PAPER

Symmetrisation schemes for global optimisation of atomic clusters

Mark T. Oakley,[a] Roy L. Johnston[a] and David J. Wales*[b]

Locating the global minima of atomic and molecular clusters can be a difficult optimisation problem. Here we report benchmarks for procedures that exploit approximate symmetry. This strategy was implemented in the GMIN program following a theoretical analysis, which explained why high-symmetry structures are more likely to have particularly high or particularly low energy. The analysis, and the corresponding algorithms, allow for approximate point group symmetry, and can be combined with basin-hopping and genetic algorithms. We report results for 38-, 75-, and 98-atom Lennard-Jones clusters, which are all multiple-funnel systems. Exploiting approximate symmetry reduces the mean time taken to locate the global minimum by up to two orders of magnitude, with smaller improvements in efficiency for LJ$_{55}$ and LJ$_{74}$, which correspond to simpler single-funnel energy landscapes.

1 Introduction

Locating the global minima of molecules with many degrees of freedom is one of the fundamental problems in computational chemistry. The optimisation of atomic or molecular clusters is a widely studied problem in this class. Clusters bound by the Lennard-Jones (LJ) potential represent a physically important set of systems, which have served as useful examples for benchmarking global optimisation algorithms. The LJ potential takes the form

$$E = 4\varepsilon \sum_{i<j} \left[\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^{6} \right], \qquad (1)$$

where ε is the equilibrium pair well depth, $2^{1/6}\sigma$ is the equilibrium pair separation, and r_{ij} is the distance between particles i and j. Here, we use reduced units with $\varepsilon = \sigma = 1$.

A wide range of global optimisation algorithms have been applied to Lennard-Jones clusters. These include basin-hopping,[1–4] genetic algorithms,[5–11] particle swarm optimisation,[12,13] a random tunnelling algorithm,[14] hierarchical global optimisation,[15] immune algorithms,[16] and dynamic lattice searching.[17–21] Most of the clusters with fewer than 1000 particles have global minimum structures that comprise an icosahedral core surrounded by an incomplete shell of atoms. For the smaller Lennard-Jones clusters, these are relatively easy to locate with any reasonable method that

includes the basin-hopping principle of steps between local minima.[2,22,23] However, there are a few cases where the global minima are much more difficult to find with an unbiased search algorithm. These structures include a truncated octahedron for LJ$_{38}$,[24] tetrahedral symmetry for LJ$_{98}$,[3] and Marks decahedra for LJ$_{75–77}$[24] and LJ$_{102–104}$,[25] and the difficulty is due to the multiple-funnel nature of the corresponding energy landscapes.[26–28] The funnel containing the icosahedral minima contains the majority of low-lying local minima, and is favoured by entropy, whereas the funnel containing the global minimum contains relatively few minima, and is separated from the structures based on icosahedral packing by a large potential energy barrier.[26,29] These examples also provide useful benchmarks for sampling global thermodynamics in systems with broken ergodicity,[27,30–32] and for rare event dynamics.[33–38]

In all of these difficult cases, the global minimum is more symmetrical than the icosahedral-based structures. It has been noted that structures with high (approximate) symmetry are expected to lie at the extremes of the potential energy distribution for local minima.[28,39,40] A principle of maximum symmetry therefore exists, namely that global minima are likely to have higher (approximate) symmetry measures, as explained in Section 2. Procedures to exploit this observation were programmed in the GMIN global optimisation code by one of us (DJW) some years ago, but the methodology has not been described before. In the present report we explain how these tools are formulated and report benchmark results for several LJ clusters with multiple funnel landscapes.

The use of symmetry improves the performance of several stochastic search algorithms,[13,41,42] either by choosing

[a] *School of Chemistry, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK*

[b] *University Chemical Laboratories, Lensfield Road, Cambridge CB2 1EW, UK. E-mail: dw34@cam.ac.uk*

This journal is © the Owner Societies 2013

Phys. Chem. Chem. Phys., 2013, **15**, 3965–3976 | 3965

high-symmetry starting points for the search or by later imposition of symmetry. The best previous published results for optimisation of Lennard-Jones clusters involve particle swarm optimisation with symmetrical structures used to initialise the search.[13] Lattice-based searches have also proved to be effective,[8,9,19–21,43,44] so long as a lattice that supports the true global minimum structure is included in the analysis. The core orbits approach described in the present contribution does not assume any particular point group symmetry, but instead searches for approximate symmetry operations in a general way. It is therefore a symmetry-biased basin-hopping procedure, but accounts for approximate symmetry and is not biased towards any particular point group. Exploiting this general symmetry bias, which is justified by the principle of maximum symmetry (described in Section 2), greatly improves the efficiency of global optimisation for the cases considered. We show that this procedure performs well when combined with both basin-hopping and with a 'Lamarckian' genetic algorithm, where the fundamental steps are also between local minima.

2 The principle of maximum symmetry

Inspection of the point group symmetries for entries in the Cambridge Cluster Database[45] supports the notion that global minima often have a significant number of symmetry elements. However, the symmetries identified in complex systems are usually only approximate.[46–48] To explain the appearance of near-symmetry in low energy structures we must therefore look beyond previous results for balanced geometries[49,50] and the epikernel principle proposed for systems subject to Jahn–Teller effects.[51,52] Although a rigorous theory is not yet available, arguments based on an idealised analysis of contributions to the total potential energy remain strongly suggestive. Here we write the total energy as a sum over contributions from a many-body expansion, involving single atom, pairwise and three-body terms, *etc.* Where quantum mechanical states approach degeneracy we would write separate series for every term in the appropriate Hamiltonian matrix.

Exact geometrical symmetry leads to exact duplications in the contributions to the total energy, while approximate symmetry leads to approximate duplications. These approximate duplications immediately provide a means to quantify the notion of near-symmetry. If the terms in the many-body expansion were drawn randomly from the same probability distribution then geometrical symmetry would be manifested as correlation. For alternative structures the distribution will actually vary. However, we would still expect the variance of the sum, which gives the total energy, to be larger when symmetry or near-symmetry is present. Symmetrical structures are then more likely to have particularly high or particularly low energy. Conversely, low-lying structures are more likely to exhibit symmetry.

More formally, denote the mean and variance of a variable, X, drawn from a probability distribution, $p(X)$, as μ and σ^2. The variance of a sum of N such variables, X_i, is then[53]

$$\mathrm{Var}\left(\sum_{i=1}^{N} X_i\right) = N\sigma^2 + N(N-1)\rho\sigma^2, \qquad (2)$$

where the correlation ρ is defined by

$$\rho\sigma^2 = \int (X-\mu)(X'-\mu)p(X,X')\mathrm{d}X\,\mathrm{d}X'. \qquad (3)$$

For $\rho = 0$ the variance is $N\sigma^2$, but for $\rho = 1$ it rises to $N^2\sigma^2$.

This analysis, and the empirical observations of symmetry in the Cambridge Cluster Database,[45] provide the motivation for combining a symmetry bias with global optimisation. The present results, which demonstrate significant improvements in efficiency for difficult cases, without bias towards a particular point group, could be regarded as justification for the procedure. In fact, methods to quantify continuous symmetry measures have been used by Avnir and coworkers,[54–59] and correlations have been found for a variety of physical properties.[60,61]

3 Exploiting approximate symmetry

Here we describe the procedures coded in GMIN[62] to treat approximate symmetry using a *core orbits* (CO) analysis. An overview of this methodology, which was employed for most of the benchmarks, is provided first, followed by the details required for implementation. An alternative approach based on the continuous symmetry measures defined by Avnir and coworkers is then described briefly in Section 3.3. This scheme is also coded in GMIN.

