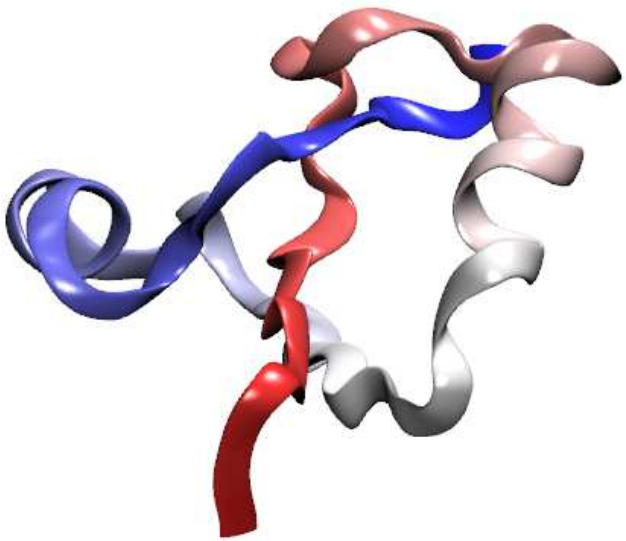
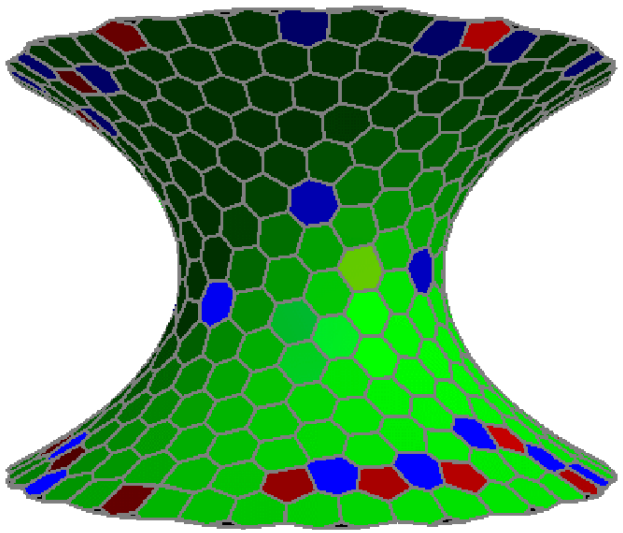
Energy Landscapes: Molecules, Nanodevices and Machine Learning

Objective: to exploit stationary points (minima and transition states) of the PES as a computational framework (*J. Phys. Chem. B*, **110**, 20765, 2006):

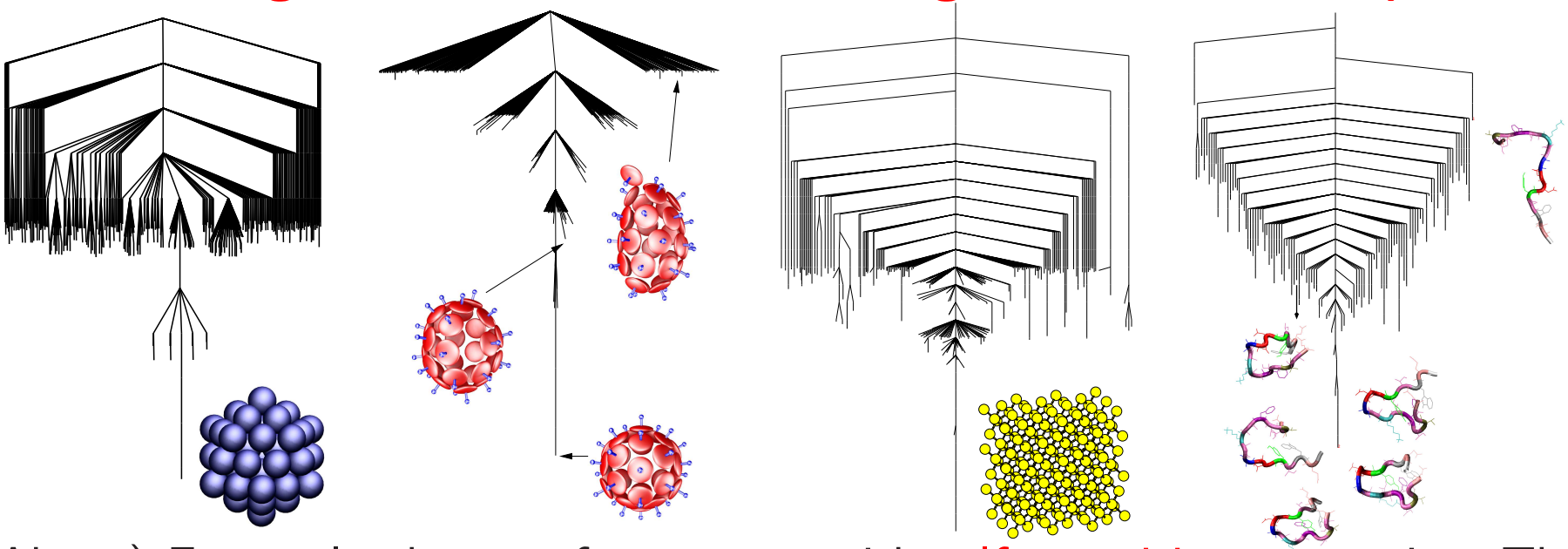
- Basin-hopping for global optimisation (*J. Phys. Chem. A*, **101**, 5111 1997)
- Basin-sampling for global thermodynamics (*J. Chem. Phys.*, **124**, 044102, 2006)
- Discrete path sampling for global kinetics (*Mol. Phys.*, **100**, 3285, 2002)

For small molecules all the relevant stationary points can be located. Larger systems, such as proteins (left) and colloids (right) require sampling.

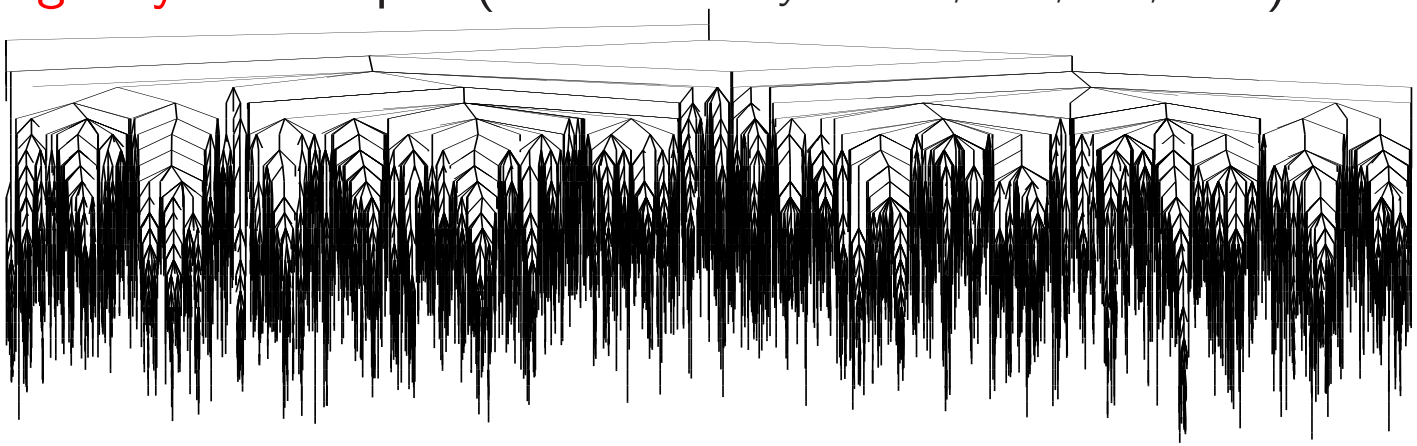




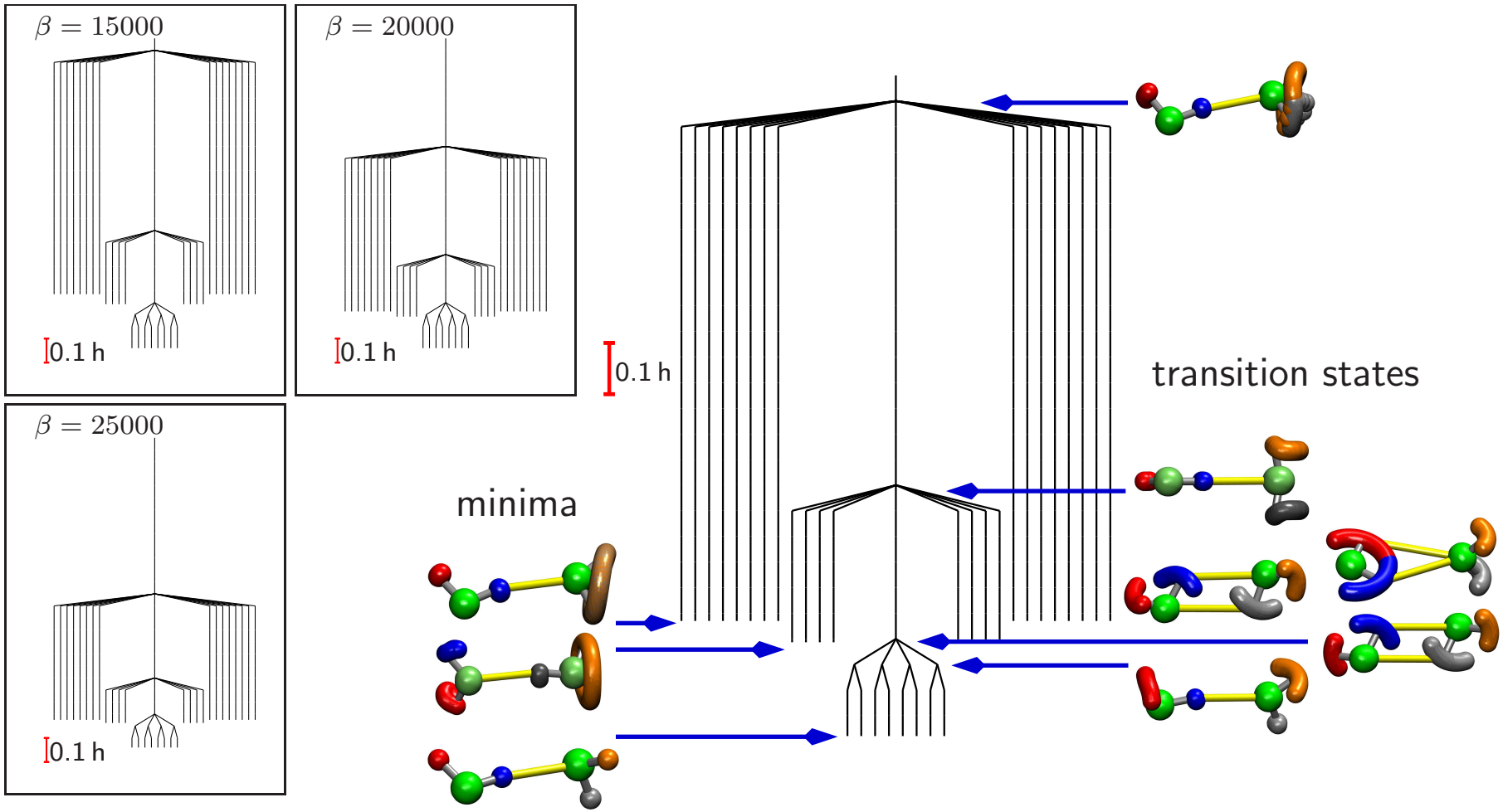
Self-Organisation is Encoded in Single Funnel Landscapes



(Above) Energy landscapes for systems with self-organising properties. The LJ_{55} cluster, an icosahedral shell, crystalline silicon, and the GB1 peptide. (Below) A glassy landscape. (*Phil. Trans. Roy. Soc. A*, **363**, 357, 2005).

