for minima search on the PES, including genetic algorithms (GA),<sup>8–13</sup> Monte Carlo simulations (MC),<sup>8,9</sup> and molecular dynamics simulations (MD).<sup>8,9</sup> The second step is usually a geometry optimization of the low-lying isomers found on the PES using higher level scoring methods, such as quantum chemistry calculations. In these two steps, the algorithms attempt to identify the global minimum at a minimal computational cost. However, the local minima and their relative energy ranking may differ using low-cost and high-cost methods. Hence, it is often necessary to include many minima in high-level calculations in search of the global minimum. Moreover, symmetrical structures have a low probability of being generated randomly due to their low entropy, and such structures can easily be missed using the standard algorithms.

In this work, we propose a method for finding the global minimum of molecular clusters in the gas phase, which attempts to overcome the above-mentioned problems (i.e., the need for a large number of expensive quantum chemistry calculations and the difficulty in identifying high symmetry clusters). The method is based on a combination of MC simulated annealing (MCSA) simulations, machine learning (ML), force field (FF) calculations, and density functional theory (DFT) or ab initio calculations. We demonstrate the merits of the MCSA-ML method and compare the results to the ABCluster results on several Lennard-Jones (LJ) clusters and four molecular clusters—Ser<sub>8</sub>(Cl<sup>-</sup>)<sub>2</sub>, H<sup>+</sup>(H<sub>2</sub>O)<sub>6</sub>, Ag<sup>+</sup>(CO<sub>2</sub>)<sub>8</sub>, and Bet<sub>4</sub>Cl<sup>-</sup>. For the latter of these clusters, the correct structure is unknown and hence we compare the experimental and simulated fragmentation patterns.

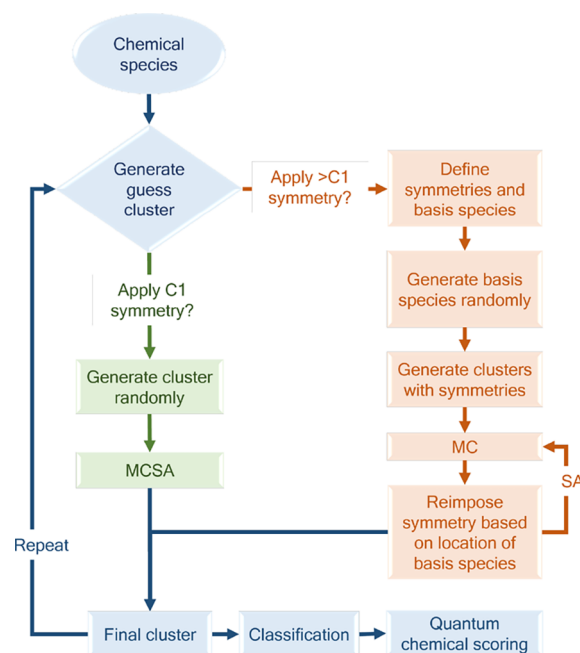
## METHODOLOGY

The goal of the current approach is to identify the global minimum of molecular clusters consisting of  $N$  chemical species (i.e., molecules and ions). The method is based on generating random clusters with and without symmetry in conjunction with MCSA to find local minima. During this stage, scoring is performed using FF calculations. The final stage is to optimize the geometries using quantum chemistry calculations (e.g., DFT). To minimize the number of quantum chemistry calculations needed, we use machine learning to group the different structures identified with MCSA and classify them into groups with similar geometries. This is achieved by machine learning employing a carefully crafted set of descriptors and classification by the  $K$ -means algorithm. The workflow is illustrated in Figure 1.

**Random Cluster Generation.** The structures of the individual molecules (herein, betaine, serine, H<sub>2</sub>O, or CO<sub>2</sub>) are geometry optimized. Subsequently, each chemical species is randomly and rigidly translated and rotated in space within a predefined volume.

**Symmetrical Cluster Generation.** Clusters with high symmetry have low entropy, so the probability of randomly identifying such molecular clusters is low. Here, we define a set of symmetry operators to generate random molecular clusters with predefined symmetries. In generating clusters with symmetry operations, the chemical species which are generated randomly will be referred to as the “basis species”.

We define the symmetries according to the algorithm described in Figure 2.

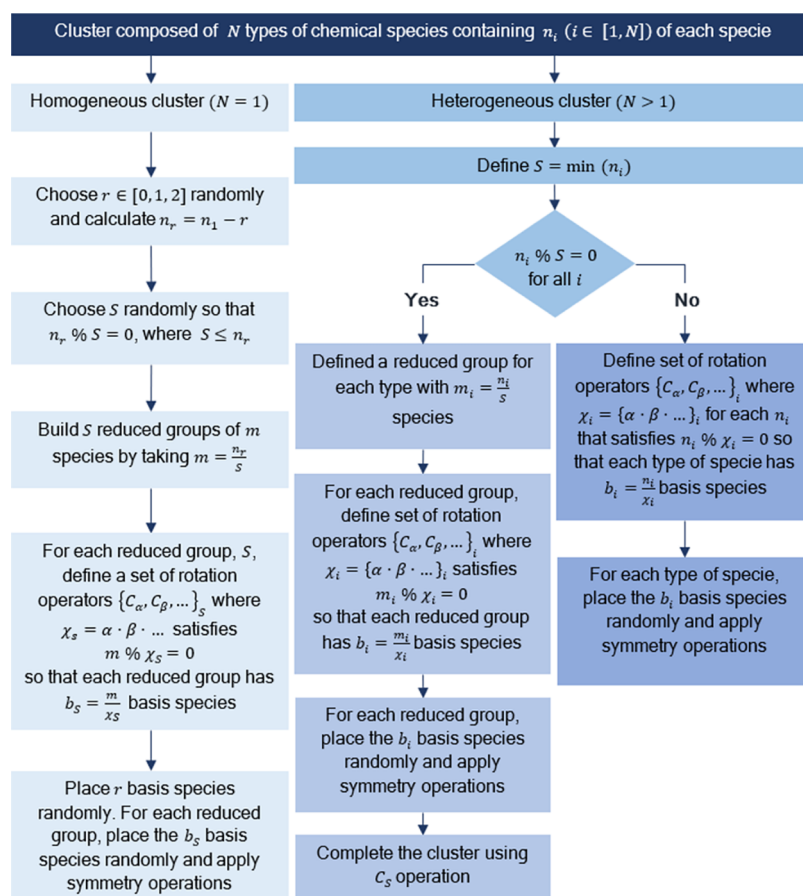


**Figure 1.** Workflow of the methodology used in the present work. “Chemical species” or “species” indicate molecules and ions.

For instance, for Ser<sub>8</sub>(Cl<sup>-</sup>)<sub>2</sub>, which is a heterogeneous cluster (i.e.,  $N = 2$ ,  $n_1 = 8$  for Ser and  $n_2 = 2$  for Cl<sup>-</sup>). We define  $S = \min(n_i) = 2$  ( $i = 1, 2$ ). Since  $n_1 \% 2 = 0$  and  $n_2 \% 2 = 0$ , we define a reduced group for each type of chemical specie,  $i$ , with  $m_i = \frac{n_i}{S}$ . I.e.,  $m_1 = \frac{8}{2} = 4$  (Ser) and  $m_2 = \frac{2}{2} = 1$  (Cl<sup>-</sup>). For each reduced group, we define a set of rotation operators. For Ser, we can define the following four sets of operations:  $\{C_1\}$ ,  $\{C_2\}$ ,  $\{C_4\}$ , and  $\{C_2, C_2\}$ . For the Cl<sup>-</sup> ions, we can define only a  $\{C_1\}$  operation. Assuming we randomly choose  $C_4$  for Ser and  $C_1$  for Cl<sup>-</sup>, we have the following basis species  $b_1 = \frac{m_1}{4} = 1$  (Ser) and  $b_2 = \frac{m_2}{1} = 1$  (Cl<sup>-</sup>). We then randomly place one Ser molecule and perform three  $C_4$  operations to generate a subcluster of four Ser molecules with symmetry  $C_4$ . Similarly, we randomly place the Cl<sup>-</sup> ion (and perform  $C_1$  symmetry), yielding a Ser<sub>4</sub>Cl<sup>-</sup> cluster. Finally, we perform a  $C_2$  symmetry operation to complete the cluster. The algorithm is entirely general and has been implemented to account for any rotational symmetry. For homogeneous clusters, for each iteration, we randomly reduced  $r \in [0, 1, 2]$  to enable more symmetry options.

