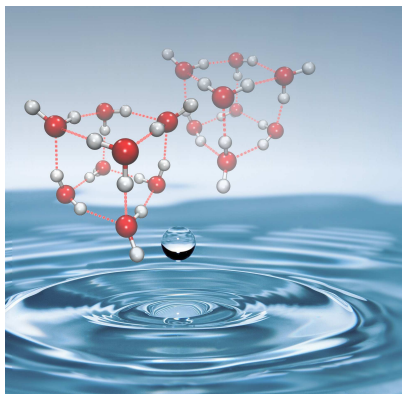
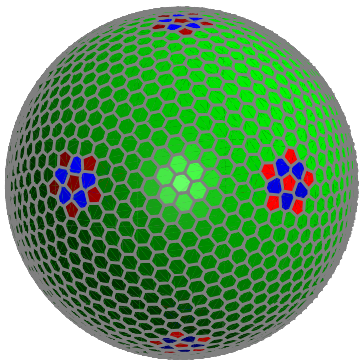
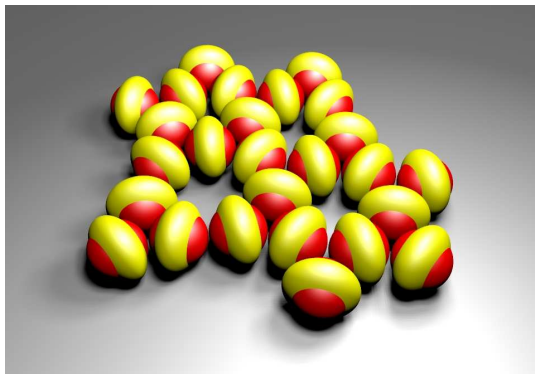
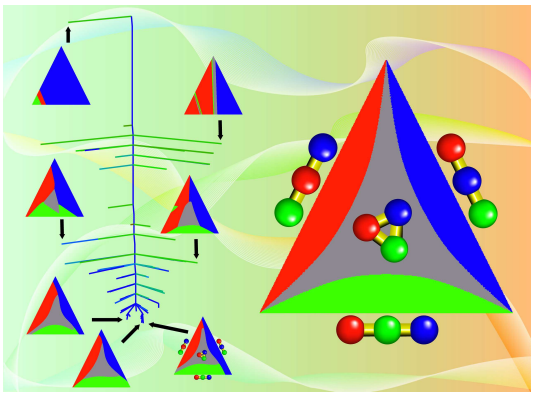
Energy Landscapes: Structure, Dynamics, and Thermodynamics

Recap: exploit stationary points (minima and transition states) as a conceptual and computational framework (*J. Phys. Chem. B*, **110**, 20765, 2006)

- Basin-hopping for global optimisation (*J. Phys. Chem. A*, **101**, 5111 1997)
- Basin-sampling for global thermodynamics (*J. Chem. Phys.*, **124**, 044102, 2006)
  Superposition Enhanced Nested Sampling (*Phys. Rev. X*, **4**, 031034, 2014)
- Discrete path sampling for global kinetics (*Mol. Phys.*, **100**, 3285, 2002)
- Coarse-grained models employ generalised local rigid bodies and ellipsoidal building blocks (*JCTC*, **9**, 4026, 2013; *Soft Matter*, **11**, 6663, 2015)

Symmetrised Basin-Hopping (*Phys. Chem. Chem. Phys.*, **15**, 3965, 2013)

Searches are biased towards geometries with a higher approximate symmetry measure, defined in terms of a cluster core and an orbit analysis.

- Justification from the *principle of maximum symmetry*.
- The bias is not specific to any particular point group or morphology.

Mean first encounter times for the global minima of LJ_N clusters averaged over 100 random starting points (Intel Xeon E5405 cpu running at 2.00 GHz):

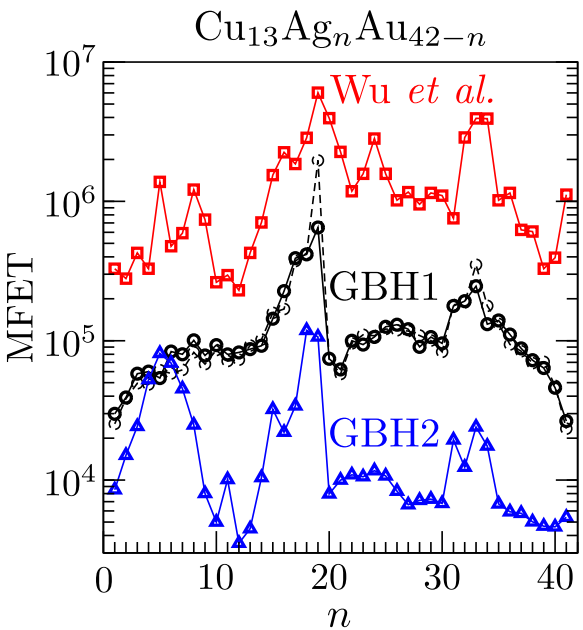
N	method	energy evaluations	minimisations	cpu time/s
38	none	185,493	1,271	4.4
38	core orbits symmetrisation	20,655	142	0.5
38	continuous symmetry measure	4,369	34	0.2
75	none	8,230,648	61,668	688.7
75	core orbits symmetrisation	50,229	338	4.2
98	none	7,017,387	48,301	1,314.3
98	core orbits symmetrisation	109,753	563	18.5

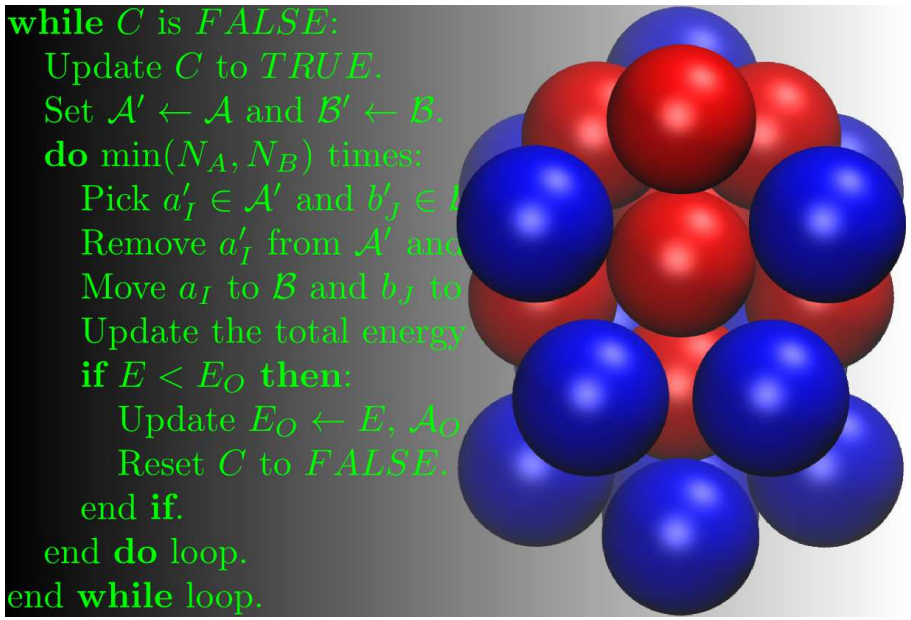
Generalised Basin-Hopping (*PRL*, **113**, 156102, 2014; *PCCP*, **17**, 902, 2016)

In GBH we focus on biminima, defined as structures where the energy cannot be lowered by interchanging inequivalent particles and requenching.

For multicomponent nanoalloy metal clusters fewer biminima exist for segregated systems with lower lattice mismatch.

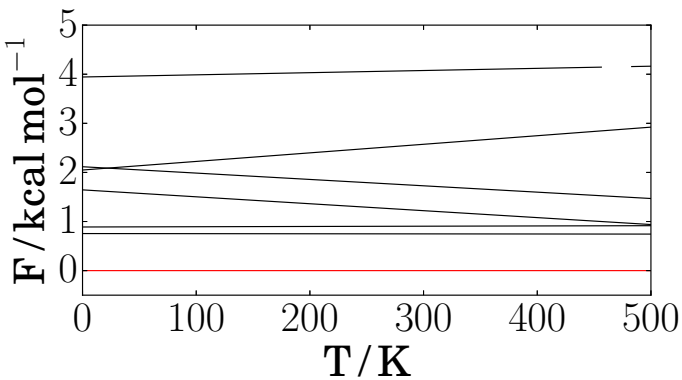
In GBH1 the biminimum property is verified by scanning the local neighbourhood, while in GBH2 a subset is considered. (*PCCP*, **17**, 28331, 2015)

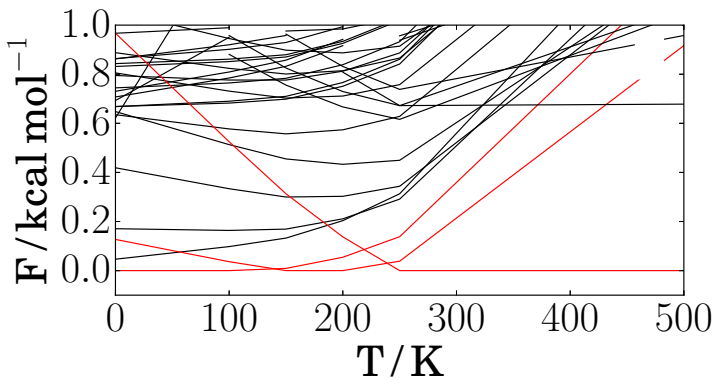


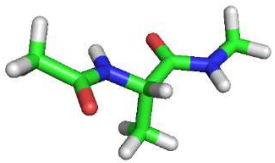


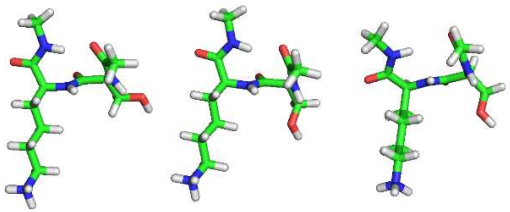
Free Energy Basin-Hopping (*Chem. Phys. Lett.*, **625**, 1, 2015)

Consider $F_{\text{new}}(T) - F_{\text{old}}(T) = V_{\text{new}} - V_{\text{old}} + k_B T \ln\left(o_{\text{new}}\bar{\nu}^{\kappa}_{\text{new}}/o_{\text{old}}\bar{\nu}^{\kappa}_{\text{old}}\right)$.









For the serine-lysine dipeptide (right) the free energy minimum changes twice with temperature, as determined by the vibrational entropy. We can exploit sparsity to calculate the vibrational partition function. (*CPL*, **685**, 288, 2017)

Using optimal Tsallis weights rather than Boltzmann can halve the mean first encounter time for the global minimum (*JCP*, **141**, 071104, 2014).

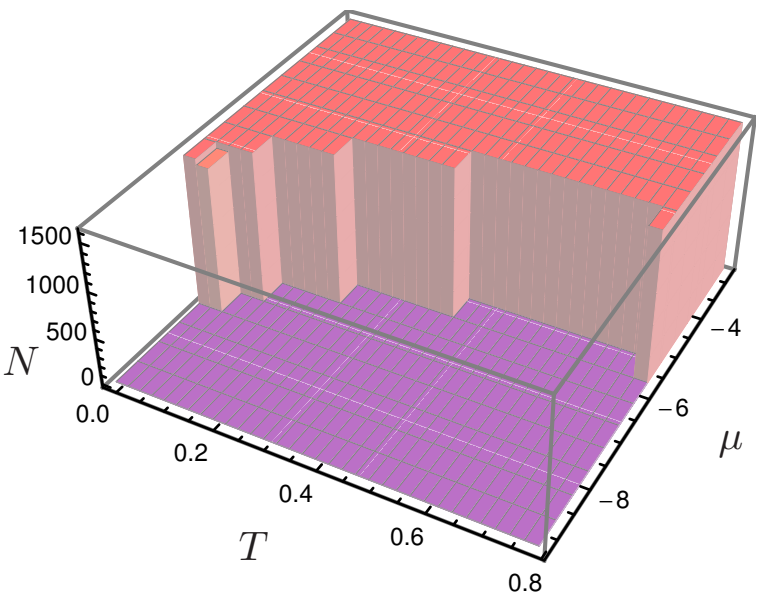
Grand and Semi-Grand Canonical Basin-Hopping (*JCTC*, **12**, 902, 2016)

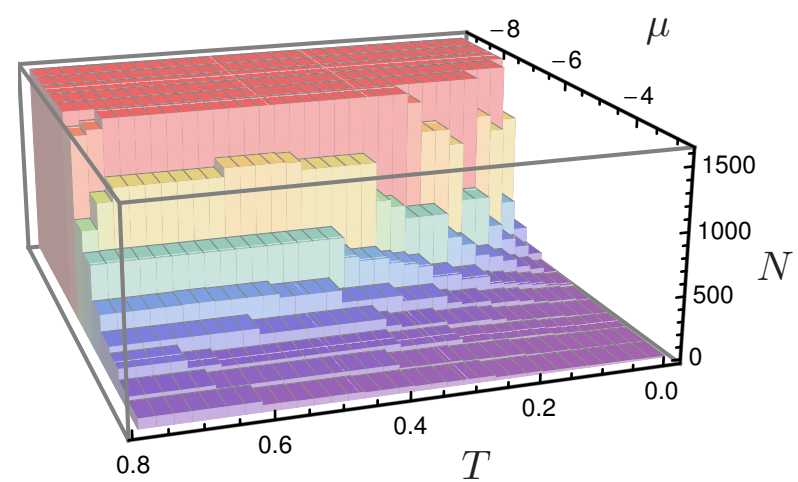
The accept/reject condition for GCBH employs the local grand potential:

$$\xi_\alpha = V_\alpha - \mu N - k_B T \ln \frac{n_\alpha \sqrt{8\pi}\, |\mathbf{I}_\alpha|^{1/2} (k_B T)^{3/2}}{\hbar^3 (\beta h \overline{\nu}_\alpha)^\kappa},$$

which includes the rigid rotor partition function, with inertia tensor $\mathbf{I}_\alpha$.

