
Exploring Energy Landscapes: From Molecular to Mesoscopic Systems

Dwaipayan Chakrabarti,^{*a,b} Halim Kusumaatmaja,^{a,c} Victor Rühle^a and David J. Wales^{*a}

We review a comprehensive computational framework to survey the potential energy landscape for systems composed of rigid or partially rigid molecules. Illustrative case studies relevant to a wide range of molecular clusters and soft and condensed matter systems are discussed.

1 Introduction

Computer simulations are now playing an important role in improving our understanding of molecular science, providing a bridge between theory and experimental observables using the tools of statistical mechanics.^{1,2} The richness of the behaviour of such systems over multiple length and time scales poses limitations on our ability to simulate them with atomistic detail to a high level of accuracy. Many of the novel computational techniques in this field are focused on addressing these limitations, either by employing advanced sampling strategies for thermodynamics³⁻¹¹ or kinetics¹²⁻¹⁹ to replace brute-force simulations, or by relying upon coarse-grained (CG) representations²⁰⁻²⁷ with fewer degrees of freedom.

An alternative approach, similar in the spirit to coarse-graining in reducing the number of degrees of freedom, is to approximate the molecules as rigid bodies. This approach depends upon a separation between the frequencies of the intermolecular modes and the intramolecular modes. A partial relaxation of this approximation involves molecular models, composed of rigid units linked by flexible segments, and allows for such a description to be adopted for complex macromolecular systems. A guided choice of rigid domains in macromolecular systems can be obtained *via* principal component analysis performed on data obtained from short molecular dynamics or Monte Carlo trajectories with all-atom representations.²⁸⁻³¹

A rigid-body description in fact offers a versatile approach, which provides useful insight beyond the molecular length scale in particle-based simulations. In particular, this representation, combined with effective potentials, can describe a rich variety of colloidal building blocks that are anisotropic in shape and/or surface chemistry.³² A wide array of building

blocks can now be produced thanks to remarkable advances in particle synthesis, thus opening up potential avenues to complex functional architectures *via* self-assembly.³²

Molecular dynamics simulations of rigid or partially rigid molecular models require equations of motion, either in generalised coordinates involving only unconstrained degrees of freedom, or using the Cartesian coordinates of all atoms satisfying constraints.³³ The former approach is commonly employed for fully rigid molecular systems, where the quaternion representation of the rotational coordinates has emerged as the description of choice owing to its robust characteristics.^{34,35} The method of constraints was originally developed for partially rigid molecules,³⁶ and is especially useful for complex macromolecular systems.³³

An alternative to standard molecular dynamics or Monte Carlo simulations is provided by focusing on stationary points of the underlying potential energy surface (PES) for the system of interest to analyse structure, thermodynamics, and kinetics.³⁷ Within this paradigm, a comprehensive computational framework employs basin-hopping global optimisation to identify stable structures,^{38,39} basin-sampling to obtain equilibrium thermodynamics relying upon the superposition approach,¹¹ and discrete path sampling for transition pathways and kinetics.^{14,15}

Basin-hopping global optimisation performs a biased random walk in the configuration space, where each step consists of a local geometry optimisation to a minimum following perturbation from the existing minimum.³⁹ In basin-sampling an approximation to the relative contributions from local minima in terms of the volumes of basins of attraction is used to obtain the potential energy density of states from a sampling of local minima.¹¹ Using discrete path sampling we obtain a connected database of local minima and transition states,⁴⁰ which defines a kinetic transition network,^{41,42} providing a platform to perform kinetic analysis^{42,43} and all the information required to visualise the energy landscape using disconnectivity graphs^{44,45}.

Efficient geometry optimisation is critical in employing the above computational framework. This energy landscape approach has recently been used in combination with a variety of rigid and partially rigid molecular models following our efforts to address the difficulties that geometry optimisation of rigid bodies is known to pose. In the present contribution we summarise the methodological developments and discuss illustrative case studies to demonstrate the utility of this

^a Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK

^b School of Chemistry, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK

^c Department of Physics, Durham University, South Road, Durham DH1 3LE, UK

* E-mail: d.chakrabarti@bham.ac.uk, dw34@cam.ac.uk

framework for systems consisting of rigid and partially rigid molecules.

The article is organised as follows. In the next section we summarise our recent methodological developments on the treatment of rigid-body rotational coordinates following a brief summary of previous work in this context, and then outline the computational approaches we undertake to explore the energy landscape. The section that follows then illustrates a wide range of applications. Finally we conclude with an outlook.

2 Methodology

2.1 Rigid-Body Rotational Coordinates: Background

Chasles' theorem, which states that the most general displacement of a rigid body can be described as a translation plus a rotation,⁴⁶ underlies our choice for a minimal set of rigid-body coordinates. While the Cartesian representation of the centre-of-mass is the usual choice for the translational coordinates, rigid-body orientation has been parameterised in various ways for geometry optimisation.^{47–52} One of the earliest schemes employs three Euler angles.⁴⁶ However, this representation has singularities at $\alpha = 0$, where α is the angle between the body-fixed and laboratory z axes, and can result in the convergence of geometry optimisation to spurious stationary points for $\alpha \rightarrow 0$.^{47,53} Moreover, the use of Euler angles was found to warrant additional precautions for characterising pathways of structural rearrangements.⁴⁷

The four-parameter unit quaternion, $Q = (q_0, q_1, q_2, q_3) \equiv (q_0, \mathbf{q})$, which provides a singularity-free description of rotation,⁵⁴ has emerged as a robust parameterisation, especially for molecular dynamics simulations.^{34,35} The elegance of quaternion algebra for manipulating rotations makes this representation attractive and it is widely used for molecular modelling.⁵⁵ However, geometry optimisation with quaternion parameters leaves us with a choice to make between two protocols; one route requires a constrained optimiser to satisfy the unit norm constraint, and the other involves unconstrained optimisation, treating all four quaternion parameters as independent variables at the expense of a scaling factor to reset the determinant of the rotation matrix to unity.⁵⁶

Alternatively, Stone and co-workers have considered a hierarchy of axis systems.^{49,50} Here, a set of rotational coordinates that is distinct from the one describing the orientation of a rigid body was used to obtain the energy derivatives.

2.2 Angle-Axis Representation

The angle-axis representation of rotation is based on Euler's theorem,⁴⁶ which states that a general displacement of a rigid body with one point fixed is a rotation about some axis. The

angle-axis representation consists of a unit vector $\hat{\mathbf{p}}$, defining the axis, and an angle θ , describing the magnitude of rotation about that axis. A three-parameter representation can be obtained in terms of the unnormalised vector $\mathbf{p} = [p_1, p_2, p_3] = \theta \hat{\mathbf{p}}$, which is often called a rotation vector, whose magnitude, θ , is the angle of rotation, and whose direction is parallel to the axis of rotation. The equivalent quaternion representation is $Q = (q_0, q_1, q_2, q_3) \equiv (q_0, \mathbf{q}) = (\cos(\theta/2), \hat{\mathbf{p}} \sin(\theta/2))$, and the reverse transformation is straightforward. The first implementation of the angle-axis scheme for geometry optimisation of rigid molecular models sacrificed efficiency for flexibility,⁵¹ and was subsequently improved using a matrix formulation.⁵⁷ This scheme provides a flexible, user-friendly interface to incorporate any system governed by site-site isotropic or anisotropic potentials, and is briefly discussed below.

For the rotation vector \mathbf{p} , Rodrigues' rotation formula⁵⁸ gives the corresponding 3×3 rotation matrix \mathbf{R} , which results in a transformation from the body-fixed frame to the space-fixed frame:

$$\mathbf{R} = \mathbf{I} + (1 - \cos\theta)\tilde{\mathbf{p}}\tilde{\mathbf{p}} + \sin\theta\tilde{\mathbf{p}}, \quad (1)$$

where \mathbf{I} is a 3×3 identity matrix, $\theta = (p_1^2 + p_2^2 + p_3^2)^{1/2}$ is the angle of rotation, $\tilde{\mathbf{p}}$ is the skew-symmetric matrix obtained from $\hat{\mathbf{p}}$:

$$\tilde{\mathbf{p}} = \frac{1}{\theta} \begin{pmatrix} 0 & -p_3 & p_2 \\ p_3 & 0 & -p_1 \\ -p_2 & p_1 & 0 \end{pmatrix}, \quad (2)$$

and a right-handed coordinate system with the right-hand rule for rotations is assumed. The product of the skew-symmetric matrix $\tilde{\mathbf{p}}$, obtained from \mathbf{p} , and any vector \mathbf{v} returns their cross product: $\tilde{\mathbf{p}}\mathbf{v} = \mathbf{p} \times \mathbf{v}$. Eq. (1) is the key to the present computational scheme.

2.3 Computing Energy Gradients and Hessians

Adopting a notation to denote derivatives of a matrix \mathbf{A} , $\partial\mathbf{A}/\partial p_k$, by \mathbf{A}_k ($k = 1, 2, 3$), one has⁵⁷

$$\begin{aligned} \mathbf{R}_k &= \frac{p_k \sin\theta}{\theta} \tilde{\mathbf{p}}^2 + (1 - \cos\theta)(\tilde{\mathbf{p}}_k \tilde{\mathbf{p}} + \tilde{\mathbf{p}} \tilde{\mathbf{p}}_k) \\ &+ \frac{p_k \cos\theta}{\theta} \tilde{\mathbf{p}} + \sin\theta \tilde{\mathbf{p}}_k. \end{aligned} \quad (3)$$

In the limit as $\theta \rightarrow 0$ the formulation in Eq. (1) reduces to the identity matrix. Analytical derivatives can be obtained in this limit by considering a Taylor expansion up to the second order terms.