**Monte Carlo Simulated Annealing.** For each molecular cluster generated randomly as described in the previous section, we performed MCSA calculations, using the CHARMM36 and CHARMM general FF (CGenFF).<sup>32,33</sup> MCSA is a standard, well-established technique and is widely used to identify the global minimum.<sup>5,34–36</sup> The cooling process during annealing allows the distribution of all configurations slowly, allowing the visit of low energy minima with increasing probability. For high symmetry clusters, at each temperature, the MC simulations were performed without preserving symmetry. At the end of each temperature step, the symmetry was enforced by replacing the nonbasis species with the basis species using the predefined symmetry operators.

**Machine Learning.** Prior to performing quantum chemistry calculations on the many molecular clusters obtained from MCSA, we statistically clustered all of the MCSA molecular



**Figure 2.** Algorithm defining symmetry operations. “Chemical species” or “species” indicate molecules and ions.

clusters based on structural similarity using the  $K$ -means algorithm, resulting in a small number of statistical clusters (SC). Hence, in subsequent quantum chemistry calculations, only a few representative structures from each SC need to be considered.

**Defining Descriptors for Molecular Cluster Characterization.** We tested several descriptors for a range of molecular clusters in an attempt to characterize the geometry of the clusters. As final descriptors, we chose the four that demonstrated the highest variance:

**Relative Ion Position.** This descriptor is a measure of the relative distribution of the same type of chemical species,  $n$ , relative to the center of mass (CM) of the entire cluster, and was therefore employed only for heterogeneous clusters. The relevant expression for  $n$  types of chemical species is

$$\text{CM}_{n,\text{relative}} = |\overrightarrow{\text{CM}}_{\text{cluster}} - \overrightarrow{\text{CM}}_n|, \text{ for all } n \quad (1)$$

Such a descriptor is helpful in detecting phase segregation.

**Dipole Moment.** The dipole moment, which is a measure of the charge distribution in the clusters, was calculated by

$$\mu = |\vec{\mu}| = \left| \sum_{i=1}^{N_{\text{atom}}} q_i \vec{r}_{c,i} \right| \quad (2)$$

where  $q_i$  is the partial charge of atom  $i$  and  $\vec{r}_{c,i}$  is the distance of atom  $i$  from the cluster centroid. The summation is over all atoms in all of the chemical species within the cluster. This descriptor provides information on the charge distribution in the cluster.

**Moment of Inertia.** Moment of inertia is calculated by

$$I = \sum_{i=1}^{N_{\text{atom}}} m_i \vec{r}_{\text{CM},i}^2 \quad (3)$$

where  $m_i$  is the mass of atom  $i$  and  $\vec{r}_{\text{CM},i}$  is the distance of atom  $i$  from the cluster's CM. This descriptor provides a measure of the distribution of the atoms and molecules in space.

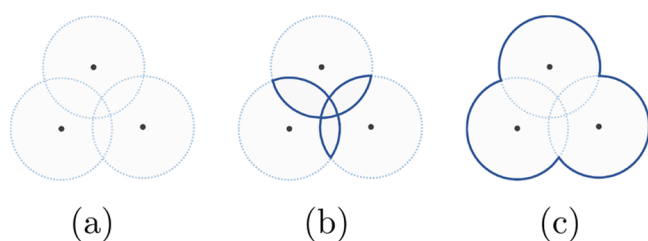
**Surface Area.** Surface area is calculated by

$$A = \sum_{i=1}^{N_{\text{atom}}} (S_i - B_i) \quad (4)$$

where  $A$  is the surface area (i.e., symmetric difference of all surface areas),  $S_i$  is a sphere around atom  $i$  with radius  $r + r_w$ , and  $r$  is the van der Waals radius of atom  $i$  and  $r_w$  equals 1.4 Å.  $B_i$  is the surface area from the intersection  $S_i$  with all the spheres<sup>37</sup> (Figure 3). This descriptor provides a measure of the shape of the atoms and molecules in space.

**Classification by  $K$ -Means.** The classification of the clusters is performed using the  $K$ -means algorithm<sup>38,39</sup> with the four descriptors described above as principal components.  $K$ -means classifies the molecular clusters to SCs according to the similarity among the numerical values of the descriptors. We also compared  $K$ -means with Spectral clustering and found that these methods gave similar results (see Figure S1 and accompanying discussion).

**Quantum Chemical Scoring.** To provide a more accurate ranking of the energy of the molecular clusters, we rescored the



**Figure 3.** (a) Spheres around atoms ( $S$ ). (b) Intersection between the spheres (bold lines) ( $B$ ). (c) Total surface area (bold lines) ( $A$ ), i.e., symmetric difference.

cluster energy using DFT or ab initio methods of the three most stable clusters from each SC obtained using the CHARMM FF.

The code scripts for the method were written in the CHARMM<sup>40</sup> and Python scripting languages. The cluster code is available upon request.

## COMPUTATIONAL DETAILS

During the MCSA simulations, the system is cooled from  $T_i$  to  $T_f$  with a cooling rate  $k$ . The temperature at each step  $s$  is

$$T(s) = T_i e^{-k \cdot s} \quad (5)$$

At each temperature,  $N_{steps}$  MC steps are performed, and the process is repeated for  $I$  iterations. All parameters are easily modified by the user, and the parameters should be customized for each system.<sup>34</sup>

The exploration of the simulated annealing parameters for various clusters followed the method outlined in ref 35. However, for the LJ clusters, we adopted the optimal parameters from ref 35 without conducting further exploration. All of the parameters used for the clusters in this paper are shown in Table 1.

**Table 1. All of the Parameters We Used for the Different Clusters**

| cluster                            | $T_i$ | $T_f$  | $k$                   | $N_{steps}$ | $I$  |
|------------------------------------|-------|--------|-----------------------|-------------|------|
| LJ6                                | 0.15  | 0.002  | $1.00 \times 10^{-2}$ | 10,000      | 20   |
| LJ7                                | 0.25  | 0.0354 | $1.00 \times 10^{-2}$ | 10,000      | 20   |
| LJ9                                | 0.22  | 0.0395 | $1.00 \times 10^{-2}$ | 10,000      | 25   |
| LJ10                               | 0.26  | 0.0869 | $1.00 \times 10^{-2}$ | 10,000      | 25   |
| LJ13                               | 0.31  | 0.0867 | $1.00 \times 10^{-2}$ | 10,000      | 30   |
| LJ19                               | 0.315 | 0.1299 | $1.00 \times 10^{-3}$ | 10,000      | 100  |
| LJ20                               | 0.38  | 0.1087 | $1.00 \times 10^{-3}$ | 10,000      | 100  |
| LJ23                               | 0.4   | 0.1385 | $1.00 \times 10^{-3}$ | 10,000      | 300  |
| LJ36                               | 0.19  | 0.118  | $1.00 \times 10^{-3}$ | 10,000      | 1000 |
| LJ38                               | 0.58  | 0.23   | $1.00 \times 10^{-6}$ | 10,000      | 2500 |
| $\text{Ser}_8(\text{Cl}^-)_2$      | 400   | 40     | $5.00 \times 10^{-1}$ | 10,000      | 700  |
| $\text{H}^+(\text{H}_2\text{O})_6$ | 100   | 1      | $5.00 \times 10^{-2}$ | 10,000      | 500  |
| $\text{Ag}^+(\text{CO}_2)_8$       | 50    | 0.5    | $5.00 \times 10^{-2}$ | 10,000      | 500  |
| $\text{Bet}_4\text{Cl}^-$          | 300   | 30     | $5.00 \times 10^{-1}$ | 10,000      | 500  |

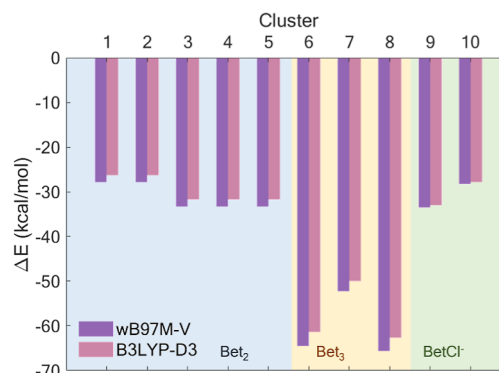
We enabled rigid translations and rotations of the chemical species as well as torsional rotations within the molecules. No cutoff was employed for nonbonded interactions.

For the classification using  $K$ -means, we determined the number of SCs using the  $k$ -elbow method.<sup>39</sup>

To choose a DFT method for this study, we compared two methods,  $w\text{B97M-V}/6\text{-31+G}^{**}$ ,<sup>41,42</sup> which is considered to be highly accurate<sup>43</sup> and  $\text{B3LYP-D3(BJ)}/6\text{-31+G}^{**}$ .<sup>42,44–47</sup> We

note that ref 43 contains a thorough comparison between a range of families of DFT methods and concluded that both  $w\text{B97M-V}$  and  $\text{B3LYP-D3(BJ)}$  are good choices for noncovalent interactions.