Blocks of conventional BH steps are employed between changes of N.



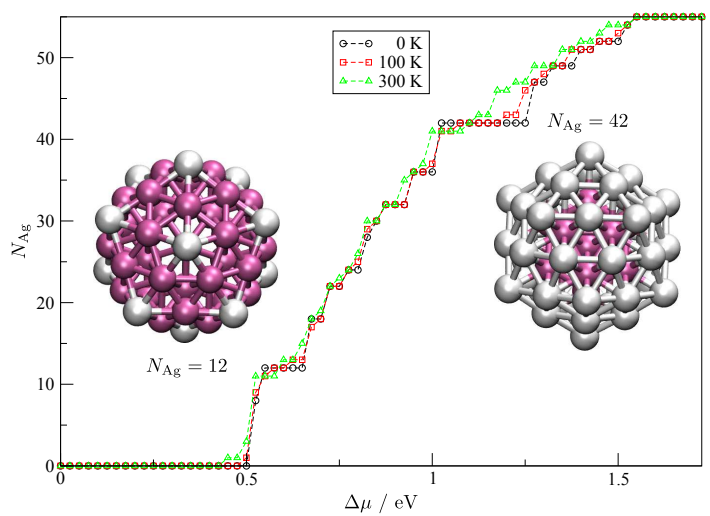


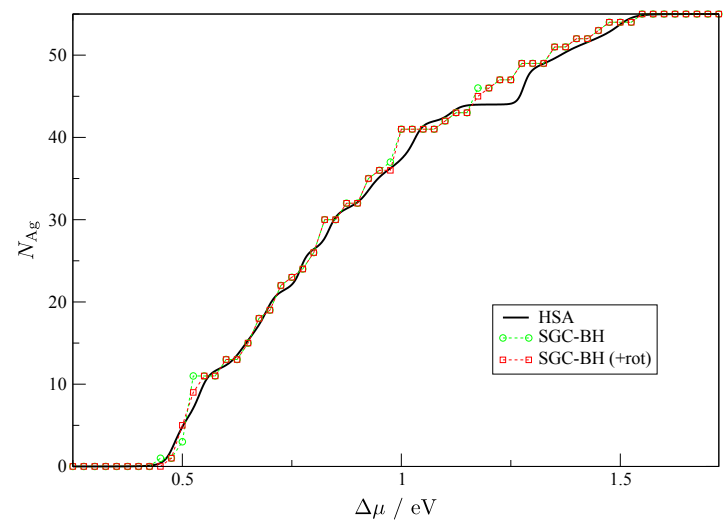
The most probable cluster size (left) corresponds to the smallest or largest size permitted. However, the least probable size can occur between the limits (right), suggesting a free energy barrier to nucleation and cluster growth.

For a binary semigrand canonical potential we consider fixed $N = N_A + N_B$ and variable N_A and N_B with potential

$$\xi_\alpha = V_\alpha - \Delta\mu N_B - k_B T \ln \frac{(k_B T)^{3/2} n_\alpha \sqrt{8\pi}\, |\mathbf{I}_\alpha|^{1/2}}{\hbar^3 (\beta h \overline{\nu}_\alpha)^\kappa}.$$

$\Delta\mu = \mu_B - \mu_A$ is the chemical potential difference; $n_\alpha = 2N_A!N_B!/o_\alpha$.





Left: the most favourable composition for icosahedral $\mathrm{Ag}_n\mathrm{Pd}_{55-n}$ as a function of $\Delta\mu$ exhibits steps at $N_\mathrm{Ag} = 12$ and $N_\mathrm{Ag} = 42$.

Right: lumping probabilities for minima with the same composition and including non-icosahedral structures smooths the steps.

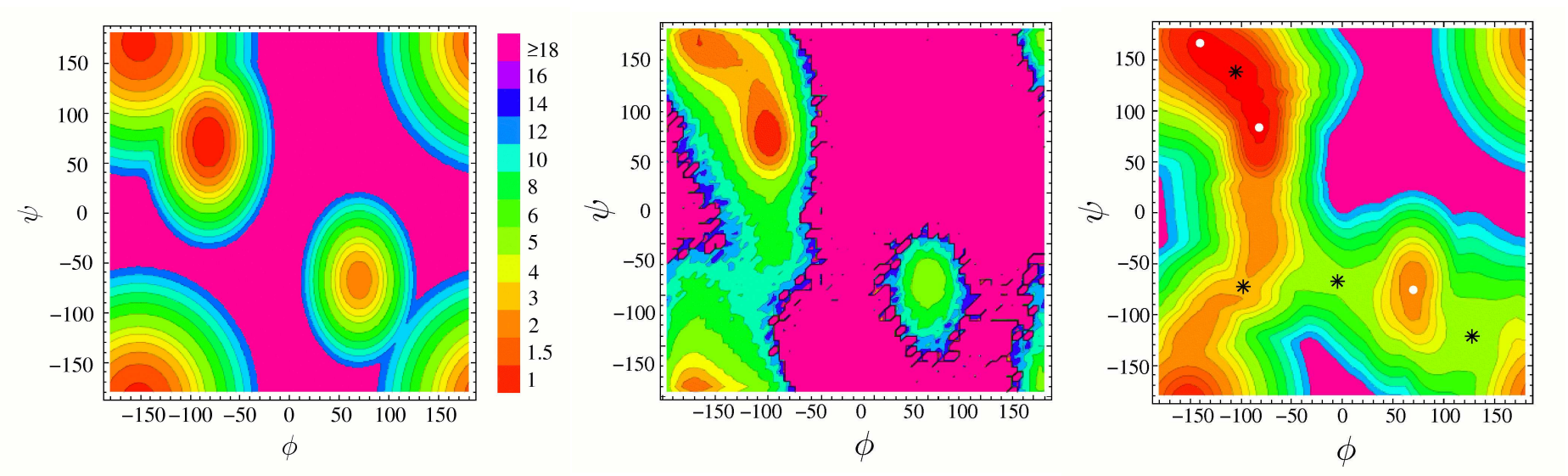
A Modified Superposition Approach (*Chem. Phys. Lett.*, **466**, 105, 2008)

Here the partition function is broken down into contributions from local minima and pathways as a function of an order parameter, a, with terms

$$Z_i(a,T) = \left(\frac{kT}{h\overline{\nu}_i}\right)^{3N-6} \frac{\exp\left(-V_i/kT\right)}{\sqrt{2\pi kTA_i}} \exp\left[-\frac{(a-a_i)^2}{2kTA_i}\right],$$

where A_i is a weighted sum of order parameter derivatives.

Free energy surfaces for alanine dipeptide (CHARMM22/vacuum) from superposition, replica exchange, and reaction path Hamiltonian superposition:



Basin-Sampling for Global Thermodynamics (*CPL*, **584**, 1, 2013)

Broken ergodicity is treated using basin-hopping, while the configuration space corresponding to high temperature is sampled by parallel tempering.

We define a two-dimensional density of states using systematic local minimisation, and couple these statistics to a model anharmonic form connected to the low-temperature limit for structures obtained via global optimisation.

The number of visits to quench potential energy bin q from instantaneous potential energy bin i in replica r is $\mathcal{N}_{iqr}$: a two-dimensional histogram.

The corresponding canonical probability distribution for replica r is

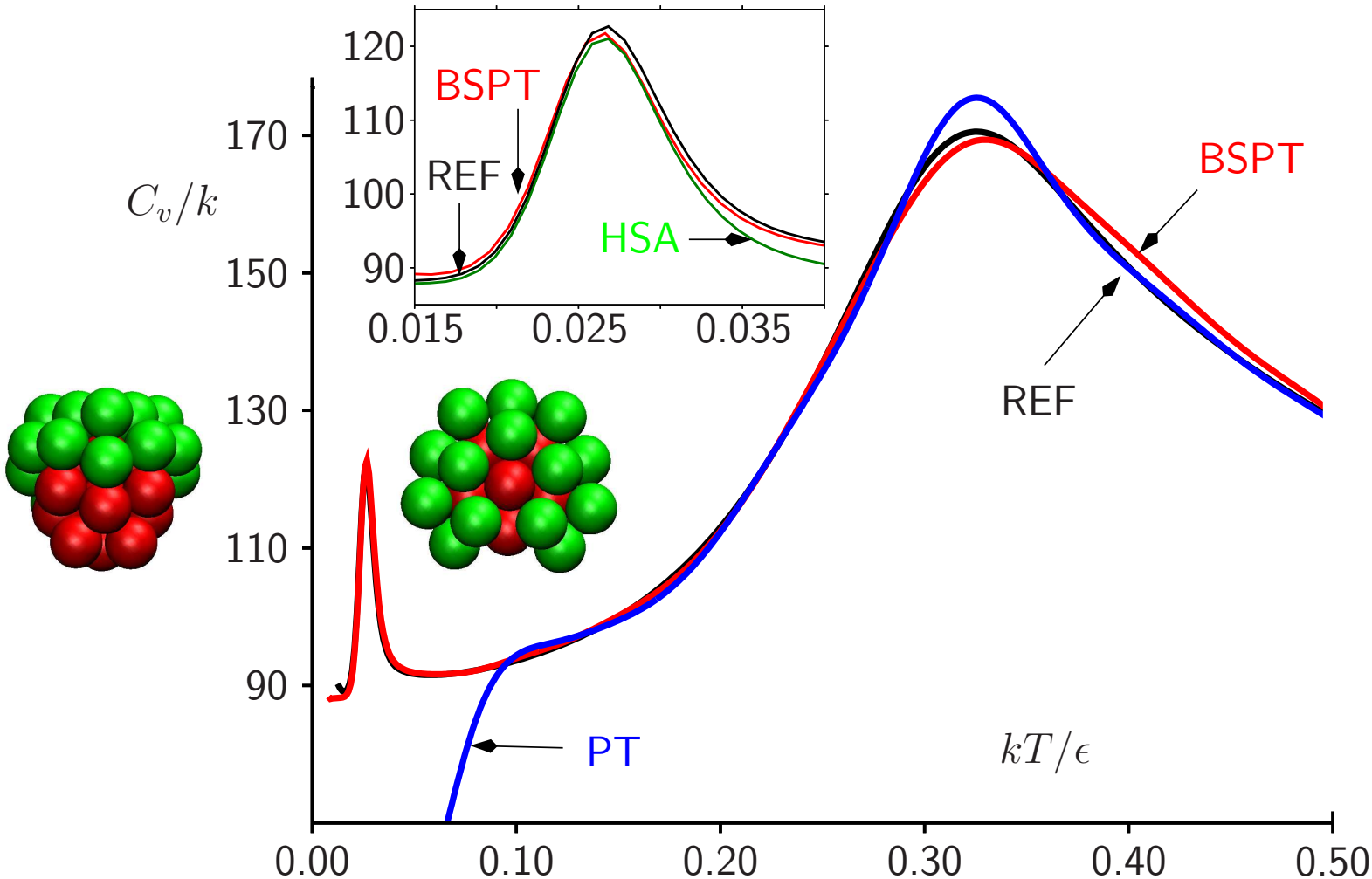
$$P(V_i^I, V_q^Q, T_r) = \mathcal{N}_{iqr}/\mathcal{N}_r \propto \Omega_c(V_i^I, V_q^Q) e^{-V_i^I/kT_r}.$$

The analytical density of states for a Morse potential suggests a model anharmonic representation for each quench bin with two fitting parameters:

$$\ln \Omega_c(V_i^I, V_q^Q) = \left[(3N-6)/2 - 1 + e^{A_q}\left(V_i^I - V_q^Q\right)\right] \ln\left(V_i^I - V_q^Q\right) + B_q.$$

Accurate thermodynamics were obtained for the solid-solid phase transition in LJ_{31} with 10^6 equilibration steps (discarded), 5×10^6 parallel tempering steps, and 5×10^6 BSPT steps using 24 replicas and quenching every 30 steps.