The second derivatives can be obtained within this computational scheme by successive application of the chain rule, where six additional 3×3 matrices of two types need to be computed, namely $\partial^2\mathbf{R}/\partial p_k^2$ and $\partial^2\mathbf{R}/\partial p_l \partial p_k$, which we denote by \mathbf{R}_{kk} and \mathbf{R}_{kl} , respectively.⁵⁷ As for the first derivatives, the limit for $\theta \rightarrow 0$ can be treated by considering Taylor expansions.

In this scheme the rigid-body coordinate information is all stored in the space-fixed frame. Terms involving the angle-axis coordinates that appear in calculating the energy and its derivatives can be obtained by the action of the rotation matrix and its derivatives on vectors or matrices defined in the body-fixed frame, as illustrated in Appendix. A significant gain in efficiency arises from precomputing these products for a given configuration, thus removing as many operations as possible from the innermost loop over each distinct pair of sites. Benchmark calculations for minimisation of $(\text{H}_2\text{O})_{55}$ clusters described by the TIP4P potential⁵⁹ showed that the present scheme in its matrix formulation gained a ten-fold speed-up over the initial implementation, and was only about 10% slower than an Euler angle scheme, which exploited the specific geometry of the water monomers.⁵⁷ The rotation matrix and its derivatives, being system-independent, can be programmed once and for all.

Even for gradient-only transition state search methods, such as hybrid eigenvector-following,^{60,61} the six degrees of freedom associated with overall translation and rotation, as well as any additional redundancies governed by symmetry, require special treatment. Analytic expressions for the corresponding Hessian eigenvectors with zero eigenvalues enable these modes to be projected out,^{62,63} or the corresponding eigenvalues to be shifted³⁷. The shifting procedure is more efficient, especially for larger systems. While it is easy to work out the eigenvectors for overall translation, finding the eigenvectors corresponding to overall rotation is more involved. We obtained these eigenvectors by considering the change in the angle-axis coordinates upon rotation about the axes of the space-fixed frame through transformations *via* quaternions, as outlined in ref. 57.

2.4 Metric Tensor

The metric tensor is essential for defining infinitesimal distances in a given coordinate system, and has been discussed for various representations of orientation space.^{34,64} Since the translational and rotational degrees of freedom have different physical dimensions, care is needed to treat them on an equal footing while mapping the PES of a rigid-body system. It is then convenient to obtain a metric tensor for the full configuration space of a rigid body using generalised coordinates without any dependence on an underlying atomistic model. By definition, the weighted metric tensor for the full configuration space of a rigid body is

$$G_{\alpha\beta} = \sum_i w_i \frac{\partial \mathbf{x}_i}{\partial q_\alpha} \cdot \frac{\partial \mathbf{x}_i}{\partial q_\beta} \quad (4)$$

with generalised coordinates $\mathbf{q} = \{\mathbf{r}, \mathbf{p}\}$, where \mathbf{x}_i is the position vector of site i in the laboratory frame, $w_i = 1$ for geometric properties, such as distances, or alternatively $w_i = m_i$

for normal mode analysis, and the sum goes over all the sites decorating the rigid body. Using angle-axis variables to represent rigid-body orientation, Rühle *et al.* have recently shown that the metric tensor can be split into three components: one that depends solely on translational degrees of freedom, $G_{\alpha\beta}^{\text{trans}}$, one for rotational degrees of freedom, $G_{\alpha\beta}^{\text{rot}}$, and a mixing term $G_{\alpha\beta}^{\text{mix}}$, with

$$G_{\alpha\beta}^{\text{trans}} = W\delta_{\alpha\beta}, \quad (5)$$

$$G_{\alpha\beta}^{\text{rot}} = \text{Tr}(\mathbf{R}_\alpha \mathbf{S} \mathbf{R}_\beta^T), \quad (6)$$

$$G_{\alpha\beta}^{\text{mix}} = (2W\mathbf{R}_\beta \mathbf{X}_w)_\alpha, \quad (7)$$

where Tr stands for the trace,

$$W = \sum_i w_i,$$

$$\mathbf{X}_w = \frac{1}{W} \sum_i w_i \mathbf{x}_i,$$

and \mathbf{S} is the weighted tensor of gyration in the reference frame of the rigid body with

$$S_{\alpha\beta} = \sum_i w_i x_{i,\alpha} x_{i,\beta}, \quad (8)$$

where $x_{i,\alpha}$ is the $\alpha (= x, y, z)$ -component of the vector \mathbf{x}_i . Note that the mixing term vanishes for any consistent choice of w_i and rigid-body mapping. The metric tensor can be used to define a coordinate independent scalar product, such as the infinitesimal distance ds , which is consistent with the atomistic description

$$ds^2 = \sum_{\alpha,\beta} dq_\alpha G_{\alpha,\beta} dq_\beta = \sum_i w_i |d\mathbf{x}_i|^2. \quad (9)$$

Rühle *et al.* have applied this result to calculate the root-mean-square gradient for energy minimisation and the kinetic energy term for normal mode analysis.⁶⁵ The metric tensor formulation is consistent with calculations performed on the full atomistic resolution while only making use of the rigid-body variables and the tensor of gyration. A measure for finite distances, which is consistent with the underlying atomistic representation, can also be obtained, and has been used to adapt nudged^{66,67} and doubly-nudged⁶⁸ elastic band methods in double-ended transition state searches for rigid-body systems.⁶⁵

2.5 Exploring the Energy Landscape

Basin-hopping (BH) global optimisation and discrete path sampling yield complementary information, and provide a framework to explore the underlying PES. The BH global optimisation approach involves a transformation of the potential energy surface, which preserves the global minimum as

well as the energies of all the local minima.³⁹ In practice, the resulting coarse-grained configuration space is explored by proposing steps, each involving a perturbation from the current minimum in both the translational and rotational coordinates for rigid bodies, followed by a local geometry optimisation to another minimum. The proposed step is accepted or rejected on the basis of a Metropolis (or other suitable) criterion based on the relative energies of the two local minima. Since the objective is to escape from the basin of attraction of the current minimum, step sizes are much larger than those typically used in Monte Carlo simulations of thermodynamic properties. A fictitious temperature is used to apply the Metropolis criterion. It is noteworthy that finding the global minimum on the PES is often a non-trivial task, especially for molecular or mesoscopic systems in the presence of anisotropic interactions.

Discrete path sampling (DPS) can be used to obtain a database of minima and the transition states that connect them.^{14,15,37,40} In this method, an initial path between two selected minima, generally consisting of a series of intervening transition states and minima, is determined by repeated use of double-ended transition state searches. The doubly-nudged⁶⁸ elastic band^{66,67,69} (DNEB) algorithm is used to identify transition state candidates, which are then accurately refined using hybrid eigenvector-following techniques.^{60,61} The two minima reached by (approximate) steepest-descent paths leaving a transition state parallel and antiparallel to the eigenvector with the unique negative eigenvalue define its connectivity. For local minimisations the limited-memory Broyden-Fletcher-Goldfarb-Shanno (LBFGS) algorithm of Liu and Nocedal is used.^{70,71} Discrete paths are then generated systematically from the initial connected path to grow the database by adding all the minima and transition states found during successive connection-making attempts for pairs of minima selected using a missing connection algorithm.⁴⁰ The discrete path that makes the largest contribution to the two-state rate constant within a steady-state approximation for the intervening minima can be extracted from the DPS database using a network formulation⁷² via Dijkstra's shortest-path algorithm⁷³ by choosing suitable edge weights.^{14,15,37,40}

3 Applications

3.1 Molecular clusters

Molecular clusters play an important role in bridging the gap between single molecules and bulk matter,⁷⁴ and especially as model systems for studying solvation.⁷⁵ In particular, water and benzene clusters have attracted special attention.^{76,77} The quest for an accurate molecular description of the liquid and solid phases of water, for which a detailed understanding of the hydrogen-bonding network is critical,⁷⁸ has

resulted in extensive experimental and theoretical studies of water clusters.^{79–82} Benzene enjoys attention as a prototypical aromatic molecule, and its spectral features make benzene clusters conducive to experimental investigation.⁷⁷ The majority of computational studies of water and benzene clusters focus on identifying low-lying minima on the potential energy surface.^{77,83–92} In fact, benzene clusters are often studied to benchmark new global optimisation techniques.^{87,88,91,93} Because of the fundamental interest in understanding the interaction of water with aromatic systems,⁹⁴ small water clusters complexed to a single benzene molecule have also been studied.^{95,96}

As an application of the present computational framework to molecular clusters, the PES for $(\text{C}_6\text{H}_6)_{13}$ was investigated in a recent study by Chakrabarti *et al.*⁹⁷ Although 13 is a magic number for benzene clusters, the symmetry of the corresponding structure is not yet settled. The consensus is on a non-crystallographic structure having a unique interior molecule. Earlier computational studies employed a number of potential energy surfaces,^{77,83,85,86,90} which were found to support low-energy structures possessing either a three-fold proper axis of rotation (C_3) or a centre of inversion (C_i), or both, producing an additional six-fold improper axis of rotation (S_6).⁹⁰ In fact, coexistence of two isomers in the experimental cluster beams was hypothesised to explain the observed spectra.^{77,90} The study by Chakrabarti *et al.* employed a relatively sophisticated atom-atom intermolecular potential,⁹⁸ which included an anisotropic repulsion term in the site-site interaction, and was parameterised via calculations from first principles.⁹⁸ Several low-lying minima for $(\text{C}_6\text{H}_6)_{13}$ were identified with basin-hopping global optimisation.⁹⁷ Six of these structures were found to have symmetry elements other than the identity, as shown in Fig. 1, belonging to the C_3 , C_i , and S_6 point groups in pairs. The mapping of the PES revealed an organisation of the low-lying region, shown in Fig. 2, where efficient relaxation to the C_3 global minimum is hindered by the presence of one or more competing minima separated by relatively high barriers. This picture is consistent with the hypothesis of coexistence for two isomers in spectroscopic experiments.