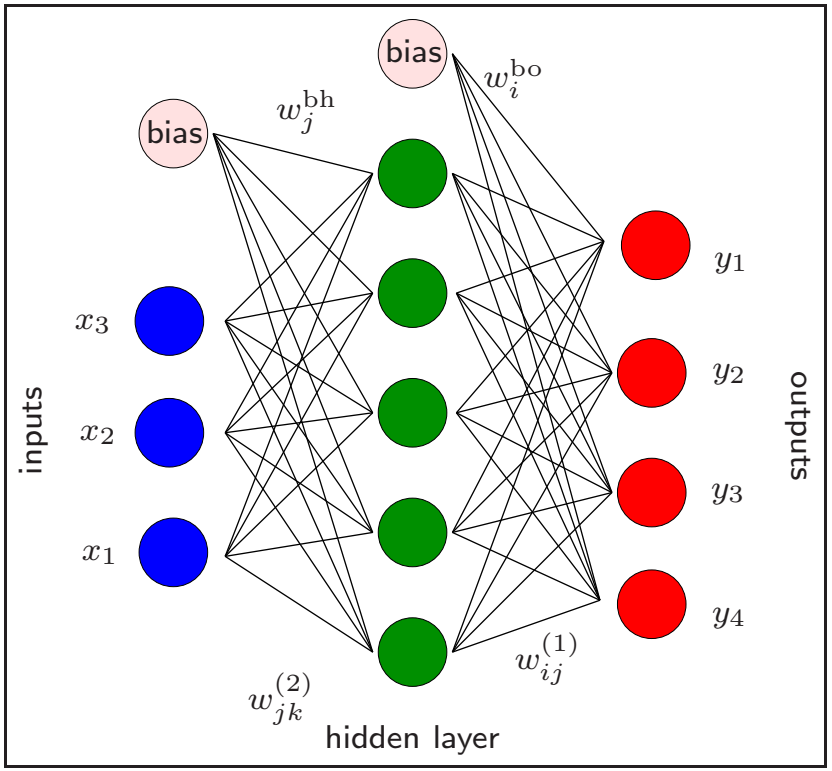


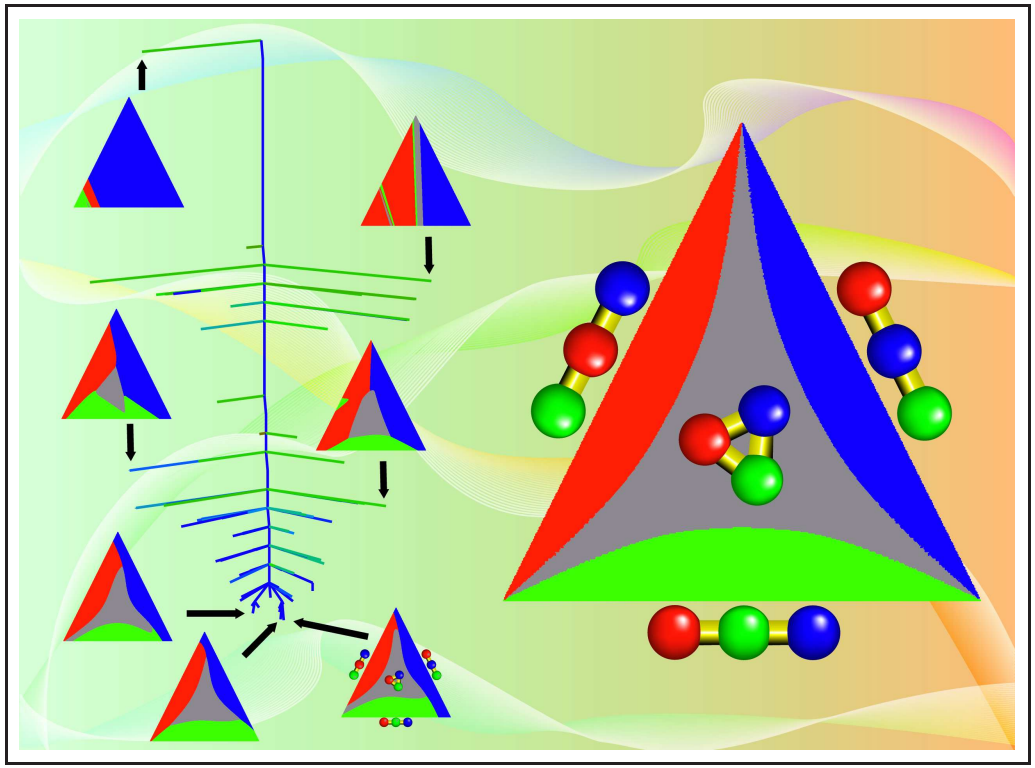
Path Integral Energy Landscapes for Water Dimer



The ring polymer landscape of $(H_2O)_2$ includes classical and delocalised minima and transition states for the MBPOL potential (results for 201 beads).

Machine Learning Landscapes (*JCP*, **144**, 124119, 2016; *CPL*, **667**, 158, 2017)

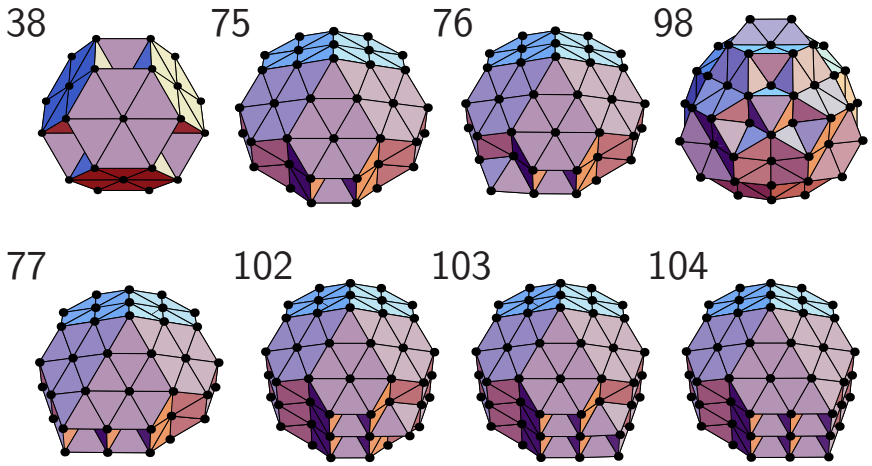




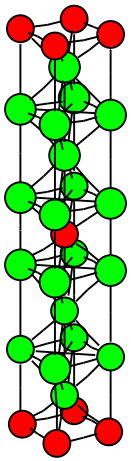
Neural network fits produce multiple solutions, defining a landscape for stationary points of the cost function for multinomial logistic regression.

In this example we predict the outcome of geometry optimisation for an atomic cluster with four distinct isomers using only the three initial bond lengths.

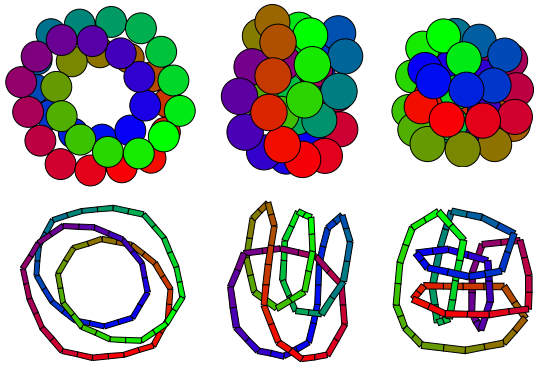
Basin-Hopping Global Optimisation (*J. Phys. Chem. A*, **101**, 5111, 1997)