In the case of betaine molecules, the benchmark calculations were performed on ten different molecular clusters ( $\text{Bet}_N$  with  $N = 2, 3$  and  $\text{Bet}_N\text{Cl}^-$  with  $N = 1$ ) which were constructed manually to include a variety of bonding features between the molecules. The results from  $\text{B3LYP-D3(BJ)}/6\text{-31+G}^{**}$  are very similar to those from  $w\text{B97M-V}/6\text{-31+G}^{**}$  (Figure 4) and allow



**Figure 4.** Energy comparison of ten clusters using  $w\text{B97M-V}$  and  $\text{B3LYP-D3(BJ)}$  with the  $6\text{-31+G}^{**}$  basis set.  $\Delta E$  is the energy difference between the complete cluster and the individual chemical species (complexation energy).

one to distinguish between the different types of molecular clusters at a lower cost. Hence, both methods are expected to allow meaningful distinctions between clusters.

For LJ clusters, we optimized the clusters using  $\text{MP2}/\text{Aug-cc-pVTZ}^{48}$  and performed single-point energy scoring using  $\text{DLPNO-CCSD(T)}/\text{Aug-cc-pVTZ}$ .<sup>48,49</sup>

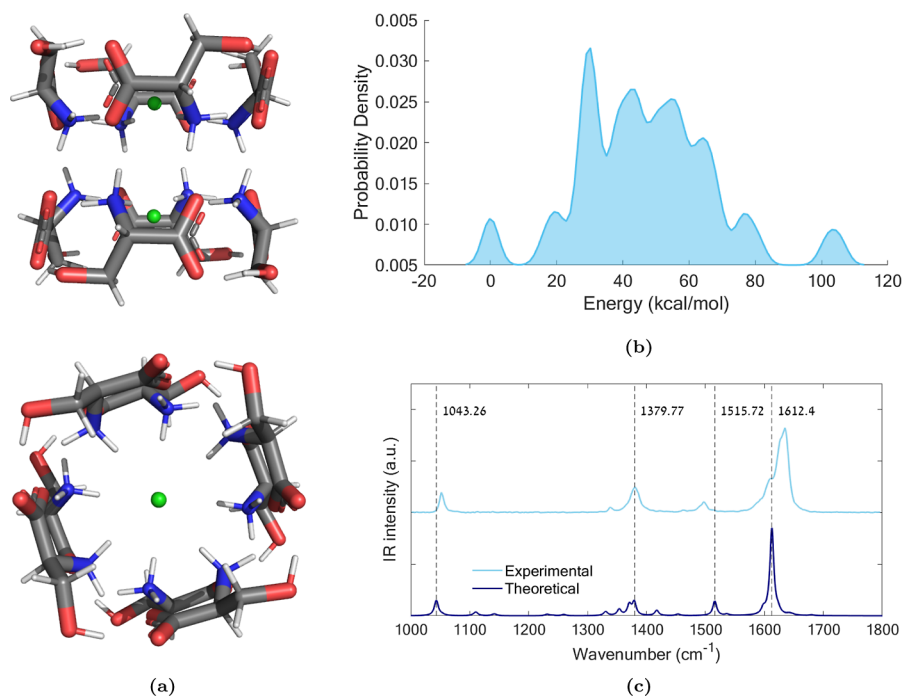
For the  $w\text{B97M-V}/6\text{-31+G}^{**}$  calculations, we used the Q-Chem program.<sup>50</sup> All other quantum chemistry calculations were performed using the Gaussian 16 program,<sup>51</sup> with the exception of  $\text{DLPNO-CCSD(T)}$  calculations which used the ORCA program.<sup>52</sup>

To summarize, for all molecular clusters, we employed the DFT level of theory ( $\text{B3LYP-D3(BJ)}$  and  $w\text{B97M-V}/6\text{-31+G}^{**}$ ) as a compromise between computational cost and accuracy. However, for the He-clusters, which are dominated by dispersion interactions, we employed coupled cluster theory.

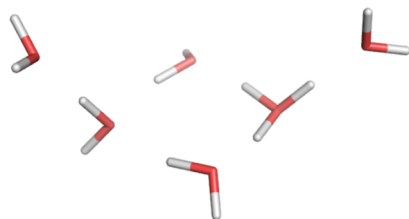
## RESULTS AND DISCUSSION

This work commenced with a set of test cases that have been studied previously and, hence, the optimal cluster structure is known. This allowed us to verify that our MCSA-ML approach can identify the lowest energy cluster and score it correctly. Subsequently, we applied our method to a cluster with an unknown structure (i.e., the  $\text{Bet}_4\text{Cl}^-$  clusters). The results obtained using our MCSA-ML approach were compared with ABCluster.<sup>24</sup> For clusters with electronic energy similar to the global minimum (GM) (i.e.,  $\text{H}^+(\text{H}_2\text{O})_6$ ,  $\text{Ag}^+(\text{CO}_2)_8$  and  $\text{Bet}_4\text{Cl}^-$ ), thermal corrections were computed, and the corresponding free energy are presented as Supporting Information, Tables S5–S7.

**Test Cases. Lennard-Jones Clusters.** As a basic toy system, we chose LJ clusters that are dominated by dispersion interactions. These clusters have been extensively studied, as



**Figure 5.** (a)  $\text{Ser}_8(\text{Cl}^-)_2$  most stable cluster (side-view on the top and top-view on the bottom). (b) Probability density, i.e., the likelihood of identifying a molecular cluster with a given energy based on the current MCSA-ML-based funneled approach. We set the energy of the most stable structure to 0 kcal/mol. (c) Comparison between the experimental IR spectrum (top) and the theoretical spectrum (bottom) of the  $\text{Ser}_8(\text{Cl}^-)_2$  cluster. The computed spectrum used  $B3LYP-D3(BJ)/6-31+G^{**}$ .



**Figure 6.** Lowest lying isomer of  $\text{H}^+(\text{H}_2\text{O})_6$  found in this work, and suggested in ref 64.

pure homoelement LJ clusters or as heteroelement clusters using different methods.<sup>30,53,54</sup> To simulate an LJ cluster, we used He atoms. We calculated the geometry of clusters of size 6, 7, 9, 10, 13, 19, 20, 23, 36, and 38. Initial scoring was performed using the standard LJ potential in CHARMM while high-level scoring was performed using  $\text{DLPNO-CCSD}(T)/\text{Aug-cc-pVTZ}/\text{MP2}/\text{Aug-cc-pVTZ}$ .

For all clusters, we obtained the GM known in the literature.<sup>55</sup> The clusters obtained using ABCluster were the known global minimum structures as well for all cluster sizes.

$\text{Ser}_8(\text{Cl}^-)_2$ . Serine clusters are widely studied due to their propensities to form clusters of size 8 and for their strong chiral preference.<sup>7</sup> The structure of the  $\text{Ser}_8(\text{Cl}^-)_2$  cluster was only recently identified and shown to possess a barrel-like structure.<sup>6</sup> This structure can be described as two layers, each composed of four serine molecules with the chloride ion at the center. In each ring, the molecules interact via hydrogen bonds between the OH and the neighboring  $\text{NH}_3^+$  and  $\text{COO}^-$  groups of the adjacent molecules. Additionally, there are hydrogen bond interactions between the molecules in the upper and lower rings, specifically between the  $\text{NH}_3^+$  groups and the  $\text{COO}^-$  groups of the neighboring molecules.

For this cluster, we used  $B3LYP-D3(BJ)/6-31+G^{**}$ , followed by rescoring of clusters within 2 kcal/mol of the global minimum using  $wB97M-V/6-31+G^{**}$ . The most stable structure obtained is shown in Figure 5a. This structure was not observed when using MCSA without symmetry, emphasizing the importance of using symmetry. The second most stable molecular cluster was less stable by  $\Delta E = 16.6$  kcal/mol according to the DFT  $B3LYP-D3(BJ)$  calculations (Figure 5b), so we can assume that the geometry described above (Figure 5a) is the dominant one.




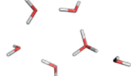
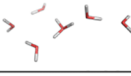



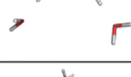
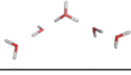


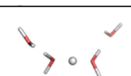
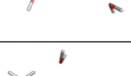
Note that the algorithm's exploration of space is not uniform, as it randomly switches between the formation of symmetric and asymmetric clusters, resulting in a biased probability distribution. Therefore, Figure 5b does not represent the true Boltzmann probability density. The key insight here is that a single structure is identified as the global minimum and that this structure is expected to be the dominant one.