As an illustration of an application to thermodynamics, Fig. 3 shows the canonical heat capacity curve for the water octamer, bound by the TIP4P potential,⁵⁹ within the harmonic superposition approximation.^{37,63,99} The canonical partition function at temperature T can be written as⁶³

$$Z(T) = \sum_i Z_i(T) \quad (10)$$

and the harmonic approximation for the vibrational density of states gives

$$Z_i(T) = \frac{n_i^* \exp(-V_i/k_B T)}{(\hbar \bar{\nu}_i/k_B T)^\kappa}. \quad (11)$$

Here Z_i is the partition function for minimum i , V_i the potential energy of minimum i , and n_i^* is the number of distinct permutation-inversion isomers, which is inversely proportional to the point group order of the minimum, \bar{v}_i is the geometric mean of the normal mode frequencies for minimum i , and κ is the number of vibrational degrees of freedom. Within this approximation the following expression⁶³ then gives the heat capacity

$$C_V = \kappa k_B + \frac{1}{k_B T^2} \left[\frac{\sum_i V_i^2 Z_i}{Z} - \left(\frac{\sum_i V_i Z_i}{Z} \right)^2 \right]. \quad (12)$$

The heat capacity curve was obtained from a database containing 6171 minima. The estimate for the position of the melting transition is $T = 211.5$ K, which is in excellent agreement with a previous value of $T = 212$ K obtained from parallel tempering simulations.¹⁰⁰ Furthermore, a small second peak appears at low temperatures, as reported previously in ref. 101. This peak corresponds to a solid/solid transition between the two lowest minima of the water octamer with D_{2d} and S_4 symmetry. Including approximate rotational densities of states from a rigid rotor description of the different local minima has little effect on the predicted thermodynamics.

3.2 Colloidal clusters

Colloidal clusters, the mesoscopic counterpart of atomic or molecular clusters, are currently an active field of research.^{102–108} Colloidal particles are large enough to be amenable to direct real space imaging, but small enough to undergo thermal Brownian motion. There has been an unprecedented growth over the last decade in the synthesis of colloidal building blocks with anisotropy in shape and/or surface chemistry.^{32,109–112} Unlike isotropic spherical colloids, which have served extensively as model systems for studying phase transitions in atomic systems,¹¹³ these novel building blocks possess directional interactions, commonly associated with molecular building blocks at a smaller length scale, raising the possibility of ‘colloidal molecules’.^{32,112} Anisotropic colloidal particles, providing scope for tuning the interparticle interactions under laboratory conditions, have proved to be promising building blocks for programmed assembly of functional architectures across length scales.^{105,114} The study of clusters of anisotropic colloids has recently received much attention, especially in the context of finite-sized superstructures and crystal precursors.^{105,108,115} The energetics and kinetics of assembly pathways for small colloidal clusters also promise to reveal fundamental insight into phenomena occurring at early stages of nucleation.¹¹⁶

The present computational framework allows us to develop a predictive understanding of the relationship between the

anisotropic features of the colloidal building blocks and the structure and symmetry of the aggregates they produce. This understanding should be very valuable for structure fabrication *via* self-assembly of tailored building blocks.³² Here we summarise our recent results for chiral colloidal clusters, focusing on designing and exploiting supracolloidal chirality. A general strategy for fabrication of helical architectures with achiral building blocks was devised in a recent study,¹¹⁷ underpinning the physics of emergent chirality at the mesoscopic scale. This study was motivated by an experiment that demonstrated a route to chiral clusters involving two competing length scales, realised with colloidal asymmetric dumbbells linked by a magnetic belt at the waist in the presence of a magnetic field.¹⁰⁵ Global optimisation identified favourable helical structures for small clusters of rigid dipolar dumbbells in the presence of an applied electric field, where the model consisted of Lennard-Jones sites representing the lobes and a point-dipole directed across the axis between the lobes.¹¹⁷ Figure 4 shows that an optimal value for the asymmetry parameter, characterised by the size ratio of the lobes, is crucial for the emergent chirality in the presence of a sufficiently strong electric field. While the dumbbells tend to align perpendicular to the field because of the dipolar interactions, steric interactions controlled by a second length scale are also important. The most intriguing aspect of this study was the demonstration of emergent chirality with oblate ellipsoids of revolution, perhaps the simplest building block that provides a realisation of two competing length scales. In the examples shown in Fig. 5, the oblate ellipsoids of revolution interact *via* two different soft anisotropic pair potentials.^{118,119} An oblate ellipsoid with idealised shape provides a coarse-grained description of discotic molecules, which tend to form columnar liquid crystals. In fact, helical order in columnar architectures is of great interest for achieving high charge-carrier mobility in discotic liquid crystals.^{120–122}

The fastest pathway that was characterised for the reversal of handedness in a helical cluster revealed some interesting features.¹²³ The energy profile for the helix-handedness inversion is shown in Fig. 6. The left- and right-handed helices have the same energy in the absence of any symmetry-breaking terms in the Hamiltonian and are enantiomers. The inversion mechanism involves a boundary between two segments of opposite handedness, which we call a defect, propagating along the helix from one end to the other. The path involves a sequence of low-lying minima as the defect hops between successive pairs of dumbbells. The two segments of opposite handedness rotate in different directions through this sequence of hopping events, marked by periodic expansion and contraction of the helical strand as a whole; the dumbbells reverse their direction of rotation as they switch from one segment to the other. The mechanism involves significant cooperativity, which couples directional rotary motion

to spring-like linear motion.¹²³ This inversion pathway, when driven appropriately, may open up new avenues for fabricating a novel class of nanoscale machines.^{124,125}

As another example, we have deduced a minimalist design rule for a Bernal spiral, *i.e.* a three-stranded helix composed of a chain of face-sharing tetrahedra,¹²⁶ using patchy colloids.¹²⁷ The target structure was realised as the global minimum for a finite-sized cluster of patchy colloids, with two different types of attractive surface sites, described as complementary patches and antipatches. For these patchy colloids, Morgan *et al.* defined a model potential, which involved an isotropic component, describing the interaction between spherical cores, and an anisotropic component governed by two types of complementary patches. A tailored spatial arrangement for three pairs of patches and antipatches, derived from the geometry of a Bernal spiral, produced the target structure as the global minimum for a range of parameter space. A minimalist design rule, which provides realistic targets for state-of-the-art experimental fabrication, was then derived by systematic removal of patches. A patch-antipatch pair offset by about 10° from the directly opposite spatial arrangement was found to be sufficient for the ground state structure to be a Bernal spiral (Fig. 7). It is noteworthy that a recent experiment¹⁰⁸ with Janus spheres observed a Bernal spiral to emerge through kinetics. The two-patch building blocks with a pair of complementary patch and antipatch are similar in spirit to Janus particles, especially for wider patch widths.

As a final example in this subsection, we present the disconnectivity graph for a 21-particle Stockmayer cluster, St_{21} , including the 1000 lowest minima.¹²⁸ The Stockmayer potential (Lennard-Jones plus point-dipole, with strength μ) has been studied extensively to model dipolar colloids. The smallest cluster where knotted structures become energetically most favoured is St_{21} .¹²⁹ A trefoil knot is the global minimum over the range $1.7 \leq \mu < 2.9$, beyond which a pair of stacked rings becomes more favourable. Figure 8 shows the disconnectivity graph for $\mu = 2.9$, exhibiting a pronounced double funnel.¹²⁸ A low-energy pathway connecting these structures was found to mostly involve rearrangement of, and particle exchange between, rings in a link.¹²⁸ It is noteworthy that the angle-axis representation of the rotational coordinates for these dipolar particles with cylindrical symmetry proved to be advantageous even though it includes a redundant degree of freedom for each particle. These redundant degrees of freedom contribute additional zero eigenvalues to the Hessian matrix. Analytical expressions for the corresponding eigenvectors were used for projection in characterising the transition states.⁵⁷

3.3 Biomolecules

To understand the functions of biomolecules, such as proteins, DNA and RNA, it is often necessary to determine their three-

dimensional structure. Methods in biomolecular structure prediction generally fall into two classes. The first class is based on bioinformatics,^{130–134} taking advantage of known correlations between sequence and tertiary structures from a large training set. An alternative class of methods employ atomistic force fields, such as CHARMM,^{135,136} AMBER,^{137,138} OPLS,^{139,140} or others, to model the interatomic and intermolecular interactions. Our aim here is to highlight the potential advantages of using rigid domains in biomolecular structure prediction combined with force-field based methods.

