



HAL
open science

Global Cluster Structure Optimization for Arbitrary Mixtures of Flexible Molecules. A Multiscaling, Object-Oriented Approach.

Johannes Manfred Dieterich, Bernd Hartke

► **To cite this version:**

Johannes Manfred Dieterich, Bernd Hartke. Global Cluster Structure Optimization for Arbitrary Mixtures of Flexible Molecules. A Multiscaling, Object-Oriented Approach.. *Molecular Physics*, 2010, 108 (03-04), pp.279-291. 10.1080/00268970903446756 . hal-00580687

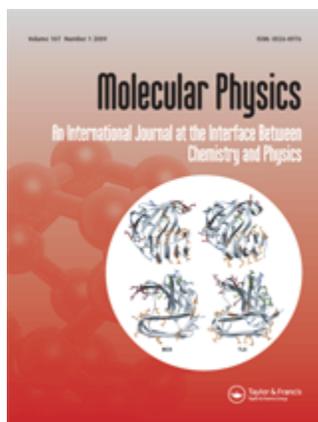
HAL Id: hal-00580687

<https://hal.science/hal-00580687>

Submitted on 29 Mar 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



**Global Cluster Structure Optimization for Arbitrary Mixtures
of Flexible Molecules.
A Multiscaling, Object-Oriented Approach.**

Journal:	<i>Molecular Physics</i>
Manuscript ID:	TMPH-2009-0263.R1
Manuscript Type:	Special Issue Paper - In honour of Prof Werner 60th birthday
Date Submitted by the Author:	21-Oct-2009
Complete List of Authors:	Dieterich, Johannes; Christian-Albrechts-University, Institute for Physical Chemistry Hartke, Bernd; Christian-Albrechts-University, Institute for Physical Chemistry
Keywords:	global optimization, genetic algorithms, cluster structures, Lennard-Jones, Kanamycin A



ARTICLE

*OGOLEM: Global Cluster Structure Optimization
for Arbitrary Mixtures of Flexible Molecules
A Multiscaling, Object-Oriented Approach.*

Johannes M. Dieterich^a and Bernd Hartke^{a*}

^a*Institut für Physikalische Chemie, Christian-Albrechts-Universität,
Olshausenstr. 40, 24098 Kiel, Germany*

(September 2009)

In practical applications, global cluster structure optimization has so far been limited largely to homogeneous clusters of atoms or small molecules, with little or no choice in the calculation of interparticle forces. We eliminate these limitations by presenting a new program suite OGOLEM that is universal by design, both in cluster composition (including arbitrarily heterogeneous clusters of complicated molecules) and in its interfaces to force calculation backends. This is demonstrated by exemplary applications in two novel fields: strongly heterogeneous Lennard-Jones clusters (ternary, quaternary, quinary), and mixed clusters of the aminoglycoside Kanamycin A with sodium cations.

Keywords: global optimization, genetic algorithms, cluster structures, evolutionary computation, Lennard-Jones clusters, Kanamycin A

1. Introduction

Atomic and molecular clusters are important systems in the physical sciences, both for fundamental research [1] and for nanotechnology applications [2]. Structures and properties of sufficiently small clusters differ drastically from the bulk, violating standard chemical intuition. Unfortunately, however, the number of local minima on the potential energy surface for a cluster of n particles increases at least **exponentially [3, 4] with n** . Therefore, unbiased and non-deterministic global structure optimization techniques are required.

To motivate our present contribution to this area, we provide a brief (and possibly biased) sketch of the recent history in this area. One of the first general algorithms employed for this task was simulated annealing (SA) [5]. As a hopefully less compute-intensive approach, genetic algorithms (GA) were first applied to the global cluster structure optimization problem in a simple form [6]. After several algorithmic improvements, most notably the phenotype approach [7] as well as directed mutations and structural niches [8], it could be shown that the resulting evolutionary algorithm (EA) scaled only cubically in the standard benchmark application of homogeneous Lennard-Jones (LJ) clusters [8]. For several **years, basin-hopping [9–11]** and this phenotype EA were the only unbiased algorithms able to find all known LJ_n minima up to $n \approx 150$. More recent developments [12–14] have pushed this limit further up to $n = 561$, by introducing dynamic

*Corresponding author. Email: hartke@phc.uni-kiel.de

1 constructions of intermediate grids for particle placement as a further key ingre-
2 dient and by smaller variations of the basic techniques. By now the Cambridge
3 Cluster Database [15] lists LJ structures up to $n = 1000$, but inclusion of a-priori
4 information on the preferred grids (icosahedral and decahedral) is still needed for
5 the upper half of this size range [16]. Also, the ultimate question at which cluster
6 size the inevitable transition from icosahedral or decahedral structures to bulk-like
7 grids (fcc or hcp) occurs is still open, although explicit investigations of LJ cluster
8 phase change behavior has progressed to impressive sizes [17].

9
10 In the wake of this algorithmic development, both EA and basin-hopping meth-
11 ods have been applied to many kinds of homogeneous atomic clusters throughout
12 the periodic system. The situation is different for heterogeneous clusters. There has
13 been some limited work on binary LJ clusters [18–20] and binary Morse clusters [21]
14 recently. Employing the Gupta potential, the Johnston group has studied several
15 bimetallic cluster systems [22], including cluster sizes that are special for the LJ
16 system, e.g., $n = 38$ [23] and $n = 98$ [24]. Literature on ternary clusters is extremely
17 limited; in fact, we could locate only one significant paper [25] which apparently
18 is starting off this line of research. (To avoid misunderstandings, we would like to
19 emphasize that there are also related but traditionally separate research areas, for
20 example inorganic chemistry style mixed metal clusters with bulky outer ligands,
21 as those featuring in Refs. [26, 27]; such systems are not of central interest here
22 since for their successful treatment a proper level of *ab initio* electronic structure
23 theory is more important than the performance of a global structure optimization
24 scheme.) Clearly, one additional challenge of heterogeneous clusters is provided by
25 the additional degrees of freedom due to exchange of atom types within otherwise
26 identical structures, and the tight interplay of this with varying structural prefer-
27 ences. Another possibly less obvious aspect is the necessity to make the transition
28 from experimental global optimization codes tuned to deal with one or very few
29 given test systems to a general production code with the ability to handle arbitrary
30 atom types with arbitrary interaction models.

31
32 Of course, the LJ potential is not an exact potential for anything, including rare
33 gases. Likewise, other empirical potentials always are only crude approximations.
34 Therefore, frequent attempts have been made to perform global cluster structure
35 optimization directly at some better level of theory (as arbitrarily selected exam-
36 ples, see Refs. [28, 29]). Such attempts have to be viewed with caution, since this
37 causes energy (and gradient) evaluations to become more expensive by several or-
38 ders of magnitude, which seriously compromises the convergence properties of any
39 global optimization scheme. It is widely recognized by now that the global search
40 should be done with a cheap model potential (or at most with a cheap semiempir-
41 ical method), with the results then being refined at a suitable higher level of
42 theory. An automatic scheme has been proposed to generate a system-specifically
43 adjusted model potential on the fly [30], but this scheme has seen only a few ap-
44 plications [31, 32] so far. In any case, this predicament indicates a pressing need
45 for any global optimization strategy to be linked up easily with a broad spectrum
46 of energy/gradient calculation methods.