3.1 Overview of the core orbits symmetrisation

The basic approach used for most of the benchmarks reported in Tables 1 and 2 involves identification of the point group for a subcomponent (core) of the cluster, denoted $\mathscr{G}_{\mathrm{core}}$. We then analyse the complete structure in terms of orbits of $\mathscr{G}_{\mathrm{core}}$. The orbits of a point group are sets of symmetry equivalent objects, atoms in the current context, which map only amongst themselves under all operations of the group. The possible dimensions of these orbits correspond to the dimensions of subgroups of the point group, dimension h_{core}, which must therefore be integer divisors of h_{core}. The symmetry analysis used here as a global optimisation tool, involves the identification of orbits of $\mathscr{G}_{\mathrm{core}}$ for atoms that lie outside the core region. The incomplete orbits that are not fully occupied by atoms are identified. Permutations of atoms outside the core to fill in the incomplete orbits are then considered as additional symmetry-based moves, which are followed by local minimisation in the usual basin-hopping framework. No symmetry constraints are imposed during the minimisations. This approach is therefore related to lattice-based approaches,[8,9,17–21,43,44] but with sites defined by incomplete orbits of $\mathscr{G}_{\mathrm{core}}$, which change throughout a run depending upon the current configuration.

This CO symmetry analysis involves four principal steps, which are summarised in Fig. 1, and outlined below. Full details are supplied in Section 3.2.

CO step (1). The first phase basically consists of choosing a suitable core of atoms, which also defines the origin of coordinates, X_{CC}, for the approximate symmetry analysis. This procedure is necessary to define point group operations for the core,

3966 | *Phys. Chem. Chem. Phys.*, 2013, **15**, 3965–3976

This journal is © the Owner Societies 2013

Table 1 Mean first encounter times for the global minimum in basin-hopping applications to Lennard-Jones clusters. The statistics correspond to averages over 100 random starting points for LJ$_{38}$, LJ$_{55}$, and LJ$_{74}$, and the effect of changing the number of starting points is illustrated for LJ$_{75}$ and LJ$_{98}$ using averages over the first 10, 20 and 50 starting points. All jobs were run until the global minimum was located: there are no failures. Atoms were randomly placed within a containing sphere of radius 3σ to generate these initial structures. The standard deviations of all values are similar to the means, and the cpu time corresponds to an Intel Xeon E5405 cpu running at 2.00 GHz. Input and output files corresponding to these results can be found on the GMIN website[62]

Sample size	Cluster size	Method	Mean first encounter time		
			Energy evaluations	Minimisations	cpu time/s
100	38	None	185 493	1271	4.4
100	38	CO symmetrisation	20 655	142	0.5
100	38	CSM symmetrisation	4369	34	0.2
100	55	None	15 733	92	0.6
100	55	CO symmetrisation	9356	103	0.5
100	74	None	50 569	329	3.5
100	74	CO symmetrisation	31 370	281	2.2
10	75	None	7 127 345	53 178	596.3
20	75	None	8 150 177	61 050	682.8
50	75	None	7 315 542	54 684	612.1
100	75	None	8 230 648	61 668	688.7
10	75	CO symmetrisation	46 272	343	3.9
20	75	CO symmetrisation	54 168	377	4.6
50	75	CO symmetrisation	50 640	322	4.2
100	75	CO symmetrisation	50 229	338	4.2
10	98	None	3 702 487	25 521	677.7
20	98	None	5 825 389	40 079	1070.9
50	98	None	6 419 745	44 211	1193.9
100	98	None	7 017 387	48 301	1314.3
10	98	CO symmetrisation	70 905	385	11.4
20	98	CO symmetrisation	81 004	435	13.3
50	98	CO symmetrisation	108 276	564	18.2
100	98	CO symmetrisation	109 753	563	18.5

Table 2 Mean first encounter times for the global minimum in genetic algorithm applications to Lennard-Jones clusters. In most cases, the statistics correspond to 100 random starting points and all searches were run until the global minimum was located. For non-symmetrised searches on LJ$_{75}$ and LJ$_{98}$, 25 starting points were used and some searches failed to locate the global minimum, as discussed in Section 4. Atoms were randomly placed within a containing sphere of radius 3σ to generate these initial structures. The standard deviations of all values are similar to the means, and the cpu time corresponds to an Intel Xeon E5405 cpu running at 2.00 GHz. Input and output files corresponding to these results can be found on the GMIN website[62]

Cluster size	Symmetry method	Population	Mean first encounter time		
			Energy evaluations	Minimisations	cpu time/s
38	None	20	404 825	2885	16.5
38	CO	10	25 365	105	1.1
55	None	20	42 059	329	2.9
55	CO	10	11 329	45	0.9
74	None	50	329 254	2561	36.1
74	CO	50	590 346	2218	76.5
75	None	100	1 515 985 966	12 254 742	176 760.8
75	CO	5	87 250	296	11.3
98	None	100	93 155 859	773 289	16 645.2
98	CO	5	178 651	555	33.6

since the centre of coordinates for the complete structure will generally not coincide with that of the core.

An iterative procedure with a distance-dependent weighting to dampen the contributions of more distant atoms is applied first. The atoms are then sorted according to their distance from X_{CC}, and the largest difference between successive values is located. The origin is translated to the centre of coordinates for the X_{core} atoms in this core set before the gap. A full orbit analysis is then performed based on this new origin using the radial distances.

CO step (2). The variation of the point group is monitored as orbits of atoms corresponding to increasing radii are added to the system. Matrix representations of the atomic mappings corresponding to point group operations are constructed.

(1) Assign a core.

(2) Analyse (approximate) point group symmetries for the core
as subsets of non-core atoms are progressively included.

(3) Identify incomplete orbits generated by non-core atoms using
the symmetry operations of the core.

(4) Propose symmetrised moves by filling in incomplete orbits
and quench the resulting structures

Fig. 1 Summary of the four principal steps involved in the core orbits symmetrisation scheme.

These operations are not generally conserved for the complete structure, and therefore constitute approximate symmetries of subsets of atoms. To help account for the approximate symmetry the transformation matrices were 'purified' using a least-squares fit, as described in Secion 3.2. The disappearance of point group operations as the number of orbits increases beyond the N_{core} set is analysed. The approximate symmetry operations that are lost in the last change of point group (usually to C_1) are employed to construct a group $\mathscr{L}$, which is used in step (3) to define incomplete orbits.

CO step (3). The operations of group $\mathscr{L}$ are applied to all the atoms outside the core to generate new orbits. Incomplete orbits, where one or more of the sites are not occupied by atoms, are identified for use in step (4).

CO step (4). In this last step we propose 'symmetrised' moves and apply local minimisation to the resulting structures. The lowest minimum encountered in this symmetrisation phase, or the initial starting geometry if it is lower, is the structure that is randomly perturbed in the next basin-hopping step. A number of possibilities were considered, of which two were selected for the current benchmarks. The first moves involve filling in incomplete orbits outside the core with one or two missing atoms. Here, the most weakly bound atoms were simply moved to the relevant sites. The second class of moves uses structures where orbits outside the core are all either completely full, or completely empty. An efficient algorithm can be formulated to enumerate these possibilities, based on direct count methods developed for vibrational densities of states. The symmetrised structures are quenched to obtain local minima, up to a maximum number of quenches for each symmetrisation phase.

The core orbits symmetrisation analysis was conducted at intervals of N_{sym} basin-hopping steps. In fact, choosing $N_{sym} = 1$ proved to be the most efficient strategy for the clusters considered in the present work. In this case, each basin-hopping step is preceded by a set of quenches based on symmetrised structures. Random Cartesian displacements were employed in the basin-hopping moves, leaving core atoms unperturbed. If symmetry elements were diagnosed for the initial minimum, the Cartesian displacements were symmetrised to preserve the corresponding point group. Since the energy minimisation does not involve any constraints the core can change during these steps. The non-core atoms are perturbed in the basin-hopping moves and

are unconstrained during minimisation, so the core can also change if a different set of core atoms is identified in the next step.

3.2 Details of the core orbits implementation

Here we provide details of the core orbits symmetrisation scheme. Readers who are more interested in the applications than in implementing the procedure can skip to Secion 3.3.