Simulations with all-atom representations tend to become prohibitively expensive for large systems. However, in many cases, biomolecules may be considered as partially rigid molecules. Local rigidifications can be guided by principal component analysis of short molecular dynamics or Monte Carlo trajectories.^{28–31} There are also methods, such as the pebble game,¹⁴¹ where graph theoretical techniques are employed to analyse bond networks, and thus rigidity, in different parts of the molecules of interest. Alternatively, groups of atoms may be frozen simply because their relative deformations are irrelevant for the problems at hand. Unlike the examples discussed above, here each rigid body is distinct and it may consist of an arbitrary number of atoms or interaction sites.

To illustrate the use of the present computational framework, benchmark calculations using the basin-hopping method were carried out by Kusumaatmaja *et al.* for two small peptides,¹⁴² the tryptophan zipper¹⁴³ (PDB: 1LE0; using the CHARMM force field^{135,136}) and chignolin¹⁴⁴ (PDB: 1UAO; using the AMBER force field^{137,138}). In both cases, the mean first encounter time to find the global minimum from random starting configurations is faster with local rigidifications, by factors of 4.2 and 2.5 for the tryptophan zipper and chignolin, respectively. Minimal groupings were used, where only peptide bonds, termini and side chain rings were considered rigid. The groupings are illustrated by the isosurfaces in Fig. 9 for the tryptophan zipper. We note that, in contrast to coarse-graining, this approach still accounts for the full atomistic interactions between the locally rigid domains, though interactions between atoms within the same rigid domain need not be computed.

The largest computational gain comes from the reduced number of basin-hopping steps required to find the global minimum, as the search space is reduced. Local minima that differ only slightly in the atomic configurations of the rigidified groups are no longer accessible. To quantify this reduction in search space, we selected 175,685 local minima from the unrigidified search, and reoptimised the configurations for the local rigidifications presented in Fig. 9. We found that the number of distinct minima decreased to 123,227, *i.e.* a reduction by approximately 30%.

Naturally, more aggressive rigidification results in further

reduction in the number of degrees of freedom, and the efficiency gain will become more important the larger the system size. Unfortunately there is always the danger of over-rigidification, and therefore biophysical or biochemical insights are necessary. To check for self-consistency, structures optimised with rigidification can be relaxed allowing for more intramolecular degrees of freedom up to a fully atomistic description. As an example of advantageous local rigidification, Mochizuki *et al.*¹⁴⁵ have exploited this computational framework to estimate the protein-ligand binding free energy of human aldose reductase. The conformational space of the protein and the protein-ligand complex can be factorised into regions close to and far from the binding pocket. The former region must be fully accounted for, while the latter can be represented by a single, but consistent, configuration for all the calculations. It was found that up to 70-80% of the protein macromolecule may be rigidified while retaining a reasonably accurate representation of the binding free energy.

4 Conclusion

We have outlined a comprehensive computational framework for surveying the underlying potential energy surface of rigid or partially rigid molecules. This approach yields useful information on the structure, thermodynamics, and kinetics, as illustrated by several applications to soft and condensed matter systems, *via* basin-hopping global optimisation,^{38,39} basin-sampling,¹¹ and discrete path sampling^{14,15}. A key element of these procedures is geometry optimisation, which is used recursively to characterise local minima and transition states on the PES. The interface for this framework with rigid and partially rigid molecular models relies upon a computational scheme that uses an angle-axis representation for the rigid-body rotational coordinates. This representation, when implemented using a matrix formulation, has proved to be robust as well as computationally efficient for gradient-based geometry optimisation, and provides a user-friendly interface.

The versatility of a rigid-body description for particle-based simulations goes beyond what is described in the preceding sections. For example, a number of theoretical and computational methodologies have attempted to address the challenges of modelling viral capsid assembly.¹⁴⁶ The formation of capsids with well-defined morphologies illustrates the remarkable ability of biological matter to self-assemble, where a protective coat for the genetic material is formed by a large number of protein subunits. Each of the protein subunits contains thousands of atoms, making particle-based simulations with all-atom representation prohibitively expensive for the timescales (milliseconds or larger) of interest. Instead the subunits are treated as rigid bodies in a wide class of coarse-grained models.¹⁴⁶ In another application of the present computational framework, Fejer *et al.* systematically manipulated

anisotropic interactions in a simple coarse-grained model, where an ellipsoid core and two repulsive LJ sites are held together rigidly.¹⁴⁷ A wide variety of complex morphologies, including shells, tubes and spirals were produced.¹⁴⁷

In the majority of the applications so far, the present framework has dealt with finite-sized systems. However, it can be extended to explore the crystal energy landscapes for small organic molecules, where reasonably accurate intermolecular potentials are available within a rigid-body approximation.^{148–150} Applications to predict crystal structures of molecular solids and polymorphism should be very valuable to crystal engineering.¹⁵¹ While the global minimum in the crystal energy landscape corresponds to the thermodynamic crystal structure, the low-lying minima constitute likely candidates for polymorphic forms. The barrier heights between low-lying minima on the crystal energy landscape are critical for a predictive understanding of polymorphism.¹⁵⁰

5 Appendix

The prescription for obtaining the first derivatives of the energy with respect to the rigid-body coordinates is outlined here for site-site isotropic potentials to highlight the salient features of the angle-axis representation of rotation in its matrix formulation. The position vector of the centre-of-mass, $\mathbf{r} = [r_1, r_2, r_3]$, and the rotation vector, \mathbf{p} , account for the six rigid-body coordinates. However, it is important to note that \mathbf{p} must be treated differently from \mathbf{r} . If the coordinates of two rigid bodies are denoted using the superscripts I and J , and the sites within each rigid body by subscripts i and j , then for site-site isotropic potentials the total energy is

$$U = \sum_I \sum_{J < I} \sum_{i \in I} \sum_{j \in J} f_{ij}(r_{ij}), \quad (13)$$

where $r_{ij} = |\mathbf{r}_{ij}| = |\mathbf{r}_i - \mathbf{r}_j|$ and $f_{ij} \equiv U_{ij}^{IJ}$ is the pair potential between sites i and j . If ζ represents one of the six coordinates of rigid body I , then the first derivative of the potential energy is

$$\frac{\partial U}{\partial \zeta} = \sum_{J \neq I} \sum_{i \in I} \sum_{j \in J} f'_{ij}(r_{ij}) \frac{\partial r_{ij}}{\partial \zeta}, \quad (14)$$

where $f'_{ij} = df_{ij}(r_{ij})/dr_{ij}$. Here

$$\frac{\partial r_{ij}}{\partial \mathbf{r}^I} = \hat{\mathbf{r}}_{ij}, \quad (15)$$

and

$$\frac{\partial r_{ij}}{\partial p_k^I} = \hat{\mathbf{r}}_{ij} \cdot \frac{\partial \mathbf{r}_{ij}}{\partial p_k^I} = \hat{\mathbf{r}}_{ij} \cdot (\mathbf{R}_k^I \mathbf{r}_i^0), \quad (16)$$

where the following relationship was used

$$\mathbf{r}_{ij} = \mathbf{r}^I + \mathbf{R}^I \mathbf{r}_i^0 - \mathbf{r}^J - \mathbf{R}^J \mathbf{r}_j^0. \quad (17)$$

Here, the vectors in the body-fixed frame are denoted by the superscript 0. The Hessian matrix for site-site isotropic potentials can easily be obtained within this scheme, which can also deal efficiently with site-site anisotropic potentials. Further details, including the treatment of anisotropic potentials can be found in ref. 57. The present framework has been implemented in the global optimisation program, GMIN, and the geometry optimisation program, OPTIM, both of which are available for use under the Gnu public software licence.^{152,153} The programs contain a variety of rigid-body intermolecular potentials, including dipolar and polarisable potentials with analytic derivatives.

Acknowledgements

The authors gratefully acknowledge support from the EPSRC and European Research Council. DC acknowledges support also through a Marie Curie Incoming International Fellowship from the European Commission and an Oppenheimer Research Fellowship from the University of Cambridge.