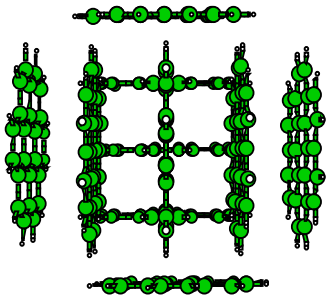
Non-icosahedral Lennard-Jones Clusters



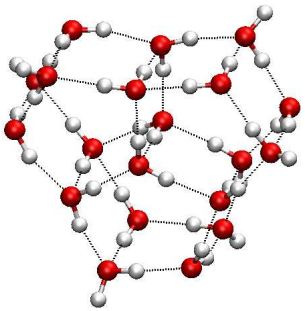
Binary LJ unit cell



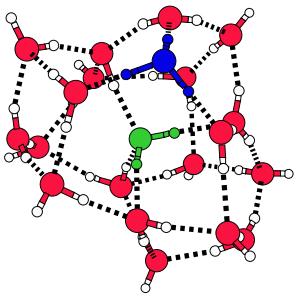
Stockmayer



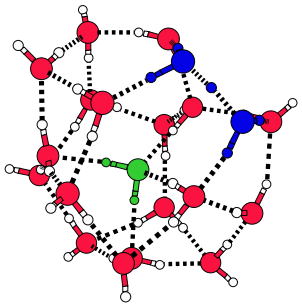
coronene$_{10}$



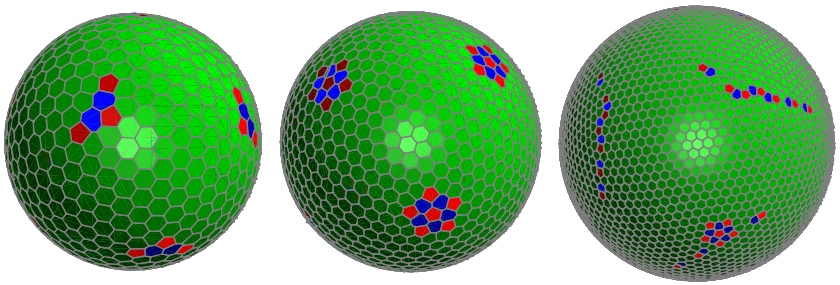
$(H_2O)_{20}$



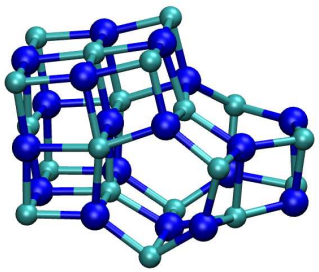
$H_3O^+(H_2O)_{20}$ Eigen



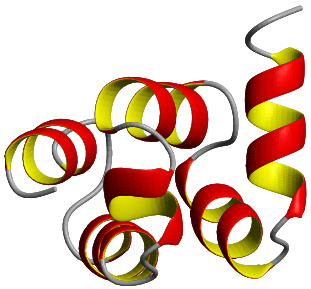
$H_3O^+(H_2O)_{20}$ Zundel



Thomson problem



$(NaCl)_{18}Na^+$



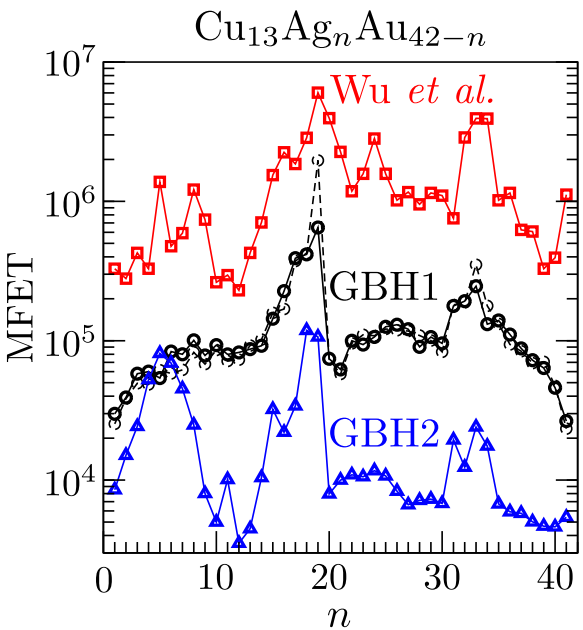
HYPA/FBP11

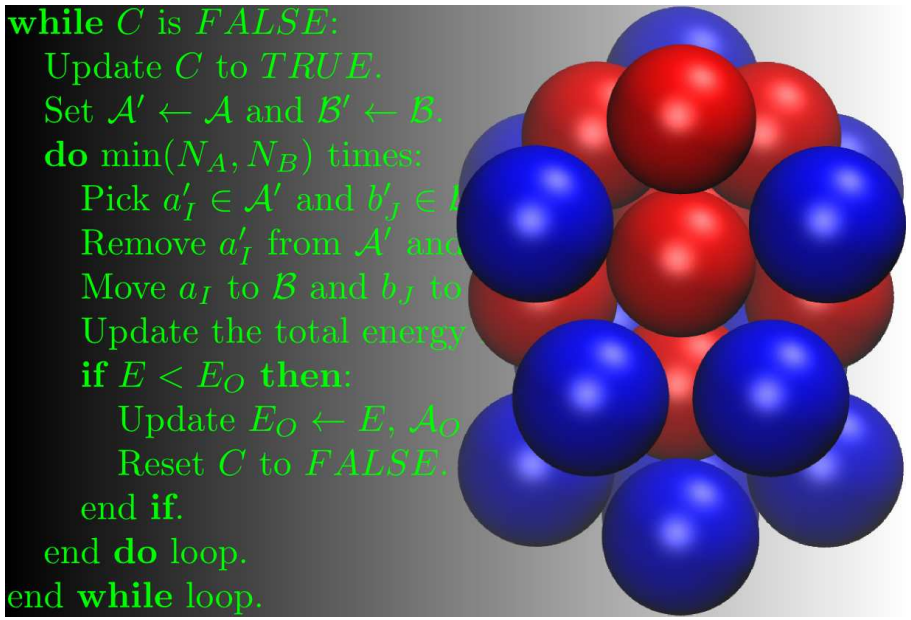
Generalised Basin-Hopping (*PRL*, **113**, 156102, 2014; *PCCP*, **17**, 902, 2016)

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)





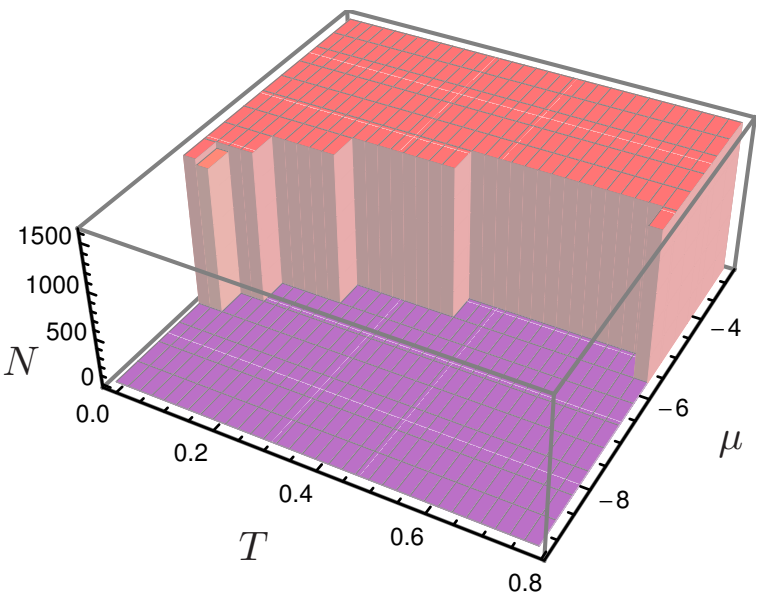
Grand and Semi-Grand Canonical Basin-Hopping (*JCTC*, **12**, 902, 2016)

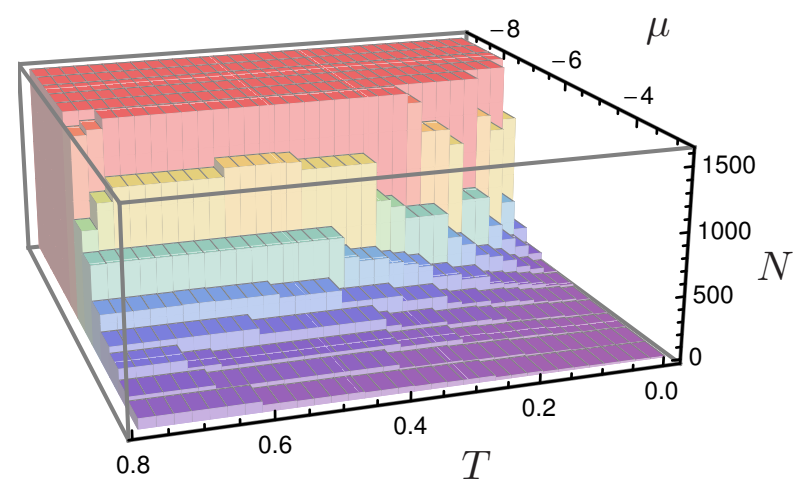
The accept/reject condition for GCBH employs the local grand potential:

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

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

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



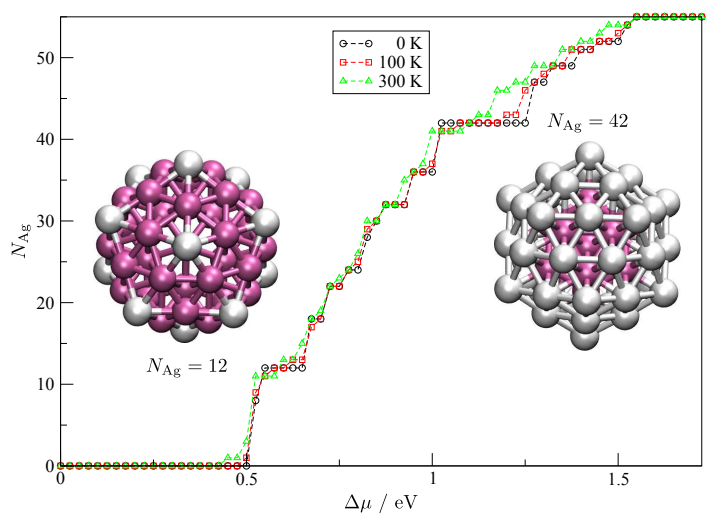


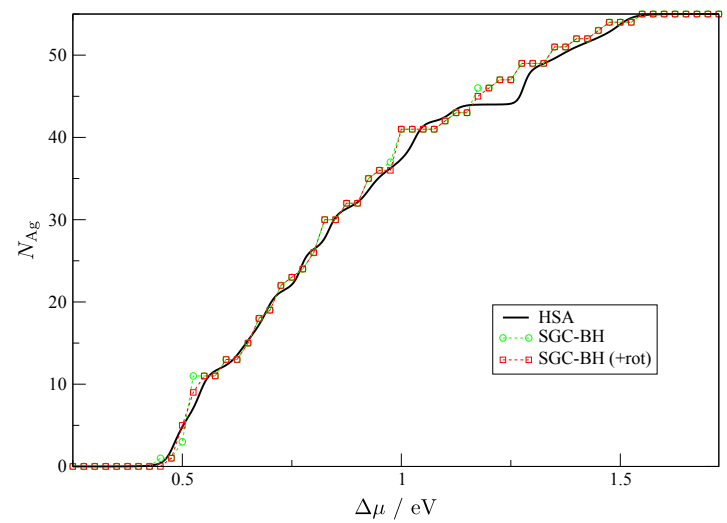
The most probable cluster size (left) corresponds to the smallest or largest size permitted. However, the least probable size can occur between the limits (right), suggesting a free energy barrier to nucleation and cluster growth.

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.