CO step (1). An iterative approach was used to shift the origin to the centre of coordinates for an initial subset of atoms. The centre of coordinates, X_{CC}, was initialised as

$$X_{CC}^0 = \frac{1}{N}\sum_{\alpha} X_{\alpha}, \tag{4}$$

where N is the number of atoms and X_{α} is the three-dimensional position vector corresponding to atom α. For iteration j we then evaluated X_{CC}^j as

$$X_{CC}^j = \frac{\sum_{\alpha} X_{\alpha}e^{-\zeta|X_{\alpha}-X_{CC}^{j-1}|}}{\sum_{\alpha} e^{-\zeta|X_{\alpha}-X_{CC}^{j-1}|}}, \tag{5}$$

where ζ is a parameter that determines how rapidly the contributions of more distant atoms are damped. Up to 200 iterations were allowed, but normally the refinement of X_{CC} was terminated by the convergence condition $|X_{CC}^j - X_{CC}^{j-1}| < 10^{-4}$ in a few cycles for typical values of ζ of order unity.

The origin was then redefined using the centre of coordinates for the subset of atoms with increasing radii up to the largest radial gap. This analysis exploits the fact that atoms belonging to a particular orbit of the point group must lie at the same distance from the centre of coordinates. We denote the distance of atom α from X_{CC} by d_{α}. If the largest gap between sorted distances d_{α} occurred between entries β and $\beta + 1$ then we set

$$X_{CC} = \frac{1}{\beta}\sum_{\alpha=1}^{\beta} X_{\alpha}. \tag{6}$$

All coordinates were then translated to move X_{CC} to the origin, and the N_{core} atoms defining X_{CC} were designated the initial core of the cluster. The atoms were then resorted according to their distance from the revised centre of coordinates, and a full orbit analysis was conducted using the updated values of d_{α}. A new orbit was assigned whenever the change between sorted distances exceeded a cutoff parameter δ_1. If the largest orbit size was one, the symmetrisation scheme terminated.

CO step (2). The variation of the point group as orbits with increasing radii were successively included was analysed in detail. Here the point group assignment routines were adapted from code originally developed for the ACES package.[63] The atoms in each of these orbits were added to a core structure containing N_{core} atoms, and point group symmetry operations for the core were identified for $N_{core} > 3$. Two more cutoffs, δ_2 and δ_3, were used here, and δ_1 was employed for further orbit analysis. To diagnose symmetry operations, the precision of the mapping for individual atoms to conserved positions was analysed.

3968 | *Phys. Chem. Chem. Phys.*, 2013, **15**, 3965–3976

This journal is © the Owner Societies 2013

For each pair of atoms, α and β, associated with site i before and after the operation, we required $|X_\alpha - X_\beta| < \delta_2$ for all sites. The parameter δ_3 was used to assess degeneracies in eigenvalues of the inertia tensor, I_γ, which were used to direct searches for higher order symmetry axes. Degeneracy was diagnosed when $|I_2 - I_3|/(I_1 + I_2 + I_3) < \delta_3$. Various self-consistency checks were applied in the point group analysis; these were found to circumvent certain failure conditions caused by rigid cutoffs. In particular, if an orbit size was found to be incompatible with the order of a rotational symmetry axis then δ_2 was decreased by a factor of ten and the point group analysis was restarted, so long as $\delta_2 > 10^{-7}$. Similarly, if degeneracy of inertia tensor eigenvalues appeared to be accidental, then δ_3 was decreased by a factor of ten down to a lower limit of 10^{-7}.

The point group symmetry will generally decrease as more orbits are included and N_{core} increases. If no point group symmetry was detected then symmetrised moves were not attempted. Otherwise we focused on the group $\mathscr{G}_{core}$, order h_{core}, for the largest number of core atoms, N_{core}, that produced a point group beyond C_1. The system was translated again so that the centre of coordinates for this core lay at the origin. All symmetry operations were then obtained in matrix form from the generators of $\mathscr{G}_{core}$.

A significant improvement in performance was obtained by refining the 3×3 matrices, s, that specified the mapping of coordinates corresponding to point group operations. For each such matrix we calculated the action of $S_{3N_{core}}$ on the current vector of coordinates, X, a column vector with $3N_{core}$ elements corresponding to the Cartesian coordinates of atom 1, X_1, then X_2 corresponding to atom 2, etc. $S_{3N_{core}}$ is the $3N_{core} \times 3N_{core}$ matrix with N_{core} copies of s along the diagonal, and zero entries elsewhere, so that

$$S_{3N_{core}} \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{pmatrix} = \begin{pmatrix} sX_1 \\ sX_2 \\ \vdots \\ sX_N \end{pmatrix}. \quad (7)$$

If $S_{3N_{core}}$ corresponds to a symmetry operation of X then its effect is to reorder the atomic coordinates. We can define a permutation operation, P_{3N}, to reverse this effect, so that

$$S_{3N_{core}} P_{3N_{core}} X = P_{3N_{core}} S_{3N_{core}} X = \begin{pmatrix} sX_{P(1)} \\ sX_{P(2)} \\ \vdots \\ sX_{P(N)} \end{pmatrix} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{pmatrix}, \quad (8)$$

where

$$P_{3N_{core}} \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{pmatrix} = \begin{pmatrix} X_{P(1)} \\ X_{P(2)} \\ \vdots \\ X_{P(N)} \end{pmatrix}, \quad (9)$$

and $P(\alpha)$ defines the mapping of atomic labels under the permutation operation. If the symmetry operation is not exact

then we can refine the matrix s by minimising the sum of squared deviations in the positions

$$T = \sum_{\alpha=1}^{N_{core}} \left| X_\alpha - s X_{P(\alpha)} \right|^2. \quad (10)$$

The solution is obtained as $s = CA^{-1}$

with $\quad C_{\beta\gamma} = \sum_{\alpha=1}^{N_{core}} [X_\alpha]_\beta [X_{P(\alpha)}]_\gamma \quad$ and

$$A_{\beta\gamma} = \sum_{\alpha=1}^{N_{core}} [X_\alpha]_\beta [X_\alpha]_\gamma, \quad (11)$$

where the subscripts after square brackets indicate vector components. To avoid singular matrices for planar geometries, where one component of every coordinate can vanish, we simply added a small constant Δ to every element of A. Setting $\Delta = 10^{-10}$ generally proved to be satisfactory.

A similar procedure can be applied for all the group operations, but in practice it proved sufficient to 'purify' the generator operations according to the above least squares problem. New symmetry operations were identified by comparing the elements for each product matrix with those already known. At least one pair of corresponding matrix elements had to differ by a parameter δ_m for two operations to be considered distinct. All the point group operations were checked for consistency with the cutoff conditions for the N_{core} atoms. The preservation or disappearance of the h_{core} symmetry operations was then tested for orbits outside the core in order of increasing atomic distance from X_{CC}. This check was terminated if only the identity operation remained, and symmetry operations detected for the complete system were recorded. Any orbits that retained all the symmetry operations of $\mathscr{G}_{core}$ were added to the core, and N_{core} adjusted accordingly. Symmetry operations preserved for the complete structure were used in subsequent geometry perturbations to produce symmetry-adapted steps, as described below.

Two more tolerance parameters were used to define the disappearance of symmetry operations for orbits outside the original N_{core} atoms. The largest deviation between atomic positions corresponding to each site before and after the symmetry operation was found, and divided by the distance of this site from the centre of coordinates. If this ratio exceeded the dimensionless cutoff δ_4 the symmetry operation was deemed to have been lost. Parameter δ_5 is a distance threshold used to identify unoccupied sites in incomplete orbits of the point group. The transformed coordinates for each symmetry operation were compared with the original positions for each orbit outside the core. If a transformed position did not lie within a distance δ_5 of any original position then this position was considered a missing site for the corresponding orbit. A distance check was performed to prevent missing sites appearing closer than δ_5, and a further check caused the procedure to terminate if the number of missing sites reached 120 (the largest possible orbit size commonly encountered).