The IR spectrum calculated theoretically for the minimum energy geometry (Figure 5a) is shown in Figure 5c and is compared to the experimentally measured IR absorption spectrum.<sup>6</sup> The similarity between the spectra indicates that the lowest energy molecular cluster is indeed the one observed experimentally, and the geometry we found is very similar to the one proposed by Seo et al.<sup>6</sup>

For this cluster, we were unable to identify the correct global minimum using ABCluster. In fact, most clusters fragment into smaller clusters.

$\text{H}^+(\text{H}_2\text{O})_6$ . Protonated water clusters in the gas phase have been extensively studied<sup>56–63</sup> due to their importance in many fields of science including atmospheric chemistry and physics. In the following, we focus on the hexamer  $\text{H}^+(\text{H}_2\text{O})_6$  cluster, which has many known isomers<sup>58–62,64</sup> and there is some debate over their ranking and the identity of the global minimum. Figure 6 shows a boat-like structure. Its “seat” is symmetrically

**Table 2. Main Water Clusters Obtained and Their Energy Ranking at the *B3LYP-D3* and *wB97M-V* Levels Using the Current Algorithm (First Two Columns)<sup>a</sup>**

| Isomer  | $\Delta E$<br><i>B3LYP-D3</i> | $\Delta E$<br><i>wB97M-V</i> | Ref. 58        |                      | Ref. 59  |              | Ref. 60 |            | Ref. 61  |              | Ref. 62 |            | Ref. 64     |             |
|---|-------------------------------|------------------------------|----------------|----------------------|----------|--------------|---------|------------|----------|--------------|---------|------------|-------------|-------------|
|   |                               |                              | S              | $\Delta E$           | S        | $\Delta E$   | S       | $\Delta E$ | S        | $\Delta E$   | S       | $\Delta E$ | S           | R           |
|    | 0.00                          | 0.21                         | -              | -                    | -        | -            | -       | -          | -        | -            | -       | -          | A           | 1           |
|    | 0.02                          | 0.18                         | V              | 0.00                 | T1       | 0.01         | 6R1     | 0.00       | VII      | 1.85         | 6V      | 1.1        | -           | -           |
|    | 0.07                          | 0.00                         | I<br>II<br>III | 0.47<br>0.58<br>0.09 | -        | -            | -       | -          | -        | -            | -       | -          | C           | 3           |
|    | 0.08                          | 0.30                         | VI             | 0.00                 | P1       | 0.00         | 6R6     | 0.42       | V        | 0.6          | 6VII    | 1.5        | B           | 2           |
|    | 0.50                          | 0.51                         | -              | -                    | T4       | 0.14         | -       | -          | -        | -            | 6III    | 1.4        | -           | -           |
|    | 0.60                          | 0.56                         | VII            | 0.31                 | T3       | 0.22         | 6R3     | 0.34       | VIII     | 2.49         | -       | -          | -           | -           |
|    | 0.92                          | 0.87                         | VIII           | 0.56                 | T2       | 0.48         | 6R2     | 0.55       | III      | 0.26         | 6IV     | 1.4        | -           | -           |
|   | 1.68                          | 1.90                         | -              | -                    | -        | -            | -       | -          | -        | -            | 6VI     | 2.5        | -           | -           |
|  | 2.00                          | 1.97                         | -              | -                    | E1       | 1.26         | 5L2     | 1.52       | VI       | 1.17         | 6I      | 1.1        | -           | -           |
|  | 2.05                          | 2.93                         | XI             | 4.01                 | -        | -            | -       | -          | -        | -            | -       | -          | -           | -           |
|  | 2.42                          | 2.54                         | -              | -                    | -        | -            | -       | -          | -        | -            | -       | -          | -           | -           |
|  | 2.53                          | 3.43                         | IX             | 1.75                 | Z1       | 1.83         | 5L1     | 2.19       | I<br>II  | 0.00<br>0.21 | 6II     | 0.0        | G           | 7           |
|  | 2.83                          | 3.40                         | -              | -                    | -        | -            | -       | -          | -        | -            | -       | -          | -           | -           |
|  | 3.40                          | 3.98                         | -              | -                    | -        | -            | -       | -          | -        | -            | -       | -          | -           | -           |
| Missed clusters   | -                             | -                            | X              | 1.39                 | E1<br>C1 | 1.26<br>0.55 | -       | -          | IV<br>IX | 0.44<br>3.26 | -       | -          | D<br>E<br>F | 4<br>5<br>6 |

<sup>a</sup>The isomers are compared to the different references where *S* is the symbol of the isomer, *R* is the energy ranking, and  $\Delta E$  is the energy relative to the global minimum in kcal/mol. Cases where the isomer does not appear in the reference are marked as “—”. The last row is for isomers that appear in the references but were not obtained here. The computational methods are *MP2/aug-cc-pVDZ* in refs 58 and 60, *CCSD(T)/aVQZ* in ref 59, *MSEVB* potential in ref 61, *B3LYP/6-31+G\** in ref 62, and neural network in ref 64.

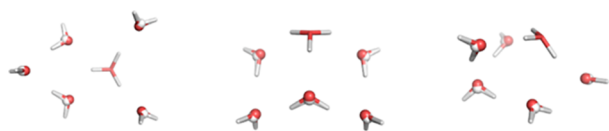


Figure 7.  $\text{H}^+(\text{H}_2\text{O})_6$  isomers obtained only by ABCluster.

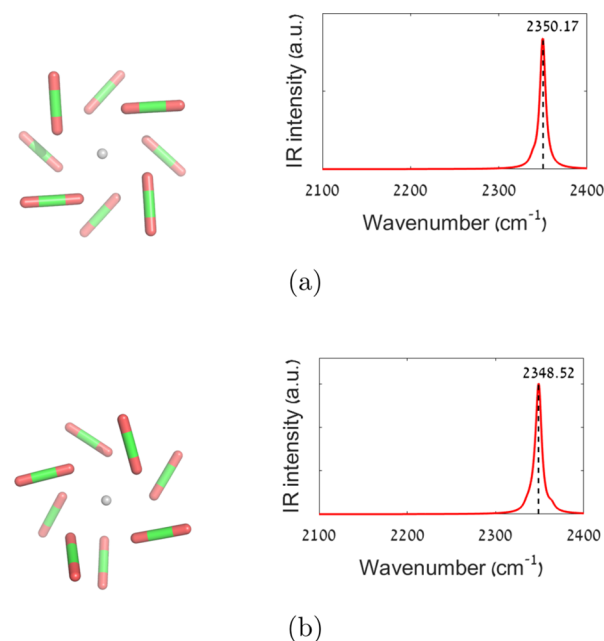


Figure 8. (a) Most stable  $\text{Ag}^+(\text{CO}_2)_8$  cluster obtained with its calculated IR spectrum. (b) Second cluster obtained with an energy difference of  $\Delta E = 0.79$  kcal/mol relative to (a), along with its calculated IR spectrum.

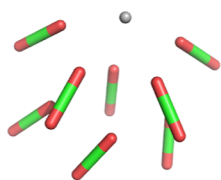


Figure 9. Lowest main  $\text{CO}_2$  cluster from ABCluster.

formed by the proton and four water molecules arranged as a rhombus, and its sides are formed by one water molecule located above the rhombus seat. We examined several previous studies on this cluster,<sup>58–62,64</sup> and only in one of these studies was this structure identified as the global minimum,<sup>64</sup> while in the others it did not appear at all. However, the second lowest energy isomer we found (with an energy of 0.02 kcal/mol above the global minimum—see Table 2) has been suggested in some previous studies<sup>58,60</sup> to be the global minimum and in ref 59 as the second lowest energy isomer being 0.01 kcal/mol above the global minimum.

For this cluster, we used  $B3LYP-D3(BJ)/6-31+G^{**}$ , followed by rescoring of clusters within 3.5 kcal/mol of the global minimum using  $\omega B97M-V/6-31+G^{**}$ . The stable clusters obtained are shown in Table 2. Note that the various local minima (LM) are very close in energy to the GM, and hence, are expected to contribute to the observed population of configurations. Free energy corrections provide additional insights and are provided in Table S5.

Most of the LM identified here appear in earlier works. There were a few isomers that the algorithm did not find and most of these are ranked among the least stable isomers. One can see that our energy rankings using  $B3LYP-D3(BJ)/6-31+G^{**}$  and  $\omega B97M-V/6-31+G^{**}$  are mostly in good agreement with the ranking in earlier works (except for ref 61 which used the empirical MSEVB potential, and ref 62 which used the  $B3LYP/6-31+G^*$  method). We treated the isomers I, II, and III in ref 58 as the same isomer and did the same for isomers I and II in ref 61.<sup>a</sup>

ABCluster was able to find the clusters we ranked as I, II, III, V, and VI, but not the other ones. However, ABCluster did find some additional structures, shown in Figure 7.