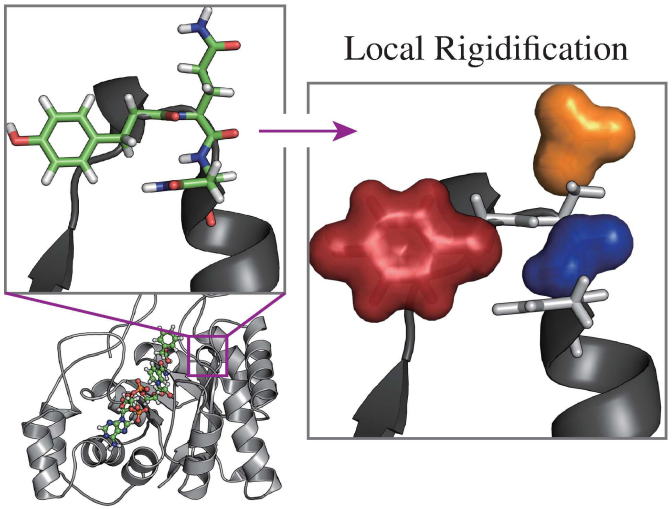
This run took 21.8 minutes compared to 110.5 hours for parallel tempering.

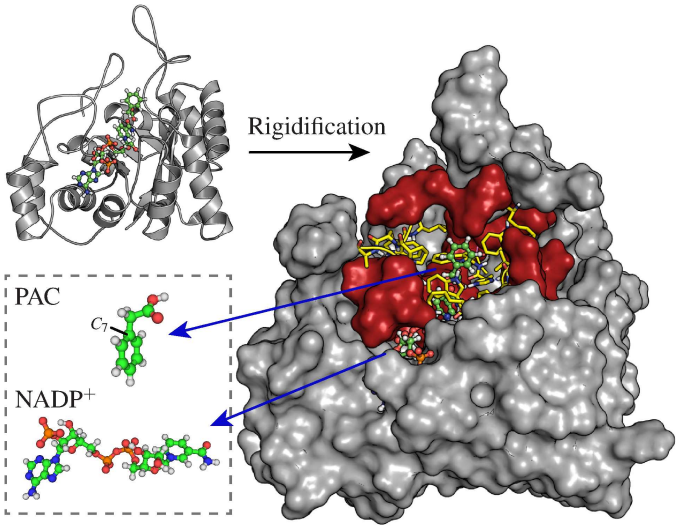


Factorised Superposition for Ligand Binding (*PCCP*, **16**, 2842, 2014)

The factorisation separates weakly coupled degrees of freedom for consistent reference conformations, based on local rigidification (*JCTC*, **8**, 5159, 2012).

For phenyl acetic acid/aldose reductase the binding free energy converges for factorisation of groups beyond 12 Å to 14 Å from the binding site.





Left: local rigidification of the TYR47 ring, the peptide bond between GLN48 and ASN49, and the amide group of GLN48 (sp^2 centre), residues 47-49.

Right: the unconstrained inner region of the complex, the locally rigidified intermediate region, and the rigidified outer region (80% of the protein).

Discrete Path Sampling (*Mol. Phys.*, **100**, 3285, 2002; **102**, 891, 2004)

Discrete path sampling constructs databases of stationary points that are relevant to global kinetics (*Int. Rev. Phys. Chem.*, **25**, 237, 2006).

No reaction coordinate is required. No projection error.

Phenomenological $A \leftrightarrow B$ rate constants can be formulated as sums over discrete paths, defined as sequences of local minima and the transition states that link them, weighted by equilibrium occupation probabilities, p_b^{eq}:

$$k_{AB}^{\mathrm{SS}} = \frac{1}{p_B^{\mathrm{eq}}} \sum_{a \leftarrow b} P_{ai_1} P_{i_1 i_2} \cdots P_{i_{n-1} i_n} P_{i_n b} \tau_b^{-1} p_b^{\mathrm{eq}} = \frac{1}{p_B^{\mathrm{eq}}} \sum_{b \in B} \frac{C_b^A p_b^{\mathrm{eq}}}{\tau_b},$$

where $P_{\alpha\beta}$ is a branching probability and C_b^A is the committor probability that the system will visit an A minimum before it returns to the B region.

The paths that make the largest contributions to k_{AB}^{SS} can be extracted using the Dijkstra or recursive enumeration algorithms, using edge weights $-\ln P_{\alpha\beta}$ (*J. Chem. Phys.*, **121**, 1080, 2004; *J. Phys. Chem. B*, **112**, 8760, 2008).

Rates from Graph Transformation (*JCP*, **124**, 234110, 2006; **130**, 204111, 2009)

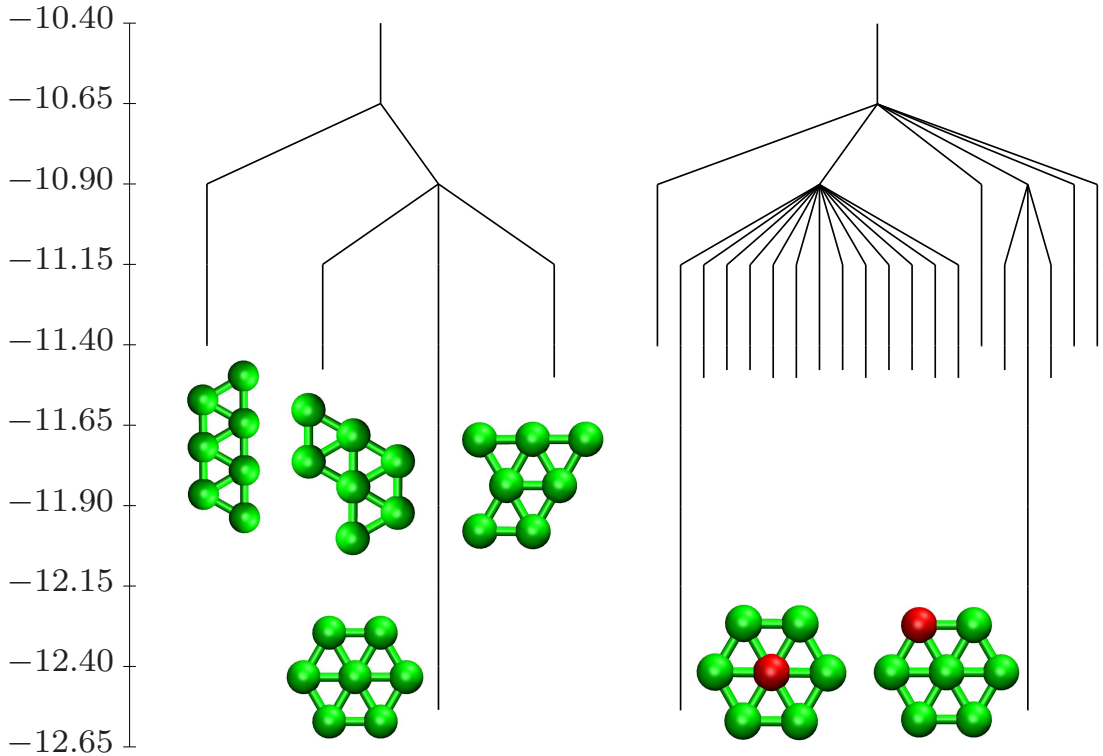
The deterministic graph transformation procedure is non-stochastic and non-iterative. Minima, x, are progressively removed, while the branching probabilities and waiting times in adjacent minima, β, are renormalised:

$$P'_{\gamma\beta} = P_{\gamma\beta} + P_{\gamma x} P_{x\beta} \sum_{m=0}^{\infty} P_{xx}^{m} = P_{\gamma\beta} + \frac{P_{\gamma x} P_{x\beta}}{1 - P_{xx}}, \qquad \tau'_{\beta} = \tau_{\beta} + \frac{P_{x\beta}\tau_{x}}{1 - P_{xx}}.$$

Each transformation conserves the MFPT from every reactant state to the set of product states with an execution time independent of temperature:

kT/K	$\Delta F_{\text{barrier}}$	N_{min}	N_{ts}	NGT/s	SOR/s	KMC/s
298	5.0	272	287	8	13	85,138
298	4.5	2,344	2,462	8	217,830	
1007	-	40,000	58,410	35	281	1,020,540
1690	-	40,000	58,410	39	122,242	

Permutational Isomerisation of LJ_7^{2D} (*Mol. Phys.*, **100**, 3285, 2002)



Disconnectivity graphs for LJ_7^{2D}. Left: permutation-inversion isomers of the four local minima are collected together. Right: one of the atoms is tagged, lowering the permutational degeneracy.

The fastest ten paths contribute about 74% of the total rate constant at $kT/\epsilon = 0.05$. Various combinations of diamond-square-diamond rearrangements make significant contributions.

Benchmarks for Landscape Exploration

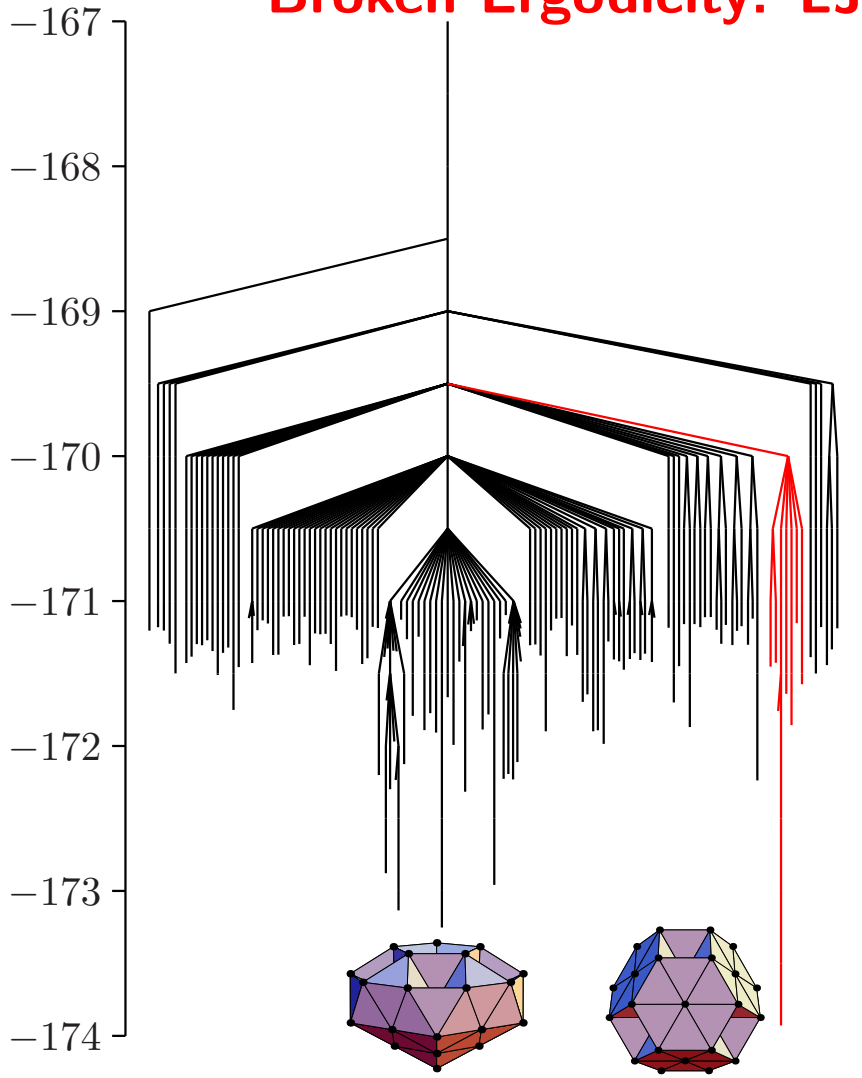
Minimisation: Nocedal's algorithm, LBFGS, with line searches removed.

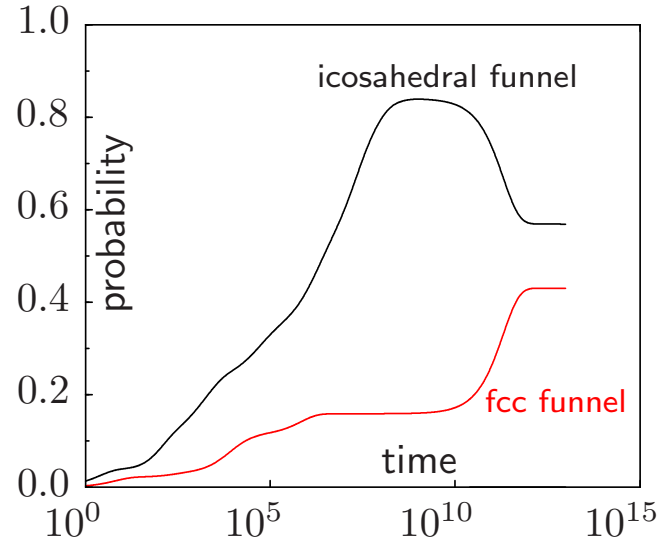
Transition states: single-ended searches use hybrid eigenvector-following (*PRB*, **59**, 3969, 1999; *JCP*, **111**, 7010, 1999; *CPL*, **341**, 185, 2001), double-ended searches use the doubly-nudged elastic band approach (*JCP*, **120**, 2082, 2004; **140**, 044115, 2014).