If symmetry elements were identified for the full structure, these were combined with the ones that were last lost when an orbit was

added to make a set $\mathscr{L}$. If there were more than 50 such operations then we just considered the last ones to be lost and used them to generate the operations for a group, $\mathscr{L}$. The symmetry operations of $\mathscr{L}$ were used to define orbits in the next phase.

CO step (3). The $N_{\text{float}} = N - N_{\text{core}}$ atoms lying outside the core at this point were identified as 'floaters'. Schemes were considered in which the most weakly bound core atoms were added to this set, but did not result in greater efficiency. We then applied all the operations of $\mathscr{L}$ to each floater atom to generate the corresponding orbit, removing duplicate sites according to the threshold parameter δ_5. If the next floater lay in an orbit generated from a previous floater, it was skipped. New orbits should be guaranteed in this way because we check that each floater does not lie in any previous new orbits and we know it does not lie in any of the core sites.

If a newly generated orbit with dimension greater than one lay completely within the non-core set of atoms defined by the original structure, we moved these atoms to the core. If all the atoms ended up in the core we returned to the normal step-taking procedure and perturbed the coordinates of each atom in the usual way.

CO step (4). The last step consists of proposing symmetrised structures and minimising them. Orbits outside the core that

In the final phase of the symmetrisation scheme, starting point geometries for quenches were proposed by distributing the N_{float} atoms in configurations where any subset of the N_{orbits} orbits generated from the floaters was precisely filled. To enumerate these possibilities we performed a convolution over the new orbits using a procedure analogous to the Beyer–Swinehart direct count algorithm for vibrational densities of states.[64,65] Here a modification was required to prevent each orbit being occupied more than once in a given state, and the enumeration could terminate if we reach the maximum number of quenches allowed per symmetry analysis, defined by the input parameter Q_{max}.

To explain the orbit enumeration algorithm we consider $r(i)$, the number of arrangements for i floater atoms that completely fill subsets of the N_{float} orbits. To calculate $r(N_{\text{float}})$ we also need to keep track of the occupation patterns for all states $r(i)$ with $i \leq N_{\text{float}}$. Let $P(i,m,j)$ be one or zero, according to whether orbit j is occupied or not for state m, corresponding to a possible occupation pattern for i floater atoms. We also require s_j, the size of orbit j, and $\sigma_j = \sum_{k=1}^{j-1} s_k$. The elements of $r(i)$ and $P(i,m,j)$ are initialised to zero for $1 \leq i \leq N_{\text{float}}$, $1 \leq m \leq m_{\text{max}}$, and $1 \leq j \leq N_{\text{orbits}}$. The parameter m_{max} specifies a maximum number of states to save for each number of floaters, to prevent memory overflow. We set $\sigma_1 = 0$ and execute the following steps:

```
for j running over the new orbits from 1 to N_float

    for i running from s_j to the smaller of N_float and s_j + σ_j

        set ξ(i), the number of new states for i floaters, equal to zero

        for m running over the r(i − s_j) states already found for i − s_j floaters

            if state m for i − s_j floaters did not use orbit j

                define a new state for i floaters from state m for i − s_j floaters plus orbit j

                increment ξ(i) by one

                record the occupation pattern for P(i, r(i) + ξ(i), k) using

                P(i − s_j, m, k) with 1 ≤ k ≤ j − 1 and set P(i, r(i) + ξ(i), j) to one.

            end if

        end loop over m

    end loop over i

    increment r(i) by ξ(i)

    set σ_{j+1} = σ_j + s_j

end loop over j
```

were identified with one or two missing atoms were first considered. The one or two most weakly bound atoms from outside the core were moved to the missing sites, and the resulting structure was energy minimised. The distance parameter δ_5 was again used to disqualify a starting configuration with two sites that were too close together. The lowest energy and configuration obtained from these quenches was saved and compared with the lowest minimum that resulted from the enumeration of non-core atoms over filled orbits described below.

Quenches were then conducted for the configurations corresponding to the $r(N_{\text{float}})$ arrangements of floaters identified, up to a maximum of Q_{max}, and the lowest minimum encountered was recorded. The symmetry analysis was performed in this fashion every N_{sym} basin-hopping steps, before a modified coordinate perturbation and quench.

The standard step-taking scheme for basin-hopping in GMIN simply involves a perturbation to each Cartesian coordinate using a displacement chosen with uniform probability from the interval $d_{\text{max}} \times [-0.5, 0.5]$. The maximum

displacement, d_{max}, is adjusted at regular intervals according to the number of steps accepted and rejected over the last interval and the target acceptance ratio. In the symmetrisation framework the core atoms identified in the most recent symmetry analysis were not perturbed. Furthermore, if symmetry operations were identified for the current structure then the perturbations were chosen to preserve these operations. The permutation $P(\alpha)$ for each atom α was identified for every symmetry operation. The initial random step, ΔX, was then symmetrised using the projection operator for the totally symmetric irreducible representation of the appropriate point group:

$$\Delta X' = \frac{1}{h}\sum_{i=1}^{h} S(i)P(i)\Delta X, \tag{12}$$

where h is the order of the group, and the $3N \times 3N$ matrices specify the transformation of coordinates and the inverse mapping of atomic labels, as above, so that $SPX = X$ for the initial configuration.

3.3 An alternative symmetrisation scheme

The alternative symmetrisation scheme coded in GMIN is based on the continuous symmetry measure (CSM) of Avnir and coworkers.[54–59] The CSM for a given point group $\mathscr{G}$, denoted $S(\mathscr{G})$, is based on the distance from the current configuration to the closest geometry with symmetry $\mathscr{G}$. The distance is calculated in $3N$-dimensional space and defines the CSM as

$$S(\mathscr{G}) = \frac{1}{N}\sum_{i=1}^{h_{\mathscr{G}}} \left| X_i - \hat{X}_i \right|^2, \tag{13}$$

where the $\hat{X}_i$ are the points in the closest configuration with point group $\mathscr{G}$. The coordinates are initially scaled so that the largest distance from the origin is one, and hence $0 \leq S(\mathscr{G}) \leq 1$. Locating the points corresponding to the closest configuration can be achieved *via* a folding and unfolding procedure,[54,56] or using an approach where the action of all the operations of $\mathscr{G}$ on X are considered.[66] The latter formulation was employed in the present work, as summarised below. Further details are given in the Appendix.

CSM step (1). All the operations of the target point group, $\mathscr{G}$, are applied to the initial configuration, X, to obtain images, Y_i. For each operation i, the permutation of atoms, $P_{3N}(i)$ that minimises the distance between X and Y_i is obtained using the shortest augmenting path algorithm.[67]

CSM step (2). For fixed permutations, $P_{3N}(i)$, the CSM is minimised with respect to the orientation of the original configuration, X. Both X and the point group operations are referred to a fixed reference axis system.

CSM step (3). The orientation of X is perturbed and the procedure in steps (1) and (2) is repeated to obtain a new CSM value. The corresponding orientation is accepted for X if the CSM is lower. This step is repeated for a specified number of cycles, or until a CSM close to zero is achieved.

CSM step (4). An average geometry was then constructed using the orientation and permutations for X that give the

smallest CSM value. This configuration can include atoms at unphysically small distances, and it was therefore relaxed through a fixed number of local basin-hopping steps accepting only downhill moves in energy.[3] The lowest minimum obtained after this local relaxation was taken as the *CSM symmetrised* structure that was perturbed in the next step of the principal basin-hopping scheme.

The CSM procedure is included here since it produced the fastest mean first encounter time for global optimisation of LJ_{38}. However, the efficiency decreases rapidly with system size, probably because some atoms collapse towards the centre of coordinates or to the principal rotation axis in our current implementation. This collapse takes the system into higher symmetry regions of configuration space that are also higher in energy (see Section 2) and less likely to be productive for global optimisation. Further refinement of the CSM symmetrisation procedure may be possible in future work. For example, a hybrid method might combine the identification of approximate symmetry in a core of atoms, described in Section 3.1, with construction of the closest configuration for a given point group using the CSM. The core orbits scheme achieves symmetrised steps in configuration space by identifying and filling in incomplete orbits outside the core, while the CSM employs a folding and unfolding procedure. Exploiting these complementary approaches might therefore be productive.