$\text{Ag}^+(\text{CO}_2)_8$ . As an additional test case, we chose the  $\text{Ag}^+(\text{CO}_2)_8$  cluster. This cluster has a measured IR spectrum and two known isomers, with an energy difference of 0.8 kcal/mol.<sup>65</sup>

For this cluster, we used  $B3LYP-D3(BJ)/6-31+G^{**}$ , followed by rescoring of clusters within 1.5 kcal/mol of the global minimum using  $\omega B97M-V/6-31+G^{**}$ .

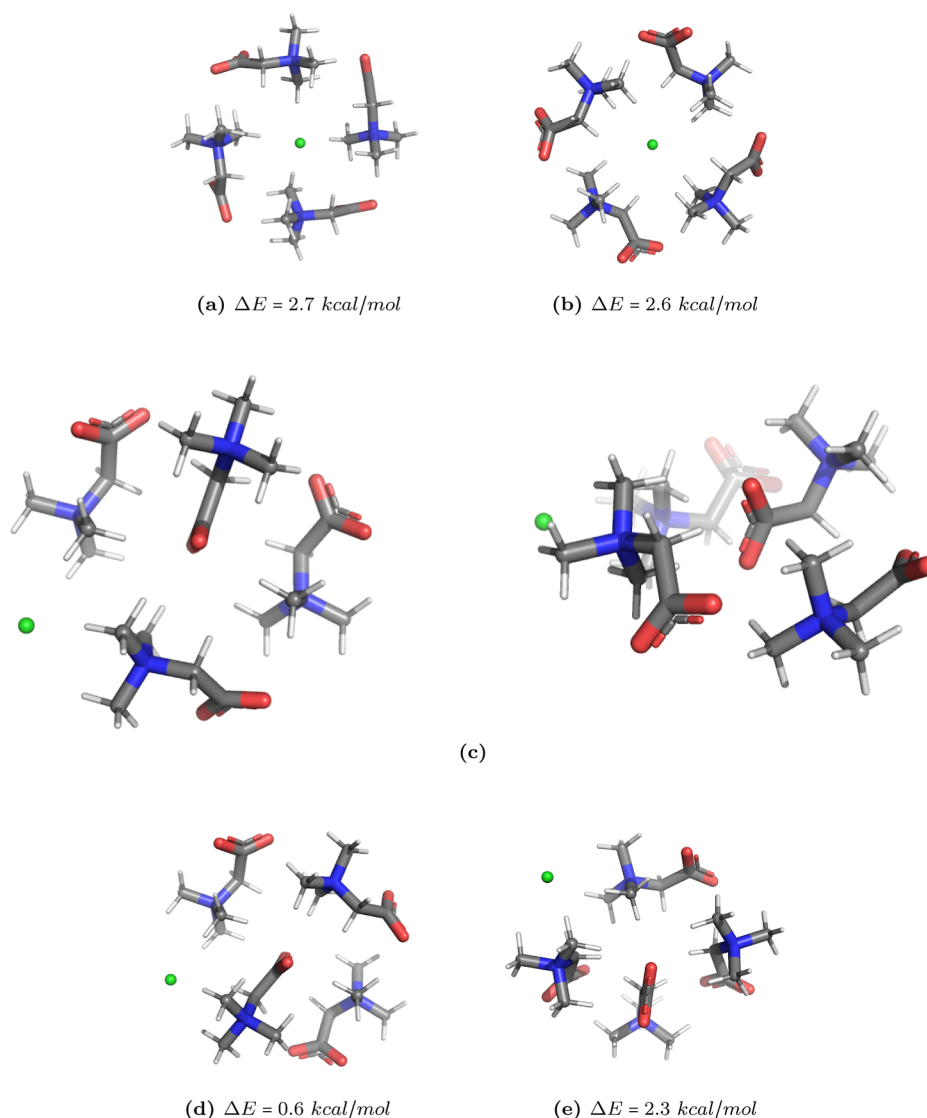
The most stable complexes obtained ( $\Delta E < 1$  kcal/mol) are shown in Figure 8, along with the calculated IR spectrum of the clusters. The most stable cluster (Figure 8a) is a two-layered complex, where each layer contains four  $\text{CO}_2$  molecules arranged in a vane shape. The silver atom is found between the two layers, in the center of the vane.

The difference in energy between the two complexes is 0.79 kcal/mol (0.67 kcal/mol using  $\omega B97M-V$ ), which is in good agreement with earlier findings.<sup>65</sup> The calculated IR spectra of both complexes also fit earlier theoretical and experimental data.<sup>65</sup>

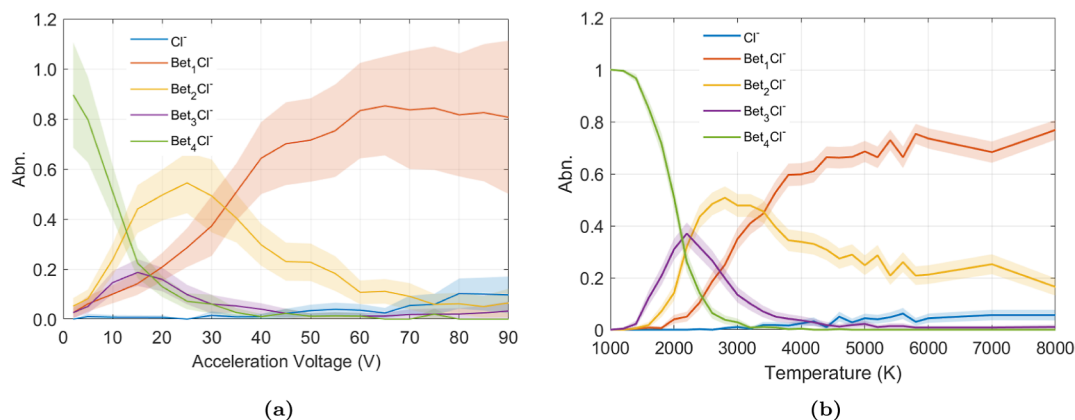
Note that the two minima are very close in energy, and this small difference persists when examining the free energy (Table S6). Hence, both structures might contribute to the observed population.

For this cluster, ABCluster generated one structure, which is shown in Figure 9, but not the lowest energy one. We obtained a similar structure, which was ranked lower (i.e., higher energy) than the global minimum by our method.

**Application to  $\text{Bet}_4\text{Cl}^-$ .** Betaine clusters are of interest due to their possible use in extending the ZITA technique. This technique allows one to determine whether the absorption of a chromophore is affected by a proximate electric field (i.e., whether the electronic transition has charge transfer character) by comparing the photoabsorption band of the bare chromophore to a chromophore complexed with a molecule with a high dipole moment. The molecule of choice in most of these studies is trimethyl glycine betaine, which is a pure zwitterion with a dipole moment of 11.9 D.<sup>1</sup> In a previous study by our group, we have shown that one can readily form clusters containing an ion of choice with a large number of neutral betaine molecules attached and that these clusters have a “magic number” with size  $N = 4$  no matter the identity of the charge.<sup>66</sup> Thus, by measuring the photoabsorption band as a function of cluster size, it may be possible to measure the shift in absorption as a function of the applied electric field. For this method, the identity of the lowest energy isomer is of crucial importance, as the electric field applied to the chromophore will be different for different isomers. For this reason, as a final test case for our approach, we chose to study the betaine tetramer, which does not have a known structure.



**Figure 10.** (a,b,d,e) show selected stable clusters, with energy differences of less than 3 kcal/mol relative to (c) using  $B3LYP-D3(BJ)/6-31+G^{**}$ . (c) Most stable cluster, top-view (left) and side-view (right). The energy ranking did not change using  $wB97M-V/6-31+G^{**}$ .

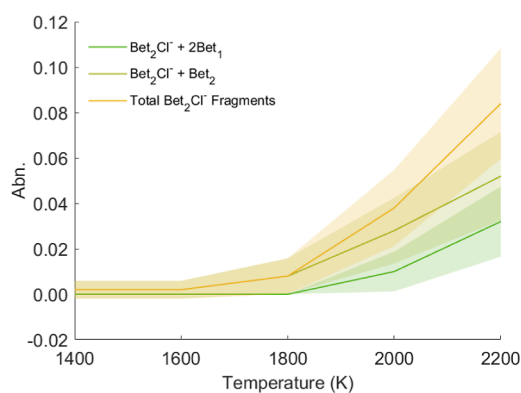


**Figure 11.** Abundance of fragmentation species as a function of (a) acceleration voltage and (b) temperature. (a) Experimental fragmentation pattern and (b) theoretical fragmentation pattern.