References

- 1 *Computer Simulations in Condensed Matter: From Materials to Chemical Biology Volume 1*, ed. M. Ferrario, G. Ciccotti and K. Binder, Springer, Berlin Heidelberg, 2006.
- 2 M. E. Tuckerman, *Statistical Mechanics: Theory and Molecular Simulation*, Oxford University Press, Oxford, 2010.
- 3 G. M. Torrie and J. P. Valleau, *J. Comput. Phys.*, 1977, **23**, 187–199.
- 4 A. P. Lyubartsev, A. A. Martinovski, S. V. Shevkunov and P. N. Vorontsov-Velyaminov, *J. Chem. Phys.*, 1992, **96**, 1776–1783.
- 5 B. A. Berg and T. Neuhaus, *Phys. Rev. Lett.*, 1992, **68**, 9–12.
- 6 E. Marinari and G. Parisi, *Europhys. Lett.*, 1992, **19**, 451–458.
- 7 J. Lee, *Phys. Rev. Lett.*, 1993, **71**, 211–214.
- 8 Y. Sugita and Y. Okamoto, *Chem. Phys. Lett.*, 1999, **314**, 141–151.
- 9 F. Wang and D. P. Landau, *Phys. Rev. Lett.*, 2001, **86**, 2050–2053.
- 10 A. Laio and M. Parrinello, *Proc. Natl. Acad. Sci. USA*, 2002, **99**, 12562–12566.
- 11 T. Bogdan, D. J. Wales and F. Calvo, *J. Chem. Phys.*, 2006, **124**, 044102.
- 12 C. Dellago, P. G. Bolhuis and P. L. Geissler, *Adv. Chem. Phys.*, 2002, **123**, 1–78.
- 13 P. G. Bolhuis, D. Chandler, C. Dellago and P. L. Geissler, *Annu. Rev. Phys. Chem.*, 2002, **53**, 291–318.
- 14 D. J. Wales, *Mol. Phys.*, 2002, **100**, 3285–3306.
- 15 D. J. Wales, *Mol. Phys.*, 2004, **102**, 891–908.
- 16 A. K. Faradjian and R. Elber, *J. Chem. Phys.*, 2004, **120**, 10880–10889.
- 17 A. M. A. West, R. Elber and D. Shalloway, *J. Chem. Phys.*, 2007, **126**, 145104.
- 18 R. J. Allen, P. B. Warren and P. R. ten Wolde, *Phys. Rev. Lett.*, 2005, **94**, 018104.
- 19 R. J. Allen, C. Valeriani and P. R. ten Wolde, *J. Phys.: Condens. Matter*, 2009, **21**, 463102.
- 20 K. Kremer and G. S. Grest, *J. Chem. Phys.*, 1990, **92**, 5057–5086.
- 21 W. Tschöp, K. Kremer, J. Batoulis, T. Burger and O. Hahn, *Acata Polym.*, 1998, **49**, 61–74.
- 22 J. C. Shelley, M. Y. Shelley, R. C. Reeder, S. Bandyopadhyay and M. L. Klein, *J. Chem. Phys. B*, 2001, **105**, 4464–4470.
- 23 S. Izvekov and G. A. Voth, *J. Chem. Phys.*, 2005, **123**, 134105.
- 24 S. D. Chao, J. D. Kress and A. Redondob, *J. Chem. Phys.*, 2005, **122**, 234912.
- 25 L. Yelash, M. Müller, W. Paul and K. Binder, *J. Chem. Theory Comput.*, 2006, **2**, 588–597.
- 26 A. Y. Shih, A. Arkhipov, P. L. Freddolino and K. Schulten, *J. Phys. Chem. B*, 2006, **110**, 3674–3684.
- 27 V. Rühle, C. Junghans, A. Lukyanov, K. Kremer and D. Andrienko, *J. Chem. Theory Comput.*, 2009, **21**, 463102.
- 28 A. Kitao and N. Go, *Curr. Opin. Struct. Biol.*, 1999, **9**, 164–169.
- 29 B. R. Brooks, D. Janežič and M. Karplus, *J. Comput. Chem.*, 1995, **16**, 1522–1542.
- 30 M. A. Balsera, W. Wriggers, Y. Oono and K. Schulten, *J. Phys. Chem.*, 1996, **100**, 2567–2572.
- 31 G. G. Maisuradze, A. Liwo and H. A. Scheraga, *J. Mol. Biol.*, 2009, **385**, 312–329.
- 32 S. C. Glotzer and M. J. Solomon, *Nature Mater.*, 2007, **6**, 557–562.
- 33 G. Ciccotti and J. P. Ryckaert, *Comput. Phys. Rep.*, 1986, **4**, 345–392.
- 34 D. J. Evans, *Mol. Phys.*, 1977, **34**, 317–325.
- 35 D. C. Rapaport, *J. Comput. Phys.*, 1985, **60**, 306–314.
- 36 J.-P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, *J. Comput. Phys.*, 1977, **23**, 327–341.
- 37 D. J. Wales, *Energy Landscapes*, Cambridge University Press, Cambridge, 2003.
- 38 Z. Li and H. A. Scheraga, *Proc. Natl. Acad. Sci. USA*, 1987, **84**, 6611–6615.

-
- 39 D. J. Wales and J. P. K. Doye, *J. Phys. Chem. A*, 1997, **111**, 5111–5116.
- 40 J. M. Carr and D. J. Wales, *J. Phys. Chem. B*, 2008, **112**, 8760–8769.
- 41 F. Noé and S. Fischer, *Curr. Opin. Struct. Biol.*, 2008, **18**, 154–162.
- 42 D. J. Wales, *Curr. Opin. Struct. Biol.*, 2010, **20**, 3–10.
- 43 D. J. Wales, *Int. Rev. Phys. Chem.*, 2006, **25**, 237–282.
- 44 O. M. Becker and M. Karplus, *J. Chem. Phys.*, 1997, **106**, 1495–1517.
- 45 D. J. Wales, M. A. Miller and T. R. Walsh, *Nature*, 1998, **394**, 758–760.
- 46 H. Goldstein, *Classical Mechanics*, Addison-Wesley, Reading, MA, 1980.
- 47 D. J. Wales and I. Ohmine, *J. Chem. Phys.*, 1993, **98**, 7257–7268.
- 48 S. L. Price, A. J. Stone and M. Alderton, *Mol. Phys.*, 1984, **52**, 987–1001.
- 49 P. L. A. Popelier and A. J. Stone, *Mol. Phys.*, 1995, **82**, 411–425.
- 50 P. L. A. Popelier, A. J. Stone and D. J. Wales, *Faraday Discuss.*, 1994, **97**, 243–264.
- 51 D. J. Wales, *Phil. Trans. R. Soc. A*, 2005, **363**, 357–377.
- 52 M. P. Allen and G. Germano, *Mol. Phys.*, 2006, **104**, 3225–3235.
- 53 A. O. Griewank, B. R. Markey and D. J. Evans, *J. Chem. Phys.*, 1979, **60**, 3449–3454.
- 54 W. R. Hamilton, *Proc. R. Irish Acad.*, 1844, **2**, 424–434.
- 55 C. F. F. Karney, *J. Mol. Graph. Mod.*, 2007, **25**, 595–604.
- 56 J.-Y. Trosset and H. A. Scheraga, *J. Comput. Phys.*, 1999, **20**, 412–427.
- 57 D. Chakrabarti and D. J. Wales, *Phys. Chem. Chem. Phys.*, 2009, **11**, 1970–1976.
- 58 R. M. Murray, Z. Li and S. S. Sastry, *A Mathematical Introduction to Robotic Manipulation*, CRC Press, Boca Raton, FL, USA, 1994.
- 59 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926–935.
- 60 L. J. Munro and D. J. Wales, *Phys. Rev. B*, 1999, **59**, 3969–3980.
- 61 Y. Kumeda, D. J. Wales and L. J. Munro, *Chem. Phys. Lett.*, 2001, **341**, 185–194.
- 62 J. Baker and W. J. Hehre, *J. Comput. Chem.*, 1991, **12**, 606–610.
- 63 D. J. Wales, *Mol. Phys.*, 1993, **78**, 151–171.
- 64 E. E. Lattman, *Acta Cryst.*, 1972, **B 28**, 1065–1068.
- 65 V. Rühle, H. Kusumaatmaja, D. Chakrabarti and D. J. Wales, *J. Chem. Theory Comput.*, 2013, **9**, DOI: 10.1021/ct400403y.
- 66 G. Henkelman and H. Jónsson, *J. Chem. Phys.*, 2000, **113**, 9978–9985.
- 67 G. Henkelman, B. P. Uberuaga and H. Jónsson, *J. Chem. Phys.*, 2000, **113**, 9901–9904.
- 68 S. A. Trygubenko and D. J. Wales, *J. Chem. Phys.*, 2004, **120**, 2082–2094.
- 69 G. Henkelman and H. Jónsson, *J. Chem. Phys.*, 1999, **111**, 7010–7022.
- 70 J. Nocedal, *Math. Comput.*, 1980, **35**, 773–782.
- 71 D. Liu and J. Nocedal, *Mathematical Programming*, 1989, **45**, 503–528.
- 72 D. A. Evans and D. J. Wales, *J. Chem. Phys.*, 2004, **121**, 1080–1090.
- 73 E. W. Dijkstra, *Numerische Mathematlk*, 1959, **1**, 269–271.
- 74 L. S. Bartell, *Annu. Rev. Phys. Chem.*, 1998, **49**, 43–72.
- 75 T. S. Zwier, *Annu. Rev. Phys. Chem.*, 1996, **47**, 205–241.
- 76 F. N. Keutsch and R. J. Saykally, *Proc. Natl. Acad. Sci. USA*, 2001, **98**, 10533–10540.
- 77 D. C. Easter, *J. Phys. Chem. A*, 2003, **107**, 7733–7742.
- 78 A. Luzar and D. Chandler, *Nature*, 1996, **379**, 55–57.
- 79 K. Liu, M. G. Brown, J. D. Cruzan and R. J. Saykally, *Science*, 1996, **271**, 62–64.
- 80 K. Liu, M. G. Brown, C. Carter, R. J. Saykally, J. K. Gregory and D. C. Clary, *Nature*, 1996, **381**, 501–503.
- 81 C. Pérez, M. T. Muckle, D. P. Zaleski, N. A. Seifert, G. C. S. B. Temelso, Z. Kisiel and B. H. Pate, *Science*, 2012, **336**, 897–901.
- 82 D. J. Wales, *J. Am. Chem. Soc.*, 1993, **115**, 11180–11190.
- 83 D. E. Williams, *Acta Cryst.*, 1980, **A36**, 715–723.
- 84 B. W. van de Waal, *Acta Cryst.*, 1981, **A37**, 762–764.
- 85 B. W. van de Waal, *J. Chem. Phys.*, 1983, **79**, 3948–3961.
- 86 F. J. Dulles and L. S. Bartell, *J. Phys. Chem.*, 1995, **99**, 17100–17106.
- 87 R. P. White and H. R. Mayne, *Chem. Phys. Lett.*, 1998, **289**, 463–468.
- 88 R. P. White, J. A. Niesse and H. R. Mayne, *J. Chem. Phys.*, 1998, **108**, 2208–2218.
- 89 C. Gonzalez and E. C. Lim, *J. Phys. Chem. A*, 2001, **105**, 1904–1908.
- 90 D. C. Easter, *J. Phys. Chem. A*, 2003, **107**, 2148–2159.
- 91 H. Takeuchi, *J. Chem. Inf. Model.*, 2007, **47**, 104–109.
- 92 A. S. Mahadevi, A. P. Rahalkar, S. R. Gadre and G. N. Sastry, *J. Chem. Phys.*, 2010, **133**, 164308.
- 93 J. L. Llanio-Trujillo, J. M. C. Marques and F. B. Pereira, *J. Phys. Chem. A*, 2011, **115**, 2130–2138.
- 94 U. Zimmerli, M. Parrinello and P. Koumoutsakos, *J. Chem. Phys.*, 2004, **120**, 2693–2699.
- 95 R. N. Pribble and T. S. Zwier, *Science*, 1994, **265**, 75–79.
-