47
48 Already the second application of GA methods to global cluster structure opti-
49 mization was on benzene clusters [33]; followed by a few other studies of this system
50 [34–36]. However, the vast majority of molecular cluster optimizations performed
51 by EA methods were concerned with water clusters, in most cases with homoge-
52 neous neutral ones [4] (see the comprehensive overview in Ref. [37]), in rarer cases
53 with water clusters containing a single different entity, e.g., an atomic ion [38–41].
54 Applications to clusters of other molecules, in particular to heterogeneous clusters
55 of more complicated, flexible molecules, seem to be lacking entirely, with one ex-
56
57
58
59
60

ception: prediction of crystal structures, either *ab initio* (see the review [42] and references cited therein) or using powder diffraction data [43, 44]. In these cases, molecules more complicated than water or benzene have been studied, but in most cases with only very few dihedral angles (if any) that were actually allowed to vary. In addition, it can be argued that crystal structure prediction is somewhat simpler than cluster structure prediction, due to the necessary limitation to the known 230 space groups which allows to treat part of the problem by exhaustive enumeration. Nevertheless, again, the employment of suitably tailored model potentials is seen as a major asset [45].

From this historical sketch, the remaining challenges for global cluster structure optimization can be identified as: (i) heterogeneity with ≥ 3 species, (ii) flexible molecules with many internal degrees of freedom, and (iii) easy access to a broad variety of methods to calculate interparticle forces, from a variety of cheap model potentials to more expensive semiempirical and *ab initio* methods, if possible also with an option to tune a given model potential to the present system under study.

We have developed a new EA program suite OGOLEM that fulfills all of these requirements. It uses the experience gained from our previous, specialized implementations [8, 39] but is written from scratch and extended to cover points (i)–(iii). Design and performance details are described in section 2. Section 3 provides application examples that highlight features (i)–(iii), namely heterogeneous atomic LJ clusters of more than three species and heterogeneous clusters of larger, more complicated molecules. Coordinate files for all clusters shown there can be obtained from the authors upon request.

2. Program Design

Nowadays, the most important feature of any program is scalability. This, in the advent of multicore processors, includes both SMP as well as MPP parallelization. Due to their inherent parallelizability, genetic algorithms are a well-suited algorithmic choice.

To fulfill this kind of multiscalability, the OGOLEM framework provides both a MPI frontend for MPP parallelization and a threading frontend for SMP parallelization omitting unnecessary MPI overhead. In both cases, a pool algorithm was used to eliminate serial bottlenecks [46].

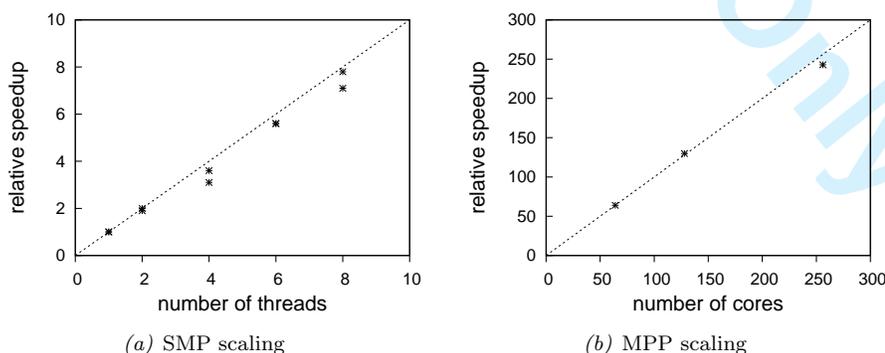


Figure 1. Scaling shown as relative speedup. Ideal linear scaling shown as line. Deviations arise from the heavy use of random numbers.

As Fig. 1 shows, the OGOLEM framework therefore provides excellent scaling both on SMP systems as well as in an MPP context. Our pool algorithm in combination with JavaTM's build-in threading mechanisms provides a straightforward

1 SMP parallelization. Therefore, the OGOLEM framework is well-equipped for the
2 even further growing importance of multicore processing.

3 Directly after scalability comes usability. This can be considered both from the
4 developers point of view, wanting a framework to allow for rapid development
5 cycles and containing clear yet flexible restraints, and from the users perspective,
6 needing easy input and flexibility in algorithmic and methodological choice. The
7 OGOLEM framework satisfies the developers needs through a rigorous object-
8 oriented programming style and well-defined but flexible interfaces. The user is
9 provided with an intuitive and easy input format and can choose between a variety
10 of methods and algorithms for her problem. These algorithms include
11

- 12 • a straightforward genotype operator,
- 13 • a phenotype operator,
- 14 • a novel packing algorithm,
- 15 • combinations of the above.

16
17
18 The *genotype*-Operator provides an unbiased *1-point-crossover*, as described in
19 Ref. [47] but acting on real-number strings. The *phenotype* operator cuts a ran-
20 domly rotated cluster using either a plane through the center of mass or, option-
21 ally, a plane randomly displaced from that one (similar to the operator described
22 in Ref. [8]).

23 The OGOLEM framework also provides a pairwise collision detection engine and,
24 as a novel development, a dissociation detection engine. The latter employs the effi-
25 cient *Warshall* algorithm[48] to build a reachability matrix. These two ingredients
26 are of crucial importance in the whole global optimization process, in particular for
27 clusters containing larger molecules of non-trivial shape: Both collision detection
28 and dissociation detection are activated before local optimization, discarding struc-
29 tures containing clashes or being dissociated. Since local optimization is responsible
30 for $\gg 90\%$ of the overall computer time [8], this erases much of the computational
31 effort otherwise wasted into local optimizations that typically show both a very
32 slow convergence behavior and non-competitive final energies.

33 The development of our novel packing algorithm was inspired by the design of
34 new scheduling algorithms in the field of computational sciences targeted at solving
35 the *NP-hard* problem of finding an optimal dynamic packing of tasks. The main
36 idea there is to order the tasks to be executed by their relative size (where the
37 size definition is crucial and might contain priorities). The biggest tasks are then
38 packed first and the smallest ones last, into the holes left by the bigger ones.
39 This is similar to the classical *knapsack* problem[49], and it can be expected that
40 optimal packing is a decisive ingredient also in global cluster structure optimization,
41 quite independent of the interparticle potential(s). Therefore, we make use of the
42 same idea and order the building blocks of the cluster by their size and pack
43 them as described above. Performance benefits arise in particular in systems of very
44 different building block sizes, as demonstrated in section 3.2. As a side effect, the
45 combination of packing with classical genetic algorithms helps to avoid premature
46 convergence of the latter, since new structures with a fresh *genome* are, with a
47 certain probability, build from scratch, potentially bringing new genes to the pool.