For this cluster, we used  $B3LYP-D3(BJ)/6-31+G^{**}$ , followed by the rescoring of clusters within 2.7 kcal/mol of the global minimum using  $wB97M-V/6-31+G^{**}$ .

**Geometries.** The most stable geometries are shown in Figure 10. The symmetrical structure shown in Figure 10a, which was also described in the previous study by our group,<sup>66</sup> is very similar to the most stable structure of the neutral tetramer,





**Figure 12.** Neutral fragments in the fragmentation of  $\text{Bet}_2\text{Cl}^-$ , obtained from 100 ps MD simulations.

where the four molecules are arranged as a square plane with the dipoles arranged in a circular fashion. The ion is located above the center of the molecules and causes a distortion such that the amine groups of each molecule are slightly elevated and point toward the ion. Interestingly, in this case a symmetry-free structures led to more stable geometries, shown as structures b–e in Figure 10. Structure b is very similar to structure a; however, the ion is located in the plane of the molecules and not above. The most stable geometry identified is shown as structure c in Figure 10. In this structure, three of the dipoles are arranged as  $\beta$ -sheets while the fourth dipole is perpendicular to the other three. Moreover, the excess ion is directly interacting with only two of the four molecules. Additional structures were identified, like antiparallel dipolar sheet structures, but these were higher in energy.

Note that the various LM have similar electronic energy to the GM and hence might contribute to observed distributions. In fact, structure c constitutes the GM on the free energy surface (Table S7).

ABCluster obtained both symmetric (i.e., the  $\text{Cl}^-$  ion in the center of the structure) and asymmetric (i.e., the  $\text{Cl}^-$  ion in the periphery of the structure) geometries (e.g., see Figure 10).

**Fragmentation of  $\text{Bet}_4\text{Cl}^-$ .** In our previous study, we measured the MS/MS (tandem mass spectrometry) spectrum of the chlorinated betaine tetramer, shown in Figure 11a.<sup>66</sup> In this earlier study, we found competition between monomer emission and cluster fission, which was not seen in the MS/MS of protonated betaine clusters (which fragment solely by monomer evaporation). This was deduced from the fact that the threshold collision voltages for the appearance of all cluster fragments are the same. Nevertheless, in MS/MS measurements only the charged fragments are observed and, therefore, it was not possible to conclusively determine whether the dimer fragment is a result of cluster fission or of a sequential evaporation of two monomers.

Here, we performed calculations of the most likely fragmentation channels using the most stable betaine cluster identified (Figure 10c). The simulations were performed using MD simulations, with 300,000 steps with a time step of 0.5 fs each, using the CHARMM FF and the CHARMM simulation package. The energy was inserted by heating. The results are shown in Figure 11b, where each data point is the average of 500 repeated MD simulations. As in the experiment, we find a similar threshold for the appearance of all three possible fragments. Overall, there is good agreement between the computed and the experimental fragmentation patterns. Importantly, the simu-

lations allow us to examine the ion-induced neutral fragments. We find that indeed cluster fission, i.e.,  $\text{Bet}_4\text{Cl}^- \rightarrow \text{Bet}_2\text{Cl}^- + \text{Bet}_2$ , is a prominent fragmentation channel, and observe a significant presence of a neutral  $\text{Bet}_2$  dimer in the fragmentation of  $\text{Bet}_2\text{Cl}^-$  (Figure 12).

## CONCLUSIONS

Herein, we presented a computational method for the prediction of structures of molecular clusters. The MCSA-ML method relies on MC annealing simulations in conjunction with a force field and quantum chemistry scoring and is augmented by machine learning classification using a set of carefully crafted descriptors. To identify low-entropy, high-symmetry clusters, the method includes symmetry operations during cluster generation. We tested the method for several Lennard-Jones clusters as toy systems and for the known molecular clusters  $\text{Ser}_8(\text{Cl}^-)_2$ ,  $\text{H}^+(\text{H}_2\text{O})_6$ , and  $\text{Ag}^+(\text{CO}_2)_8$ . Finally, we applied the method to a cluster with an unknown structure,  $\text{Bet}_4\text{Cl}^-$ . We estimate the correctness of the predicted global minimum structure of  $\text{Bet}_4\text{Cl}^-$  by computing the fragmentation patterns of the clusters using MD simulations. The simulated fragmentation pattern matches the experimental one well, lending support to the proposed structure. Additionally, the MD simulations of the fragmentation of the global minimum cluster reinforce the existence of a cluster fission channel  $\text{Bet}_4\text{Cl}^- \rightarrow \text{Bet}_2\text{Cl}^- + \text{Bet}_2$ . In comparison to results from other published works and using available software (e.g., ABCluster), we demonstrated a superior ability to predict clusters with high symmetry, and similar abilities to predict clusters with asymmetrical structures. In summary, we have presented MCSA-ML, which is a new method for structure prediction of molecular clusters. The main advantage of this method compared to earlier approaches<sup>8–22</sup> is the combined use of symmetry-based structure generation, ML-based clustering of structures using unique descriptors, and funneled energy scoring. The optimization progress can be easily monitored due to the use of ML-clustering of structures, as the number and variance of clustered structures can be easily followed. This combined strategy makes it an efficient approach for modeling molecular clusters in the gas phase.

## ASSOCIATED CONTENT

### Data Availability Statement

The source code of the algorithm and all input files are available free of charge at <https://github.com/rothmic/MCSA-ML-Global-Optimization.git>.

### Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acsomega.3c07600>.

Supporting Information includes a comparison of the *k*-means and spectral clustering algorithms (Figure S1), Cartesian coordinates of the stable molecular clusters (Tables S1–S4) and free energy corrections for  $\text{H}^+(\text{H}_2\text{O})_6$ ,  $\text{Ag}^+(\text{CO}_2)_8$  and  $\text{Bet}_4\text{Cl}^-$  clusters (Tables S5–S7) (PDF)

## AUTHOR INFORMATION

### Corresponding Authors

Yoni Toker – Department of Physics, Bar-Ilan University, Ramat-Gan 5290002, Israel; Institute of Nanotechnology and Advanced Materials, Bar-Ilan University, Ramat-Gan 5290002, Israel; [orcid.org/0000-0003-1809-7659](https://orcid.org/0000-0003-1809-7659); Email: [yonitoker@gmail.com](mailto:yonitoker@gmail.com)

Dan T. Major – Department of Chemistry and Institute of Nanotechnology and Advanced Materials, Bar-Ilan University, Ramat-Gan 5290002, Israel; [orcid.org/0000-0002-9231-0676](https://orcid.org/0000-0002-9231-0676); Email: majort@mail.biu.ac.il

## Author

Michal Roth – Department of Physics, Bar-Ilan University, Ramat-Gan 5290002, Israel; Institute of Nanotechnology and Advanced Materials, Bar-Ilan University, Ramat-Gan 5290002, Israel; [orcid.org/0000-0003-0910-9865](https://orcid.org/0000-0003-0910-9865)

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acsomega.3c07600>

## Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

The authors thank Jongcheol Seo for sharing the information shown in Figure 5c.

## ADDITIONAL NOTE

<sup>a</sup>The two references do not include Cartesian coordinates of the isomers, and based on the figures we were unable to distinguish between them.

