# CLUSTER: Searching for Unique Low Energy Minima of Structures Using a Novel Implementation of a Genetic Algorithm

René P. F. Kanters* and Kelling J. Donald*

Department of Chemistry, Gottwald Center for the Sciences, University of Richmond, Richmond, Virginia 23173, United States

Ⓢ Supporting Information

ABSTRACT: A new flexible implementation of a genetic algorithm for locating unique low energy minima of isomers of clusters is described and tested. The strategy employed can be applied to molecular or atomic clusters and has a flexible input structure so that a system with several different elements can be built up from a set of individual atoms or from fragments made up of groups of atoms. This CLUSTER program is tested on several systems, and the results are compared to computational and experimental data from previous studies. The quality of the algorithm for locating reliably the most competitive low energy structures of an assembly of atoms is examined for strongly bound Si—Li clusters, and ZnF_2 clusters, and the more weakly interacting water trimers. The use of the nuclear repulsion energy as a duplication criterion, an increasing population size, and avoiding mutation steps without loss of efficacy are distinguishing features of the program. For the Si—Li clusters, a few new low energy minima are identified in the testing of the algorithm, and our results for the metal fluorides and water show very good agreement with the literature.
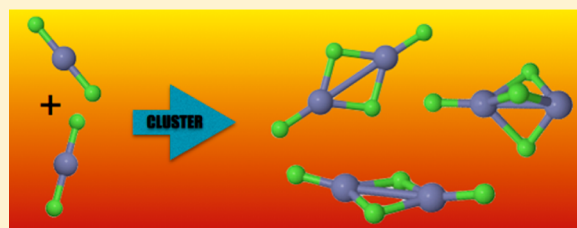
I. INTRODUCTION

Structure prediction in chemistry, even for clusters with a relatively small number of atoms, remains a challenge. It can be difficult and time-consuming to identify and systematically examine structures, locate local minima, and confidently posit a global minimum. This is especially problematic for atomic and weakly bound molecular clusters. As the number of atoms increase, the multidimensional potential energy surface that needs to be sampled becomes rapidly more complex, and the calculations needed to locate low energy structures become increasingly time-consuming. A few different methods based on the so-called genetic algorithm (GA) have been shown to be quite powerful in searching for global minima for both atomic and molecular clusters. These methodologies typically involve a gradual optimization of a fixed-sized population of structures by transferring structural information (genetic data) from an initial population to successive generations of candidate structures, thus increasing the probability of retaining favorable structural features in a systematic move toward a population containing the global minimum.[1] Improvements to GA methods over the past several years include built-in procedures to guarantee that any structure that ends up in the final population is a local minimum. This is done typically by a combination of geometry optimizations at reliable computational levels and vibrational frequency analyses and imposing criteria for entering the final populations that include the absence of imaginary frequencies.[2−6] Such an approach has given rise to the *ab initio* gradient embedded genetic algorithm (GEGA), developed by Alexandrova et al.[6] Their approach was applied initially to atomic clusters only, where it proved to be "exceptionally successful in

finding the global and the low-energy local minima"[7] and was extended in ref 7 to treat molecular clusters as well.

The GA approach applied by Deaven and Ho to atomic clusters[1] maintains a, possibly quite small, population of candidate structures that are relaxed to the nearest local minimum by either conjugate-gradient minimization or molecular dynamics quenching. The energy of each of these structures is then used as a criterion for fitness affecting the probability of each structure to be chosen as a parent and mated to generate a new generation of structures. The mating operator involves combining atoms that lie above a random plane passing through the center of one parent cluster with those below a similar plane in the second parent, ensuring that the resulting offspring has the correct number of atoms. It was found that in some cases it was necessary to apply mutations, e.g., displacing atoms to random extents or using the more involved strategy of searching for any adjacent minimum on the potential energy surface. Thereafter, the relaxed structure is included in the new population if the energy is lower by some energy difference, δE, compared to any candidate in the population, so as to avoid filling the population with almost identical low-energy candidates.

The implementation of the genetic algorithm by Alexandrova and Boldyrev in the GEGA program incorporated the added feature of carrying out a vibrational frequency analysis on stationary points encountered during the relaxation of candidate structures. When a saddle point is encountered and confirmed in this way, the first imaginary frequency mode is

followed until a local minimum is found. In that approach, a new generation of the same number of structures as in the, possibly large, initial population is generated using a similar mating method to that employed by Deaven and Ho, treating molecular units in the parent as unbreakable entities. The best-fit group is selected, therefore, so as to reduce the population back to its original size. After a mating cycle, 30% of the least favorable structures are submitted to a mutation procedure based on random kicks to the fragments as a whole and are optimized as well. The GEGA is considered converged if the same lowest-energy species remains the leading candidate for 20 generations. The GEGA program, which is written in C++, has been developed for use with the Gaussian suite of programs[8] for geometry optimizations and vibrational frequency analyses and has found extensive use in the past several years.[9]

An interesting difference in the approaches in refs 1 and 6 is the use of single matings by Deaven and Ho as opposed to the creation of new generations by Alexandrova. In the latter case, the proposed population size is at least $5N$ but not less than 40, where N is the number of atoms in an atomic cluster. This initial population is the starting point for at least 20 subsequent generations due to the convergence criterion in the algorithm, which sums up to a sequence of optimization and frequency calculations of no less than $21 \times 40 = 840$ (or as much as $21 \times 5N$) candidate structures, ignoring the calculations on mutated structures, all of which can be quite time-consuming.

In this paper we introduce a structural search algorithm, implemented in the new CLUSTER program. The program is written in the Python language and utilizes methodologies that are analogous to the GEGA approach but avoids the creation of large generations and provides significant additional levels of flexibility as well. The efficacy of our approach is tested in this work by comparing the results generated using the CLUSTER program with the available computational and experimental results for a small but diverse group of systems that have been examined in the literature before.