48 Available methods include interfaces to

- 49 • AMBER,
- 50 • DFTB+,
- 51 • MNDO,
- 52 • MOLPRO,
- 53 • MOPAC,
- 54 • NAMD,

- Orca and
- Tinker

and additional internal force fields. This includes force fields for homoatomic and heteroatomic Lennard-Jones clusters and for ionic clusters with the additional ability of system-specific global reparametrization. **With some standard force fields, optimization of internal degrees of freedom could conceivably lead to force field definition gaps (bond dissociation, change of atom type) or to clashes. The former problem can be avoided by user input, which necessarily has to include bond definitions and lower/upper bond length limits; for the latter problem, re-usage of the collision detection engine is the obvious solution.**

The choice of method is made simplistic yet transparent for the user by specifying program, method and, if applicable, further information like basis set or solvation model in a single input line. To give some examples, this starts with choosing AMBER as a backend, of course requiring an AMBER-style force field being available for this system, which can be used for bigger systems of biological relevance as shown later in this paper. Then there is, for smaller systems where a better description is needed, the possibility to choose MNDO's semiempirical methods. With MNDO's ability of using COSMO-style solvation for the system, this extends the possible input by an easy addition of the `cosmo` keyword to enable automatic water solvation. For even finer needs, e.g., the Orca interface provides combinations of quantum chemistry methods, like different DFT-functionals as well as *ab-initio* methods, and basis sets.

Through this selection of different program packages and methods therein, the user is likely to find a method for her system that provides the best compromise between accuracy and computational expense. Additionally, the user can (but does not need to) control every part of the global optimization in detail.

Completed are these features with platform-independence. Employing a pure Java™ approach, we can ensure platform-independence in principle, ranging from embedded systems to supercomputers, regardless of architecture and operating system. The only limitations arise from (i) the ratio between computational power and problem size concerning the architecture, e.g., MP2/aug-cc-pVTZ optimizations on embedded systems and (ii) the availability of, often binary, program packages on the specific platform, e.g., Microsoft Windows™ or the FreeBSD™ operating system.

We distribute the OGOLEM framework under a classical *four-clause* BSD-license to provide a high degree of freedom.

3. Applications

3.1. Mixed Lennard-Jones Clusters

The pairwise Lennard-Jones potential

$$v_{ij} = 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad (1)$$

provides a tolerably good description for clusters of homoatomic rare-gas atoms and can easily be extended to heteroatomic clusters using the Lorentz-Berthelot

1 mixing rules

$$2 \quad 3 \quad 4 \quad 5 \quad 6 \quad 7 \quad 8 \quad 9 \quad 10 \quad 11 \quad 12 \quad 13 \quad 14 \quad 15 \quad 16 \quad 17 \quad 18 \quad 19 \quad 20 \quad 21 \quad 22 \quad 23 \quad 24 \quad 25 \quad 26 \quad 27 \quad 28 \quad 29 \quad 30 \quad 31 \quad 32 \quad 33 \quad 34 \quad 35 \quad 36 \quad 37 \quad 38 \quad 39 \quad 40 \quad 41 \quad 42 \quad 43 \quad 44 \quad 45 \quad 46 \quad 47 \quad 48 \quad 49 \quad 50 \quad 51 \quad 52 \quad 53 \quad 54 \quad 55 \quad 56 \quad 57 \quad 58 \quad 59 \quad 60$$

$$\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j} \quad (2)$$

$$\sigma_{ij} = \frac{\sigma_i + \sigma_j}{2} \quad (3)$$

to obtain the mixed parameters ϵ_{ij} and σ_{ij} from their homoatomic equivalents. These rules have proven their applicability, e.g., in review [50].

As mentioned in section 1, one of the major problems with combinations of various atomic types is the exchange of atom types. To target this, we use an explicit *XChange*-Operator, providing a significant speedup in finding the global minimum over non-explicit exchange algorithms. For all computations in this section, we combine the *XChange*-Operator with an unbiased *genotype*-algorithm.

For homoatomic Lennard-Jones (LJ) clusters, there are, with few exceptions, well-defined and recurring structural units. One interesting question therefore is whether changes in composition will yield structural transitions at a given cluster size or not. As an example we looked into clusters composed of 19 LJ atoms. The homoatomic LJ cluster has a double **icosahedral** shape in the global minimum, as depicted in Fig. 2. In the following three heteroatomic cases, we keep the total

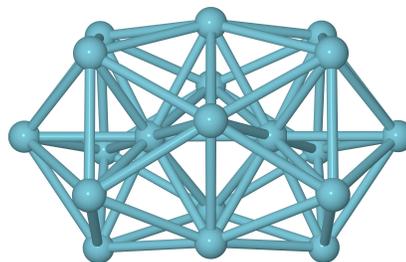


Figure 2. An Ar_{19} cluster having a double icosahedral shape. This and all following cluster structure figures were generated using Jmol [51] and POV-Ray [52].

number of 19 atoms but start to modify the cluster composition. In all calculations, we rank our pool by energy, meaning that “rank 0” is our best guess for the global minimum, “rank 1” our best guess for the next-higher local minimum, etc. To all our experience, the low-energy end of such a ranked pool provides a very reliable impression of the true energy spectrum. Of course, the high-energy end will not be representative at all, due to the small size of the pool compared to the vast search space. Thus, looking at ranks 0–10, we are very confident that we are not missing any important structures within this energy range.

Our introductory example for mixed LJ clusters is a simple substitution of two argon atoms with xenon atoms, i.e., $\text{Ar}_{17}\text{Xe}_2$. This exchange modifies the composition only slightly, and the difference in pair potential characteristics (well depth and minimum-energy distance) between Ar and Xe is not large. Since there is a large energy difference between the double icosahedron and other cluster shapes in the homoatomic case, it is not surprising that this type of exchange does not have a big impact on the shape of the cluster: As presented in Fig. 3, the global minimum of $\text{Ar}_{17}\text{Xe}_2$ still is a double icosahedron, only slightly distorted with respect to the homoatomic case, and retaining much of its symmetry. Obviously, instead of substituting the two interior argon atoms (the centers of the two icosahedra), the two xenon atoms are on the cluster surface. This can be rationalized

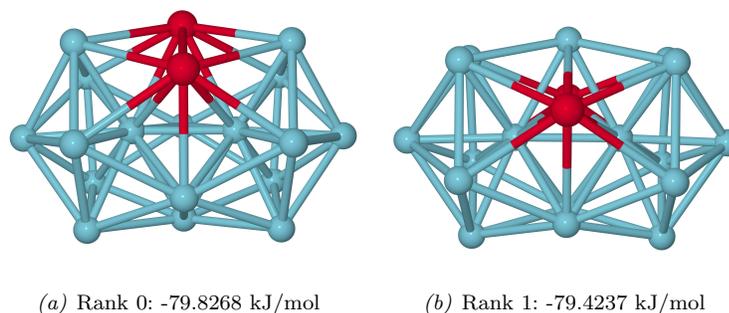


Figure 3. Minimum structures of the $\text{Ar}_{17}\text{Xe}_2$ cluster having a slightly distorted double icosahedral shape.

by remembering the fact that if an icosahedron is built from perfect tetrahedra, small gaps remain on the outer surface. In other words, the larger xenon atoms are more easily accommodated on the surface than in the interior. As in nanoalloy clusters [22], another important issue is atom type mixing vs. segregation. Here, interestingly, the two xenon atoms stay next to each other, probably because this allows for the strain induced by the size difference to be compensated by bending the long axis of the double icosahedron which only has a negligible impact on most of the nearest-neighbor pair distances.