## REFERENCES

- (1) Shikata, T. Dielectric relaxation behavior of glycine betaine in aqueous solution. *J. Phys. Chem. A* **2002**, *106*, 7664–7670.
- (2) Stockett, M. H.; Boesen, M.; Houmoller, J.; Brøndsted Nielsen, S. Accessing the Intrinsic Nature of Electronic Transitions from Gas? Phase Spectroscopy of Molecular Ion/Zwitterion Complexes. *Angew. Chem.* **2017**, *129*, 3544–3549.
- (3) Toker, Y.; Langeland, J.; Gruber, E.; Kjær, C.; Nielsen, S. B.; Andersen, L. H.; Borin, V. A.; Schapiro, I. Counterion-controlled spectral tuning of the protonated Schiff-base retinal. *Phys. Rev. A* **2018**, *98*, 043428.
- (4) Malar, E. J.; Divya, P. Structural Stability in Dimer and Tetramer Clusters of L-Alanine in the Gas Phase and the Feasibility of Peptide Bond Formation. *J. Phys. Chem. B* **2018**, *122*, 6462–6470.
- (5) Scutelnic, V.; Perez, M. A.; Marianski, M.; Warnke, S.; Gregor, A.; Rothlisberger, U.; Bowers, M. T.; Baldauf, C.; von Helden, G.; Rizzo, T. R.; Seo, J. The Structure of the Protonated Serine Octamer. *J. Am. Chem. Soc.* **2018**, *140*, 7554–7560.
- (6) Seo, J.; Warnke, S.; Pagel, K.; Bowers, M. T.; von Helden, G. Infrared spectrum and structure of the homochiral serine octamer-dichloride complex. *Nat. Chem.* **2017**, *9*, 1263–1268.
- (7) Schalley, C. A.; Weis, P. Unusually stable magic number clusters of serine with a surprising preference for homochirality. *Int. J. Mass Spectrom.* **2002**, *221*, 9–19.
- (8) Heiles, S.; Johnston, R. L. Global optimization of clusters using electronic structure methods. *Int. J. Quantum Chem.* **2013**, *113*, 2091–2109.
- (9) Srivastava, R. Application of Optimization Algorithms in Clusters. *Front. Chem.* **2021**, *9*, 9.
- (10) Marques, J. M.; Pereira, F. B.; Llanio-Trujillo, J. L.; Abreu, P. E.; Alberti, M.; Aguilar, A.; Pirani, F.; Bartolomei, M. A global optimization perspective on molecular clusters. *Philos. Trans. R. Soc., A* **2017**, *375*, 20160198.
- (11) Xiao, Y.; Williams, D. E. Genetic algorithm: a new approach to the prediction of the structure of molecular clusters. *Chem. Phys. Lett.* **1993**, *215*, 17–24.
- (12) Roberts, C.; Johnston, R. L. Investigation of the structures of MgO clusters using a genetic algorithm. *Phys. Chem. Chem. Phys.* **2001**, *3*, 5024–5034.
- (13) Llanio-Trujillo, J. L.; Marques, J. M.; Pereira, F. B. An evolutionary algorithm for the global optimization of molecular clusters: Application to water, benzene, and benzene cation. *J. Phys. Chem. A* **2011**, *115*, 2130–2138.
- (14) Zhang, J.; Dolg, M. Global optimization of clusters of rigid molecules using the artificial bee colony algorithm. *Phys. Chem. Chem. Phys.* **2016**, *18*, 3003–3010.
- (15) Elm, J. Toward a Holistic Understanding of the Formation and Growth of Atmospheric Molecular Clusters: A Quantum Machine Learning Perspective. *J. Phys. Chem. A* **2021**, *125*, 895–902.
- (16) Pickard, C. J.; Needs, R. Ab initio random structure searching. *J. Condens. Matter Phys.* **2011**, *23*, 053201.
- (17) Doye, J. P.; Wales, D. J. Polytetrahedral clusters. *Phys. Rev. Lett.* **2001**, *86*, 5719–5722.
- (18) Zilka, M.; Dudenko, D. V.; Hughes, C. E.; Williams, P. A.; Sturniolo, S.; Franks, W. T.; Pickard, C. J.; Yates, J. R.; Harris, K. D.; Brown, S. P. Ab initio random structure searching of organic molecular solids: assessment and validation against experimental data. *Phys. Chem. Chem. Phys.* **2017**, *19*, 25949–25960.
- (19) Chou, J.; Hsing, C.; Wei, C.; Cheng, C.; Chang, C. Ab initio random structure search for 13-atom clusters of fcc elements. *J. Condens. Matter Phys.* **2013**, *25*, 125305.
- (20) Tan, L.; Pickard, C. J.; Yu, K.; Sapelkin, A.; Misquitta, A. J.; Dove, M. T. Structures of CdSe and CdS nanoclusters from ab initio random structure searching. *J. Phys. Chem. C* **2019**, *123*, 29370–29378.
- (21) Lu, Z.; Zhu, B.; Shires, B. W.; Scanlon, D. O.; Pickard, C. J. Ab initio random structure searching for battery cathode materials. *Chem. Phys.* **2021**, *154*, 174111.
- (22) McMahon, J. M. Ground-state structures of ice at high pressures from ab initio random structure searching. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2011**, *84*, 220104.
- (23) Wales, D. J. Symmetry, near-symmetry and energetics. *Chem. Phys. Lett.* **1998**, *285*, 330–336.
- (24) Zhang, J.; Dolg, M. ABCluster: the artificial bee colony algorithm for cluster global optimization. *Phys. Chem. Chem. Phys.* **2015**, *17*, 24173–24181.
- (25) Zhang, J.; Glezakou, V.-A.; Rousseau, R.; Nguyen, M.-T. NWPEsSe: An Adaptive-Learning Global Optimization Algorithm for Nanosized Cluster Systems. *J. Chem. Theory Comput.* **2020**, *16*, 3947–3958.
- (26) Dieterich, J. M.; Hartke, B. Improved Cluster Structure Optimization: Hybridizing Evolutionary Algorithms with Local Heat Pulses. *Inorganics* **2017**, *5*, 64.
- (27) Yañez, O.; Báez-Grez, R.; Inostroza, D.; Rabanal-León, W. A.; Pino-Rios, R.; Garza, J.; Tiznado, W. AUTOMATON: A Program That Combines a Probabilistic Cellular Automata and a Genetic Algorithm for Global Minimum Search of Clusters and Molecules. *J. Chem. Theory Comput.* **2019**, *15*, 1463–1475.
- (28) Rodríguez-Fernández, R.; Pereira, F. B.; Marques, J. M.; Martínez-Núñez, E.; Vázquez, S. A. GAFit: A general-purpose, user-friendly program for fitting potential energy surfaces. *Comput. Phys. Commun.* **2017**, *217*, 89–98.
- (29) Wales, D. J. GMIN A Program for Finding Global Minima and Calculating Thermodynamic Properties from Basin-Sampling. <http://www-wales.ch.cam.ac.uk/GMIN/>.
- (30) Wales, D. J.; Doye, J. P. K. Global Optimization by Basin-Hopping and the Lowest Energy Structures of Lennard-Jones Clusters Containing up to 110 Atoms. *J. Phys. Chem. A* **1997**, *101*, 5111–5116.
- (31) Oakley, M. T.; Johnston, R. L.; Wales, D. J. Symmetrisation schemes for global optimisation of atomic clusters. *Phys. Chem. Chem. Phys.* **2013**, *15*, 3965–3976.
- (32) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D. J. Charmm general force field: A force field for drug-like molecules compatible with the Charmm all-atom additive biological force fields. *J. Comput. Chem.* **2010**, *31*, 671–690.
- (33) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell, A. D. J. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **2017**, *14*, 71–73.