- 96 C. J. Gruenloh, J. R. Carney, C. A. Arrington, T. S. Zwier, S. Y. Fredericks and K. D. Jordan, *Science*, 1997, **276**, 1678–1681.
- 97 D. Chakrabarti, T. S. Totton, M. Kraft and D. J. Wales, *Phys. Chem. Chem. Phys.*, 2011, **13**, 21362–21366.
- 98 T. S. Totton, A. J. Misquitta and M. Kraft, *J. Chem. Theory Comput.*, 2010, **6**, 683–695.
- 99 M. R. Hoare and P. Pal, *J. Cryst. Growth*, 1972, **17**, 77–96.
- 100 J. Hernández-Rojas, B. S. Gonzalez, T. James and D. J. Wales, *J. Chem. Phys.*, 2006, **125**, 224302.
- 101 M. A. C. Pablo Nigra and S. Kais, *J. Chem. Phys.*, 2001, **115**, 2621–2628.
- 102 V. N. Manoharan, M. T. Elsesser and D. J. Pine, *Science*, 2003, **301**, 483–487.
- 103 F. Baletto and R. Ferrando, *Rev. Mod. Phys.*, 2005, **77**, 371–423.
- 104 L. Hong, A. Cacciuto, E. Luijten and S. Granick, *Nano Lett.*, 2006, **6**, 2510–2514.
- 105 D. Zerrouki, J. Baudry, D. Pine, P. Chaikin and J. Bibette, *Nature*, 2008, **455**, 380–382.
- 106 G. Meng, N. Arkus, M. P. Brenner and V. N. Manoharan, *Science*, 2010, **327**, 560–563.
- 107 J. A. Fan, C. Wu, K. Bao, J. Bao, R. Bardhan, N. J. Halas, V. N. Manoharan, P. Nordlander, G. Shvets and F. Capasso, *Science*, 2010, **328**, 1135–1138.
- 108 Q. Chen, J. K. Whitmer, S. Jiang, S. C. Bae, E. Luijten and S. Granick, *Science*, 2011, **331**, 199–202.
- 109 S.-M. Yang, S.-H. Kim, J.-M. Lim and G.-R. Yi, *J. Mater. Chem.*, 2008, **18**, 2177–2190.
- 110 S. Sacanna, W. T. M. Irvine, P. M. Chaikin and D. J. Pine, *Nature*, 2010, **464**, 575–578.
- 111 A. B. Pawar and I. Kretzschmar, *Macromol. Rapid Commun.*, 2010, **31**, 150–168.
- 112 F. Li, D. P. Josephson and A. Stein, *Angew. Chem. Int. Ed.*, 2011, **50**, 360–388.
- 113 V. J. Anderson and H. N. W. Lekkerkerker, *Nature*, 2002, **416**, 811–815.
- 114 Q. Chen, S. C. Bae and S. Granick, *Nature*, 2011, **469**, 381–384.
- 115 D. J. Kraft, R. Ni, F. Smallenburg, M. Hermes, K. Yoon, D. A. Weitz, A. van Blaaderen, J. Groenewold, M. Dijkstra and W. K. Kegel, *Proc. Natl. Acad. Sci. USA*, 2012, **109**, 10787–10792.
- 116 R. W. Perry, G. Meng, T. G. Dimiduk, J. Fung and V. N. Manoharan, *Faraday Discuss.*, 2012, **159**, 211–234.
- 117 D. Chakrabarti, S. N. Fejer and D. J. Wales, *Proc. Natl. Acad. Sci. USA*, 2009, **106**, 20164–20167.
- 118 L. Paramonov and S. N. Yaliraki, *J. Chem. Phys.*, 2005, **123**, 194111.
- 119 M. A. Bates and G. R. Luckhurst, *J. Chem. Phys.*, 1996, **104**, 6696–6709.
- 120 D. Adam, P. Schuhmacher, J. Simmerer, L. Häussling, K. Siemensmeyer, K. H. Eitzbachi, H. Ringsdorf and D. Haarer, *Nature*, 1994, **371**, 141–143.
- 121 L. SchmidtMende, A. Fechtenkötter, K. Müllen, E. Moons, R. H. Friend and J. D. MacKenzie, *Science*, 2001, **Science**, 1119–1122.
- 122 V. Percec, M. Glodde, T. K. Bera, Y. Miura, I. S. I, K. D. Singer, V. S. Balagurusamy, P. A. Heiney, I. Schnell, A. R. A, H. W. Spiess, S. D. Hudson and H. Duan, *Nature*, 2002, **419**, 384–387.
- 123 D. Chakrabarti and D. J. Wales, *Soft Matter*, 2011, **7**, 2325–2328.
- 124 G. A. Ozin, I. Manners, S. Fournier-Bidoz and A. Arsenault, *Adv. Mater.*, 2005, **17**, 3011–3018.
- 125 K. Miwa, Y. Furusho and E. Yashima, *Nature Chem.*, 2010, **2**, 444–449.
- 126 J. D. Bernal, *Proc. R. Soc. Lond. A*, 1964, **280**, 299–322.
- 127 J. W. R. Morgan, D. Chakrabarti, N. Dorsaz and D. J. Wales, *ACS Nano*, 2013, **7**, 1246–1256.
- 128 J. D. Farrell, C. Lines, J. D. Shepherd, D. Chakrabarti, M. A. Miller and D. J. Wales, *Soft Matter*, 2013, **9**, 5407–5416.
- 129 M. A. Miller and D. J. Wales, *J. Phys. Chem. B*, 2005, **109**, 23109–23112.
- 130 M. Rehmsmeier, P. Steffen, M. Hehsmann and R. Giegerich, *RNA*, 2004, **10**, 1507–1517.
- 131 E. Gasteiger, A. Gattiker, C. Hoogland, I. Ivanyi, R. D. Appel and A. Bairoch, *Nucleic Acids Res.*, 2003, **31**, 3784–3788.
- 132 L. A. Kelley, R. M. MacCallum and M. J. Sternberg, *J. Mol. Biol.*, 2000, **299**, 501–522.
- 133 Y. Zhang, *BMC Bioinformatics*, 2008, **9**, 40.
- 134 K. Bryson, L. J. McGuffin, R. L. Marsden, J. J. Ward, J. S. Sodhi and D. T. Jones, *Nucleic Acids Res.*, 2005, **33**, W36–W38.
- 135 B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan and M. Karplus, *J. Comput. Chem.*, 1983, **4**, 187217.
- 136 B. R. Brooks, C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caffisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York and M. Karplus, *J. Comput. Chem.*, 2009, **30**, 15451614.
- 137 J. W. Ponder and D. A. Case, *Adv. Prot. Chem.*, 2003, **66**,

-
- 27–85.
- 138 D. Case, T. Darden, T. Cheatham, III, C. Simmerling, J. Wang, R. Duke, R. Luo, R. Walker, W. Zhang, K. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A. Goetz, I. Kolossvai, K. Wong, F. Paesani, J. Vanicek, R. Wolf, J. Liu, X. Wu, S. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D. Roe, D. Mathews, M. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko and P. Kollman, *Amber 12*, University of California, San Francisco, 2012.
- 139 W. L. Jorgensen and J. Tirado-Rives, *J. Am. Chem. Soc.*, 1988, **110**, 1657–1666.
- 140 W. L. Jorgensen, D. S. Maxwell and J. Tirado-Rives, *J. Am. Chem. Soc.*, 1996, **118**, 11225–11236.
- 141 D. J. Jacobs and M. F. Thorpe, *Phys. Rev. Lett.*, 1995, **75**, 4051–4054.
- 142 H. Kusumaatmaja, C. S. Whittleston and D. J. Wales, *J. Chem. Theory Comput.*, 2012, **8**, 5159–5165.
- 143 A. G. Cochran, N. J. Skelton and M. A. Starovasnik, *Proc. Nat. Acad. Sci. USA*, 2001, **98**, 5578–5583.
- 144 S. Honda, K. Yamasaki, Y. Sawada and H. Morii, *Structure*, 2004, **12**, 1507–1518.
- 145 K. Mochizuki, C. S. Whittleston, S. Somani, H. Kusumaatmaja and D. J. Wales, *Communicated for publication*, 2013.
- 146 M. F. Hagan, *Adv. Chem. Phys.*, 2013, **155**, in press.
- 147 S. N. Fejer, D. Chakrabarti and D. J. Wales, *ACS Nano*, 2010, **4**, 219–228.
- 148 S. M. Woodley and R. Catlow, *Nature Mater.*, 2008, **7**, 937–946.
- 149 J. Pillardy, Y. A. Arnautova, C. Czaplewski, K. D. Gibson and H. A. Scheraga, *Proc. Natl. Acad. Sci. USA*, 2001, **98**, 12351–12356.
- 150 S. L. Price, *Acc. Chem. Res.*, 2009, **42**, 117–126.
- 151 G. R. Desiraju, *Angew. Chem. Int. Ed.*, 2007, **46**, 8342–8356.
- 152 D. J. Wales and T. V. Bogdan, GMIN: A program for finding global minima and calculating thermodynamic properties from basin-sampling; <http://www-wales.ch.cam.ac.uk/GMIN>.
- 153 D. J. Wales, OPTIM: A program for optimising geometries and calculating reaction pathways; <http://www-wales.ch.cam.ac.uk/OPTIM>.