4 Results

We have tested the performance of the symmetrisation routines in conjunction with the basin-hopping (BH) algorithm[1,22,23] and the Birmingham Cluster Genetic Algorithm (GA).[68] Both of these algorithms are implemented in the GMIN global optimisation program.[62] The basin-hopping procedures in GMIN provide a variety of different step-taking and accept/reject strategies. In each case the overall procedure involves hops between local minima obtained following geometry optimisation after each structural perturbation.[2] The minimisation step enables large moves in configuration space to be proposed and accepted, and effectively removes downhill barriers from the potential energy surface. Occupation probability distributions for structures corresponding to different morphologies are also broadened, which is important for multi-funnel potential energy landscapes.[29,69] Similarly, the 'Lamarckian' genetic algorithm performs a local minimisation on all offspring and mutant structures before the selection step.

We have evaluated the performance of the symmetrisation algorithms for three Lennard-Jones clusters with nonicosahedral global minima: LJ_{38}, LJ_{75} and LJ_{98} (Fig. 2–4). We have also tested the symmetrisation algorithm on some clusters with global minima based on icosahedral packing. LJ_{55} is an interesting case because the Mackay icosahedron[70] corresponds to a 'magic number' for a variety of clusters, and represents an easy global optimisation target. The global minimum for LJ_{74} is an icosahedral core surrounded by a incomplete shell of atoms, and is included as an example of a 'typical' Lennard-Jones cluster in this size range.

The performance of the search algorithms is measured by the mean time to the first encounter of the global minimum. Both search algorithms employ operators to restart searches that appear to have stagnated. These restarts ensure that, given enough time, all searches eventually find the global minimum. For most of the benchmarks, the mean times for the first encounter of the global minimum are recorded over 100

Fig. 2 Structures of the global minimum and lowest minimum based on icosahedral packing for LJ$_{38}$. Atoms are coloured by distance from the centre of the cluster from dark (core) to light (surface). Atoms in the same orbit are the same shade.

(a) -173.928427ε (b) -173.252378ε

Fig. 3 Structures of the global minimum and lowest minimum based on icosahedral packing for LJ$_{75}$. Atoms are coloured by distance from the centre of the cluster from dark (core) to light (surface). Atoms in the same orbit are the same shade.

(a) -397.492331ε (b) -396.282249ε

Fig. 4 Structures of the global minimum and lowest minimum based on icosahedral packing for LJ$_{98}$. Atoms are coloured by distance from the centre of the cluster from dark (core) to light (surface). Atoms in the same orbit are the same shade.

(a) -543.665361ε (b) -543.642957ε

independent searches from random starting positions. When these clusters are analysed using the genetic algorithm without symmetrisation, the mean first encounter times are long enough that it is impractical to wait for 100 searches to complete. In these cases, we performed 25 searches capped at 5000 generations and calculated the mean first encounter time as the total time taken in all searches divided by the number of times the global minimum was found. Each optimisation run was performed on a single core of an Intel Xeon E5405 CPU with a clock speed of 2.0 GHz. We also report run times in terms of the number of minimisations and energy/gradient evaluations, both of which are independent of the processor used.
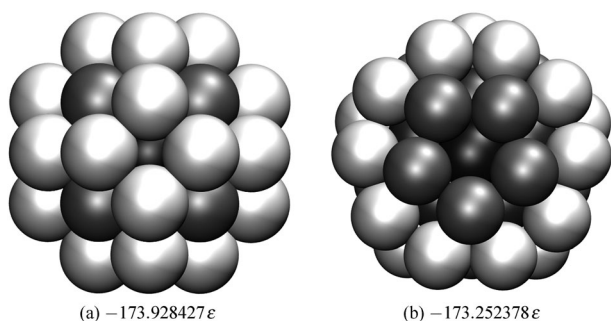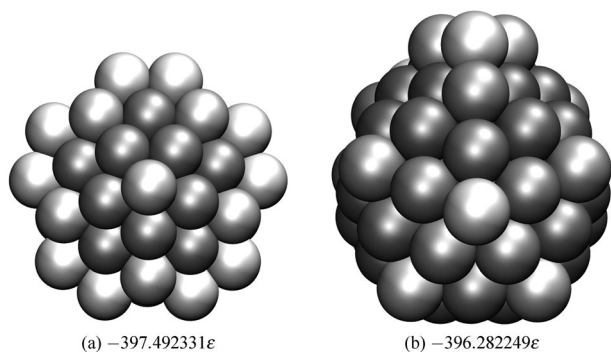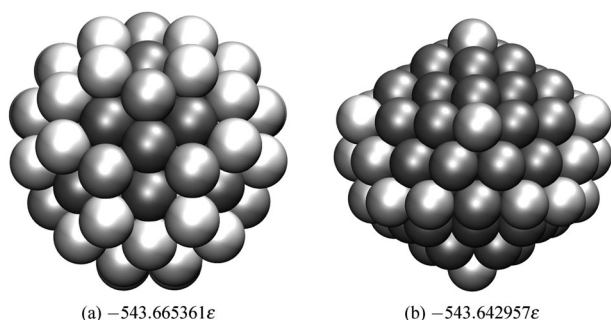
4.1 Basin-hopping

The basin-hopping algorithms in GMIN include an operator that restarts the search with a new random structure if there has been no improvement to the energy of the system after a fixed number of steps.[71] This procedure is specified by the *NEWRESTART* keyword in GMIN. A larger value for this parameter produces more extensive local exploration of single funnels, whereas more frequent restarts favour the exploration of new regions of configurational space. The configuration of the lowest minimum encountered since the last reseeding is added to a cyclic taboo list,[72] which is used to avoid revisiting regions that have recently been searched.[71] This procedure is controlled by the *AVOID* keyword in GMIN, which specifies the length of the taboo list and the distance threshold, which triggers reseeding if the current minimum is too close to any member of the list. The metric employed here was the Euclidean distance in $3N$-dimensional space, minimised with respect to overall translation, rotation, and permutation-inversion operations. This distance optimisation involves a shortest augmenting path algorithm, as described elsewhere.[67] Standard procedures to move weakly bound or surface atoms[1] were also employed; example input and output files will be made available from the Cambridge Cluster Database[45] to ensure that our benchmarks are reproducible.

For each test system we present results for the mean first-encounter time (MFET) averaged over 100 randomised starting points, where the atoms were randomly dispersed throughout a sphere of radius 3σ. The MFET exhibits a small dependence on the radius of this container, which should be accounted for in comparing different approaches. The initial coordinates will be included together with the other GMIN input and output files on the GMIN website, to facilitate future comparisons.

All BH runs were continued until the global minimum was encountered, to provide proper statistics. For LJ$_{75}$ and LJ$_{98}$ Table 1 includes results for the mean first encounter time averaged over the first 10, 20 and 50 starting points. The averages exhibit significant fluctuations for sample sizes of 10 and 20, and could be somewhat misleading, particularly for the runs that do not use symmetrised moves. The values obtained for 100 random starting points are probably sufficient to produce useful statistics for comparison between different methods, while the results for a sample size of 50 provide a

reasonable guide. The discussions below will focus on the averages obtained over 100 runs.

The truncated octahedral global minimum LJ_{38} is the easiest of the examples involving multiple-funnel potential energy landscapes. Without the use of symmetrisation, BH requires 1271 minimisations to find the global minimum, which corresponds to 4.4 s of cpu time (Table 1). Using the core orbits symmetrisation approach reduces this effort to 142 minimisations in 0.5 s, while the CSM symmetrised moves lower these values to 34 minimisations and 0.2 s. However, this system is the only one of the three multi-funnel landscapes considered where the CSM approach is competitive. We aim to investigate the CSM framework further in future work.