This is confirmed by comparing to the lowest local minimum (rank 1 in Fig. 3): There, the two xenon atoms are separated but otherwise in very similar positions and surroundings. For the rank 0 structure, the long axis is bent by 5.6° ; for the rank 1 structure, this value is 3.2° , implying that a larger portion of the size-difference strain has to be compensated by other distortions, leading to a greater percentage of pair distances being slightly away from their optimal values, which in turn leads to a worse overall energy. It has to be pointed out, however, that these effects are very small, since the total energy difference is only 0.5 kJ/mol, or half a percent of the overall binding energy. This is a good indicator for the performance of the OGOLEM suite, and also a reminder of the importance of atom type exchange operators, as already noticed in the literature [22].

Of course, we can expect the energy landscape to get more complicated with every inclusion of more atom types. Considering the raw amount of possible compositions and their combinatorial explosion, already for ternary clusters any attempt at a complete treatment would be beyond the scope of this article. For this reason, we restrict the discussion to a few exemplary results, trying to give explanations why certain cluster structures are superior to others and which computational challenges are important in these systems.

Looking into the quaternary LJ cluster $\text{Ar}_2\text{Kr}_5\text{Xe}_5\text{Ne}_7$, we can spot three different types of structures, within the best ten structures of our genetic pool and less than 0.5 kJ/mol apart. These three structures are depicted in Fig. 4; all of them show essentially the same distorted double icosahedral structure. The atom type distribution in the global minimum structure appears to be “messy” at first sight; it can, however, be rationalized a posteriori: One of the two icosahedra in the double icosahedral structure has the composition $\text{Ar}_2\text{Kr}_5\text{Xe}_5\text{Ne}_1$, with the biggest Xe atoms on one end, a middle layer of middle-sized Kr atoms, and the smallest Ar and Ne atoms at the other end. In effect, this brings all but one Xe atom into the central plane of the double icosahedron; there, they can participate in filling in the tetrahedral-packing gaps in both icosahedra. Then, the remaining six Ne atoms have two extreme options for forming the second icosahedron: They could either arrange around the single Ne atom end of the first icosahedron, or around

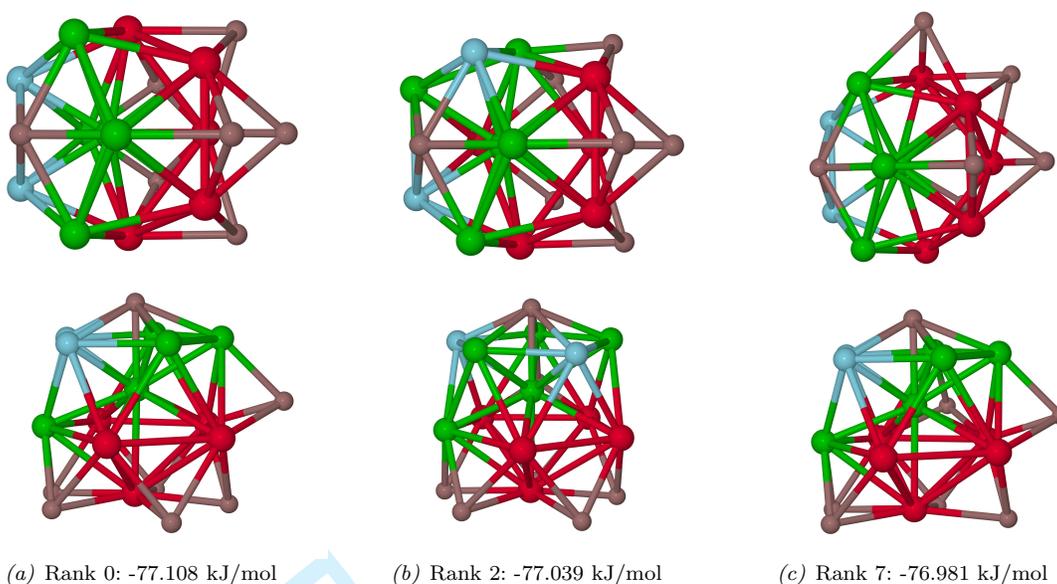


Figure 4. Geometries of the $\text{Ar}_2\text{Kr}_5\text{Xe}_5\text{Ne}_7$ cluster, each displayed in two different views (top, bottom).

the Xe atom end. The first option has the advantage that the size discrepancies at the border between the two icosahedra are minimized; the second option allows for energetically more favorable Ne-Xe contacts. It turns out that this second option wins, but at the expense of the system being unable to close this Ne icosahedron. Instead, the Ne atom that should sit at the far end of the long axis moves into a second geometrical layer, capping a Xe-Xe-Kr triangular face (i.e., in effect, five Ne-Ne interactions at distances much longer than the optimal one are replaced by one Ne-Xe and one Ne-Kr interaction at optimal distances).

The other two structures shown in Fig. 4 turn out to be very small variations of this rank 0 structure just described. From the rank 0 to the rank 7 structure, one additional Ne atom “jumps” from the Ne icosahedron into the second layer; no other changes occur, and the total energy decreases only by 0.13 kJ/mol (0.16%). Despite of these minor differences, disturbingly, six other structures are ranked in-between these two, sometimes even with larger structural changes. As an example for these cases, we show the rank 2 structure. It differs from the rank 0 one in two respects: (i) the second-layer Ne atom moves to a different triangular face (but of the same composition, Xe-Xe-Kr, so the effect of this hop is very small), and (ii) one Ar-Kr pair switches its identity, leading to segregation of the two formerly neighboring Ar atoms, and to several changes in the numbers of nearest-neighbor contacts per atom type pairs. Nevertheless, the overall energy change is even smaller than for the single-atom hop from rank 0 to rank 7. As a side note, the overall energy favours the most symmetric structure out of these three closely lying possibilities. This discussion further documents the considerably increased difficulty of the mixed LJ problem and the utmost importance of unbiased but efficient search tools.