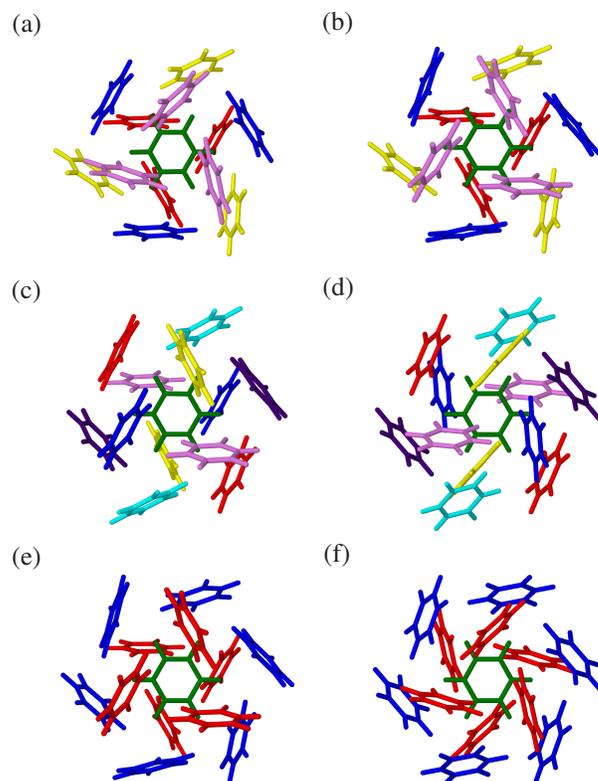


Fig. 1: Structures of six low-lying minima for $(C_6H_6)_{13}$. Distinct pairs of structures with C_3 (a,b), C_i (c,d), and S_6 (e,f) symmetry are shown with the central benzene molecule (green) is in the plane of the page and molecules belonging to the same orbit of the point group identified by the same colour. The structures in (a), (c), and (e) have lower energies than (b), (d), and (f), respectively. The structure in (a) corresponds to the global minimum. Reproduced from Ref. 97.

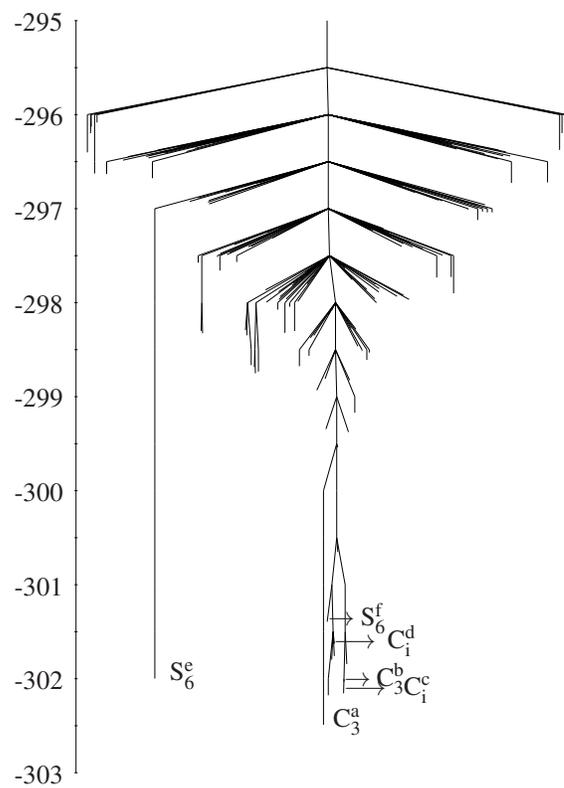


Fig. 2: Disconnectivity graph for the $(\text{C}_6\text{H}_6)_{13}$ cluster. The lowest 200 minima are shown, highlighting the six low-lying minima shown in Fig. 1. Superscripts are used to distinguish between structures of the same symmetry, following the labels for the panels of Fig. 1. The unit of energy is kJ mol^{-1} on the vertical axis. The graph groups enantiomers together, where applicable. Reproduced from Ref. 97.

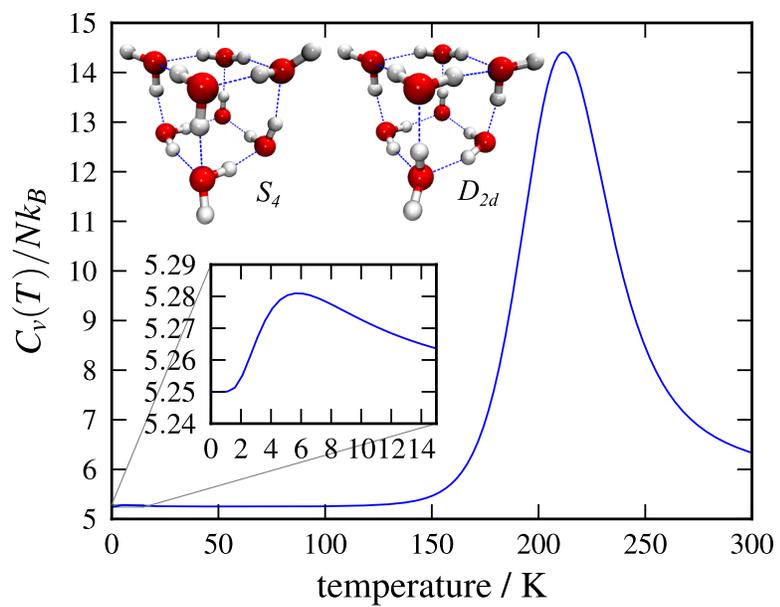


Fig. 3: Canonical heat capacity curve for the water octamer, described by the TIP4P pair potential.

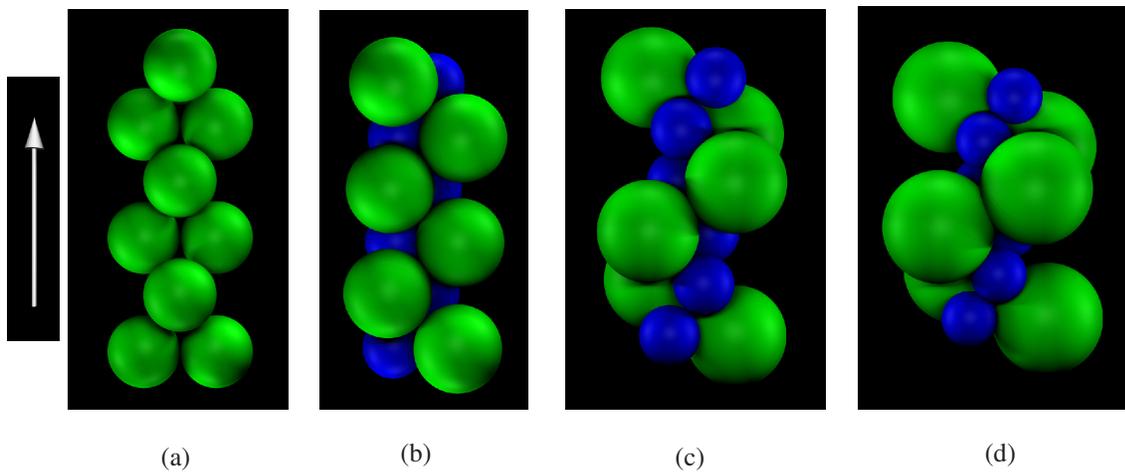
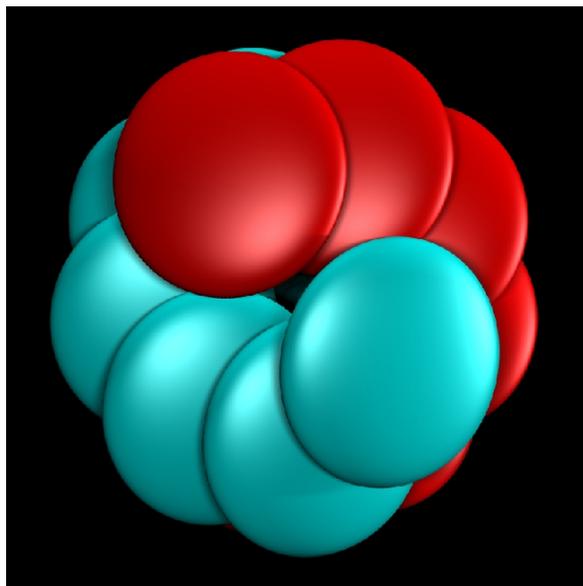
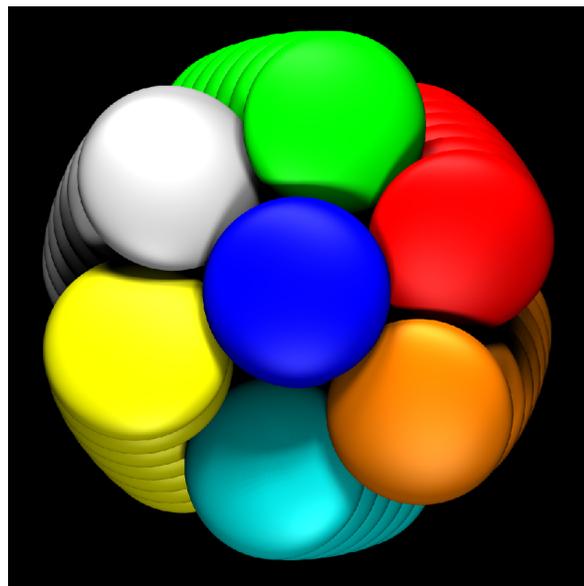