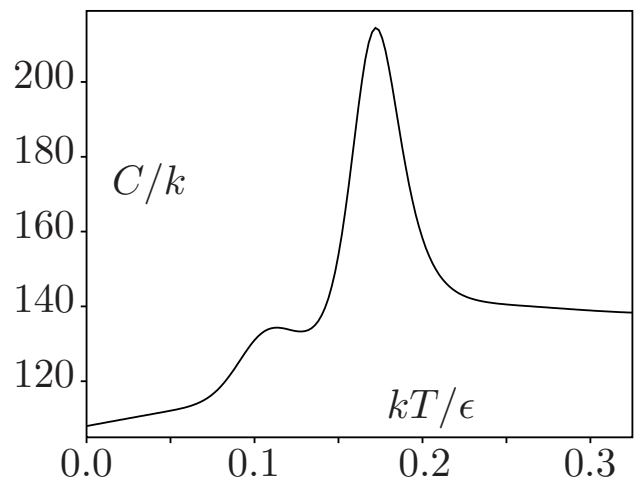
The `GMIN` (global optimisation), `OPTIM` (transition states and pathways) and `PATHSAMPLE` (discrete path sampling) programs are available from the Cambridge Landscape Database under the Gnu General Public License.

- Interfaces to many electronic structure codes are included.
- Current svn tarball image: http://www-wales.ch.cam.ac.uk
- http://www-wales.ch.cam.ac.uk/tsbenchmarks.html Peptide examples
- http://theory.cm.utexas.edu/benchmarks/index.html OptBench test suite
- https://github.com/wales-group/examples Curated examples

Broken Ergodicity: LJ_{38} (*Phys. Rev. E*, **60**, 3701, 1999)







LJ_{38} exhibits a double funnel due to competition between icosahedral and truncated octahedral morphologies. The interconversion rate for Ar_{38} is calculated as $55\,s^{-1}$ at 14 K where a solid-solid transition occurs.

THE JOURNAL OF CHEMICAL PHYSICS **135**, 034108 (2011)

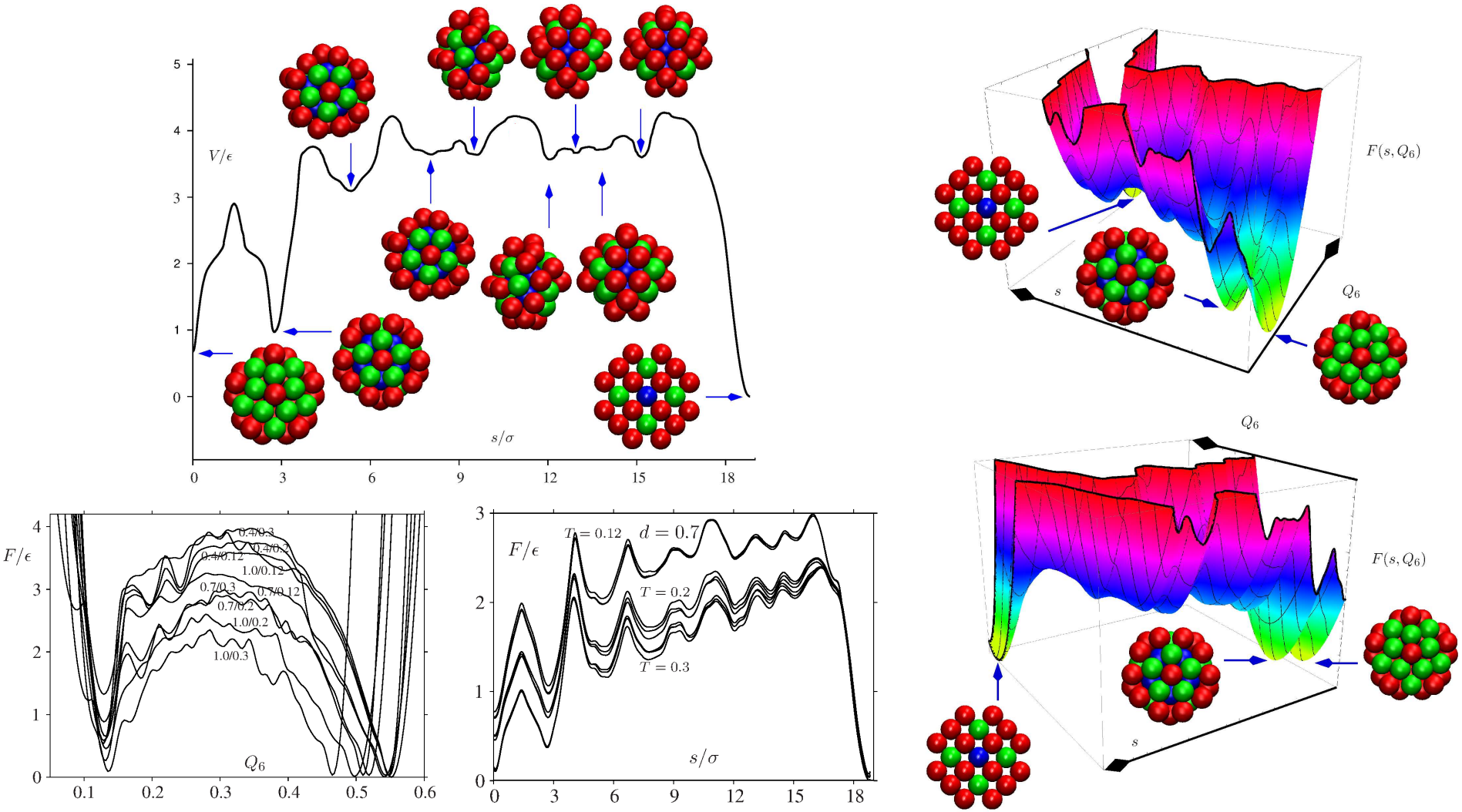
Simulating structural transitions by direct transition current sampling: The example of LJ_{38}

Massimiliano Picciani,[1,a)] Manuel Athènes,[1] Jorge Kurchan,[2] and Julien Tailleur[3]
[1]*CEA, DEN, Service de Recherches de Métallurgie Physique, F-91191 Gif-sur-Yvette, France*
[2]*CNRS; ESPCI, 10 rue Vauquelin, UMR 7636 PMMH, 75005 Paris, France*
[3]*School of Physics of Astronomy, SUPA, University of Edinburgh, The King's Buildings, Mayfield Road, EH9 3JZ Edinburgh, United Kingdom*

(Received 2 March 2011; accepted 21 June 2011; published online 20 July 2011)

Another attempt to study the transitions between the two funnels of LJ_{38} relies on the use of transition path sampling.[33] Because of the number of metastable states separating the two main basins, the traditional shooting and shifting algorithm failed here, despite previous success for smaller LJ clusters.[39] The authors thus developed a two-ended approach which manages to successfully locate reaction paths between the two basins: they started from a straight trial trajectory linking the two minima, and obtained convergence towards trajectories of energies similar to those obtained in the discrete path sampling approach.[33] Although the authors point out the lack of ergodicity in the sampling within their approach and the sensitivity on the "discretization" of the trajectories, this is nevertheless a progress and the main drawback remains the high computational cost (the work needed 10^5 h of central processing unit (cpu) time) to obtain such converged trajectories. In contrast, the simulations we present below required less than 10^2 h of cpu time.

Free Energy Profiles for an LJ_{38} Pathway (*JCP*, **142**, 130901, 2015.)

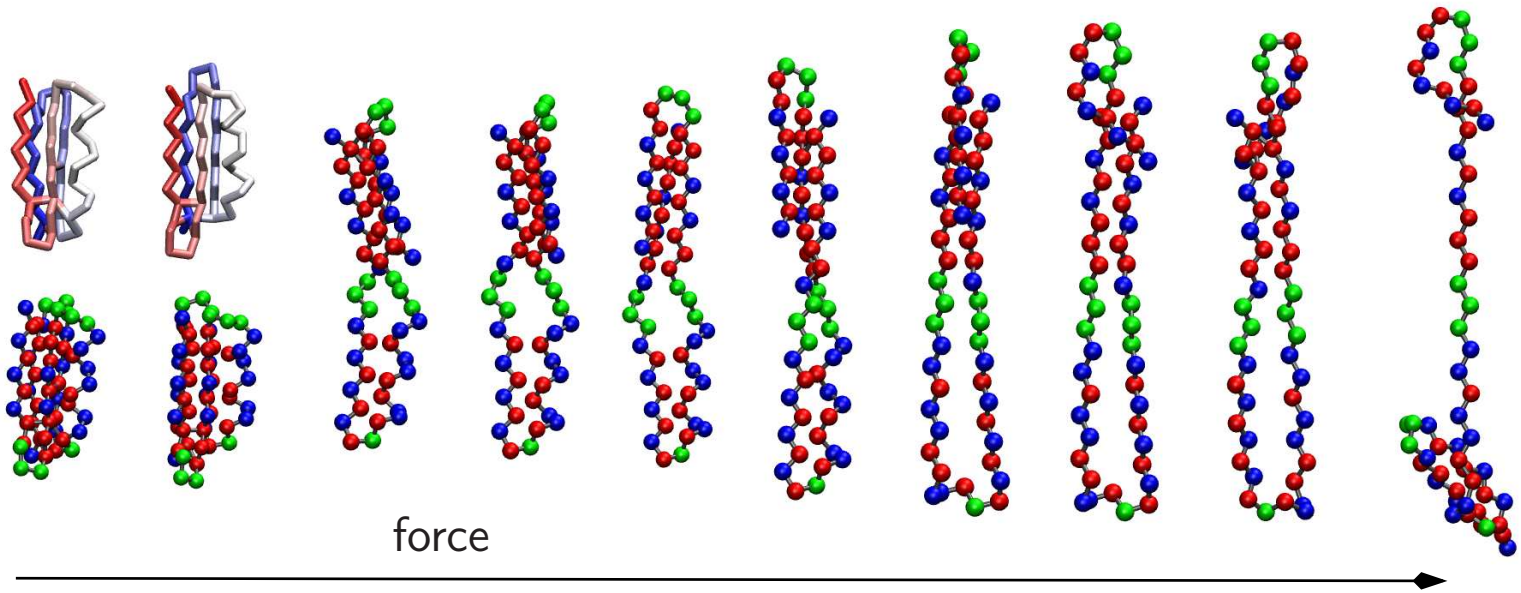


Projection onto the bond order parameter Q_6 averages over surface reorganisation and other mechanistic details. However, the profile based on pathways defined geometrically faithfully reflects the underlying barriers.

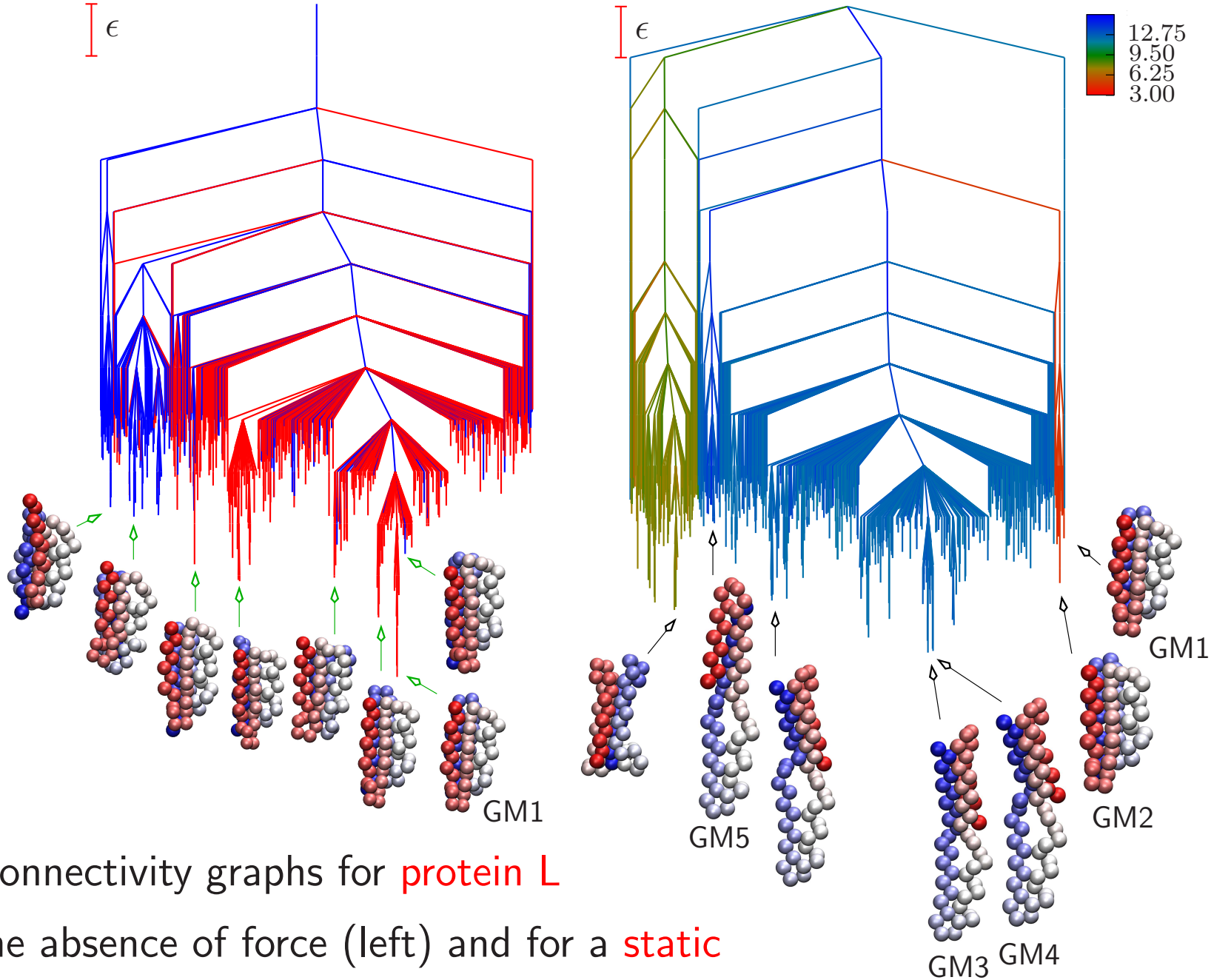
Folding and Pulling for Protein L and Protein G

Folding pathways and the evolution of the energy landscape as a function of static force have been analysed for protein L and protein G using a sequence-dependent BLN model. B=hydrophobic, L=hydrophilic, and N=neutral.