II. METHODS

The job of the CLUSTER program is to locate reliably the lowest energy structures on the potential energy surfaces for chemical species using a genetic algorithm that builds and automatically updates a population of candidate structures that are verified to be local minima for the system in question. Candidates consist of fragments, which may be a single atom or a group of atoms, making up a cluster with the required composition. Each fragment is assigned a name in the input file, and the identity of the fragment is determined by this name. Fragments are considered identical if they have the same name. For ease of use, the CLUSTER program has a set of predefined, built-in element and fragment libraries, but user-defined libraries can be used as well. The element radii are taken from the WebElements Web site,[10] and the fragments are ORCA[11] optimized using PBE/TZVPP[12] and VDW06 model chemistries.[13]

For geometry optimization and vibrational frequency calculations, the CLUSTER program has been developed to interface conveniently with several popular and some more recent computational programs: the ADF program,[14] Gaussian 03 and 09,[8] ORCA,[11] and MOPAC.[15] Since the CLUSTER program needs to be able to generate input files for these various programs, and given the different types of calculations required for the search procedure, we have built into the

software the added flexibility of user managed templates for the generation of the input files for the various software options, in case the built-in templates do not allow for the possible complexity that one might need in the input files.

After the initial population is generated, it is continuously updated during a sequence of matings in which the acceptable candidates from each mating may be immediately added to the population thus allowing new candidates to be used as soon as possible in the creation of subsequent candidates. These updates can result in (i) the replacement of existing candidates in the population if they are considered duplicates or (ii) in the addition of new candidates if the new species are considered to be unique.

The program terminates when the requested amount of matings has been performed or a certain number of matings did not change the makeup of the population.

1. Population Initialization. Since the initial population must consist of verified minima in the potential energy surface, it is computationally expensive and time-consuming to generate the population with a large number of candidates that meet that criterion at the outset. Allowing the population size to grow during the course of the search enables us to start with a small initial population of minimally one candidate. This requirement in the CLUSTER program is an improvement on previous strategies.

Importantly, the initial population can be the population from an earlier run, i.e., a restart of the procedure to continue more matings, a user-defined set of structures that is first reoptimized before further use, or (possibly in combination with one of the previous populations) a set of newly built and optimized candidates. For the latter, the program supports 1D, 2D, or 3D cluster build procedures. The 1, 2, or 3 reflects the dimensionality of the random displacement vector used in positioning each fragment, which is randomly rotated around its center. So, '1', '2', and '3' here are not necessarily the dimensionality of the final structure, unless it consists of only atomic fragments. Before positioning a fragment, it is randomly rotated. The sequence in which the fragments are positioned is itself random as well. The translation distance along the displacement vector is such that the newly placed fragment atoms will possibly touch but not interpenetrate one or more atoms in the previously placed fragments thus avoiding interatomic distances that are too small, which could cause computational problems. In order to build structures where the atoms do not touch, one may assign an extra radius in a fragment definition, redefine the atomic radii for elements, and/or provide a scale factor—a multiplication factor—for the translation vector.

2. Optimization. There is no guarantee that a geometry optimization will converge within the maximum number of optimization steps requested in the input file, so the optimization of a candidate is implemented as a maximum number of cycles consisting of a geometry optimization, frequency calculation followed by a possible reoptimization. Every structure whose geometry is to be optimized is added to a library of structures, so the user can inspect some of the intermediate structures that are on the path of the optimization.

If the geometry is not converged after a fixed maximum number of these (re)-optimization and frequency calculation cycles, the candidate is discarded. Candidates may also be discarded if the computational program runs into problems, e.g., due to wave function convergence issues, preventing it from calculating the energy of the system, or if it does not

B

dx.doi.org/10.1021/ct500744k | *J. Chem. Theory Comput.* XXXX, XXX, XXX–XXX
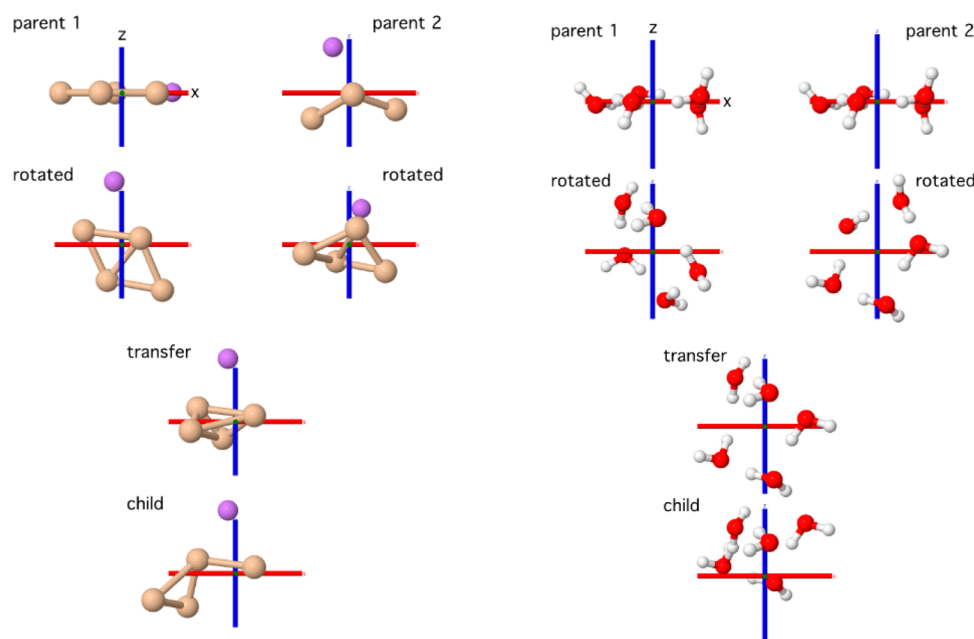
Figure 1. Two examples of successful mating attempts. On the left for a heteroatomic cluster, Si$_4$Li, where two Si atomic fragments from parent 2 are transferred and in the final child positioned to avoid interatomic distances that are too small. On the right-hand side, three water molecules are transferred in a $(H_2O)_5$ cluster. In this case the parents happen to be identical. The fragments transferred from parent 2 are moved, as a whole, closer to the water molecules from parent 1 (child) because the water molecule is not defined to have an extra radius. All images are projected along the y-axis onto the xz-plane with the z-axis pointing up in the plane of the paper.

report the frequencies in a frequency calculation. These discarded candidate structures are also saved separately so the user may inspect them later.