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

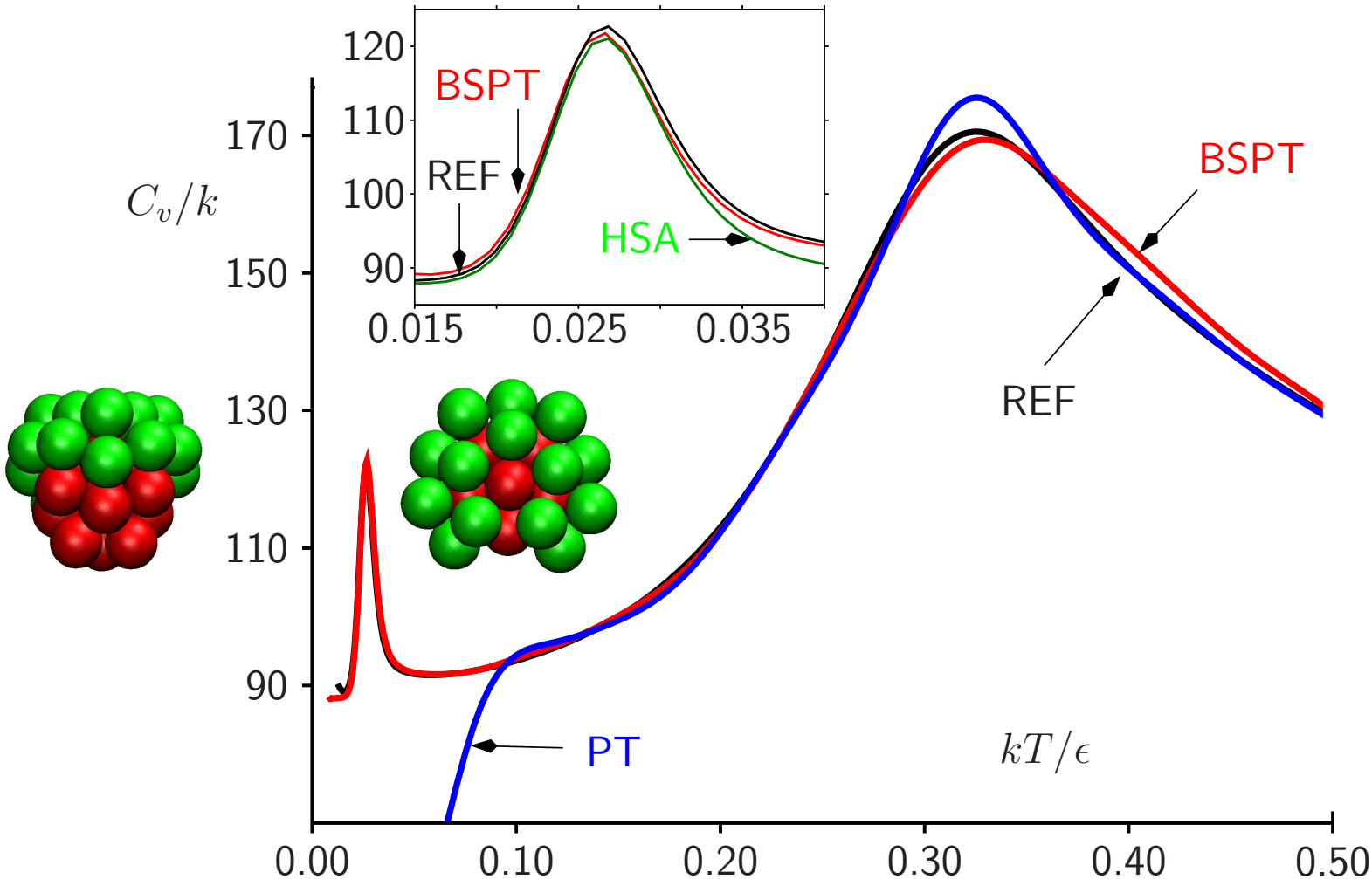
$$P(V_i^I, V_q^Q, T_r) = \mathcal{N}_{iqr}/\mathcal{N}_r \propto \Omega_c(V_i^I, V_q^Q) e^{-V_i^I/kT_r}.$$

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

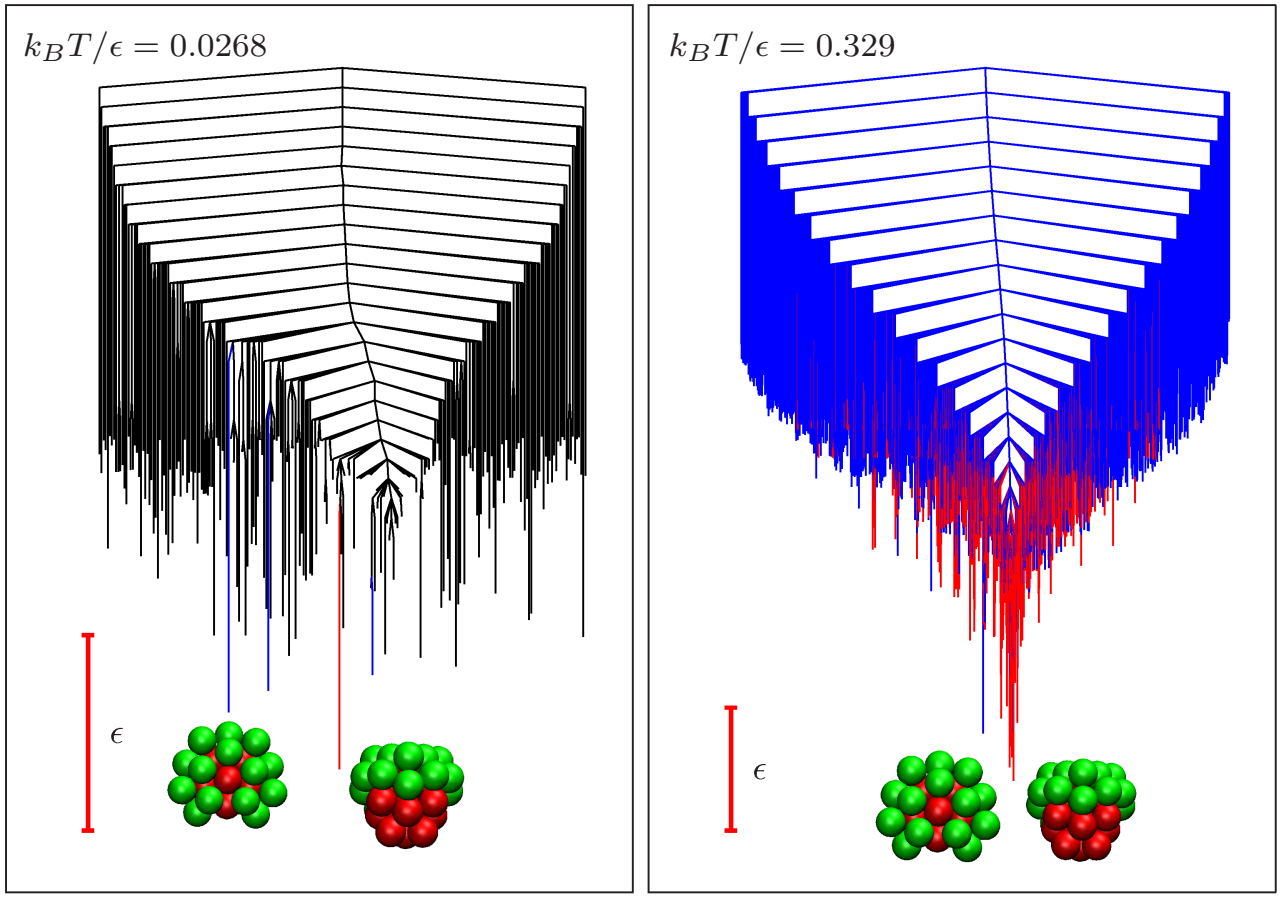
$$\ln \Omega_c(V_i^I, V_q^Q) = \left[(3N-6)/2 - 1 + e^{A_q}\left(V_i^I - V_q^Q\right)\right] \ln\left(V_i^I - V_q^Q\right) + B_q.$$

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.



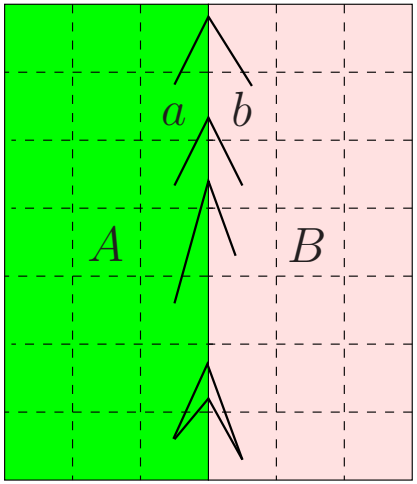
Assigning Heat Capacity Features (*Phys. Rev. E*, **95**, 030105R, 2017)

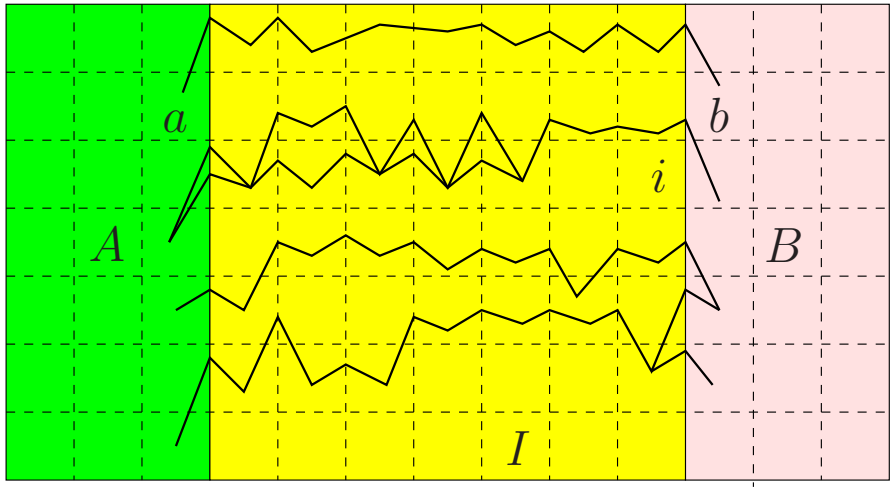


Contributions to C_V can be decomposed as sums over local minima with positive and negative occupation probability gradients, $g_\gamma(T) = \partial p_\gamma(T)/\partial T$:

$$C_V = \kappa k_B + k_B T^2 \sum_{\gamma}^{g_\gamma(T)>0} g_\gamma(T)^2/p_\gamma(T) + k_B T^2 \sum_{\gamma}^{g_\gamma(T)<0} g_\gamma(T)^2/p_\gamma(T)$$

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





no intervening minima

$$\frac{p_a(t)}{p_{a'}(t)} = \frac{p_a^{\rm eq}}{p_{a'}^{\rm eq}} \qquad \dot{p}_i(t) = 0 \qquad \frac{p_b(t)}{p_{b'}(t)} = \frac{p_b^{\rm eq}}{p_{b'}^{\rm eq}}$$

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, $p_b^{\rm eq}$:

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

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.