Protein L forms the N-terminal hairpin 1 first, followed by the C-terminal hairpin 2. The order is reversed for protein G, with an early intermediate.



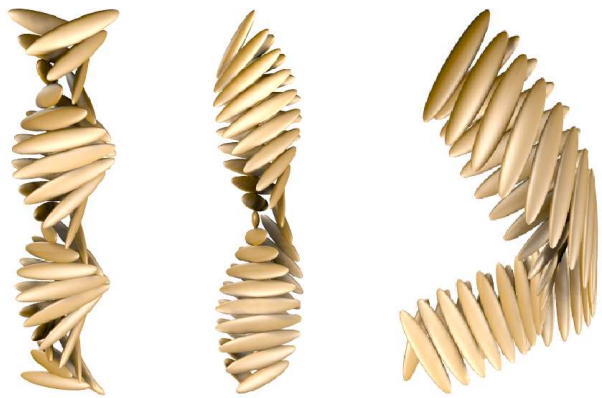
Distinct global minima for protein L as a function of increasing static pulling force applied between beads 10 and 32. (*J. Phys. Chem. B*, **136**, 8394, 2012)



Disconnectivity graphs for protein L in the absence of force (left) and for a static pulling force (right) applied between beads 10 and 32.

Helical Bilayers From Frustrated Building Blocks (*JPCB*, **117**, 7918, 2013)

helical fibre morphologies



bilayer filaments

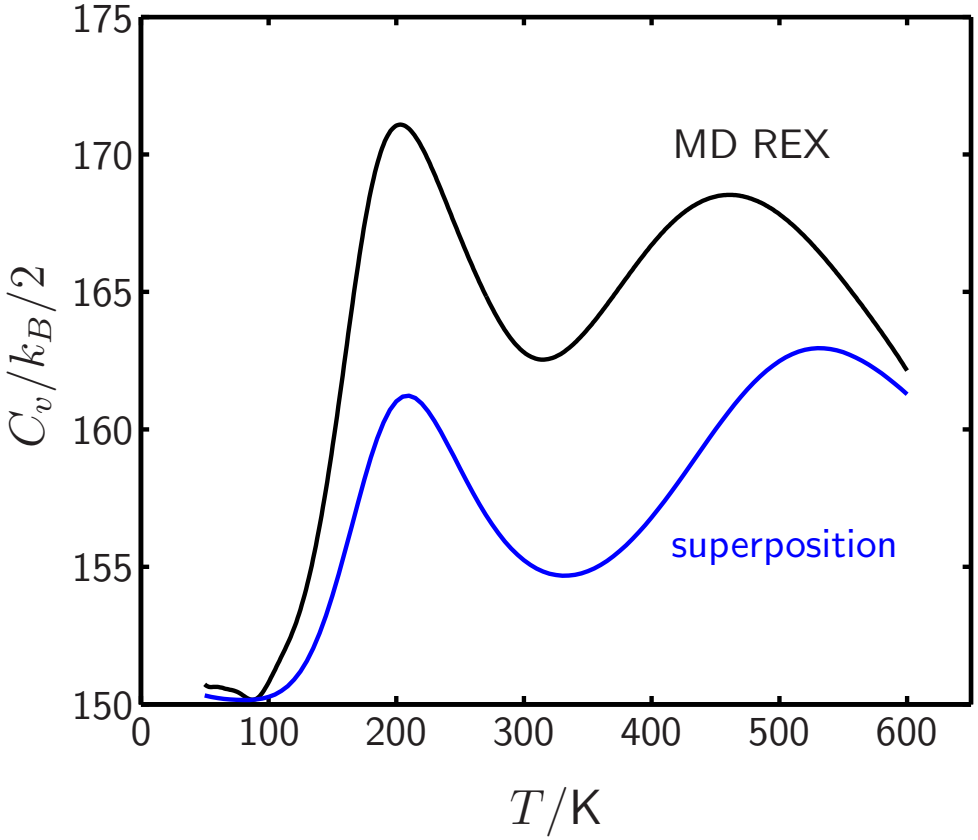
magnetic bilayers

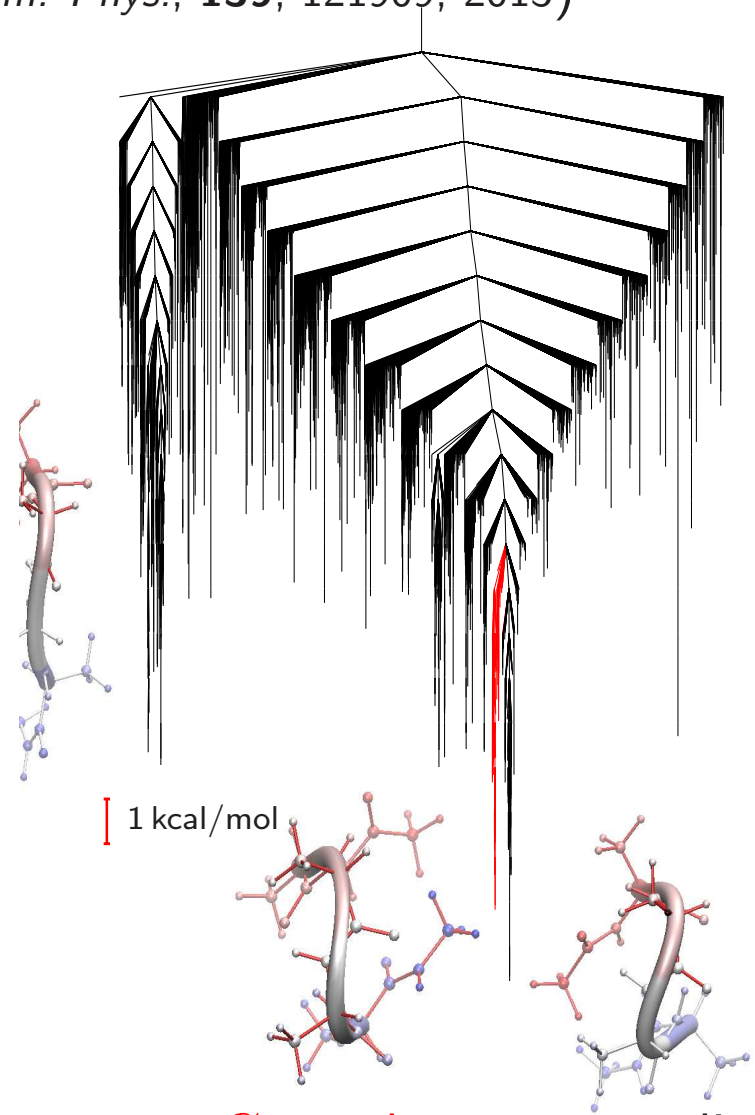
Left: introduction of a cytochrome domain into an amyloid fibre can change the morphology from twisted to spiral ribbons and induce systematic kinking.

Centre: rigid building blocks consisting of two ellipsoids can reproduce these structures, which are also observed for Bauhinia seedpods.

Right: the structure depends mostly on the internal geometry of the building blocks, rather than details of the potential. The design principles extend to macroscopic helices formed from elliptical magnets.

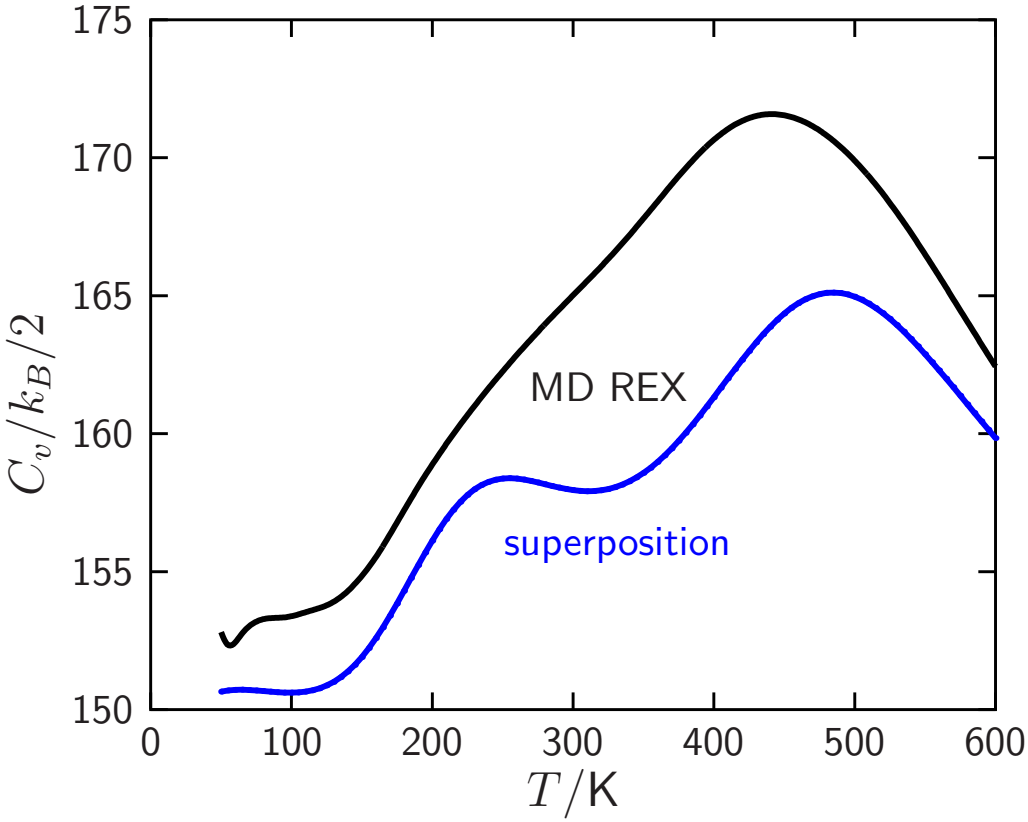
Thermodynamics for Ala_4 (*J. Chem. Phys.*, **139**, 121909, 2013)

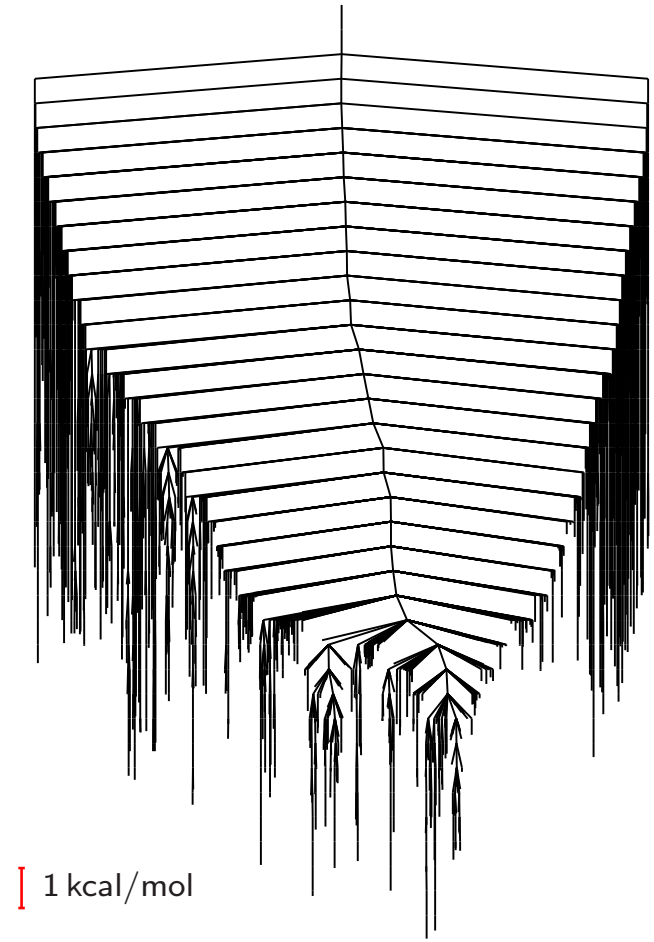




Ala_4 in vacuum (charmm27) has a low temperature C_v peak, corresponding to the hundred or so lowest minima in the disconnectivity graph. The high temperature peak corresponds to the finite system analogue of melting.