In order to avoid duplicate efforts of unnecessary additional optimizations or frequency calculations, every stationary point found is stored and saved in a library of converged structures so that any new stationary point may be checked against this library to determine whether it has been encountered before. This is an important test that has been missing to the best of our knowledge in previous GA based search methods. The option is provided, nonetheless, to check for duplicates against the candidates in the population of local minima only, all stationary points, or to not check at all. The criteria for uniqueness of the geometry are based on the energy, E, calculated by the computational program as well as the nuclear repulsion energy, R, calculated by the CLUSTER program itself. The latter quantity (R) is computed directly as the sum of the unique pairwise coulomb repulsion energy of the N charged nuclei in the cluster according to the following equation (using atomic units)

$$R = \sum_{i=1}^{i=N} \sum_{i<j} \frac{q_i q_j}{r_{ij}}$$

(1)

where q_i and q_j are the nuclear charges of atoms i and j, and r_{ij} is the distance between the two atoms. If both values are within the specified resolution values δE and δR, respectively, compared to an earlier encountered stationary point, the stationary point is considered a duplicate and is added to a library of duplicate candidates. More specifically, candidate 1 is considered a duplicate of candidate 2 if $E_1 > E_2$ and $E_1 - E_2 < \delta E$ as well as $|R_1 - R_2| < \delta R$. We want to mention, however, that this criterion for duplicate detection does not distinguish between optical isomers. If the stationary point is considered a

duplicate, it is added to a library of duplicate candidates, and no further calculation on the candidate is performed.

If a stationary point is considered to be unique as well as a local minimum, i.e., if all of its vibrational frequencies are above a specified cutoff value (for which the default is zero), it is added to a library of viable candidates. It is also added to the population if the energy is within a predefined window above the lowest energy structure in the existing population, after which duplicates in the population are removed, thus allowing the new candidate to replace an existing candidate.

If the stationary point was not a minimum, the lowest energy frequency vibrational mode may be followed in one *or both* directions in an attempt to find one or two minima. This optional bifurcation in the search methodology is still limited by the maximum number of optimization starts, so as to avoid being trapped in an infinite loop. The bifurcation that can happen at this stage is the reason that the optimization of a single candidate may result in more than one new candidate in the population.

All libraries mentioned as well as the population itself are saved after every change and stored as multistep xyz files[16] so that they may be easily visualized using programs such as Jmol[17] or ChemCraft.[18]

3. Mating Procedure. To create a child, two parents are randomly chosen from the population using the energy as a criterion for fitness. For each parent the center of every fragment is located. The center of these points is defined as the center for the parent. Each parent is then randomly rotated around that center. The child is initially a clone of the first parent. The part of the clone that is above the xy-plane is preserved; each of the remaining fragments is replaced with an identical one from below the xy-plane in parent 2, if possible. This may fail to transfer any fragments if the clone and parent 2 have different fragments below the xy-plane. If that occurs, the random rotation and transfer is attempted again, up to a

predefined maximum number of times, after which optionally a new 3D candidate may be created instead or the mating is aborted.

If it was possible to transfer fragments during the mating (transfer in Figure 1), they are treated as a whole and translated along the vector to its geometric center such that, keeping any possibly additional fragment radii in mind, one or more atoms will possibly touch but not interpenetrate atoms from the nontransferred fragments (child in Figure 1). Two examples of successful mating attempts are shown in Figure 1.

4. Mutation Procedure. Genetic algorithm based methods for finding stable molecular structures typically implement several genetic operators. At a minimum, these will include a crossover, or so-called recombination or mating, routine similar to the procedure we have just described for the CLUSTER program. A "mutation" operator, a mechanism for distorting a candidate, is often included as well. The CLUSTER program in its current form does not employ mutation operators for several reasons:

(i) For our type of candidate structures a mutation is a geometry modification, i.e. a random kick, of fragments or atoms in such a way that the mutation is still a candidate that a computational program can work with. This means that care needs to be taken to avoid creating interatomic distances that are too small. But that is by no means a trivial requirement to satisfy for three-dimensional structures with possibly large numbers of atoms.

(ii) The computational program optimizes all atomic positions so mutations of atomic positions are automatically considered but admittedly not random ones.

(iii) A form of mutation is inherent in the mating methodology by virtue of its random rotation of the parents so that even when mating the same parents multiple times, different offspring will likely result. And finally,

(iv) in a system with molecular fragments, due to the geometry optimization of the candidates the relative atomic positions within the fragment change. This may be considered a mutation of the fragment, which may be used in future mating procedures.

III. RESULTS AND DISCUSSION

To test the efficacy of the CLUSTER program several systems already discussed in the literature were selected for consideration. For this analysis, we decided to use primarily the ADF software[14] as our quantum mechanical partner program of choice. The ADF2013 program has a rather generous default energy criterion for geometry convergence (1×10^{-3} Hartrees) for our purposes, so we opted for the tighter criterion of 1×10^{-4} Hartrees.

1. Si_4Li and Si_4Li_2 Binary Clusters. The results of the GEGA calculations for Si_4Li_n ($n = 1-7$) binary clusters using the B3LYP method in combination with the Stuttgart-Dresden pseudopotentials (SDD) as implemented in the Gaussian 03 suite of programs were recently reported.[19] The structures were further optimized in that work using the B3LYP/def2-TZVPP model chemistry which afforded the zero point energy (ZPE) correction as well as the structures for the single point energy calculations at the CCSD(T)/def2-TZVPP level. The paper did not provide details about the settings for the GEGA program, but based on the original recommendation for the population size to be used,[6] at least 840 candidate structures must have been optimized. In our calculations we started with a build of

two 1D, five 2D, and 20 3D candidates followed by only 30 matings, for a minimum of 57 candidates in all.