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

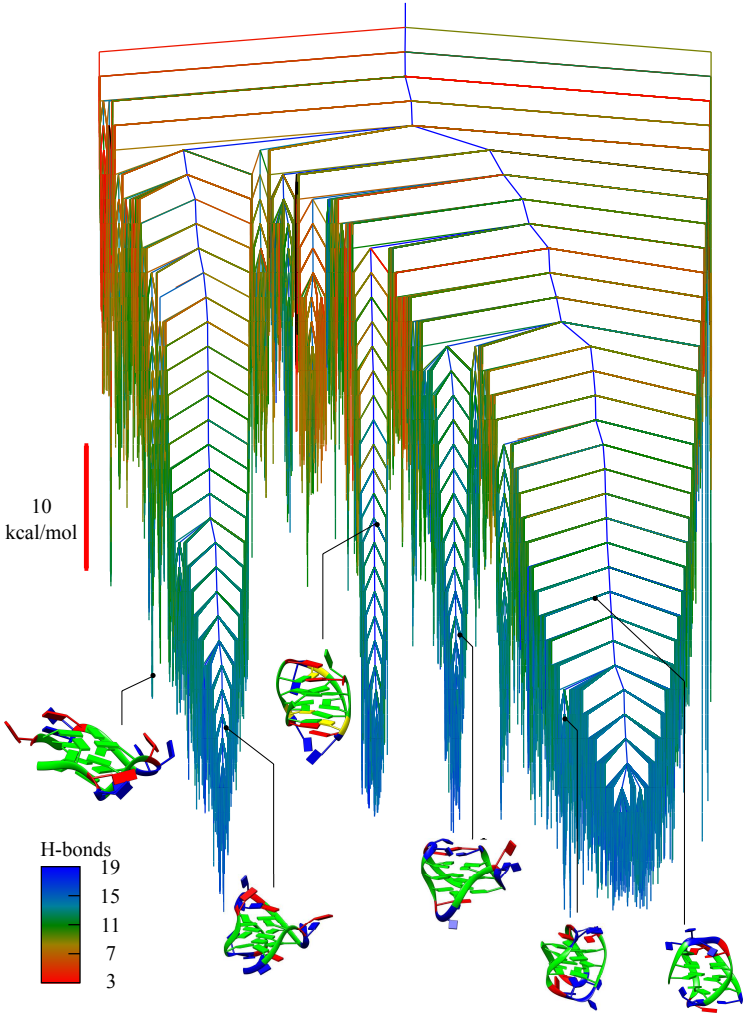
The deterministic graph transformation procedure is non-stochastic and non-iterative. Minima, x, are progressively removed, while the branching probabilities and waiting times in adjacent minima, β, are renormalised:

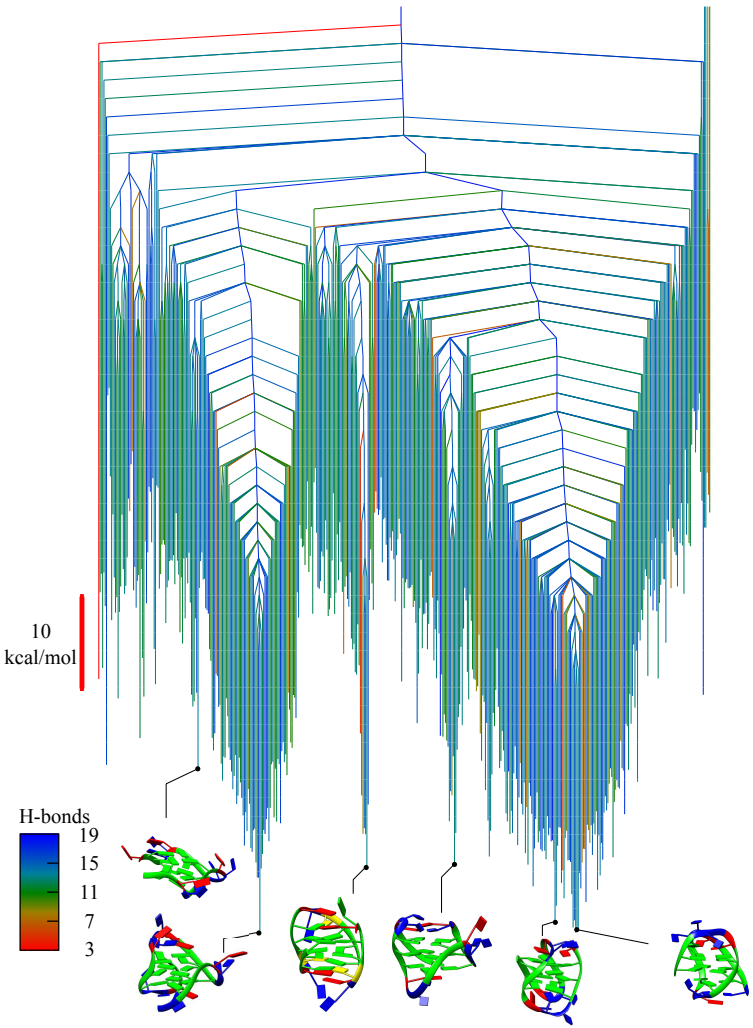
$$P'_{\gamma\beta} = P_{\gamma\beta} + P_{\gamma x} P_{x\beta} \sum_{m=0}^{\infty} P_{xx}^{m} = P_{\gamma\beta} + \frac{P_{\gamma x} P_{x\beta}}{1 - P_{xx}}, \qquad \tau'_{\beta} = \tau_{\beta} + \frac{P_{x\beta}\tau_{x}}{1 - P_{xx}}.$$

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	

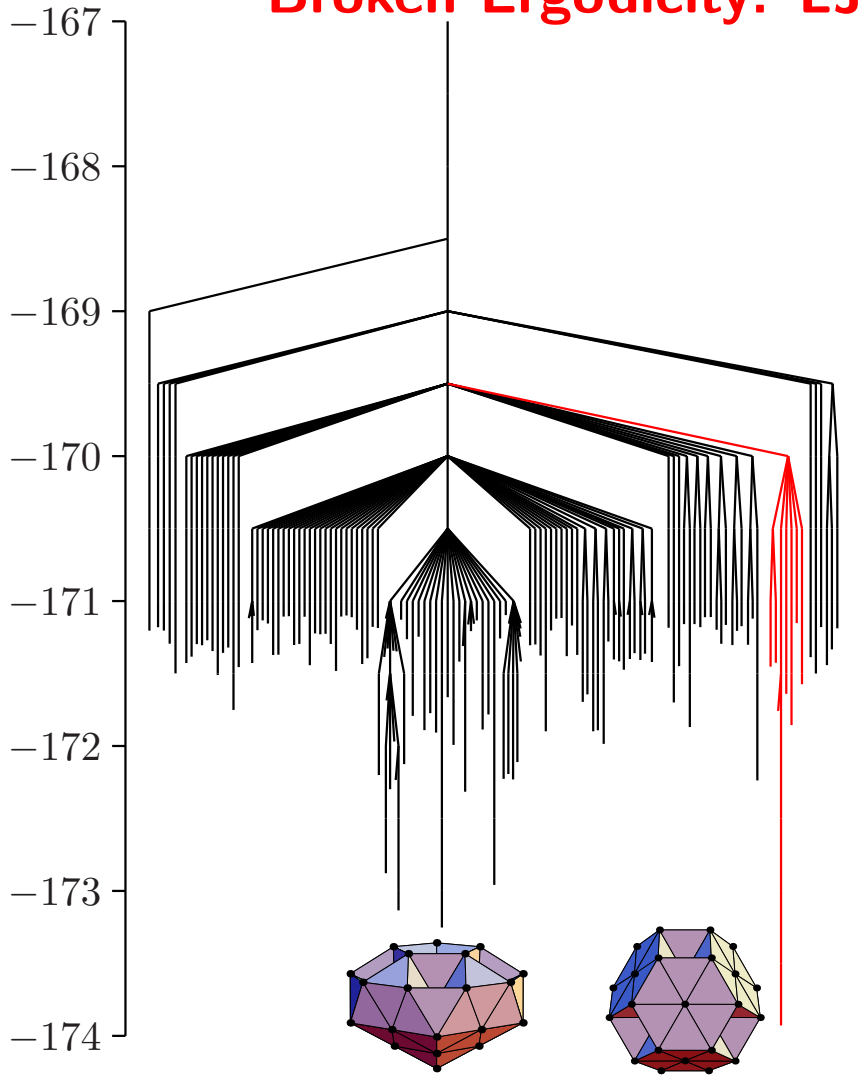
Landscapes for a DNA four-fold telomere repeat: $A(G_3TTA)_3G_3$