As an example representing the class of quinary LJ clusters we picked the $\text{Ar}_5\text{Kr}_5\text{Xe}_5\text{Ne}_2\text{He}_2$ cluster. As before, Fig. 5 displays three typical structures from the top ten in our pool. In comparison to the quaternary case, some patterns start to emerge. Again, all of these three quinary structures are almost identical, with only very few atom hops or atom type exchanges, as discussed below. Furthermore, in spite of the fairly different atomic composition, there are many similarities between these quinary structures and the quaternary ones discussed above. First of

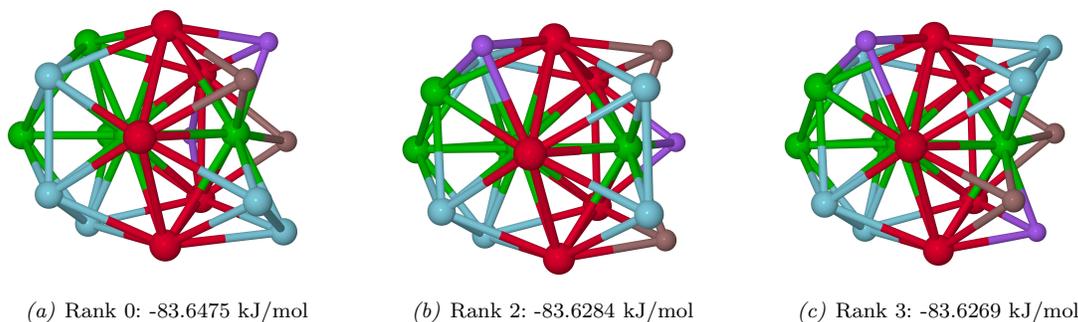


Figure 5. Geometries of the $\text{Ar}_5\text{Kr}_5\text{Xe}_5\text{Ne}_2\text{He}_2$ LJ cluster.

all, the general cluster framework is the same: It is again a double icosahedron, with one long-axis cap missing due to an atom having hopped into a second layer. Also, in both composition cases, the largest atom type (Xe) is arranged preferentially on the middle equator, joining the two icosahedra. The remaining atom types are distributed such that one icosahedron preferentially contains larger atoms, the other one smaller atoms (despite the fact that more symmetrical distributions could be realized, in particular in this quinary case). Finally, the atom hopping into the second layer is of the smallest type (here a He atom, above a Ne atom) and sits on a Xe-Xe-Kr triangular face in all cases.

Also the differences between the quinary structures shown in Fig. 5 are qualitatively very similar to those discussed for the quaternary case. From the rank 0 structure to the rank 2 structure, an argon gets exchanged with a krypton, splitting the three adjacent argon atoms in the larger-atom icosahedron into a 2+1 segregation. This is accompanied by the outer He atom hopping to a different Xe-Xe-Kr triangular face. From the rank 2 to the rank 3 structure, the geometrical change is even smaller: It merely consists of turning one of the icosahedra relative to the other one around the long axis by 72° (one fifth of a full turn around this five-fold axis; i.e., in the homoatomic case, these two clusters would be identical). For this particular atom type distribution, this rotation operation does not change the count of nearest-neighbor interactions for the different atom-type pairs. Hence, the overall energy difference arises only from non-nearest neighbor interactions. This is in line with the fact that this energy change is very small (0.0015 kJ/mol or 0.0018%). Again, disturbingly, the seemingly larger change upon going from the rank 0 to the rank 2 structure has an even smaller overall effect on the energy.

Nevertheless, as Doye, Miller and Wales [53] have shown, for 19 LJ atoms (in the homogeneous case) there is a particularly large energy difference between the (double) icosahedral global minimum and the best clusters of the other structural prototypes (decahedral and fcc). Therefore, as already mentioned above, it is not surprising that we have not found a fundamental structural change (away from the double icosahedron) even in the quaternary and quinary cases shown above. At other total atom counts, however, the difference between the structural prototypes is minimal, in particular at those famous sizes where the global minimum structure of small LJ clusters is not of the dominating icosahedral type anymore. The first such size is 38 atoms [53]. Therefore, for LJ_{38} , we expect to see major structural changes already for comparatively small changes in cluster composition. As a reminder, for the homogeneous case, the global minimum structure is a fcc packed one (see Fig. 6), despite almost all of the search space being dominated by icosahedral structures. For this reason, the fcc global minimum of LJ_{38} is non-trivial to find already in the homogeneous case. Therefore, we cannot yet present

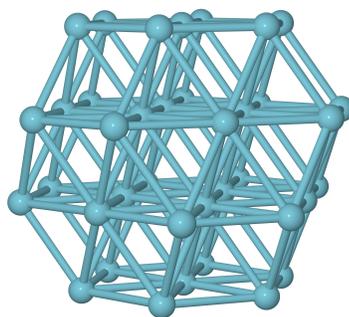


Figure 6. The global minimum of the Ar_{38} cluster, a *fcc* packing.

final results in this first report. Nevertheless, our preliminary results do show an obvious tendency towards icosahedral structures upon going to mixed LJ clusters with a total of 38 atoms. As a typical example, we show our best candidate for the global minimum of $\text{Ar}_{30}\text{Xe}_4\text{Kr}_4$ in Fig. 7. As to be expected from the fairly

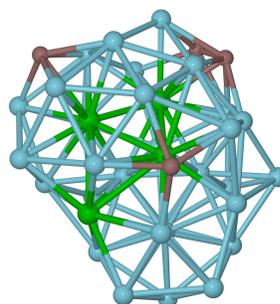


Figure 7. The global minimum of the $\text{Ar}_{30}\text{Xe}_4\text{Kr}_4$ cluster, clearly not *fcc*.

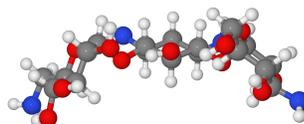
intricate discussion for the various LJ_{19} cases above, this structure is difficult to analyze. Nevertheless, the presence of (approximately) five-fold axes is strikingly obvious, ruling out the possibility that this could be a cut-out from an *fcc* crystal. Actually, at least one pattern recognized above for LJ_{19} of various compositions can be found here again: The largest atoms (Xe) preferentially sit at the boundary between (complete or partial) icosahedral motifs. Further discussion is referred to future publications of this ongoing work.

3.2. *Kanamycin A Aggregation*

In recent experimental studies in the institute for dermatology at the university of Kiel [54], larger aggregates of compounds related to Kanamycin A (KA) with certain physiological cations have been implicated as pathogen-associated agents, leading to a non-inflammatory innate immune response of the human skin [55]. Apparently, already for the parent compound KA, with the exception of a few isolated hints [56–58], nothing is known about molecular aggregates, neither experimentally nor theoretically. Therefore, we are currently investigating clusters of this compound class. Full details will be published elsewhere [59]. Here we show first results on small clusters of the parent compound KA, to demonstrate the capabilities of the OGOLEM suite in dealing with more complicated species containing internal degrees of freedom with non-trivial influence on the outer molecular shape and properties.