Thermodynamics for Ala_4 in Vacuum: AMBER

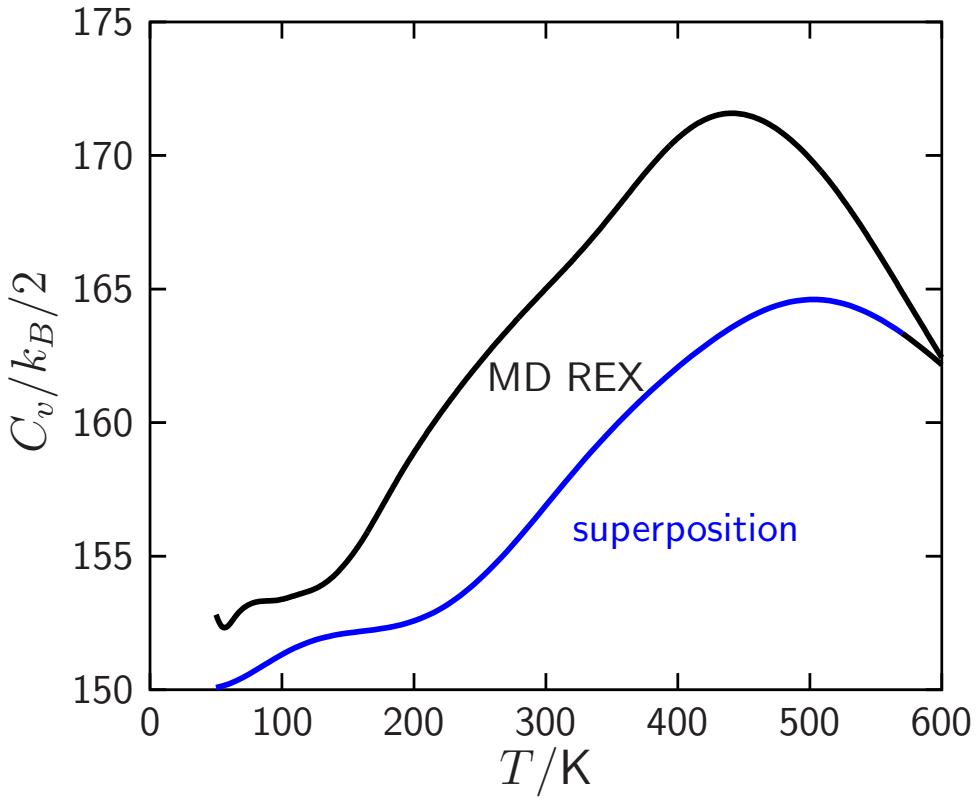


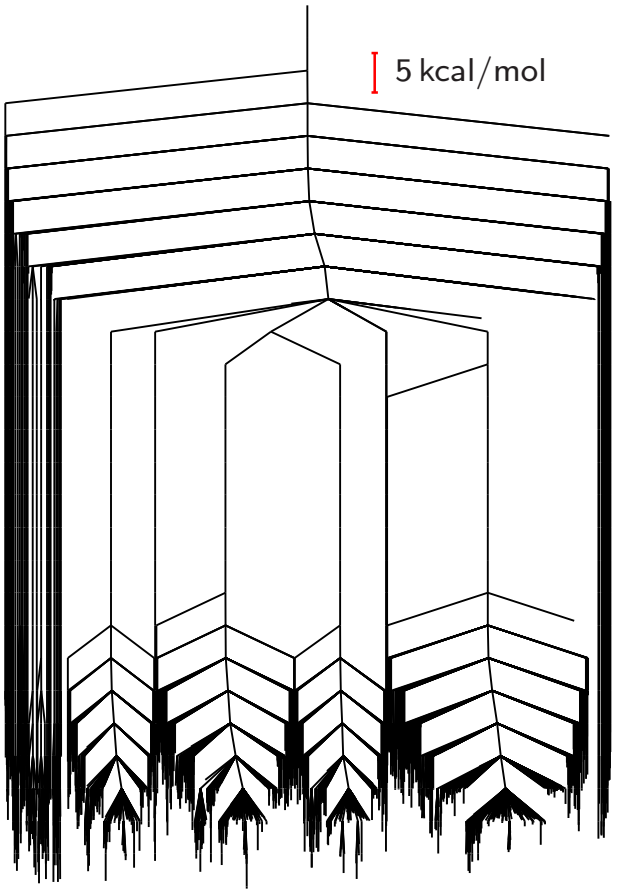


Replica exchange (REX) and superposition results for the heat capacity C_v of ala_4 in vacuum (amber99sb) appear similar to CHARMM.

(*J. Chem. Phys.*, **139**, 121909, 2013)

Thermodynamics for Ala_4 in Vacuum: AMBER





The global minimum for this potential has a mixture of L and D amino acids. The landscape separates into regions with different L/D composition, spearated by barriers of order 90 kcal/mol. (*J. Chem. Phys.*, **139**, 121909, 2013)

Symmetry and Energetics (*Chem. Phys. Lett.*, **285**, 330, 1998; **294**, 262, 2008)

'The symmetry characteristic of a phenomenon is the maximal symmetry compatible with the existence of the phenomenon' (Pierre Curie)

'The perfection of mathematical beauty is such... that whatsoever is most beautiful and regular is also found to be most useful' (D'Arcy Thompson)

However, many systems do not exhibit their highest possible symmetry.

We can write the total energy as a sum over contributions from a many-body expansion, involving single atom, pairwise and three-body terms, etc.

If these terms are drawn from the same distribution then geometrical symmetry (degeneracies) would be manifested as correlation.

The variance is larger when correlation is present. Symmetrical structures are therefore more likely to have particularly high or particularly low energy. Low-lying structures are therefore likely to exhibit symmetry.

More formally, denote the mean and variance of a variable, X, drawn from probability distribution, $p(X)$, as μ and σ^2.

The variance of a sum of N such variables, X_i, is then

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

where the correlation ρ is defined by

$$\rho\sigma^2 = \int (X-\mu)(X'-\mu)p(X,X')dXdX'.$$

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

Open question: how should we treat approximate geometrical symmetry, for example, in large biomolecules or condensed matter?

Open question: is it possible to build a more rigorous theory, and make connections with concepts such as designability?

Connecting Dynamics and Thermodynamics (*Science*, **293**, 2067, 2001)

The organisation of a PES is governed by its stationary points, where Taylor expansions provide local descriptions in terms of Hessian matrices.

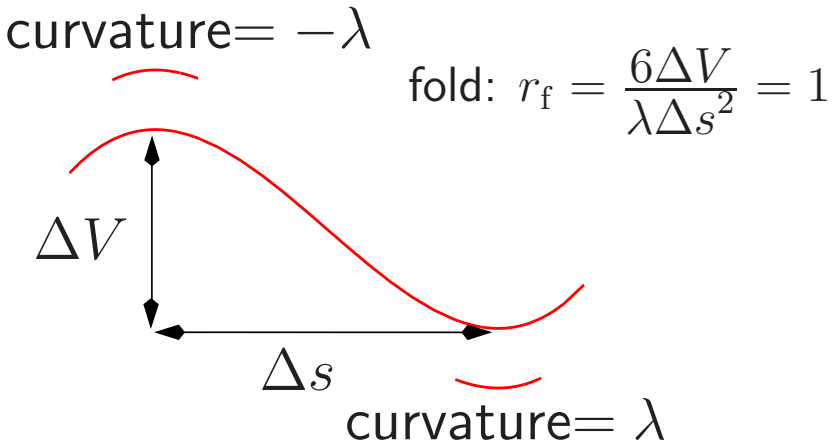
The organisation of families of PES's as a function of parameters in the potential is determined by the stationary points that possess additional zero Hessian eigenvalues, known as non-Morse points.

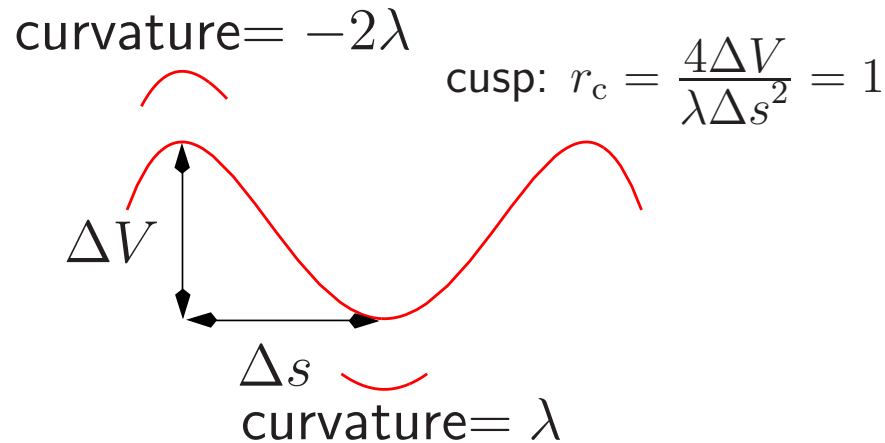
Catastrophe theory provides a local representation of the PES around non-Morse points as a function of both atomic coordinates and parameters.

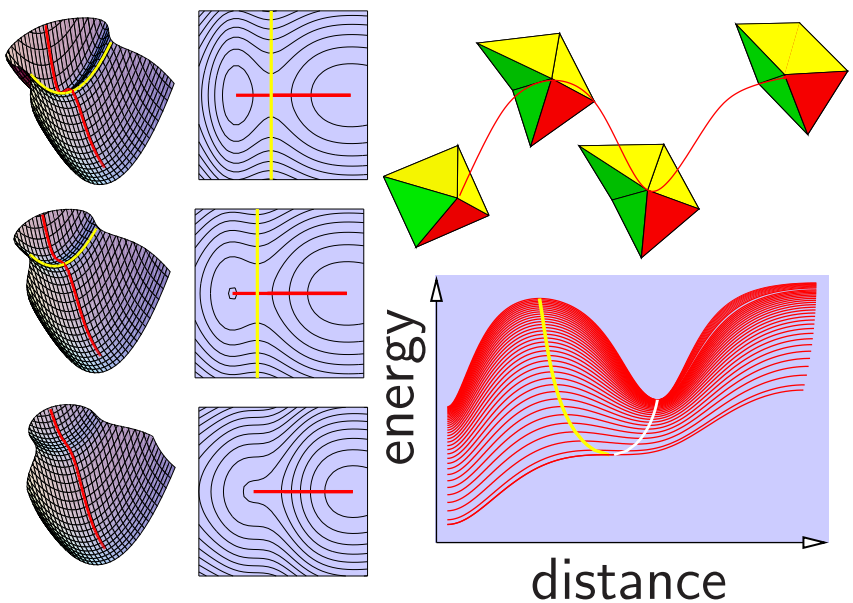
The splitting lemma reduces the dimensionality to the essential variables, while transversality guarantees that the resulting classifications are universal.

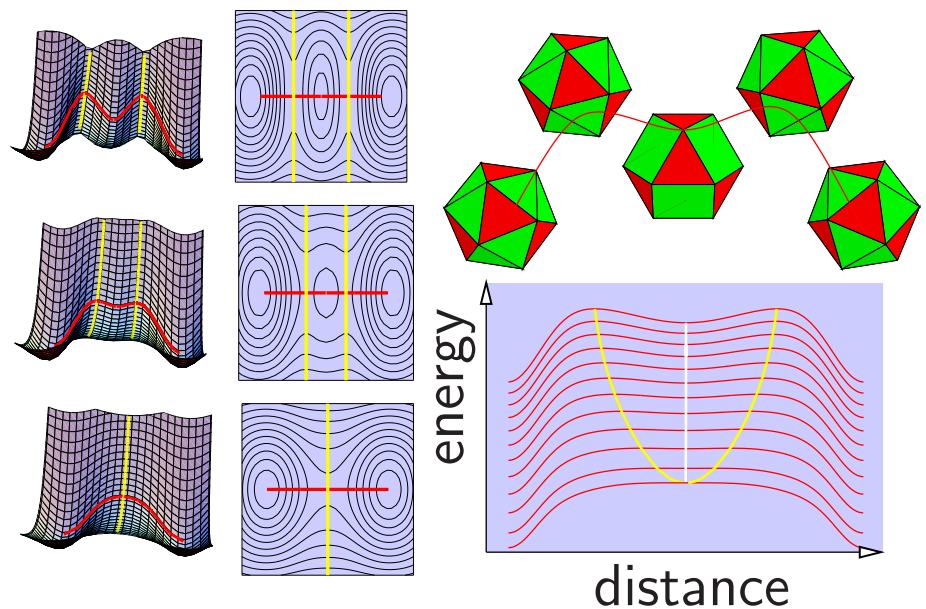
The simplest one-parameter catastrophes are the fold, $f(x) = \frac{1}{3}x^3 + ax$, and the symmetrical cusp, $f(x) = \frac{1}{4}x^4 + \frac{1}{2}ax^2$.