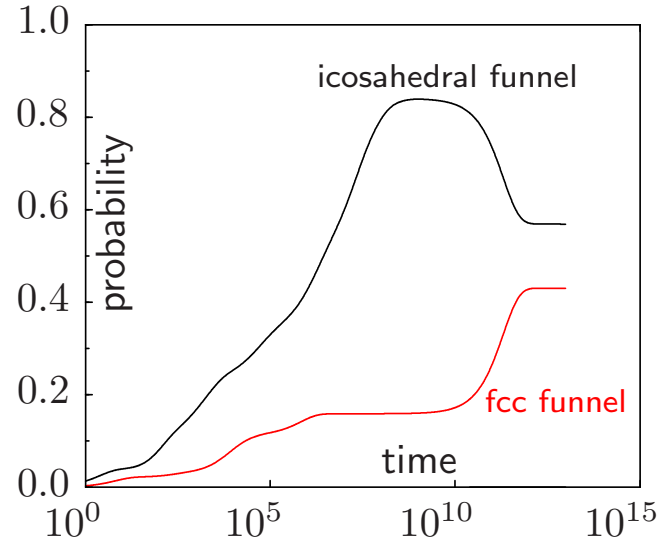


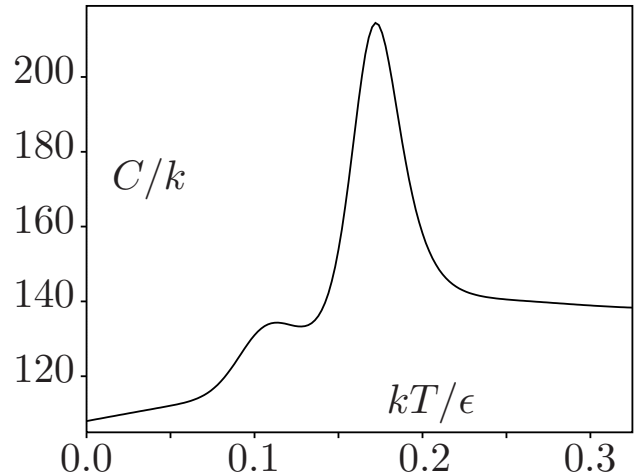


G-quadruplexes, stabilised by quartets of guanine bases, decrease the activity of telomerase, which maintains the length of telomeric repeats. Misfunctioning is associated with numerous cancer conditions. (*JCP*, **147**, 152715, 2017)

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)

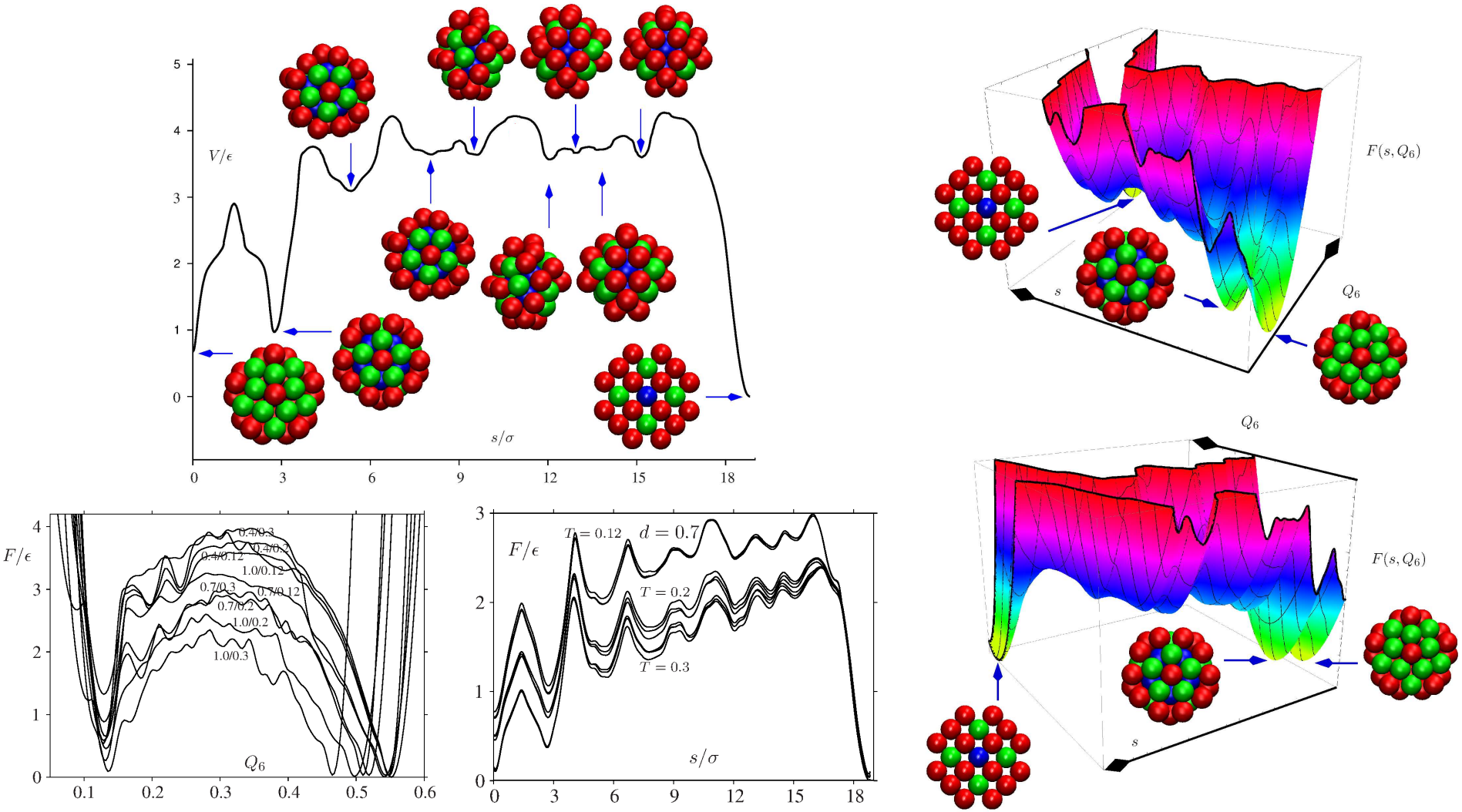
Simulating structural transitions by direct transition current sampling: The example of LJ_{38}

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)

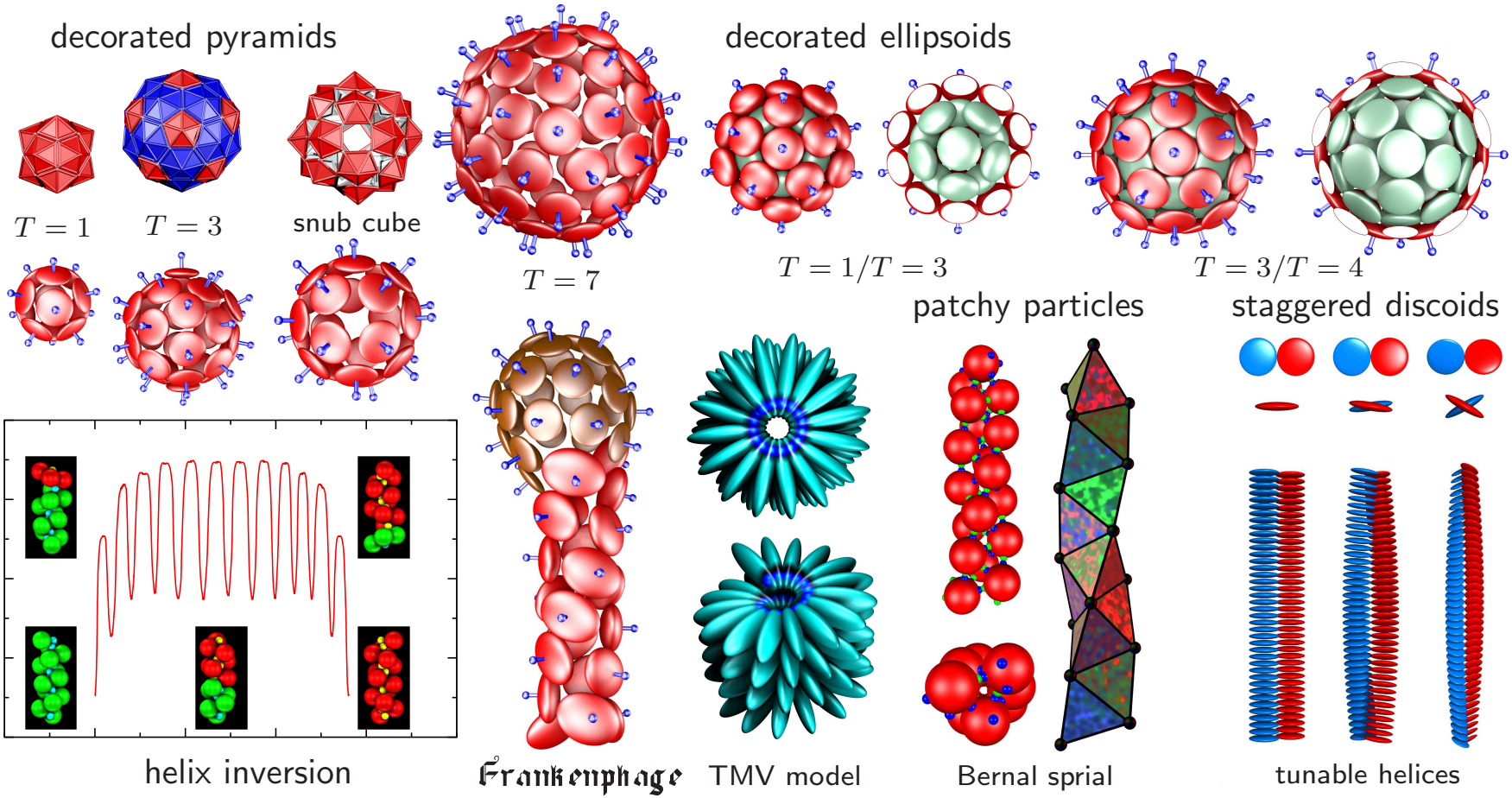
Another attempt to study the transitions between the two funnels of LJ_{38} relies on the use of transition path sampling.[33] Because of the number of metastable states separating the two main basins, the traditional shooting and shifting algorithm failed here, despite previous success for smaller LJ 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.

Coarse-Grained Models (*PCCP*, **11**, 1970, 2009; *ACS Nano*, **4**, 219, 2010)



The angle-axis formulation provides a particularly convenient framework for mesoscopic modelling, using both decorated rigid bodies and ellipsoids.

All the terms involving angle-axis coordinates can be obtained by the action of a rotation matrix and its derivatives, which are system-independent.