- (34) Ledesma, S.; Avina, G.; Sanchez, R. *Simulated Annealing*; Tan, C. M., Ed.; IntechOpen, 2008; Chapter 20.
- (35) Karabin, M.; Stuart, S. J. Simulated annealing with adaptive cooling rates. *Chem. Phys.* **2020**, *153*, 114103.
- (36) Bernardi, R. C.; Melo, M. C.; Schulten, K. Enhanced sampling techniques in molecular dynamics simulations of biological systems. *Biochim. Biophys. Acta, Gen. Subj.* **2015**, *1850*, 872–877.
- (37) Wodak, S. J.; Janin, J. Analytical approximation to the accessible surface area of proteins. *Proc. Natl. Acad. Sci. U.S.A.* **1980**, *77*, 1736–1740.
- (38) Macqueen, J. Some methods for classification and analysis of multivariate observations. *Berkeley Symposium on Mathematical Statistics and Probability*, 1967; Vol. 1, pp 281–297.
- (39) Syakur, M. A.; Khotimah, B. K.; Rochman, E. M.; Satoto, B. D. Integration K-Means Clustering Method and Elbow Method for Identification of the Best Customer Profile Cluster. *IOP Conf. Ser.: Mater. Sci. Eng.* **2018**, *336*, 012017.
- (40) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem.* **1983**, *4*, 187–217.
- (41) Mardirossian, N.; Head-Gordon, M.  $\omega$ B97M-V A computationally optimized, range-separated hybrid, meta-GGA density functional with VV10 nonlocal correlation. *J. Chem. Phys.* **2016**, *144*, 214110.
- (42) Nagy, B.; Jensen, F. *Reviews in Computational Chemistry*; John Wiley & Sons, Ltd, 2017; Chapter 3, pp 93–149.
- (43) Martin, J. M.; Santra, G. Empirical Double-Hybrid Density Functional Theory: A ‘Third Way’ in Between WFT and DFT. *Isr. J. Chem.* **2020**, *60*, 787–804.
- (44) Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A: At, Mol, Opt. Phys.* **1988**, *38*, 3098–3100.
- (45) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785–789.
- (46) Miehlisch, B.; Savin, A.; Stoll, H.; Preuss, H. Results obtained with the correlation energy density functionals of becke and Lee, Yang and Parr. *Chem. Phys. Lett.* **1989**, *157*, 200–206.
- (47) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comput. Chem.* **2011**, *32*, 1456–1465.
- (48) Woon, D. E.; Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. IV. Calculation of static electrical response properties. *Chem. Phys. Phys.* **1994**, *100*, 2975–2988.
- (49) Liakos, D. G.; Neese, F. Is it possible to obtain coupled cluster quality energies at near density functional theory cost? Domain-based local pair natural orbital coupled cluster vs modern density functional theory. *J. Chem. Theory Comput.* **2015**, *11*, 4054–4063.
- (50) Shao, Y.; Gan, Z.; Epifanovsky, E.; Gilbert, A. T. B.; Wormit, M.; Kussmann, J.; Lange, A. W.; Behn, A.; Deng, J.; Feng, X.; Ghosh, D.; Goldey, M.; Horn, P. R.; Jacobson, L. D.; Kaliman, I.; Khaliullin, R. Z.; Kuś, T.; Landau, A.; Liu, J.; Proynov, E. I.; Rhee, Y. M.; Richard, R. M.; Rohrdanz, M. A.; Steele, R. P.; Sundstrom, E. J.; Woodcock, H. L.; Zimmerman, P. M.; Zuev, D.; Albrecht, B.; Alguire, E.; Austin, B.; Beran, G. J. O.; Bernard, Y. A.; Berquist, E.; Brandhorst, K.; Bravaya, K. B.; Brown, S. T.; Casanova, D.; Chang, C. M.; Chen, Y.; Chien, S. H.; Closser, K. D.; Crittenden, D. L.; Diedenhofen, M.; DiStasio, R. A.; Do, H.; Dutoi, A. D.; Edgar, R. G.; Fatehi, S.; Fusti-Molnar, L.; Ghysels, A.; Golubeva-Zadorozhnaya, A.; Gomes, J.; Hanson-Heine, M. W.; Harbach, P. H.; Hauser, A. W.; Hohenstein, E. G.; Holden, Z. C.; Jagau, T. C.; Ji, H.; Kaduk, B.; Khistyayev, K.; Kim, J.; Kim, J.; King, R. A.; Klunzinger, P.; Kosenkov, D.; Kowalczyk, T.; Krauter, C. M.; Lao, K. U.; Laurent, A. D.; Lawler, K. V.; Levchenko, S. V.; Lin, C. Y.; Liu, F.; Livshits, E.; Lochan, R. C.; Luenser, A.; Manohar, P.; Manzer, S. F.; Mao, S. P.; Mardirossian, N.; Marenich, A. V.; Maurer, S. A.; Mayhall, N. J.; Neuscammann, E.; Oana, C. M.; Olivares-Amaya, R.; O'Neill, D. P.; Parkhill, J. A.; Perrine, T. M.; Peverati, R.; Prociuk, A.; Rehn, D. R.; Rosta, E.; Russ, N. J.; Sharada, S. M.; Sharma, S.; Small, D. W.; Sodt, A.; Stein, T.; Stück, D.; Su, Y. C.; Thom, A. J.; Tsuchimochi, T.; Vanovschi, V.; Vogt, L.; Vydrov, O.; Wang, T.; Watson, M. A.; Wenzel, J.; White, A.; Williams, C. F.; Yang, J.; Yeganeh, S.; Yost, S. R.; You, Z. Q.; Zhang, I. Y.; Zhang, X.; Zhao, Y.; Brooks, B. R.; Chan, G. K.; Chipman, D. M.; Cramer, C. J.; Goddard, W. A.; Gordon, M. S.; Hehre, W. J.; Klamt, A.; Schaefer, H. F.; Schmidt, M. W.; Sherrill, C. D.; Truhlar, D. G.; Warshel, A.; Xu, X.; Aspuru-Guzik, A.; Baer, R.; Bell, A. T.; Besley, N. A.; Chai, J. D.; Dreuw, A.; Dunietz, B. D.; Furlani, T. R.; Gwaltney, S. R.; Hsu, C. P.; Jung, Y.; Kong, J.; Lambrecht, D. S.; Liang, W.; Ochsenfeld, C.; Rassolov, V. A.; Slipchenko, L. V.; Subotnik, J. E.; et al. Advances in molecular quantum chemistry contained in the Q-Chem 4 program package. *Mol. Phys.* **2015**, *113*, 184–215.
- (51) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 16 Revision C.01*; Gaussian Inc: Wallingford CT, 2016.
- (52) Neese, F.; Wennmohs, F.; Becker, U.; Riplinger, C. The ORCA quantum chemistry program package. *J. Chem. Phys.* **2020**, *152*, 224108.
- (53) Hoare, M. R.; Pal, P. Statics and Stability of Small Cluster Nuclei. *Nat. Phys. Sci.* **1971**, *230*, 5–8.
- (54) Woods, B. S.; Acioli, P. H. Drag Assisted Simulated Annealing Method for Geometry Optimization of Molecules. *Procedia Comput. Sci.* **2015**, *51*, 1878–1886.
- (55) Wales, D. J.; Doye, J. P. K.; Dullweber, A.; Hodges, M. P.; Calvo, F. Y. N. F.; Hernández-Rojas, J.; Middleton, T. F. Cambridge Cluster Database. <http://www-wales.ch.cam.ac.uk/CCD.html> (accessed 2023-09-03).
- (56) Shin, J.-W.; Hammer, N. I.; Diken, E. G.; Johnson, M. A.; Walters, R. S.; Jaeger, T. D.; Duncan, M. A.; Christie, R. A.; Jordan, K. D. Infrared Signature of Structures Associated with the  $H^+$  ( $H_2O$ ) $_n$  ( $n = 6$  to 27) Clusters. *Science* **2004**, *304*, 1137–1140.
- (57) Choi, T. H.; Jordan, K. D. Application of the SCC-DFTB Method to  $H^+$ ( $H_2O$ ) $_6$ ,  $H^+$ ( $H_2O$ ) $_21$ , and  $H^+$ ( $H_2O$ ) $_22$ . *J. Phys. Chem. B* **2010**, *114*, 6932–6936.
- (58) Kumar, R.; Christie, R. A.; Jordan, K. D. A Modified MSEVB Force Field for Protonated Water Clusters. *J. Phys. Chem. B* **2009**, *113*, 4111–4118.
- (59) Heindel, J. P.; Yu, Q.; Bowman, J. M.; Xantheas, S. S. Benchmark Electronic Structure Calculations for  $H_3O^+(H_2O)_n$ ,  $n = 0–5$ , Clusters and Tests of an Existing 1,2,3-Body Potential Energy Surface with a New 4-Body Correction. *J. Chem. Theory Comput.* **2018**, *14*, 4553–4566.
- (60) Jieli, M.; Aida, M. Classification of OH Bonds and Infrared Spectra of the Topology-Distinct Protonated Water Clusters  $H_3O^+ + (H_2O)_n - 1$  ( $n \leq 7$ ). *J. Phys. Chem. A* **2009**, *113*, 1586–1594.
- (61) Christie, R. A.; Jordan, K. D. Finite Temperature Behavior of  $H^+$ ( $H_2O$ ) $_6$  and  $H^+$ ( $H_2O$ ) $_8$ . *J. Phys. Chem. B* **2002**, *106*, 8376–8381.
- (62) Jiang, J.-C.; Wang, Y.-S.; Chang, H.-C.; Lin, S. H.; Lee, Y. T.; Niedner-Schatteburg, G.; Chang, H.-C. Infrared Spectra of  $H^+$ ( $H_2O$ ) $_5–8$  Clusters: Evidence for Symmetric Proton Hydration. *J. Am. Chem. Soc.* **2000**, *122*, 1398–1410.
- (63) Hodges, M. P.; Wales, D. J. Global minima of protonated water clusters. *Chem. Phys. Lett.* **2000**, *324*, 279–288.
- (64) Kondati Natarajan, S.; Morawietz, T.; Behler, J. Representing the potential-energy surface of protonated water clusters by high-dimensional neural network potentials. *Phys. Chem. Chem. Phys.* **2015**, *17*, 8356–8371.

(65) Zhao, Z.; Kong, X.; Yang, D.; Yuan, Q.; Xie, H.; Fan, H.; Zhao, J.; Jiang, L. Reactions of Copper and Silver Cations with Carbon Dioxide: An Infrared Photodissociation Spectroscopic and Theoretical Study. *J. Phys. Chem. A* **2017**, *121*, 3220–3226.

(66) Ben Yaacov, A.; Chen, L.; Musbat, L.; Nihamkin, M.; Kaiser, A.; Toker, Y. Clusters of betaine with positive and negative ions: Evidence for the betaine tetramer being magic. *J. Chem. Phys.* **2019**, *151*, 184303.