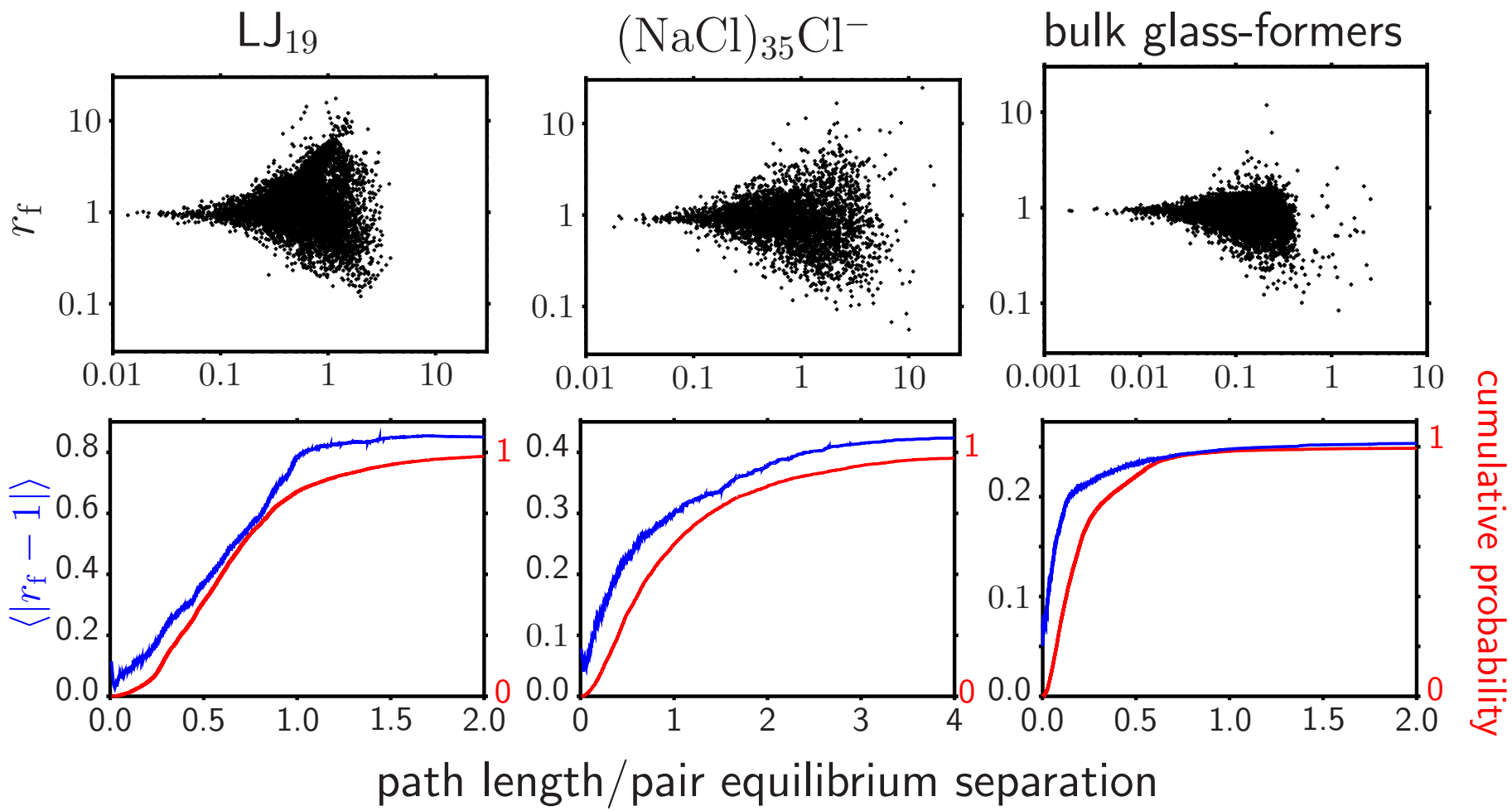
Geometries of the fold and cusp catastrophes.





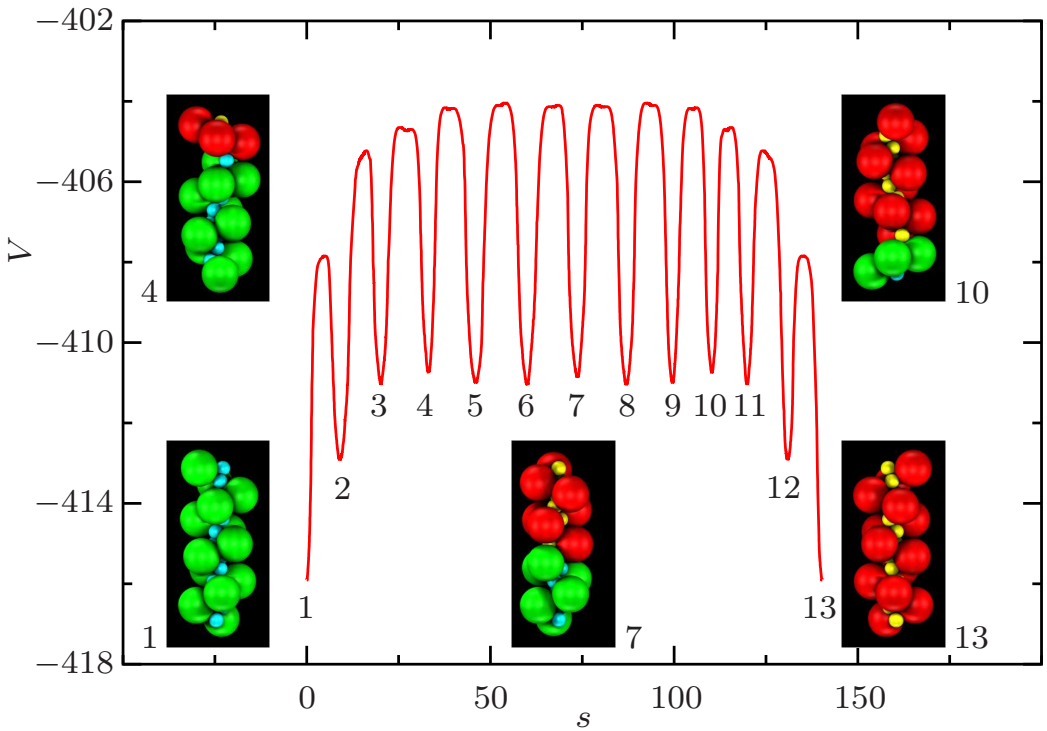






For systems with a fixed potential we effectively have a snap-shot of parameter space. On average, r_{f} remains close to unity for many pathways in both model clusters and bulk, providing an explanation for Hammond's postulate.

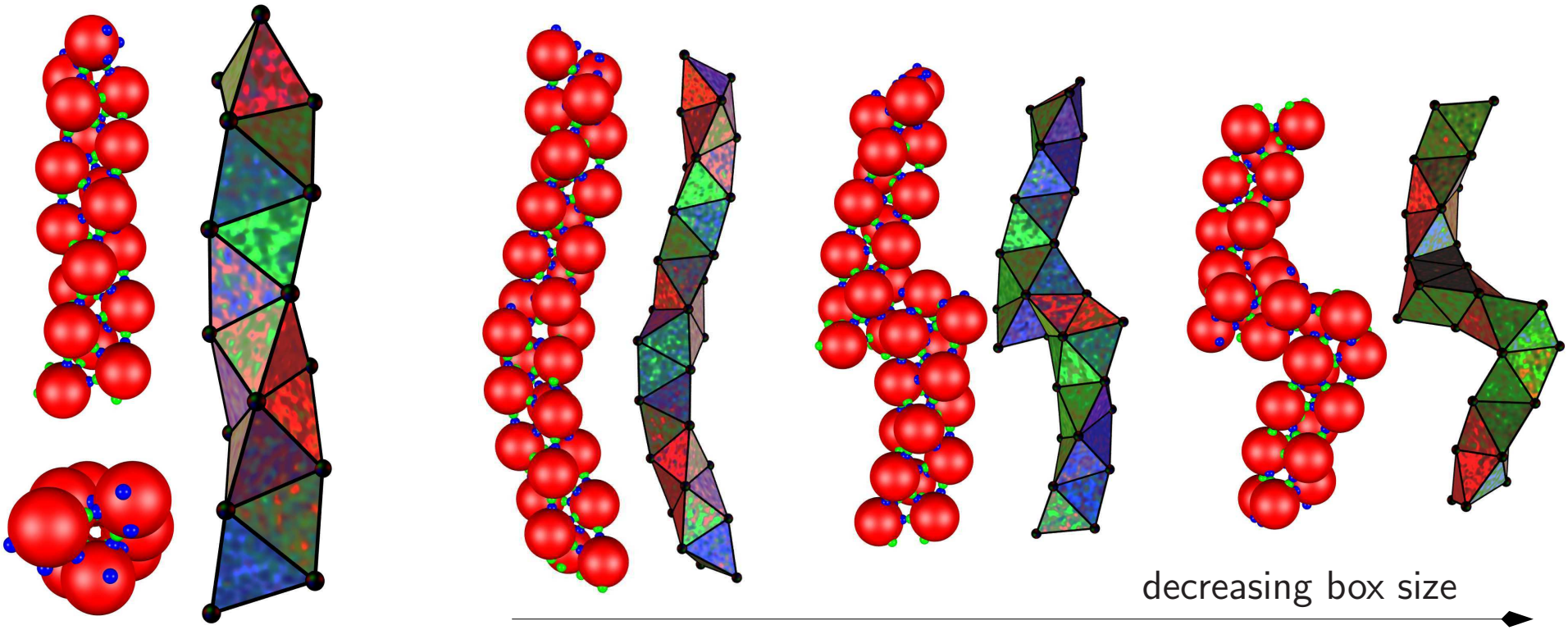
A Nanodevice (*Soft Matter*, **7**, 2325, 2011)



Coupled linear and rotary motion has been characterised for a helix composed of 13 asymmetric dipolar dumbbells in the presence of an electric field.

The helix changes handedness as the boundary between segments propagates along the strand via successive steps that switch the dumbbells.

Designing a Bernal Spiral (*ACS Nano*, **7**, 1246, 2013)

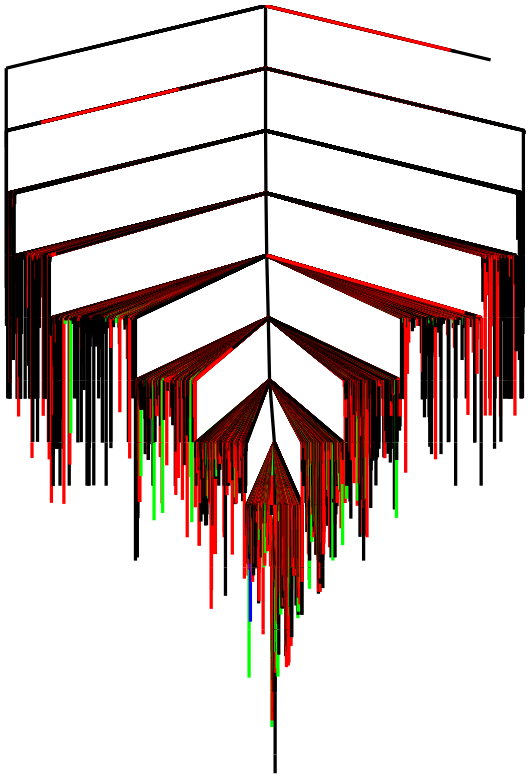


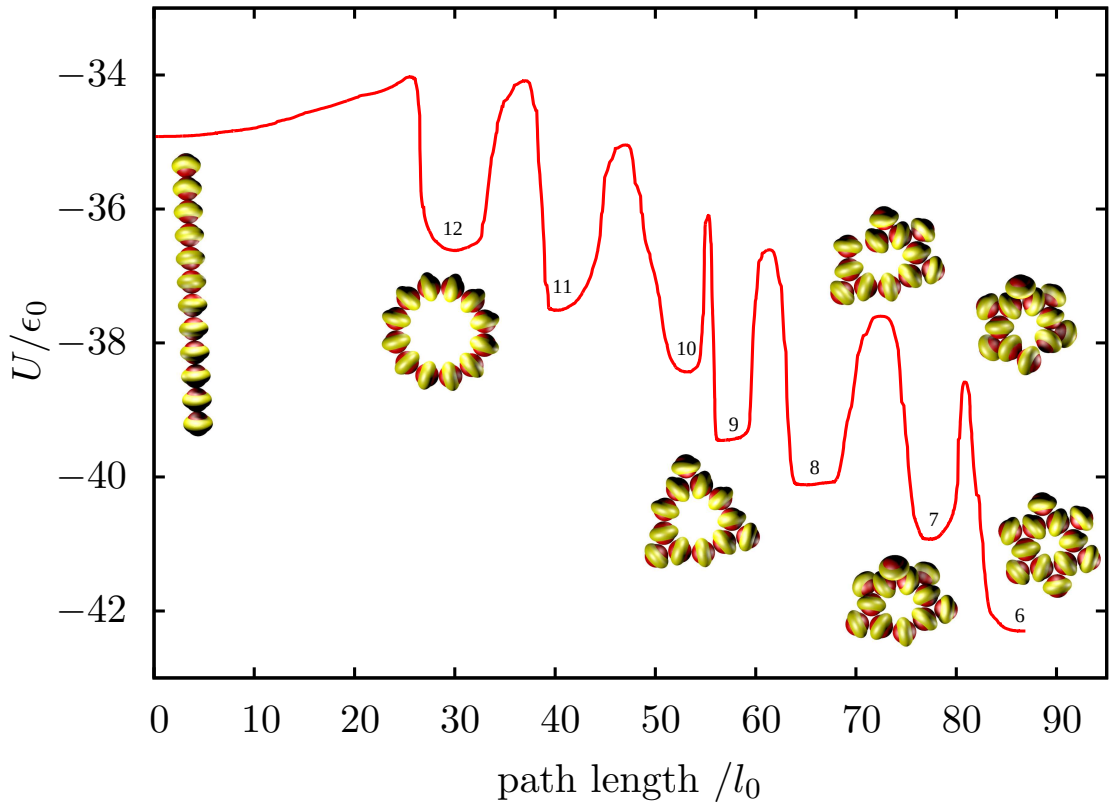
The simplest building blocks that support a Bernal spiral as the global minimum involve a single patch-antipatch pair offset by about 10° from linearity.