1st derivatives: $\mathbf{R}_k \equiv \dfrac{\partial \mathbf{R}}{\partial p_k} = \dfrac{p_k \sin\theta}{\theta}\,\widetilde{\mathbf{p}}^2 + (1-\cos\theta)(\widetilde{\mathbf{p}}_k\widetilde{\mathbf{p}} + \widetilde{\mathbf{p}}\widetilde{\mathbf{p}}_k) + \dfrac{p_k\cos\theta}{\theta}\widetilde{\mathbf{p}} + \sin\theta\,\widetilde{\mathbf{p}}_k$, with $\widetilde{\mathbf{p}}_1 = \dfrac{1}{\theta^3}\begin{pmatrix} 0 & p_1p_3 & -p_1p_2 \\ -p_1p_3 & 0 & p_1^2-\theta^2 \\ p_1p_2 & \theta^2-p_1^2 & 0 \end{pmatrix}$

2nd derivatives : $\mathbf{R}_{kk} \equiv \dfrac{\partial^2 \mathbf{R}}{\partial p_k^2} = \dfrac{2p_k\sin\theta}{\theta}(\widetilde{\mathbf{p}}_k\widetilde{\mathbf{p}} + \widetilde{\mathbf{p}}\widetilde{\mathbf{p}}_k) + \left(\dfrac{p_k^2\cos\theta}{\theta^2} - \dfrac{p_k^2\sin\theta}{\theta^3} + \dfrac{\sin\theta}{\theta}\right)\widetilde{\mathbf{p}}^2$

$$+\ (1-\cos\theta)(2\widetilde{\mathbf{p}}_k^2 + \widetilde{\mathbf{p}}_{kk}\widetilde{\mathbf{p}} + \widetilde{\mathbf{p}}\widetilde{\mathbf{p}}_{kk}) + (-\frac{p_k^2\sin\theta}{\theta^2} - \frac{p_k^2\cos\theta}{\theta^3} + \frac{\cos\theta}{\theta})\widetilde{\mathbf{p}} + \frac{2p_k\cos\theta}{\theta}\widetilde{\mathbf{p}}_k + \sin\theta\,\widetilde{\mathbf{p}}_{kk},$$

and $\mathbf{R}_{kl} \equiv \dfrac{\partial^2 \mathbf{R}}{\partial p_k p_l} = \dfrac{p_k\sin\theta}{\theta}(\widetilde{\mathbf{p}}_l\widetilde{\mathbf{p}} + \widetilde{\mathbf{p}}\widetilde{\mathbf{p}}_l) + (\dfrac{p_kp_l\cos\theta}{\theta^2} - \dfrac{p_kp_l\sin\theta}{\theta^3})\widetilde{\mathbf{p}}^2 + \dfrac{p_l\sin\theta}{\theta}(\widetilde{\mathbf{p}}_k\widetilde{\mathbf{p}} + \widetilde{\mathbf{p}}\widetilde{\mathbf{p}}_k)$

$$+\ (1-\cos\theta)(\widetilde{\mathbf{p}}_{kl}\widetilde{\mathbf{p}} + \widetilde{\mathbf{p}}_k\widetilde{\mathbf{p}}_l + \widetilde{\mathbf{p}}_l\widetilde{\mathbf{p}}_k + \widetilde{\mathbf{p}}\widetilde{\mathbf{p}}_{kl}) - (\frac{p_kp_l\sin\theta}{\theta^2} + \frac{p_kp_l\cos\theta}{\theta^3})\widetilde{\mathbf{p}} + \frac{p_k\cos\theta}{\theta}\widetilde{\mathbf{p}}_l + \frac{p_l\cos\theta}{\theta}\widetilde{\mathbf{p}}_k + \sin\theta\,\widetilde{\mathbf{p}}_{kl}.$$

Denote positions in the body-fixed frame by superscript 0. For rigid bodies I and J with sites i and j defining site-site isotropic potentials U_{ij}^{IJ} the potential energy is

$$U = \sum_I \sum_{J<I} \sum_{i\in I} \sum_{j\in J} f_{ij}(r_{ij}), \qquad \text{where} \qquad r_{ij} = |\mathbf{r}_{ij}| = |\mathbf{r}_i - \mathbf{r}_j| \qquad \text{and} \qquad f_{ij} \equiv U_{ij}^{IJ} \qquad \text{so that}$$

$$\frac{\partial U}{\partial \zeta} = \sum_{J\neq I}\sum_{i\in I}\sum_{j\in J} f'_{ij}(r_{ij})\frac{\partial r_{ij}}{\partial \zeta}, \quad \text{where} \quad f'_{ij} = \frac{df_{ij}(r_{ij})}{dr_{ij}}, \quad \frac{\partial r_{ij}}{\partial \mathbf{r}^I} = \hat{\mathbf{r}}_{ij}, \quad \frac{\partial r_{ij}}{\partial p_k^I} = \hat{\mathbf{r}}_{ij}\cdot\frac{\partial \mathbf{r}_{ij}}{\partial p_k^I} = \hat{\mathbf{r}}_{ij}\cdot(\mathbf{R}_k^I\mathbf{r}_i^0), \quad \mathbf{r}_{ij} = \mathbf{r}^I + \mathbf{R}^I\mathbf{r}_i^0 - \mathbf{r}^J - \mathbf{R}^J\mathbf{r}_j^0.$$

$$\frac{\partial^2 U_{ij}^{IJ}}{\partial r_k^I \partial r_l^J} = f_2(r_{ij})r_{ij,k}r_{ij,l}\epsilon_{IJ} + f_1(r_{ij})\delta_{kl}\epsilon_{IJ},$$

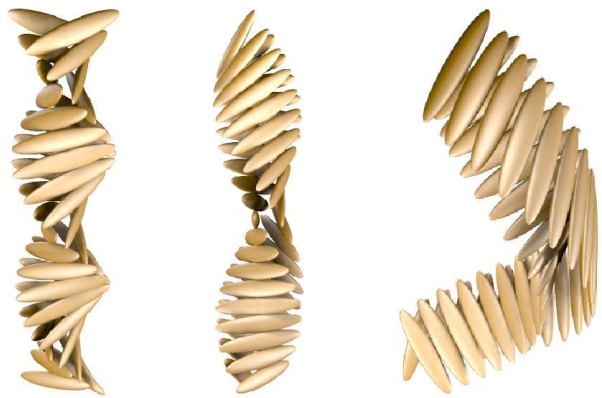
$$\frac{\partial^2 U_{ij}^{IJ}}{\partial p_k^I \partial p_l^J} = f_2(r_{ij})(\mathbf{r}_{ij}\cdot\mathbf{R}_k^I\mathbf{r}_i^0)(\mathbf{r}_{ij}\cdot\mathbf{R}_l^I\mathbf{r}_i^0)\delta_{IJ} - f_2(r_{ij})(\mathbf{r}_{ij}\cdot\mathbf{R}_k^I\mathbf{r}_i^0)(\mathbf{r}_{ij}\cdot\mathbf{R}_l^J\mathbf{r}_j^0)(1-\delta_{IJ}) + f_1(r_{ij})(\mathbf{R}_k^I\mathbf{r}_i^0)\cdot(\mathbf{R}_l^I\mathbf{r}_i^0)\delta_{IJ}$$
$$-f_1(r_{ij})(\mathbf{R}_k^I\mathbf{r}_i^0)\cdot(\mathbf{R}_l^J\mathbf{r}_j^0)(1-\delta_{IJ}) + f_1(r_{ij})(\mathbf{r}_{ij}\cdot\mathbf{R}_{kl}^I\mathbf{r}_i^0)\delta_{IJ},$$

$$\frac{\partial^2 U_{ij}^{IJ}}{\partial r_k^I \partial p_l^J} = f_2(r_{ij})(\mathbf{r}_{ij}\cdot\mathbf{R}_l^I\mathbf{r}_i^0)r_{ij,k}\delta_{IJ} - f_2(r_{ij})(\mathbf{r}_{ij}\cdot\mathbf{R}_l^J\mathbf{r}_j^0)r_{ij,k}(1-\delta_{IJ}) + f_1(r_{ij})[\mathbf{R}_k^I\mathbf{r}_i^0]_l\delta_{IJ} - f_1(r_{ij})[\mathbf{R}_l^J\mathbf{r}_j^0]_l(1-\delta_{IJ}).$$

where $f_1(r_{ij}) = f'_{ij}(r_{ij})/r_{ij}$, $f_2(r_{ij}) = f'_1(r_{ij})/r_{ij}$, $\epsilon_{IJ} = 1$ for $I = J$ and $\epsilon_{IJ} = -1$ for $I \neq J$, and δ_{IJ} is the Kronecker delta.

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

helical fibre morphologies



bilayer filaments

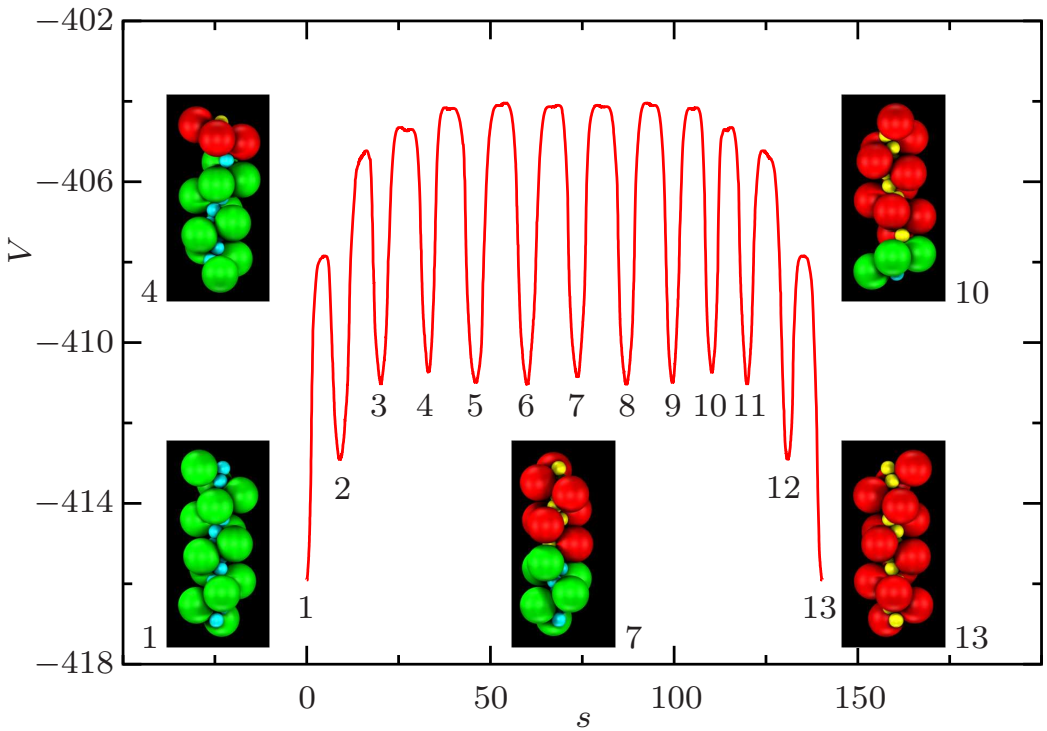
magnetic bilayers

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.