The basin-hopping results starting from random geometries without symmetrised moves involve a factor of 60 fewer steps than reported elsewhere for basin-hopping schemes.[73] The origin of this discrepancy presumably lies within the details of the different implementations. The mean number of local minimisations without symmetrised moves is about the same as for a basin-hopping framework based on molecular dynamics moves with feedback and escape steps.[11] The mean number of function calls reported in the latter work is about 774 000, while the number of energy/gradient evaluations in the present work is about 185 000 (Table 1). These results are likely to reflect the performance of the minimisation algorithms employed. Using order parameters based on common neighbour analysis[74] can improve on unbiased basin-hopping by about a factor of four for LJ_{38}. Differences in performance of less than an order of magnitude are probably not very significant, since additional tuning for specific systems can usually produce improvements on this scale. However, combining some of these approaches with symmetrised moves might result in still faster hybrid schemes.

Without symmetrised steps it requires an average of 6.2×10^4 minimisations, corresponding to 902.6 s of cpu time, to locate the decahedral global minimum of LJ_{75}. Reseeding the search if there is no improvement in the energy after around 70 basin-hopping steps is the optimum strategy here. Using symmetrisation reduces the mean first encounter time by over two orders of magnitude to 4.2 s. The global minimum of LJ_{98} with tetrahedral symmetry requires a similar computational effort without symmetrisation, requiring an average of 4.8×10^4 basin-hopping steps, corresponding to 1314.3 s of cpu time. The optimal value for the minimum reseeding interval is about 100 BH steps. With symmetrisation, the mean first encounter time is 18.5 s, which is a 70-fold improvement. These mean first encounter times for LJ_{75} and LJ_{98} with symmetrised moves correspond to much smaller minimum reseeding intervals of 5 and 10 BH steps, respectively. In general, we would expect this parameter to scale roughly with the average number of BH steps required to locate the global minimum.

For LJ_{55} the global minimum is located rapidly whether symmetrisation is used or not. When using symmetrisation the optimal value for the *NEWRESTART* parameter is around 60 within the GMIN setup employed. The mean first encounter times are comparable to the 100 to 200 basin-hopping steps

required by other optimisation algorithms.[2,11,13] For LJ_{74} the mean first encounter times are optimised for *NEWRESTART* values of around 75 and 90 BH steps with and without symmetrisation, respectively. Symmetrisation produces a speedup of about 50% in this case, which is much less significant than for the multiple-funnel landscapes corresponding to LJ_{38}, LJ_{75} and LJ_{98}.

The use of symmetrisation generally provides between one and two orders of magnitude increase in efficiency for the multifunnel cases and can provide a modest improvement for simpler landscapes. A longer interval between possible reseeding operations works best if symmetrised moves are not used. If the global minimum is not known in advance then a larger value for this parameter is probably advisable.

4.2 Genetic algorithm

The GA considered here uses one of two versions of the Deaven–Ho cut and splice method to generate offspring.[5] The first, as used in previous optimisations of Lennard-Jones clusters,[11,75] is to cut along a plane passing through the centre of the cluster and splice two equally sized fragments. In the second, the plane does not have to pass through the centre of the cluster, allowing splicing between two unevenly-sized spherical caps. The second strategy is optimal for all searches except for the non-symmetrised searches on LJ_{38}. Mutants are constructed by duplicating an existing parent or offspring structure and displacing each Cartesian coordinate of every atom by a distance $<d_{max}$.

The genetic algorithm is elitist, with the parent structures remaining in the population until they are replaced by more stable offspring or mutant structures. Thus, the fitness of the worst member of the population always improves or remains the same after every generation. A duplicate predator is employed to maintain the diversity of the population.[76] If a generation of the GA leads to no improvement in the mean energy of the population the search has stagnated. When this situation occurs, a new population of random structures is generated and a new epoch begins. No information is transferred from one epoch to the next and each epoch is effectively an independent search. The efficiency of the GA is affected by the size of the population. We have performed searches with population sizes of 5, 10, 20, 50 and 100 individuals and the optimum population for each search is shown in Table 2.

The fastest published optimisations of LJ_{38} with a genetic algorithm took 2000 energy minimisations.[75] Our non-symmetrised GA requires an average of 2884 minimisations to locate the global minimum, which is comparable to these results. Symmetrisation of the GA reduces the required search time by a factor of approximately ten (Table 2). The global minimum of LJ_{55} is rapidly located without any symmetrisation, requiring 324 minimisations. This result is improved to 45 minimisations when symmetrisation is used. For both of these clusters, the time needed for the symmetrised GA is similar to that for symmetrised basin-hopping. For LJ_{74}, symmetrisation leads to little change in the number of energy minimisations required to locate the global minimum. Of the clusters in this study, this is the only one where the symmetrised GA is substantially slower than symmetrised BH.

This journal is © the Owner Societies 2013

Phys. Chem. Chem. Phys., 2013, **15**, 3965–3976 | **3973**

The global minimum of LJ_{75} is known to be relatively difficult to locate with a GA,[11] and we only find it in one of the 25 runs of the non-symmetrised GA. This result represents one hit in 2000 independent epochs of the GA requiring a total of 49 hours of cpu time. Hartke's evolutionary algorithm,[7] which employs the concept of niches to maintain the population diversity, is the most successful algorithm of this class for LJ_{75}. The niches are based on the surface coverage for two-dimensional projections of the cluster structure. This method tends to add clusters with different symmetries into the population, which is clearly advantageous for this system. When symmetrisation is included in the GA, the mean first encounter time drops to 11.3 s, which is about three times longer than the symmetrised basin-hopping result. The fastest searches for LJ_{75} are obtained with a population size of five. However, the mean first encounter time is not very sensitive to the population size and searches with 100 members are slower by a factor of less than two.

Without symmetrisation, the GA performs better on optimisation of LJ_{98} than for LJ_{75}, but still only locates the global minimum in 12/25 searches. This result corresponds to one hit in every 77 epochs, requiring an average of 5 hours of cpu time. With symmetrisation, the mean first encounter times are similar to the results for symmetrised basin-hopping. As with LJ_{75}, the fastest searches were obtained with a population size of five, but searches with larger populations are not much slower.

The fastest published optimisation of the high-symmetry Lennard-Jones clusters of which we are aware was obtained by the CALYPSO particle swarm optimisation algorithm.[13] The benchmarks presented here represent a small improvement over CALYPSO for the optimisation of LJ_{38} (600 minimisations), and a larger improvement for the optimisation of LJ_{75} (2900 minimisations). Both the symmetrised basin-hopping and genetic algorithms locate the global minimum of LJ_{98} with an average run time well under a minute without failures, whereas CALYPSO only found this structure in 3/10 reported attempts.[13]

5 Conclusions

The use of approximate symmetry substantially improves the efficiency of global optimisation for the atomic clusters we have considered with multifunnel potential energy surfaces.[26–28] Here we have exploited the principle of maximum symmetry[28,39,40] to bias the searches towards geometries with a higher symmetry measure, defined in terms of a cluster core. This theory is outlined in Section 2, and explains why symmetrised moves provide a more productive way to explore the configuration space. The bias is general, and not specific to a particular point group for the *core orbits* procedure, which corresponds to most of the benchmarks we have presented. This approach involves a number of additional parameters, mostly related to geometrical cutoffs, but these are likely to be relatively transferable between related systems.

For LJ_{38}, symmetrisation reduces the time required to find the global minimum by more than a factor of ten. For LJ_{75} and LJ_{98} basin-hopping searches that do not use the symmetrisation

algorithm require an average of around 60 000 to 50 000 steps for 100 random starting points, and searches using the GA take much longer. When symmetrisation is used, the global minima of both clusters are located about 85 and 180 times faster for LJ_{98} and LJ_{75}, respectively, corresponding to 563 and 338 BH steps, on average. For clusters with simpler energy landscapes, such as LJ_{55} and LJ_{74}, symmetrisation provides more modest improvements in run time. In systems with global minima that have no exact symmetry elements the likely speedup will depend on the complexity of the underlying landscape, and the degree of approximate symmetry. Further tests will be needed to gauge these effects in future work, but our results for LJ_{74} suggest that the core orbits procedure will generally be beneficial, even for global minima without any formal symmetry.