1a. Si_4Li Binary Clusters. A CLUSTER calculation for the binary Si_4Li system using ADF with the LDA/TZP model chemistry with a large core for all atoms, using $\delta E = 5 \times 10^{-4}$ Hartrees (0.314 kcal/mol) and $\delta R = 1$ Hartree, resulted in a final population of six candidate structures (see Table 1), which

Table 1. Results of the Si_4Li CLUSTER Search Using ADF's LDA/TZP(Large) Computational Method[c]

Label	Structure[a]	LDA/TZP(Large)		CCSD(T)/def2-TZVPP//B3LYP/def2-TZVPP		
		ΔE	R	Δ(E+ZPE)	R	Δ(E+ZPE)
1.1		0.00	279.8	0.00	278.6	—
1.2		0.04	276.9	0.26	276.7	0.00
1.3		2.96	282.2	5.35	281.0	5.08 [b]
1.4		—	—	—	—	5.20
1.5		4.47	279.5	6.14	278.7	—
1.6		6.88	279.4	8.75	278.2	—
1.7		41.28	244.6	—	—	—

[a]Representative geometry (slight differences in geometry may exist at different levels of theory). [b]The coordinates with a reported Δ(E+ZPE) of 5.9 kcal/mol actually corresponded to the 1.3 structure. Their reported image and energy corresponds to structure 1.5. It was not recalculated here since the coordinates were not available. [c]All ΔE values are in kcal/mol, and the R values are in Hartree units. The right-hand column energies are recalculated based on the geometric information provided by Osorio et al.[19]

were all found during the initial building stage. The effect of the matings was such that in one case a nearly identical but slightly lower energy structure was found which replaced the one that was found earlier. In the final population the two lowest energy structures differed by only 0.04 kcal/mol, which is significantly less than δE. Both structures were retained, however, because the difference in R was larger than δR. This situation shows clearly the importance of the δR criterion as an additional test of the equivalence of structures in the population.

In order to compare the energies of these candidates with the literature, Gaussian 09 CCSD(T)/def2-TZVPP//B3LYP/def2-TZVPP calculations were performed.[20] These retained the five lowest energy structures only, because the calculations for the highest energy candidate failed in the optimization stage. Comparison of the final geometries with the three reported by Osorio et al. shows that a new lower energy structure, 1.1, more stable by 0.26 kcal/mol including the ZPE correction, was found. The second lowest energy structure, 1.2, found by the CLUSTER program was the same as the lowest energy structure in that work. The reported third lowest energy structure, 1.4 here, was not captured in the final population. A CLUSTER run performing a reoptimization of the literature structures at the LDA/TZP(Large) level showed that structure 1.4 optimizes to

a stationary point with a lowest frequency of −13.2 cm^{-1} and would be discarded from the search. Thus, the level of theory used may affect the final structures that are located since for some methods the low energy stationary points may emerge with small negative frequencies and are likely to be discarded. In addition to the new lowest energy structure, two new minima, 1.5 and 1.6, were found using the CLUSTER program as well.

Comparing the ΔEs from the LDA/TZP(Large) calculations and the Δ(E+ZPE) from the CCSD(T)/def2-TZVPP// B3LYP/def2_TZVPP calculations shows that the differences are 2.4 kcal/mol or less, suggesting that the LDA/TZP(Large) methodology may be a sufficient level of theory for the CLUSTER program for these kinds of systems for the generation of a reliable set of final structures at least. Moreover, the nuclear repulsion factors, which are quite sensitive to differences in internuclear distances, are only slightly reduced in the B3LYP/ def2-TZVPP calculations, thus suggesting close similarity in the optimized geometries obtained by these two methodologies.

1b. Si$_4$Li$_2$ Binary Clusters. Following up on our identification of a lower energy structure for the Si$_4$Li cluster compared to the previous results from the GEGA calculations, the Si$_4$Li$_2$ system was also investigated. We used the same initial settings for the search, except that δR was set to 2 Hartree. The results that we obtained for the final population using the LDA/TZP(Large) method whose final population was further calculated at the CCSD(T)/def2-TZVPP//B3LYP/def2-TZVPP level of theory are presented in Table 2.

The initial build procedure yielded four of the seven structures found after the 30 matings allowed. The run was extended for an additional 50 matings because structure 2.3 was not encountered. These additional matings slightly lowered the energies for structures 2.1, 2.5, and 2.8 and added structures 2.3 and 2.10. The fact that new minimum energy structures were found during the additional matings suggests that the amount of structures built initially was too limited, but it demonstrates as well that the mating procedure is robust enough to fill in the gaps. Unfortunately, one can never say conclusively, especially in the absence of any experimental data, whether all of the competitive geometries within a given energy window above the lowest energy structure have been captured, or if the lowest energy isomer is the global minimum.

In this case as well, the CLUSTER program found a lower energy structure in 2.1. Since the literature structures were based on a GEGA search that used B3LYP/SDD, a search using Gaussian 09 software and the B3LYP/SDD model chemistry was performed. This CLUSTER search did find structure 2.1 but with an energy of about 3.0 kcal/mol higher than structure 2.2.

Structure 2.7 was not found in the CLUSTER calculation. An LDA/TZP(Large) geometry optimization of 2.7 yielded a converged structure with a negative frequency of −33.1 cm^{-1}, which, when followed, optimized to structure 2.1. On the other hand, the B3LYP/def2-TZVPP optimization of 2.8 converged to the same geometry as 2.6. Thus, we find in this case again that the theoretical method used affected the outcome since the different methods predicted different isomers to be local minima.

The relative energy ordering of the structures is slightly different for the LDA calculations than for the CCSD(T) calculations: the order of structures 2.3 and 2.4 as well as 2.5 and 2.6 are reversed. This is not too surprising since the relative energy differences between these pairs of structures are quite small.