One of the main questions with this system is the role of physiological ions in

1 the aggregation process. As a prelude to future studies, we therefore look into the
2 differences arising from different KA:cation ratios in small clusters. All calculations
3 shown here were carried out with NAMD[60] and AMBERs GAFF[61, 62] force
4 fields as a backend, and with the packing operator in 1:1 mixture with a classical
5 *genotype* algorithm. As a starting point, a pre-optimized KA geometry using
6 Orca's B3LYP/TZVP[63, 64] was taken (see Fig. 8). Qualitatively, there are
7
8
9



10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
Figure 8. The starting point, a B3LYP/TZVP optimized monomer structure.

no structural differences between this DFT structure and the GAFF-optimized ones occurring during our global optimization runs. This gives some indication that this force field description is not totally inadequate. We do not claim, however, that this is the optimal or even a reliable description. A thorough inquiry into this question will be presented elsewhere [59]. For our present purposes, a possibly reasonable description is fully adequate. All claims made in this section should be taken as valid only within the particular force field description used here.

As an introductory example, we show the aggregate of two KAs with a single sodium cation. This is not only the smallest cluster possible of the described kind but there are also experimental indications supporting the hypothesis that this may be an important building block for clusters of bigger size. As can be seen in Fig 9, the two KAs encapsulate the sodium cation in a manner similar to two chelate ligands, coordinating it with eight oxygen atoms within 3.2Å distance.

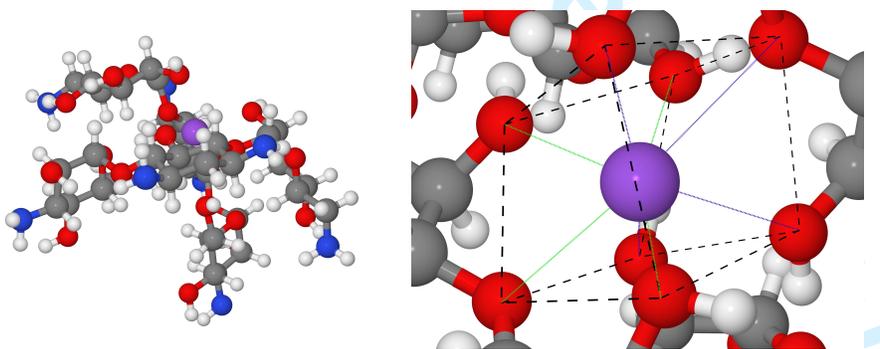


Figure 9. The aggregate KA_2Na^+ and a close-up of the coordinative region, indicating its tetragonal-prismatic form.

In the KA-Na-KA closest contact area, not all Na-O distances are equal, instead they range from 2.4 Å to 3.11 Å. Since the next nearest possible coordination site is much farther away (4.2 Å), it does seem to be sensible to classify this as an eightfold coordination, in the form of a distorted tetragonal prism. While this is usually perceived as over-coordination for the comparatively small Na cation in aqueous biochemistry [65], it has been observed in zeolites [66] and inorganic compounds [67]. Of course, as indicated above, it could also be a force field artifact. This will be checked in ongoing studies.

Looking into larger clusters, we would like to present KA tetramers both with a single sodium cation and with two sodium cations, showing differences arising from the addition of a single sodium cation to the system. As can be seen from Fig. 10 and the close-up provided there, the rank 0 structure for KA_4Na^+ again coordinates the sodium cation with eight oxygens within 3.2 Å distance and with the same structural pattern. This similarity with the KA dimer could be a first support for the suspicion that KA_2Na^+ could be a building block in larger KA-cation clusters.

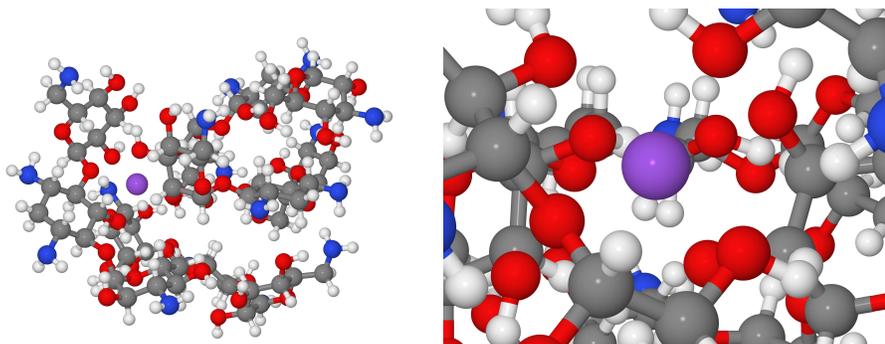


Figure 10. The aggregate KA_4Na^+ (Energy: 237.305 kJ/mol) and a close-up of the coordinative region.

The KA_4Na^+ system turns out to be very stable: In all our global optimization runs, not a single dissociated structure was found. This changes drastically upon addition of a second sodium cation, generating $KA_4Na_2^{2+}$. With two sodium cations (and neither counterions nor explicit or implicit solvent), the KA tetramer has a rather high tendency to dissociate. In fact, without employing our new packing algorithm as additional operator, it is very hard to find any non-dissociated structures at all. Even with packing, enabled the fraction of dissociated structures in a typical genetic pool after global optimization typically is at or above 50%, as shown in Fig. 11. Nevertheless, with packing as essential ingredient, we can effi-

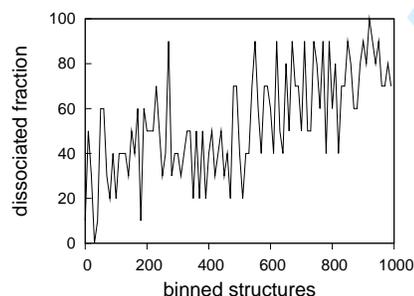
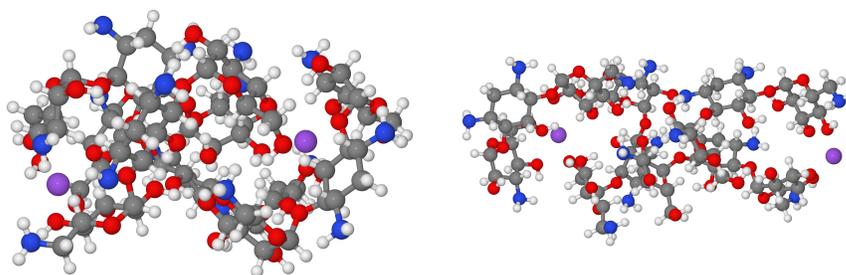


Figure 11. Dissociation ratios as a function of the binned structures of the genetic pool.

ciently find structures for $KA_4Na_2^{2+}$ that are not only fully associated but also very compact. One of them is our best candidate for the global minimum structure of this system (Fig. 12a). Even in this global minimum candidate structure, there are some indications that four KA molecules have difficulties in holding two positively charged sodium cations at a distance of only 9.6 Å. If we assume that the eight-fold coordination described for KA_2Na^+ above is ideal, this coordination should also be possible to attain in $KA_4Na_2^{2+}$, since this cluster simply is $(KA_2Na^+)_2$. Clearly, however, the coordination surroundings of both sodium cations are different in our rank 0 structure: Both are only sixfold coordinated. Apparently, the