Left: Alternative views of a chiral Bernal spiral consisting of 18 particles.

Right: compressed spirals (30 particles, periodic boundaries) exhibit supercoiling or breaks, which resemble structures seen in confocal microscopy.

Kagome Structures (*Soft Matter*, **11**, 6663, 2015)





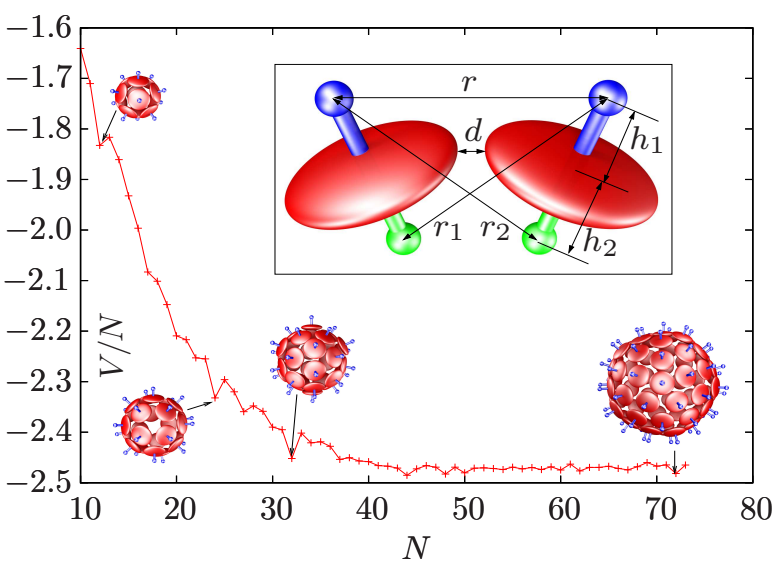
Energetically stabilised Kagome structures were designed using soft anisotropic triblock Janus particles.  This unconstrained model predicts that sedimentation effects enhance the stability.

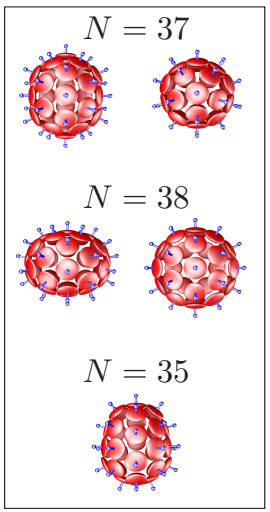
Rearrangements between competing structures are highly cooperative.

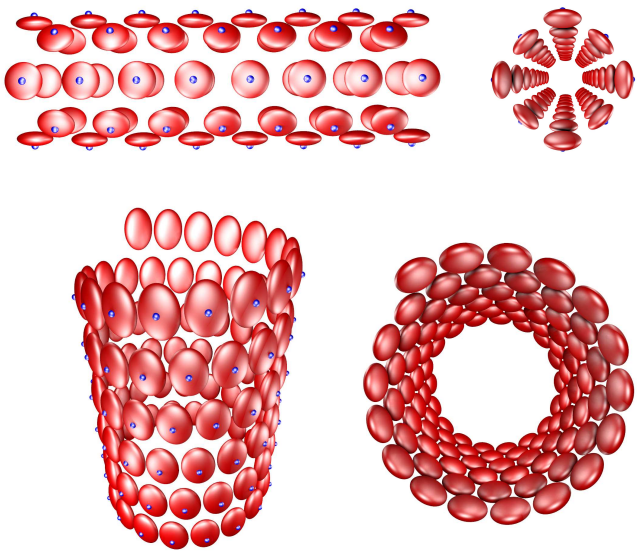
Emergent Behaviour from Simple Models (*ACS Nano*, **4**, 219, 2010)

Adding two repulsive axial Lennard-Jones sites to an ellipsoidal core produces remarkably versatile building blocks. Oblate ellipsoids favour shells, while stronger repulsion for the longer semiaxis produces tubes and spirals.

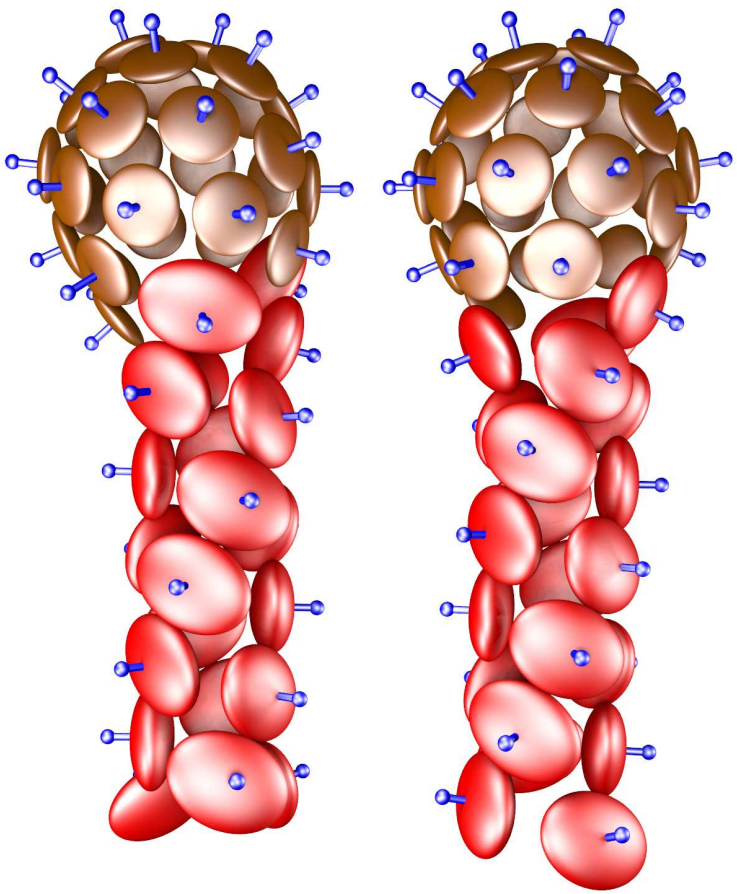
Global minima for the oblate ellipsoids include icosahedra for $N = 12$, 32 and 72 ($T = 1, 3$ and 7), the snub cube observed for polyoma virus capsids at $N = 24$, and conical, biaxial, prolate, and oblate shells at other sizes.

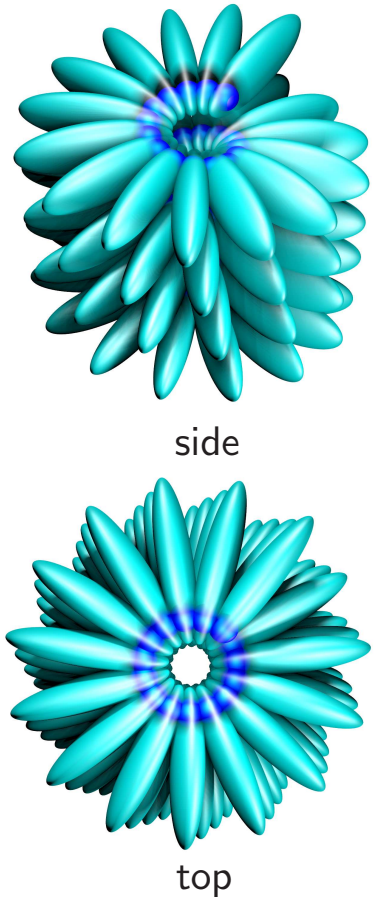






Modelling Mesoscopic Structures (*ACS Nano*, **4**, 219, 2010)

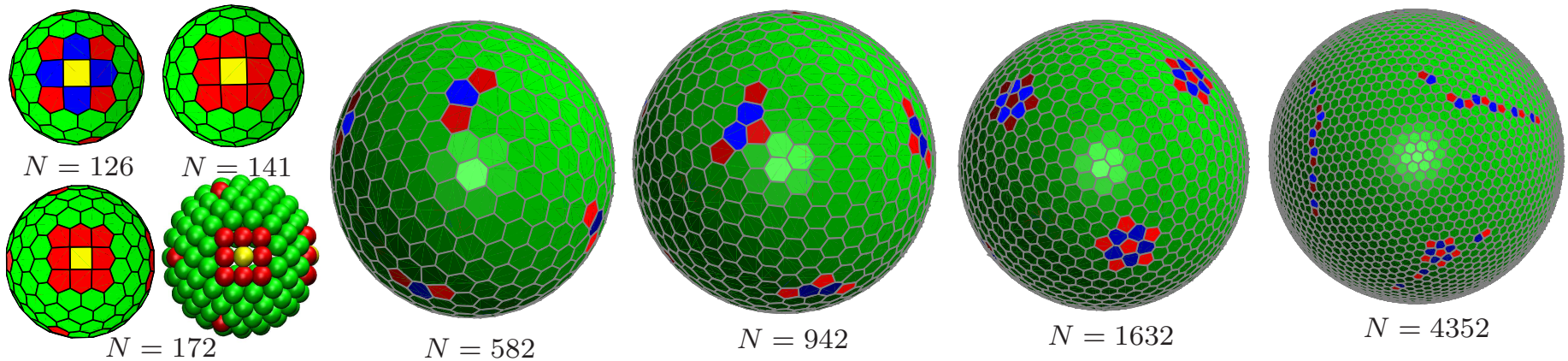




Mixing ellipsoidal building blocks that favour shells and tubes produces structures with distinct head and tail regions (left): the Frankenphage.

Particles with a Lennard-Jones site buried in the ellipsoid assemble into a spiral structure (right) with parameters similar to tobacco mosaic virus.

The Thomson Problem (*Phys. Rev. B*, **74**, 212101, 2006; **79**, 224115, 2009)



Long-ranged potential: $V = \sum_{i<j} 1/|\mathbf{r}_i - \mathbf{r}_j|$ with $|\mathbf{r}_i| = 1$. Twelve five-coordinate particles (disclinations) enable a spherical system to obey Euler's rule for the disclination charge.

Pentagon patches, extended dislocations (scars), twinned defects, rosettes, and embryonic grain boundaries occur in larger systems.

Structures provide models for spherically constrained systems: multielectron bubbles in superfluid helium, cell surface layers, 'colloidosomes', colloidal silica microspheres, superconducting films, lipid rafts deposited on vesicles.

Thermodynamic Properties from the Superposition Approach

In the superposition approach we write the global partition function as a sum over local minima. The weight for minimum α includes the number of distinguishable permutation-inversion isomers, N_α^{PI}, which cannot be superimposed by rigid body rotations when the atoms are labelled.

This factor is related to the order of the rigid molecule point group, o_α, as

$$N_\alpha^{\mathrm{PI}} = \left(2 \prod_\beta N_\beta! \right) / o_\alpha \equiv P / o_\alpha. \tag{1}$$

$Z(T)$ is decomposed into contributions from each local minimum, $Z_\alpha(T)$, which are identical for all the corresponding permutation-inversion isomers:

$$Z(T) = \sum_\alpha N_\alpha^{\mathrm{PI}} Z_\alpha(T) = P \sum_{\alpha=1}^{N^{\mathrm{st}}} Z_\alpha(T)/o_\alpha. \tag{2}$$

Higher point group symmetry acts to reduce the thermodynamic weight.

Regrouping Stationary Point Databases

Lumping local minima together (recursively) if they are separated by low barriers or fast rates reduces the dimension of the kinetic transition network (*J. Chem. Phys.*, **123**, 234901, 2005; *J. Chem. Phys.*, **121**, 1080, 2004). It also provides a self-consistent definition of products and reactants.

The occupation probability and free energy of a group of minima, J are

$$p_J^{\mathrm{eq}}(T) = \sum_{j \in J} p_j^{\mathrm{eq}}(T) \quad \text{and} \quad F_J(T) = -kT \ln \sum_{j \in J} Z_j(T),$$

and the free energy of the transition states connecting J and L is then

$$F_{LJ}^{\dagger}(T) = -kT \ln \sum_{(lj)^{\dagger}} Z_{lj}^{\dagger}(T), \qquad l \in L,\ j \in J,$$

$$\text{with} \quad k_{LJ}^{\dagger}(T) = \sum_{(lj)^{\dagger}} \frac{p_j^{\mathrm{eq}}(T)}{p_J^{\mathrm{eq}}(T)} k_{lj}^{\dagger}(T) = \frac{kT}{h} \exp\left[-\frac{\left(F_{LJ}^{\dagger}(T) - F_J(T) \right)}{kT} \right].$$