Table 2. Results of the Si$_4$Li$_2$ CLUSTER Search Using ADF's LDA/TZP(Large) Computational Method, Sorted by CCSD(T) Energies^c

Label	Structure ^{a}	LDA/TZP(Large)		CCSD(T)/def2-TZVPP//B3LYP/def2-TZVPP		
		ΔE	R	Δ(E+ZPE)	R	Δ(E+ZPE)
2.1		0.00	313.9	0.00	312.5	—
2.2		4.57	310.7	2.43	309.5	0.00
2.3 ^{b}		6.40	310.0	2.79	309.3	0.36
2.4		6.08	310.6	3.24	309.8	0.82
2.5		7.32	313.6	4.03	312.5	—
2.6		6.89	308.2	4.89	305.8	2.46
2.7		—	—	—	—	6.58
2.8		10.20	310.3	optimized to 2.6		—
2.9		16.51	306.3	12.82	305.3	—
2.10 ^{b}		31.33	281.6	26.42	280.4	—

^{a}Representative geometry (slight differences in geometry may exist at different levels of theory). ^{b}These structures were found after extending the calculation with an additional 50 matings. ^{c}All ΔE values are in kcal/mol, and the R values are in Hartree units. The right-hand column energies are recalculated based on the geometric information provided by Osorio et al.[19]

2. Zinc Difluoride Dimers and Trimers. As a test for more tightly bound molecular clusters, the dimers, and the trimers of ZnF$_2$ were investigated. All quantum calculations were performed with the ADF program using the LDA method with a TZP basis set and a large core.

2a. Zn$_2$F$_4$. The group 12 dihalide dimers $(MX_2)_2$ have been investigated in the past few years because of the increasing role of relativistic effects going down group 12 from Zn to Hg,[21−23] and because of the apparent links between the bonding in the rigid linear monomers and the bonding preferences in the extended solids.[24] Computational and experimental results are in agreement that the ZnF$_2$ monomer is linear and that the dimer is a D_{2h} $(ZnF_2)_2$ rhombus with a terminal F atom bonded to each of the Zn atoms as well (Figure 2.i).

To mimic the approach used in ref 21 and to assess how well and how efficiently the CLUSTER program would reproduce the set of solutions found therein, a population library was created containing the 11 possible dimer geometries depicted in Figure 2.

This library was used as a starting population for the CLUSTER search where the initial geometries were first optimized and the remaining structures used in 30 matings. As might have been expected, the initial optimization of the seeded population retained only the D_{2h} structure (generated from isomers i, ii, vii, viii, ix, and x) and the C_{3v} isomer (iii). Isomers iv and v
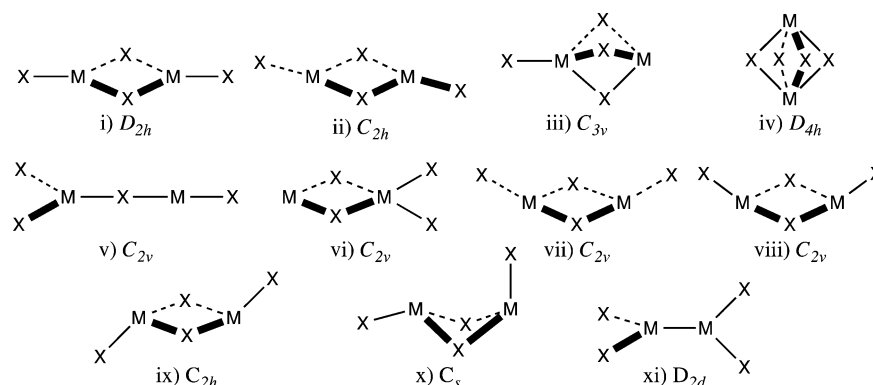
Figure 2. Possible dimer structures as considered in ref 21.

optimized to stationary points that were not minima. Isomer iv failed to yield a stationary point after two more times of the default 50 geometry optimizations and was aborted. Inspection of this aborted structure, however, showed that it was close to the D_{2h} geometry as well. Following the largest imaginary frequency in both directions for isomer v gave the D_{2h} structure. Isomer vi did not achieve wave function convergence and was aborted. Isomer xi generated a stationary point that was also found in the isomer iv optimization that retained imaginary frequencies and was aborted. All but one mating, which failed to achieve wave function convergence, optimized to the same D_{2h} geometry, with no duplicate of the C_{3v} isomer.

Deciding on Fragments. A cluster with the Zn_2F_4 chemical formula may be built using different sets of fragments. Three approaches that we considered were (1) using two linear ZnF_2 monomers with a Zn−F bond distance of 1.727 Å, (2) using one ZnF_2 monomer, one Zn and two F atoms, and (3) using two Zn atoms and four F atoms, i.e., a heteroatomic cluster. All of the searches were initiated by building two 1D, five 2D and 20 3D candidates and followed by 30 matings, using a $\delta E = 5 \times 10^{-4}$ Hartrees and a $\delta R = 2$ Hartree. Due to random procedures used in the CLUSTER program different runs may give different results, so five runs were executed for each. In order to process all viable candidates found in the 15 CLUSTER runs, the viable candidate libraries were concatenated into a single library and filtered using a δR of 5 Hartree and without an energy window, so that all unique viable candidates were retained, even the ones with a very high energy. A subset of the results from those rounds of calculations is presented in Table 3 (the full population file is available in the Supporting Information).

Two Monomers. The CLUSTER calculations using two monomers only found the lowest energy D_{2h} dimer, 3.1 in Table 3, but not the C_{3v} one, 3.3. The D_{2h} geometry was found during the 1D build stage while 2D builds, 3D builds, and matings only marginally lowered its energy. The fact that only the D_{2h} isomer was found may be because the F−Zn−F angles of 132°−140° present in both D_{2h} and C_{3v} isomers can be relatively easily obtained from the initial 180° bond angle. The C_{3v} isomer, 3.3, also requires an F−Zn−F angle closer to 90°, but that amount of bending is quite costly when starting with a bent fragment from a D_{2h} parent structure at the mating stage and even more so for a linear ZnF_2 monomer unit at the building stage. One of the five runs encountered the curious structure 3.2 in Table 3, which is quite similar to 3.1. In that structure, the singly bonded fluorine atoms are located above the Zn_2F_2 plane, with Zn−Zn−F angles of about 170°.