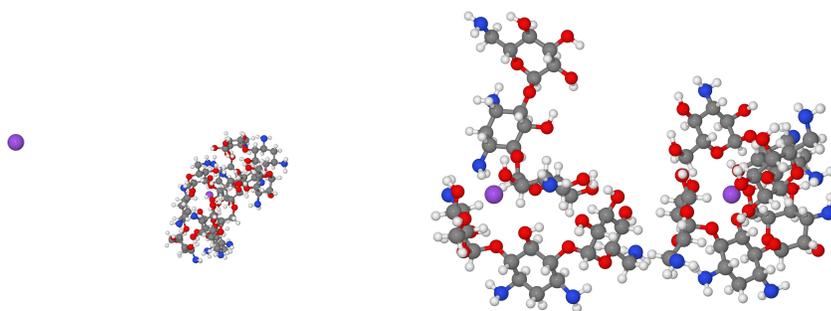
(a) Rank 0: 253.1852 kJ/mol

(b) Rank 255: 353.4694 kJ/mol

Figure 12. Fully associated structures of the $\text{KA}_4\text{Na}_2^{2+}$ cluster.

less-than-optimal sodium coordination is partially compensated by favorable KA-KA non-bonding contacts. As support for this claim we also show another type of fully associated structure in Fig. 12b. There, KA_4 formation is obviously very loose, and at the same time the sodium coordination clearly is deficient. As a result, the total energy is even worse than for many dissociated structures (shown in the following paragraph).

Apparently, there are two basic modes for $\text{KA}_4\text{Na}_2^{2+}$ to dissociate. One is loss of an isolated sodium cation, leaving a stable KA_4Na^+ behind. The other one is dissociation into two KA_2Na^+ building blocks. Typical candidates for both classes are shown in Fig. 13. It is not surprising but typical that the $\text{KA}_4\text{Na}^+ + \text{Na}^+$ case



(a) Rank 110: 332.0423 kJ/mol

(b) Rank 480: 386.1038 kJ/mol

Figure 13. Dissociated structures of the $\text{KA}_4\text{Na}_2^{2+}$ system.

has a better energy than the $\text{KA}_2\text{Na}^+ + \text{KA}_2\text{Na}^+$ case. Association of KA with itself happens via multiple hydrogen bonds, which overcompensates the loss of 5–8 Na-O coordinative contacts. In fact, $\text{KA}_4\text{Na}^+ + \text{Na}^+$ occur frequently in our pools, while $\text{KA}_2\text{Na}^+ + \text{KA}_2\text{Na}^+$ cases are hard to find.

Preliminary results for $\text{KA}_4\text{Na}_3^{3+}$ indicate that despite our packing algorithm we are unable to find non-dissociated clusters for this system. An obvious but still speculative rationalization for this finding could be that four KA molecules in compact formation simply do not provide enough room and a sufficiently high number of coordination contacts to keep three sodium cations within a pair distance of 9 Å, which is the minimum sodium atom pair distance observed in $\text{KA}_4\text{Na}_2^{2+}$.

4. Summary and outlook

After identifying particle type heterogeneity, larger flexible molecules, and variability on calculating interparticle forces as the remaining challenges in global cluster structure optimization, we have presented our OGOLEM program that is designed to meet all these challenges, resulting in a general and user-friendly framework for diverse practical applications.

As examples demonstrating these abilities we have first proceeded beyond the standard benchmark of homogeneous LJ clusters and the beginning research on binary systems, by making first steps into the area of ternary, quaternary and quinary LJ clusters. Depending on cluster size, we find both cases where the qualitative cluster structure is (almost) unchanged compared to the homogeneous case and others where even small changes in atom type composition lead to drastic structural transitions. Many of the observed effects occur within very narrow energy windows, re-emphasizing the need for particle type exchange operators, incorporated into OGOLEM.

As second demonstration system, we have examined small clusters of KA with sodium cations, as an example for heterogeneous clusters containing particles of very different kind, including a large, flexible species. Here it turned out that another new ingredient, a packing algorithm, is vital for an efficient treatment.

With these exemplary applications, we have demonstrated the ability of OGOLEM to treat complicated systems in global cluster structure optimization. This opens up a plethora of possible future applications. As already indicated in section 3.2, the KA aggregation results shown here only are the very first steps into a complex project. Further ongoing applications include Sb/Ge vanadates featuring curious arrangements of vanadate polyhedra, with possible future use as molecular magnets, and the influence of small structural variations on micelle building propensities in glycosylated lipids. Work on these and other systems is in progress in our lab.

Acknowledgements

On this occasion, both authors would like to thank Prof. Dr. Hans-Joachim Werner for his generous support. **Both authors are also grateful to Prof. Dr. Jens-Michael Schröder for his enthusiastic initiative in starting a collaboration on KA aggregates.** A large computer time grant at the HLRN supercomputer center (Berlin/Hannover) is gratefully acknowledged.

References

- [1] D. J. Wales, *Energy Landscapes* (Cambridge University Press, Cambridge, 2003).
- [2] A. M. Bittner, *Surf. Sci. Reports* **61**, 383 (2006).
- [3] M. R. Hoare and J. McInnes, *Adv. Phys.* **32** (1983) 791.
- [4] B. Hartke, *Z. Phys. Chem.* **214**, 125 (2000).
- [5] S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi, *Science* **220**, 671 (1983).
- [6] B. Hartke, *J. Phys. Chem.* **97**, 9973 (1993).
- [7] D. M. Deaven, and K. M. Ho, *Phys. Rev. Lett.* **75**, 288 (1995).
- [8] B. Hartke, *J. Comput. Chem.* **20**, 1752 (1999).
- [9] Z. Li, and H. A. Scheraga, *Proc. Natl. Acad. Sci. USA* **84**, 6611 (1987).
- [10] D. J. Wales, and J. P. K. Doye, *J. Phys. Chem. A* **101**, 5111 (1997).
- [11] D. J. Wales, and H. A. Scheraga, *Science* **285**, 1368 (1999).
- [12] J. Lee, I. H. Lee, and J. Lee, *Phys. Rev. Lett.* **91**, 080201 (2003).
- [13] X. G. Shao, L. J. Cheng, and W. S. Cai, *J. Comput. Chem.* **25**, 1693 (2004).
- [14] H. Takeuchi, *J. Chem. Inf. Model.* **46**, 2066 (2006); **47**, 104 (2007).
- [15] <http://www-wales.ch.cam.ac.uk/CCD.html>
- [16] X. L. Yang, W. S. Cai, and X. G. Shao, *J. Comput. Chem.* **28**, 1427 (2007); **29**, 1772 (2008).