Fig. 4: Global minima for clusters of six dipolar dumbbells in the presence of an applied electric field, illustrating the importance of the asymmetry of the building blocks for this field-driven assembly to produce a helix. Here the asymmetry parameter α , characterised by the size ratio of the lobes of the dumbbells, is increased from 1 to 2 for dipole moment $\mu_D = 1$ and electric field $E = 5$, with the latter parameters in reduced units:¹¹⁷ (a) $\alpha = 1$; (b) $\alpha = 1.43$; (c) $\alpha = 1.67$; (d) $\alpha = 2$. The arrow corresponds to the field direction. Helices are observed in (c) and (d). Reproduced from Ref. 117.



(a)



(b)

Fig. 5: Global minima for clusters of N axially symmetric discoids bound by the oblate ellipsoids of revolution that interact *via* two different soft anisotropic pair potentials. (a) $N = 13$ with the Paramonov-Yaliraki potential¹¹⁸ and (b) $N = 49$ with a modified Gay-Berne pair potential.¹¹⁹ Different colours are used to distinguish between the stacks. Reproduced from Ref. 117.

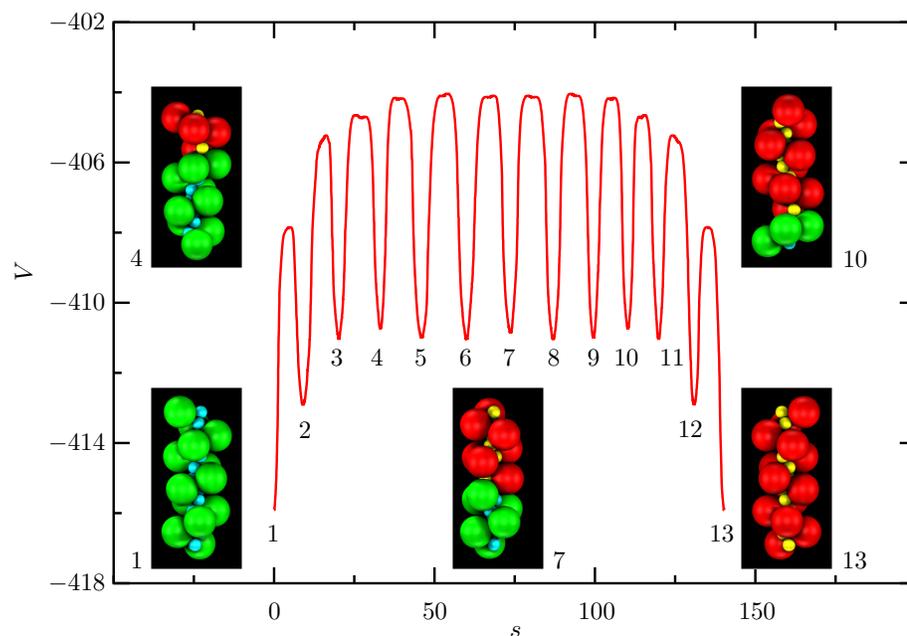


Fig. 6: Potential energy profile for a supracolloidal helix along the fastest pathway identified between enantiomers. The potential energy, V , is shown as a function of the integrated Cartesian path length along the pathway, s . The structures of some of the low-lying minima, which are numbered along the pathway, are superimposed on the energy profile, including the global minima at either end. The two segments of opposite handedness are distinguished by different colours. The right-handed segment is shown in green and cyan and the left-handed one in red and yellow. The helical axis for the global minima is parallel to the static electric field. V and s are in reduced units. Reproduced from Ref. 123 with permission from The Royal Society of Chemistry.

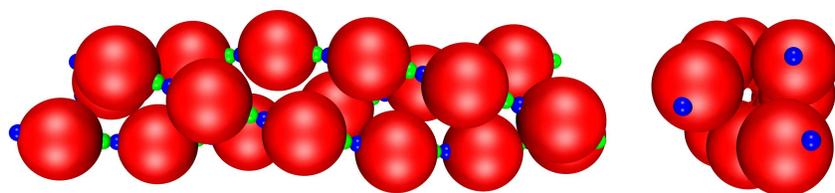


Fig. 7: Global minimum for 18 patchy colloids, each decorated with a patch-antipatch pair offset by roughly 10° from the directly opposite spatial arrangement. (a) Side view; (b) top view. Reprinted with permission from Ref. 127. Copyright 2013 American Chemical Society.

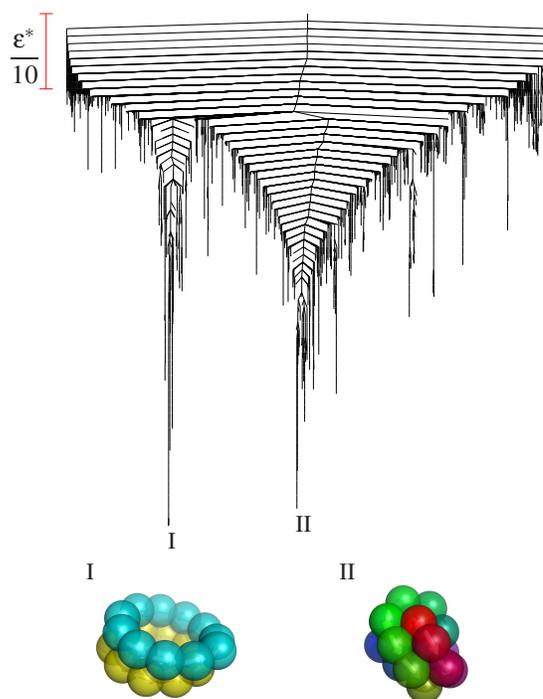


Fig. 8: Disconnectivity graph for St_{21} with a reduced dipole moment value of $\mu = 2.90$. Branches to the 1000 lowest energy minima are shown. Here, ϵ^* is the full Stockmayer pair energy for parallel head-to-tail dipole vectors at the corresponding value of dipole moment. The two structures correspond to the stacked ring (I) and trefoil knot (II) minima. Reproduced from Ref. 128 with permission from The Royal Society of Chemistry.

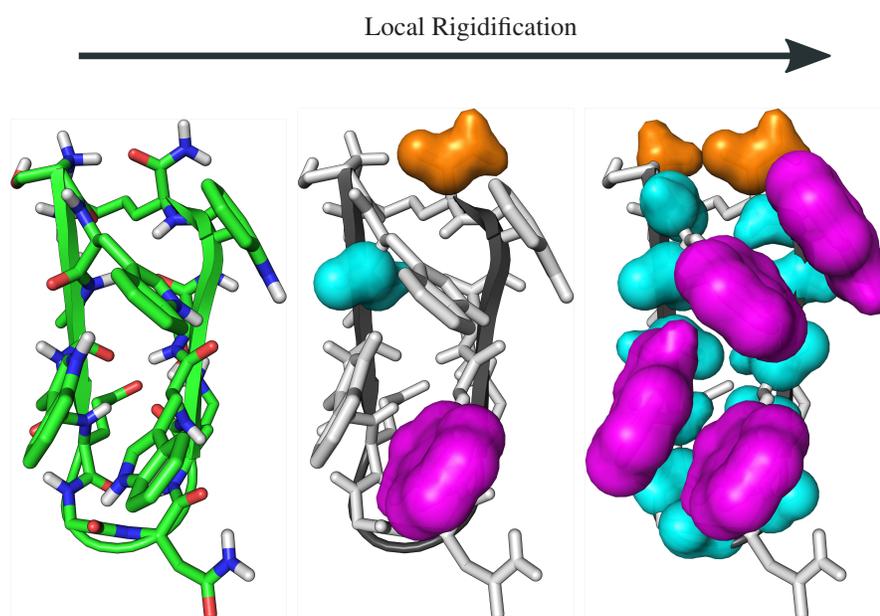


Fig. 9: Global minimum configuration of the tryptophan zipper (trpzip 1, PDB: 1LE0). From left to right, an increasing number of atoms is locally rigidified during global optimisation. The magenta, cyan, and orange isosurfaces correspond to the tryptophan side chain rings, peptide bonds, and end groups.