Table 3. Combined Results of Zn_2F_4 CLUSTER Calculations Using ADF's LDA/TZP(Large) Computational Method Focusing on Structures Less than 125 kcal/mol above the Lowest Energy Structure[b]

Code	ΔE	R	Structure	Fragments used [a]		
				2 ZnF₂	ZnF₂ Zn 2 F	2 Zn 4 F
3.1	0.00	757.0		5	5	5
3.2	0.52	758.3		1	0	0
3.3	22.81	822.8		0	4	1

[a]The values reflect the number of runs out of five in which the structure was encountered. [b]All ΔE values are in kcal/mol, and the R values are in Hartree units.

One Monomer and Three Atoms. Combining a linear ZnF_2 monomer with one Zn and two F atoms allows us to sample a larger area of the potential energy surface and returns certain local minima of structures that are not found by simply combining two monomers. The CLUSTER program encountered the D_{2h} structure in all five runs during the 1D and the 2D build stages. In four out of five runs it also located the C_{3v} structure but only during the 3D build or mating stages. Some higher energy local minima of almost 130 kcal/mol above that of 3.1 were found too, but those are so high in energy as to be of no interest to us here. Of the three approaches used, this one (as we will confirm presently) performed the best, since it found both of the previously established stable low energy isomers more often in the five CLUSTER runs compared to the other approaches (see Table 3).

Atoms Only. Building up the Zn_2F_4 system as an atomic cluster only (from six individual atom units) generated the C_{3v} isomer in only one of the five runs. All of the runs found the D_{2h} structure with the first ones in each run encountered during the build stage, be it 1D, 2D, or 3D. For the one run in which 3.3 was found, it was generated during the 3D building stage. The matings only resulted in duplicates of 3.1. Again, higher energy local minima were found that are too high in energy to be of interest to us in this work.

It is disconcerting that the C_{3v} isomer is so hard to find using the random build and mating procedures implemented in the

Table 4. Combined Results of Zn_3F_6 CLUSTER Calculations Using ADF's LDA/TZP(Large) Computational Method Focusing on Structures Less than 130 kcal/mol above the Lowest Energy Structure[b]

Code	ΔE	R	Structure	Fragments used [a]							
				$3ZnF_2$	$2ZnF_2$ Zn 2F	D_{2h} ZnF_2	C_{3v} ZnF_2	D_{2h} Zn 2F	C_{3v} Zn 2F	ZnF_2 2Zn 4F	3Zn 6F
4.1	0.00	1557.1		5	5	5	5	5	5	4	4
4.2	0.36	1560.1		0	0	1	0	0	0	0	0
4.3	4.78	1759.7		5	5	5	5	5	5	3	4
4.4	7.52	1661.6		0	0	0	1	0	0	0	0
4.5	13.55	1627.5		5	5	5	4	5	4	4	4
4.6	15.78	1534.9		3	4	1	0	0	0	3	2
4.7	20.10	1809.9		0	0	0	1	0	0	0	0
4.8	25.04	1650.0		1	1	0	5	0	5	3	0

[a]The values reflect the number of the runs out of five in which the structure was encountered. [b]All ΔE values are in kcal/mol, and R values are in Hartree units.

CLUSTER program. One of the reasons for this may be that the area of the sampling space that minimizes to the D_{2h} symmetry is large but small for the C_{3v} isomer. That latter is apparently better sampled by using fragments that do not have any *a priori* geometry enforced. Another possibility is that the geometry optimization procedure implemented in the computational program, which determine a certain step size to be used when trying to find a minimum, may more easily step out of shallow and narrow minima, with low transition energy barriers to alternative stable solutions.

2b. Zn_3F_6. For the trimer system, eight building procedures were considered, i.e., by combining three monomers, two monomers with one Zn and two F atoms, one monomer with two Zn and four F atoms, one of each of the two optimized dimers with either a monomer or with a Zn and 2 F atoms, and finally by combining the atoms individually as units in the system, i.e., treating the system as an atomic cluster. Due to the large amount of approaches used, the results of each build strategy are not discussed in detail. Again, each build method was run five times, after which all viable candidates libraries found in the 40 CLUSTER runs were concatenated into a single library and was filtered using, in this case, a somewhat higher δR of 10 Hartree without an energy window, so that all unique viable candidates were retained.

A subset of the results from this trimer study presented in Table 4 (the full population file is available in the Supporting Information). Isomers 4.2, 4.4, and 4.7 were all encountered in only one of the 40 runs, suggesting that these minima are hard to capture. In all of those cases, the approach of adding a monomer to one of the dimer isomers was the only way in which the novel isomers were located. The other three isomers with relative energies less than 15 kcal/mol (4.1, 4.3, and 4.5)

were found by all of the build approaches used but not necessarily in every run. Closer inspection of the duplicates reported when filtering the viable candidates from all 40 runs shows that when isomer 4.1 was found, it was encountered during the build stage, except in the case where the cluster was built as a pure atomic cluster. This suggests that the clustering procedure does benefit by using larger fragments even though the variety of initial built structures is limited by having less fragments overall. One might consider increasing the number of candidates to build before starting the mating procedure, but the mating procedure can have similar effects as building the candidates with larger fragments to begin with because of the amount of fragments transferred.

Of the isomers that were encountered only once, 4.2—a distortion of 4.1—and 4.4 were both generated during the 3D build stage, while 4.7 was the result of a mating.

3. Weakly Interacting Molecular Clusters. Recently the results of applying the CLUSTER program to ionized radicals of two weakly interacting phenylacetylene molecules were presented as part of a larger project.[25] Here we will present the results of the CLUSTER program applied to water trimers, $(H_2O)_3$. The initial searches were performed five times by building two 1D, five 2D, and 20 3D clusters using H_2O monomers, followed by 30 matings using MOPAC's PM7 semiempirical method with the keyword PRECISE and a δR of 1 Hartree. The viable populations were combined and refiltered with a δE of 0.0008 Hartree (~0.5 kcal/mol). The resulting nine isomers were then reoptimized, using the LDA/TZP-(Large) method and a δR of 1 Hartree.