- 1 [17] E. G. Noya, and J. P. K. Doye, *J. Chem. Phys.* **124**, 104503 (2006).
- 2 [18] J. P. K. Doye, and L. Meyer, *Phys. Rev. Lett.* **95**, 063401 (2005).
- 3 [19] S. M. Cleary, and H. R. Mayne, *Chem. Phys. Lett.* **418**, 79 (2006).
- 4 [20] V. K. de Souza, and D. J. Wales, *J. Chem. Phys.* **130**, 194508 (2009).
- 5 [21] D. Parodi, and R. Ferrando, *Phys. Lett. A* **367**, 215 (2007).
- 6 [22] R. L. Johnston, *Dalton Trans.*, 4193 (2003).
- 7 [23] L. O. Paz-Borbón, R. L. Johnston, G. Barcaro, and A. Fortunelli, *J. Chem. Phys.* **128**, 134517 (2008).
- 8 [24] L. O. Paz-Borbón, T. V. Mortimer-Jones, R. L. Johnston, A. Posada-Amarillas, G. Barcaro, and
 9 A. Fortunelli, *Phys. Chem. Chem. Phys.* **9**, 5202 (2007).
- 10 [25] A. A. Dzhurakhalov, I. Atanasov, and M. Hou, *Phys. Rev. B* **77**, 115415 (2008).
- 11 [26] P. D. Harvey, A. Eichhöfer, and D. Fenske, *J. Chem. Soc., Dalton Trans.*, 3901 (1998).
- 12 [27] C. Schrodt, F. Weigend, and R. Ahlrichs, *Z. Anorg. Chem.* **628**, 2478 (2002).
- 13 [28] K.-M. Ho, A. A. Shvartsburg, B. Pan, Z.-Y. Lu, C.-Z. Wang, J. G. Wacker, J. L. Fye, and M. F. Jarrold,
 14 *Nature* **392**, 582 (1998).
- 15 [29] I. Rata, A. A. Shvartsburg, M. Horoi, T. Frauenheim, K. W. M. Siu, and K. A. Jackson, *Phys. Rev. Lett.*
 16 **85**, 546 (2000).
- 17 [30] B. Hartke, *Chem. Phys. Lett.* **258**, 144 (1996).
- 18 [31] B. Hartke, *Theor. Chem. Acc.* **99**, 241 (1998).
- 19 [32] B. Hartke, H.-J. Flad, and M. Dolg, *Phys. Chem. Chem. Phys.* **3** 5121 (2001).
- 20 [33] Y. Xiao, and D. E. Williams, *Chem. Phys. Lett.* **215**, 17 (1993).
- 21 [34] W. J. Pullan, *J. Chem. Inf. Comput. Sci.* **37**, 1189 (1997).
- 22 [35] R. P. White, J. A. Niesse, and H. R. Mayne, *J. Chem. Phys.* **108**, 2208 (1998).
- 23 [36] H. Takeuchi, *J. Chem. Inf. Model.* **47**, 104 (2007).
- 24 [37] H. Takeuchi, *J. Chem. Inf. Model.* **48**, 2226 (2008).
- 25 [38] P. Chaudhury, R. Saha, and S. Bhattacharyya, *Chem. Phys.* **270**, 277 (2001).
- 26 [39] F. Schulz, and B. Hartke, *Chem. Phys. Chem.* **3**, 98 (2002).
- 27 [40] F. Schulz, and B. Hartke, *Theor. Chem. Acc.* **114**, 357 (2005).
- 28 [41] B. S. González, J. Hernández-Rojas, and D. J. Wales, *Chem. Phys. Lett.* **412**, 23 (2005).
- 29 [42] S. M. Woodley, and R. Catlow, *Nature Materials* **7**, 937 (2008).
- 30 [43] Z. Zhou, V. Siegler, E. Y. Cheung, S. Habershon, K. D. M. Harris, and R. L. Johnston,
 31 *Chem. Phys. Chem.* **8**, 650 (2007).
- 32 [44] S. Y. Chong, and M. Tremayne, *Chem. Commun.*, 4078 (2006).
- 33 [45] M. A. Neumann, F. J. J. Leusen, and J. Kendrick, *Angew. Chem. Int. Ed.* **47**, 2427 (2008).
- 34 [46] B. Bandow, and B. Hartke, *J. Phys. Chem. A* **110**, 5809 (2006).
- 35 [47] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning* (Kluwer Academic
 36 Publishers, Cambridge, 1989).
- 37 [48] S. Warshall, *J. ACM* **9**, 11 (1962).
- 38 [49] F. Diedrich, R. Harren, K. Jansen, R. Thöle, and H. Thomas, *JCST* **23**, 749 (2008).
- 39 [50] A. White, DSTO Aeronautical and Maritime Research Laboratory, DSTO-TN-0302 (2000).
- 40 [51] Jmol: an open-source Java viewer for chemical structures in 3D. <http://www.jmol.org/>
- 41 [52] POV-Ray - The Persistence of Vision Raytracer www.povray.org
- 42 [53] J. P. K. Doye, M. A. Miller, and D. J. Wales, *J. Chem. Phys.* **111**, 8417 (1999).
- 43 [54] J.-M. Schröder, *personal communication*, (2007).
- 44 [55] K. Wehkamp, L. Schwichtenberg, J.-M. Schröder, and J. Harder, *J. Invest. Dermatol.* **126**, 121 (2006).
- 45 [56] N. D'Amelio, E. Gaggelli, N. Gaggelli, E. Molteni, M. C. Baratto, G. Valensin, M. Jezowska-Bojczuk,
 46 and W. Szczepanik, *Dalton Trans.*, 363 (2004).
- 47 [57] X. Q. Lu, M. Zhang, J. W. Kang, X. Q. Wang, L. Zhuo, and H. D. Liu, *J. Inorg. Biochem.* **98**, 582
 48 (2004).
- 49 [58] M. Kopaczynska, M. Lauer, A. Schulz, T. Wang, A. Schaefer, and J. H. Fuhrhop, *Langmuir* **20**, 9270
 50 (2004).
- 51 [59] J. M. Dieterich, U. Gerstel, B. Hartke, and J.-M. Schröder, *manuscript in preparation*.
- 52 [60] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel,
 53 L. Kale, and K. Schulten. *J. Comput. Chem.*, **26**, (2005) 1781.
- 54 [61] D. A. Case, T. A. Darden, T. E. Cheatham, III, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo,
 55 K. M. Merz, B. Wang, D. A. Pearlman, M. Crowley, S. Brozell, V. Tsui, H. Gohlke, J. Mongan,
 56 V. Hornak, G. Cui, P. Beroza, C. Schafmeister, J. W. Caldwell, W. S. Ross, and P. A. Kollman (2004),
 57 AMBER 8, University of California, San Francisco.
- 58 [62] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, *J. Comput. Chem.* **25**, 1157
 59 (2004).
- 60 [63] Orca: an ab initio, DFT and semiempirical electronic structure package.
- [64] A. Schaefer, H. Horn, and R. Ahlrichs, *J. Chem. Phys.* **97**, 2571 (1992).
- [65] S. Varma, D. Sabo, and S. B. Rempe, *J. Molec. Biol.* **376**, 13 (2008).
- [66] A. Martucci, A. Alberti, M. D. Guzman-Castillo, F. Di Renzo, and F. Fajula, *Micropor. Meso-*
por. Mater. **63**, 33 (2003).
- [67] S. Ledain, A. Leclaire, M. M. Borel, and B. Raveau, *J. Solid State Chem.* **129**, 298 (1997).