The performance of symmetrised basin-hopping and the symmetrised GA are comparable, and both locate the non-icosahedral global minima significantly faster than other unbiased algorithms. We would expect the same improvement for basin-hopping variants that include the critical minimisation phase, but employ alternative step-taking or accept/reject procedures.[11,43,73,77–81]

Appendix: details of the CSM-based symmetrisation approach

CSM step (1)

For a given target point group, $\mathscr{G}$, dimension $h_{\mathscr{G}}$, we first construct the images of the $3N$-dimensional scaled coordinate vector, $\boldsymbol{Y}(i) = \boldsymbol{S}_{3N}(i)\boldsymbol{X}$, for each point group operation, i. Matrix representations of the operations were programmed as data for all the standard point groups. For each image vector $\boldsymbol{Y}(i)$ the shortest augmenting path algorithm was then applied to find the permutation that minimised the distance to $\boldsymbol{X}$,[67] which we can represent by the matrix $\boldsymbol{P}_{3N}(i)$. $\hat{\boldsymbol{X}}_i$ was then calculated as

$$\hat{\boldsymbol{X}}_i = \frac{1}{h_{\mathscr{G}}}\sum_{i=1}^{h_{\mathscr{G}}} \boldsymbol{P}_{3N}(i)\boldsymbol{S}_{3N}(i)\boldsymbol{X}, \qquad (14)$$

where $h_{\mathscr{G}}$ is the dimension of group $\mathscr{G}$. In fact, the shape measure $S(\mathscr{G})$ will only agree with the distance calculated from the folding/unfolding procedure[54,56] for a particular subset of the possible permutations.[66] The unrestricted permutational minimisation of the distance to $\boldsymbol{X}$ for each operation was employed in the present work in the absence of an efficient scheme for constrained optimisation. This approach is sufficient to provide more symmetrical candidate structures for global optimisation. However, it did not prove to be competitive with the core orbits approach for the larger clusters considered in the present work.

CSM step (2)

Fixing the permutations, $\boldsymbol{P}_{3N}(i)$, the CSM was then minimised with respect to the orientation of the original configuration, $\boldsymbol{X}$, treating the N atoms as a rigid body. Angle-axis coordinates were used to describe the orientation of $\boldsymbol{X}$, and analytic first

3974 | *Phys. Chem. Chem. Phys.*, 2013, **15**, 3965–3976

This journal is © the Owner Societies 2013

derivatives with respect to the three angular variables were employed. Here we exploited the same framework described in our recent investigations of coarse-grained models that treat decorated rigid bodies within the angle-axis framework.[82,83] Minimisation in this three-dimensional space was achieved with a modified version of the limited-memory Broyden–Fletcher–Goldfarb–Shanno (LBFGS) algorithm,[84,85] which is the same technique employed for all the other geometry optimisations in the present work. Our implementation therefore uses point group operations referred to a fixed reference frame, and rotates the coordinates X to minimise $S(\mathscr{G})$, rather than transforming the operations themselves.

CSM step (3)

The CSM value that results from the optimisation described above corresponds to a local minimum with respect to the orientation of the initial configuration for a fixed set of permutation matrices, $P_{3N}(i)$. To find the lowest possible CSM we implemented a separate basin-hopping procedure, taking steps in the angle-axis variables that define the orientation of the initial configuration, X. The optimal permutations were recalculated after these displacements, and the CSM minimised again in this three-dimensional space. Here we accepted downhill moves only[3] for a maximum number of basin-hopping steps, terminating if the CSM fell below a value of 10^{-6}.

CSM step (4)

The rotation matrix, c, corresponding to the orientation of X with the lowest CSM value was then used to construct an average geometry as

$$\hat{X} = \frac{1}{h_{\mathscr{G}}} \sum_{i=1}^{h_{\mathscr{G}}} P_{3N}(i) S_{3N}(i) C_{3N} X, \qquad (15)$$

where C_{3N} is the matrix with N copies of c along the diagonal, and zeros elsewhere. $\hat{X}$, was then rescaled to undo the distance normalisation used to define the CSM, and taken as the starting point for a fixed number of basin-hopping steps, N_{CSM}, in the full Cartesian coordinate space, accepting only downhill moves in energy.[3] This phase of local optimisation of the energy serves to relax the structure specified by $\hat{X}$, which may correspond to a high energy configuration. Local relaxation is required because the averaging procedure that minimises the CSM and defines $\hat{X}$ allows atoms to collapse towards the principal rotation axis, or to the origin, which can result in unphysically small interatomic distances. The lowest minimum obtained at the end of this phase specifies the proposed move for the *CSM symmetrisation* procedure implemented in the present work.

The collapse of atoms towards symmetry axes or the origin in the above procedure arises when points do not occur in sets that match the dimensions of orbits of the chosen point group. Our averaging procedure then results in orbits of dimension one. Nevertheless, CSM symmetrisation provides the most efficient scheme for the LJ_{38} cluster (using point group C_i). For the larger systems considered we have not yet been able to locate effective parameters, and the CSM results are not competitive. It may be necessary to implement a constrained permutational optimisation, and change the averaging to avoid collapse of atoms from incomplete orbits, for the CSM symmetrisation to prove effective in larger systems.

References

1 D. J. Wales and J. P. K. Doye, *J. Phys. Chem. A*, 1997, **101**, 5111–5116.
2 D. J. Wales and H. A. Scheraga, *Science*, 1999, **285**, 1368–1372.
3 R. H. Leary and J. P. K. Doye, *Phys. Rev. E*, 1999, **60**, R6320–R6322.
4 R. H. Leary, *J. Global Opt.*, 2000, **18**, 367–383.
5 D. M. Deaven, N. Tit, J. R. Morris and K. M. Ho, *Chem. Phys. Lett.*, 1996, **256**, 195–200.
6 M. D. Wolf and U. Landman, *J. Phys. Chem. A*, 1998, **102**, 6129–6137.
7 B. Hartke, *J. Comput. Chem.*, 1999, **20**, 1752–1759.
8 D. Romero, C. Barrón and S. Gómez, *Comput. Phys. Commun.*, 1999, **123**, 87–96.
9 Y. Xiang, H. Jiang, W. Cai and X. Shao, *J. Phys. Chem. A*, 2004, **108**, 3586–3592.
10 X. Shao, Y. Xiang and W. Cai, *Chem. Phys.*, 2004, **305**, 69–75.
11 S. E. Schönborn, S. Goedecker, S. Roy and A. R. Oganov, *J. Chem. Phys.*, 2009, **130**, 144108.
12 S. T. Call, D. Y. Zubarev and A. I. Boldyrev, *J. Comput. Chem.*, 2007, **28**, 1177–1186.
13 J. Lv, Y. Wang, L. Zhu and Y. Ma, *J. Chem. Phys.*, 2012, **137**, 084104.
14 H. Jiang, W. Cai and X. Shao, *Phys. Chem. Chem. Phys.*, 2002, **4**, 4782–4788.
15 S. V. Krivov, *Phys. Rev. E*, 2002, **66**, 025701.
16 X. Shao, L. Cheng and W. Cai, *J. Chem. Phys.*, 2004, **120**, 11401–11406.
17 X. Shao, L. Cheng and W. Cai, *J. Comput. Chem.*, 2004, **25**, 1693–1698.
18 X. Yang, W. Cai and X. Shao, *J. Comput. Chem.*, 2007, **28**, 1427–1433.
19 L. Cheng, W. Cai and X. Shao, *ChemPhysChem*, 2007, **8**, 569–577.