As shown in Table 5, the lowest energy isomer found, 5.1, is equivalent to that identified already in the literature (see ref 26 and references therein). Only after careful analysis of the

Table 5. Results of the $(H_2O)_3$ CLUSTER Search Using MOPAC and ADF's LDA/TZP(Large) Computational Method[c]

Label	Structure [a]	PM7		LDA/TZP(large)	
		ΔE	R	ΔE	R
5.1 [b]	<image>	0.00	87.4	0.00	89.3
5.2	<image>	1.49	87.2	1.30	88.9
5.3	<image>	3.80	79.2	optimized to 5.1 ccw	
5.4	<image>	4.74	78.4	14.65	76.3
5.5	<image>	4.96	87.0	optimized to 5.1 cw	
5.6	<image>	5.02	83.8	optimized to 5.1 ccw	
5.7	<image>	5.53	84.6	optimized to 5.1 ccw	
5.8	<image>	5.78	78.5	14.61	76.9
5.9	<image>	6.13	82.5	optimized to 5.1 ccw	

[a]Representative geometry (slight differences in the geometry may exist at different levels of theory). [b]This structure has a mirror image that should have the same energy and repulsion resolution. Because of this, structures that converged to 5.1 have a cw (clockwise) or ccw (counterclockwise) designation on the right in the table reflecting the direction in which the bridging OH bonds in the ring point when looking from the side that has the two dangling hydrogens. [c]All ΔE values are in kcal/mol, and R values are in Hartree units.

converged structures library was it found, however, that isomers 5.4 and 5.8 were considered to be duplicates when they really were not, thus suggesting that a smaller δR value was necessary. Refiltering the viable candidates with a δR of 0.5 Hartree, less than 1% of the R values encountered, allowed both of them to remain in the final filtered population.

A complication in this system stems from the fact that isomers that form a triangular arrangement in which all three water molecules contribute a hydrogen to H-bonds that make up the edge, e.g., 5.1 and 5.2, can exist in two stereoisomeric forms that should have the same energy and nuclear repulsion and will thus be considered duplicates. Inspection of the duplicates and converged libraries show that both isomers for 5.1 are indeed encountered and that a slight difference in the final energy due to the geometry convergence settings may determine which stereoisomer ends up in the final population. For instance, the MOPAC isomer for 5.1 was replaced by the stereoisomer that ADF obtained when optimizing isomer 5.3. It is important for users of this and other structure searching protocols and regular *ab initio* optimization programs, as well, to recognize and consider specifically cases where isoenergetic stereoisomers are possible. The CLUSTER program will record these alternatives in the libraries of viable or duplicate structures.

The results suggest that the bonding patterns that one may encounter in clusters of three water molecules may be divided in two general types: triangular and bent. The triangular clusters have, besides possible stereoisomers, differences in how many water molecules donate H atoms involved in the H-bonds along the edges of the ring. In the lowest energy isomers, 5.1 and 5.2, for example, each water molecule participates as an H donor. In the higher energy isomers, 5.5, 5.6, 5.7, and 5.9, however, only two of the three water molecules act as H donors, the other water molecule supplying two lone pairs only, leaving both of its hydrogen atoms dangling outside the ring. The energies of these four isomers span a range of about 1.2 kcal/mol and have very similar geometries; the only difference is a small change in where the plane through water molecule with both dangling hydrogens intersects the opposite edge of the triangle, suggesting that these structures might easily interconvert and may constitute shallow local minima. It is interesting therefore that the DFT optimization of these isomers all result in isomers that belong to the first set, 5.1, thus showing another case were the level of computational rigor affects which isomers are considered stable. Yet the MOPAC results suggest that having a single water molecule provide both hydrogens for hydrogen bonding with another molecule donating two lone pairs for the bonds, e.g. structure 5.5, is less favorable than having each water molecule in the ring system contribute in both fashions, e.g., structure 5.1.

Finally, focusing on the three bent open-chain isomers found at the PM7 level, 5.3, 5.4, and 5.8, we see that the difference between them lies in how many H bond donors the central water molecule provides. The lowest energy isomer, with a single H bond donor from the central atom, optimizes at the LDA/TZP(large) model chemistry to the cyclic isomer, 5.1, while the other cases remain as open-chains but with higher relative energies of 14.6 kcal/mol.

IV. CONCLUSIONS

The CLUSTER program presented in this work generates libraries of local minima on the potential energy surface of clusters consisting of fragments of atoms and or molecules. Capturing the population of optimal candidates is an advantage of the approach used in the program. It has the benefit of creating only a small library of unique structures, without limiting the space sampled. The addition of the nuclear repulsion energy as a second criterion in the uniqueness test in addition to the total energy of the system allows for a finer discrimination between structures that are nearly iso-energetic but structurally distinguished, e.g., the Si_4Li clusters 1.1 and 1.2.

Although we allowed only a rather small number of candidates to be generated in a CLUSTER run, for example (27 initial builds followed by 30 matings), only the Si_4Li_2 case required an extension to the number of matings to capture the same structures that were reported previously in the literature, and we managed to locate in the process, as well, some additional low energy solutions, which in two cases were lower than any of those reported previously.

In several cases the matings that followed the build stage did not add new structures to the population but found only marginally lower energy versions of basically the same geometry. This suggests that in some systems creating a large enough number of candidates in the build procedure and forgoing matings may suffice. One is provided with the option to do so in the software input file if a test of that sort is desirable, though we do not necessarily recommend it.

The ZnF_2 calculations suggest that if one studies oligomers of n molecular units that result in tightly bound systems, the search may benefit from using different approaches in how the cluster is built, e.g., by using n-1 molecular fragments in combination with the atomic version of one of the fragments so as to allow for more flexibility and a better sampling of the potential energy surface. For $n > 2$ one may also consider using an (n-1)mer optimized geometry as one fragment to which a monomer is added.