20 X. Shao, X. Yang and W. Cai, *J. Comput. Chem.*, 2008, **29**, 1772–1779.

21 X. Wu, W. Cai and X. Shao, *Chem. Phys.*, 2009, **363**, 72–77.

22 Z. Li and H. A. Scheraga, *Proc. Natl. Acad. Sci. U. S. A.*, 1987, **84**, 6611.

23 Z. Li and H. A. Scheraga, *THEOCHEM*, 1988, **179**, 333.

24 J. P. K. Doye, D. J. Wales and R. S. Berry, *J. Chem. Phys.*, 1995, **103**, 4234–4249.

25 J. P. K. Doye and D. J. Wales, *Chem. Phys. Lett.*, 1995, **247**, 339–347.

26 D. J. Wales, M. A. Miller and T. R. Walsh, *Nature*, 1998, **394**, 758–760.

27 J. P. K. Doye, M. A. Miller and D. J. Wales, *J. Chem. Phys.*, 1999, **110**, 6896–6906.

28 D. J. Wales, *Energy Landscapes*, Cambridge University Press, Cambridge, 2003.

29 J. P. K. Doye, D. J. Wales and M. A. Miller, *J. Chem. Phys.*, 1998, **109**, 8143–8153.

30 J. P. Neirotti, F. Calvo, D. L. Freeman and J. D. Doll, *THEOCHEM*, 2000, **112**, 10340–10349.

31 F. Calvo, J. P. Neirotti, D. L. Freeman and J. D. Doll, *J. Chem. Phys.*, 2000, **112**, 10350–10357.

32 D. D. Frantz, *J. Chem. Phys.*, 2001, **115**, 6136–6157.

33 M. A. Miller, J. P. K. Doye and D. J. Wales, *J. Chem. Phys.*, 1999, **110**, 328–334.

34 D. J. Wales, *Mol. Phys.*, 2002, **100**, 3285–3305.

35 D. J. Wales, *Mol. Phys.*, 2004, **102**, 891–908.

36 G. Adjanor, M. Athenes and F. Calvo, *Eur. Phys. J. B*, 2006, **53**, 47–60.

37 F. Calvo, *Phys. Rev. E*, 2010, **82**, 046703.

38 M. Picciani, M. Athenes, J. Kurchan and J. Tailleur, *J. Chem. Phys.*, 2011, **135**, 034108.

39 D. J. Wales, *Chem. Phys. Lett.*, 1998, **285**, 330–336.

40 D. J. Wales, *Chem. Phys. Lett.*, 1998, **294**, 262.

41 S. E. Wheeler, P. v. R. Schleyer and H. F. Schaefer, *J. Chem. Phys.*, 2007, **126**, 104104.

42 C. J. Pickard and R. J. Needs, *J. Phys.: Condens. Matter*, 2011, **23**, 053201.

43 L. Cheng, Y. Feng, J. Yang and J. Yang, *J. Chem. Phys.*, 2009, **130**, 214112.

44 X. Lai, R. Xu and W. Huang, *Sci. China: Chem.*, 2011, **54**, 985–991.

45 D. J. Wales, J. P. K. Doye, A. Dullweber, M. P. Hodges, F. Y. Naumkin, F. Calvo, J. Hernández-Rojas and T. F. Middleton, *The Cambridge Cluster Database*, URL http://www.wales.ch.cam.ac.uk/CCD.html, 2001.

46 S.-J. Chen and K. A. Dill, *J. Chem. Phys.*, 1996, **104**, 5964.

47 M. E. Kellman, *J. Chem. Phys.*, 1996, **105**, 2500.

48 K. Yue and K. Dill, *Proc. Natl. Acad. Sci. U. S. A.*, 1995, **92**, 146.

49 J. Leech, *Mathematical Gazette*, 1957, **41**, 81.

50 D. J. Wales, *J. Am. Chem. Soc.*, 1990, **112**, 7908–7915.

51 A. Ceulemans, D. Beyens and L. G. Vanquickenborne, *J. Am. Chem. Soc.*, 1984, **106**, 5824.

52 A. Ceulemans and L. G. Vanquickenborne, *Struct. Bonding*, 1989, **71**, 125.

53 J. Pitman, *Probability*, Springer-Verlag, New York, 1993.

54 H. Zabrodsky, S. Peleg and D. Avnir, *J. Am. Chem. Soc.*, 1992, **114**, 7843.

55 P. W. Fowler, *Nature*, 1992, **360**, 626.

56 H. Zabrodsky, S. Peleg and D. Avnir, *J. Am. Chem. Soc.*, 1993, **115**, 8278.

57 H. Zabrodsky, S. Peleg and D. Avnir, *J. Am. Chem. Soc.*, 1993, **115**, 11656.

58 H. Zabrodsky and D. Avnir, *J. Am. Chem. Soc.*, 1995, **117**, 462.

59 O. Katzenelson, H. Zabrodsky Hel-Or and D. Avnir, *Chem.–Eur. J.*, 1996, **2**, 174.

60 V. Buch, E. Gershgoren, H. Zabrodsky Hel-Or and D. Avnir, *Chem. Phys. Lett.*, 1995, **247**, 149.

61 D. R. Kanis, J. S. Wong, T. J. Marks, M. A. Ratner, H. Zabrodsky, S. Keinan and D. Avnir, *J. Phys. Chem.*, 1995, **99**, 11061.

62 D. J. Wales, *GMIN: A program for finding global minima and calculating thermodynamic properties from basin-sampling*, http://www-wales.ch.cam.ac.uk/GMIN/.

63 J. F. Stanton, J. Gauss, J. D. Watts, W. J. Lauderdale and R. J. Bartlett, *Int. J. Quantum Chem.*, 1992, **44**, 879–894.

64 T. Beyer and D. F. Swinehart, *Commun. ACM*, 1973, **16**, 379.

65 S. E. Stein and B. S. Rabinovitch, *J. Chem. Phys.*, 1973, **58**, 2438.

66 M. Pinsky, C. Dryzun, D. Casanova, P. Alemany and D. Avnir, *J. Comput. Chem.*, 2008, **29**, 2712–2721.

67 D. J. Wales and J. M. Carr, *J. Chem. Theory Comput.*, 2012, **8**, 5020–5034.

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

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

70 A. L. Mackay, *Acta Crystallogr.*, 1962, **15**, 916.

71 D. J. Wales and T. Head-Gordon, *J. Phys. Chem. B*, 2012, **116**, 8394–8411.

72 D. Cvijovic and J. Klinowski, *Science*, 1995, **267**, 664–666.

73 S. Goedecker, *J. Chem. Phys.*, 2004, **120**, 9911–9917.

74 G. Rossi and R. Ferrando, *J. Phys.: Condens. Matter*, 2009, **21**, 084208.

75 V. A. Froltsov and K. Reuter, *Chem. Phys. Lett.*, 2009, **473**, 363–366.

76 G. A. Cox, T. V. Mortimer-Jones, R. P. Taylor and R. L. Johnston, *Theor. Chem. Acc.*, 2004, **112**, 163–178.

77 U. H. E. Hansmann and L. T. Wille, *Phys. Rev. Lett.*, 2002, **88**, 068105.

78 M. Iwamatsu and Y. Okabe, *Chem. Phys. Lett.*, 2004, **399**, 396–400.

79 S. Shanker and P. Bandyopadhyay, *J. Phys. Chem. A*, 2011, **115**, 11866–11875.

80 S. Kazachenko and A. J. Thakkar, *Chem. Phys. Lett.*, 2009, **476**, 120–124.

81 S. Roy, S. Goedecker, M. J. Field and E. Penev, *J. Phys. Chem. B*, 2009, **113**, 7315–7321.

82 D. J. Wales, *Philos. Trans. R. Soc., A*, 2005, **363**, 357–377.

83 D. Chakrabarti and D. J. Wales, *Phys. Chem. Chem. Phys.*, 2009, **11**, 1970–1976.

84 J. Nocedal, *Mathematics of Computation*, 1980, **35**, 773–782.

85 D. C. Liu and J. Nocedal, *Math. Prog.*, 1989, **45**, 503–528.