Weakly interacting molecular clusters can have an abundance of local minima making it nearly impossible to categorically state that the global minimum in the potential energy surface is found in a given search. The population of structures found by the CLUSTER program allows one to find bonding patterns in local minima, which may help to direct the user in systematically building a library of candidates that the CLUSTER program can reoptimize as well as use for further matings.

Finally, in several systems calculated here, it was observed that the computational method chosen affected which minima are, or even can be, found, i.e., a geometry that is a minimum at one level of theory may not be a minimum at a different level of theory. This is well understood, so one is encouraged, therefore, to employ the implementation of the genetic algorithm outlined in this work and other protocols for exploring potential energy surfaces with different model chemistries where it might be crucial to assess the quality of the final population. The CLUSTER program has been coded to partner with a number of common quantum chemical software packages that offer the facilities for the exploration of potential energy surfaces with a wide range of the most efficient and the most reliable model chemistries on the market today.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

The libraries of structures reported in this work, the source code, and the manual for the CLUSTER program. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Authors
*E-mail: rkanters@richmond.edu.
*E-mail: kdonald@richmond.edu.

Notes
The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) Deaven, D. M.; Ho, K. M. *Phys. Rev. Lett.* **1995**, *75*, 288−291.
(2) Wales, D. J.; Doye, J. P. K. *J. Phys. Chem. A* **1997**, *101*, 5111−5116.
(3) Johnston, R. L. *Dalton Trans.* **2003**, 4193−4207.
(4) Ivanov, A. S.; Boldyrev, A. I. *Phys. Chem. Chem. Phys.* **2012**, *14*, 15943−15952.
(5) Heiles, S.; Johnston, R. L. *Int. J. Quantum Chem.* **2013**, *113*, 2091−2109.
(6) Alexandrova, A. N.; Boldyrev, A. I. *J. Chem. Theory Comput.* **2005**, *1*, 566−580.
(7) Alexandrova, A. N. *J. Phys. Chem. A* **2010**, *114*, 12591−12599.
(8) Gaussian, Inc.: Wallingford, CT. http://www.gaussian.com (accessed 10/13/2014).
(9) (a) Perez-Peralta, N.; Contreras, M.; Tiznado, W.; Stewart, J.; Donald, K. J.; Merino, G. *Phys. Chem. Chem. Phys.* **2011**, *13*, 12975−12980. (b) Contreras, M.; Osorio, E.; Ferraro, F.; Puga, G.; Donald, K. J.; Harrison, J. G.; Merino, G.; Tiznado, W. *Chem.—Eur. J.* **2013**, *19*, 2305−2310.
(10) Covalent radius. URL http://www.webelements.com/periodicity/covalent_radius/ (accessed May 5, 2012).
(11) Neese, F. The ORCA program system. *WIREs Comput. Mol. Sci.* **2012**, *2*, 73−78.
(12) Schaefer, A.; Horn, H.; Ahlrichs, R. *J. Chem. Phys.* **1992**, *97*, 2571.
(13) Grimme, S. *J. Comput. Chem.* **2006**, *27*, 1787−1799.
(14) (a) te Velde, G.; Bickelhaupt, F. M.; van Gisbergen, S. J. A.; Fonseca Guerra, C.; Baerends, E. J.; Snijders, J. G.; Ziegler, T. *J. Comput. Chem.* **2001**, *22*, 931. (b) Fonseca Guerra, C.; Snijders, J. G.; te Velde, G.; Baerends, E. J. *Theor. Chem. Acc.* **1998**, *99*, 391. (c) ADF, SCM, Theoretical Chemistry, Vrije Universiteit, Amsterdam, The Netherlands, http://www.scm.com (accessed 10/13/2014).
(15) (a) Stewart, J. J. P. MOPAC, Stewart Computational Chemistry. Web: http://OpenMOPAC.net (accessed 10/13/2014). (b) Maia, J. D. C.; Carvalho, G. A. U.; Mangueira, C. P.; Santana, S. R.; Cabral, L. A. F.; Rocha, G. B. *J. Chem. Theory Comput.* **2012**, *8*, 3072−3081.
(16) XYZ file format is described at the Open Babel Web site. http://openbabel.org/docs/2.3.0/FileFormats/XYZ_cartesian_coordinates_format.html (accessed 10/13/2014).
(17) Jmol: an open-source Java viewer for chemical structures in 3D. http://www.jmol.org/ (accessed 10/13/2014).
(18) ChemCraft: http://www.chemcraftprog.com (accessed 10/13/2014).
(19) Osorio, E.; Villalobos, V.; Santos, J. C.; Donald, K. J.; Merino, G.; Tiznado, W. *Chem. Phys. Lett.* **2012**, *522*, 67−71.
(20) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09, Revision A.02*; Gaussian, Inc.: Wallingford, CT, 2009.
(21) Donald, K. J.; Hargittai, M.; Hoffmann, R. *Chem.—Eur. J.* **2009**, *15*, 158−177.
(22) (a) Ruberto, R.; Pastore, G.; Özen, A. S.; Akdeniz, Z.; Tosi, M. P. *Eur. Phys. J. D* **2012**, *66*, 229:1−8. (b) Özen, A. S.; Akdeniz, Z.; Ruberto, R.; Pastore, G.; Tosi, M. P. *Phys. Lett. A* **2014**, *378*, 431−433.
(23) Hargittai, M. *Chem. Rev.* **2000**, *100*, 2233−2301.
(24) Kaupp, M.; von Schnering, G. *Inorg. Chem.* **1994**, *33*, 2555−2564 and references therein.
(25) Momoh, P. O.; Attah, I. K.; El-Shall, M. S.; Kanters, R. P. F.; Pinski, J.; Abrash, S. A. *J. Phys. Chem. A* **2014**, *118*, 8251−8263 DOI: 10.1021/jp5010488.
(26) Maheshwary, S.; Patel, N.; Sathyamurthy, N.; Kulkarni, A. D.; Gadre, S. R. *J. Phys. Chem. A* **2001**, *105*, 10525−10537.