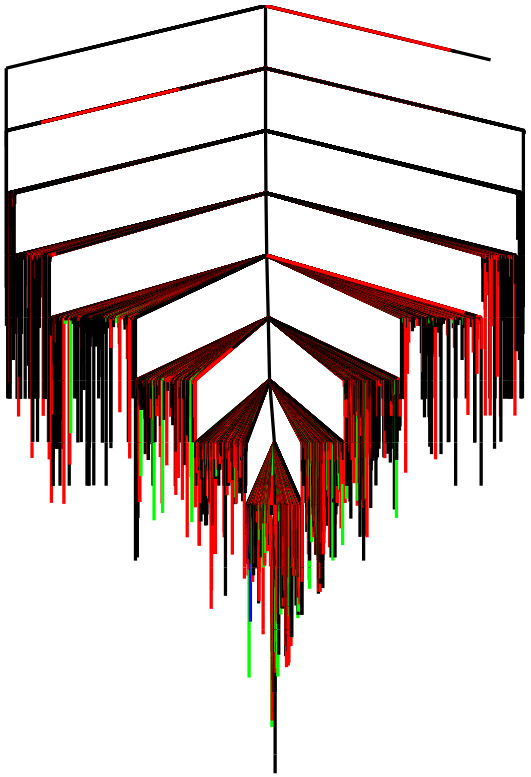
A Nanodevice (*Soft Matter*, **7**, 2325, 2011)

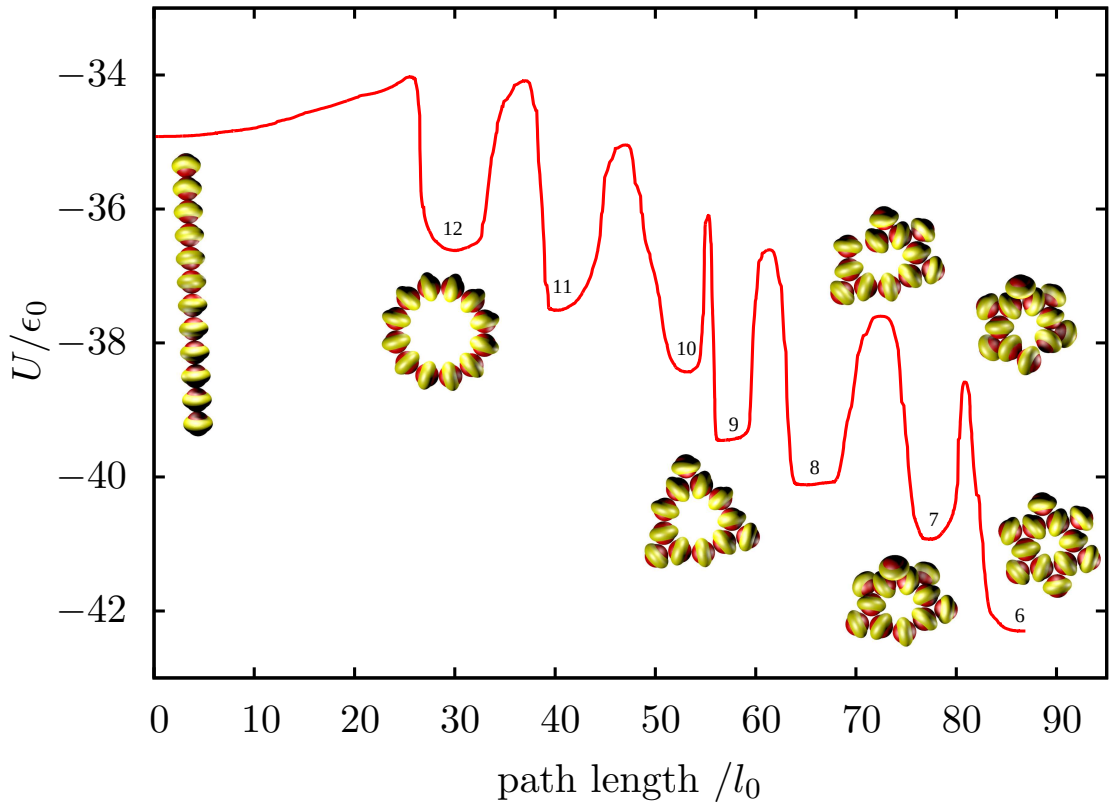


Coupled linear and rotary motion has been characterised for a helix composed of 13 asymmetric dipolar dumbbells in the presence of an electric field.

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

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

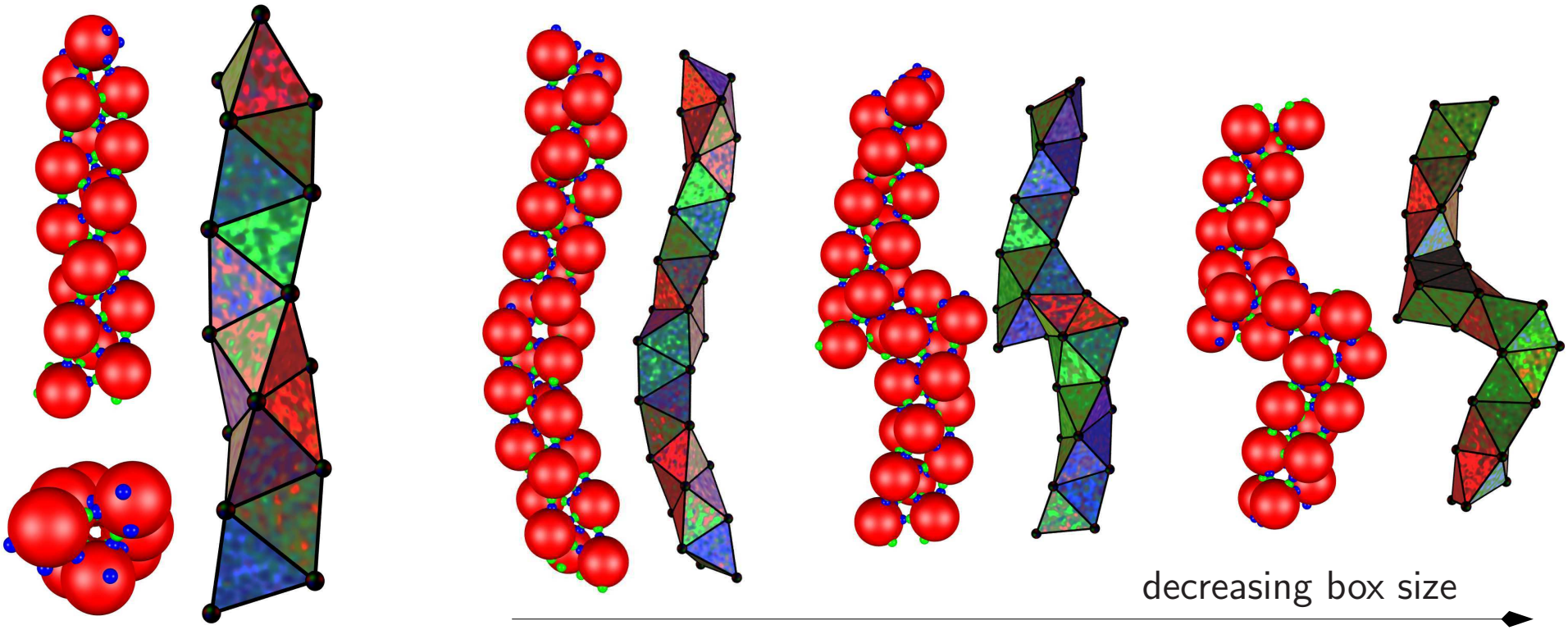




Energetically stabilised Kagome structures were designed using soft anisotropic triblock Janus particles.  This unconstrained model predicts that sedimentation effects enhance the stability.

Rearrangements between competing structures are highly cooperative.

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.

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.

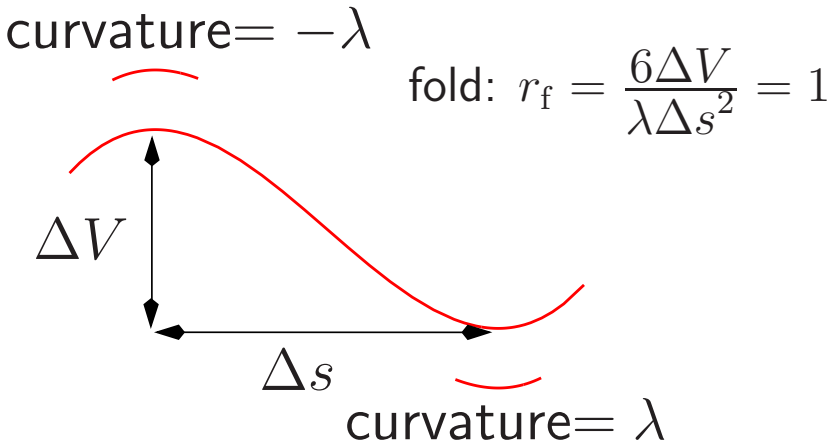
The organisation of families of PES's as a function of parameters in the potential is determined by the stationary points that possess additional zero Hessian eigenvalues, known as non-Morse points.

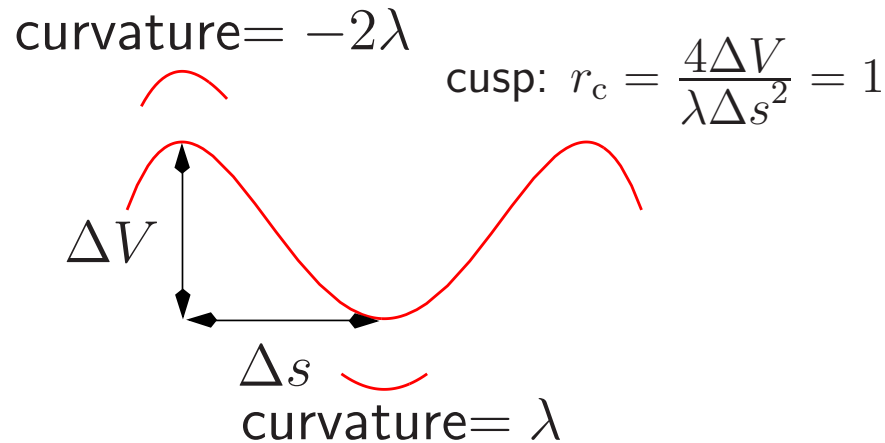
Catastrophe theory provides a local representation of the PES around non-Morse points as a function of both atomic coordinates and parameters.

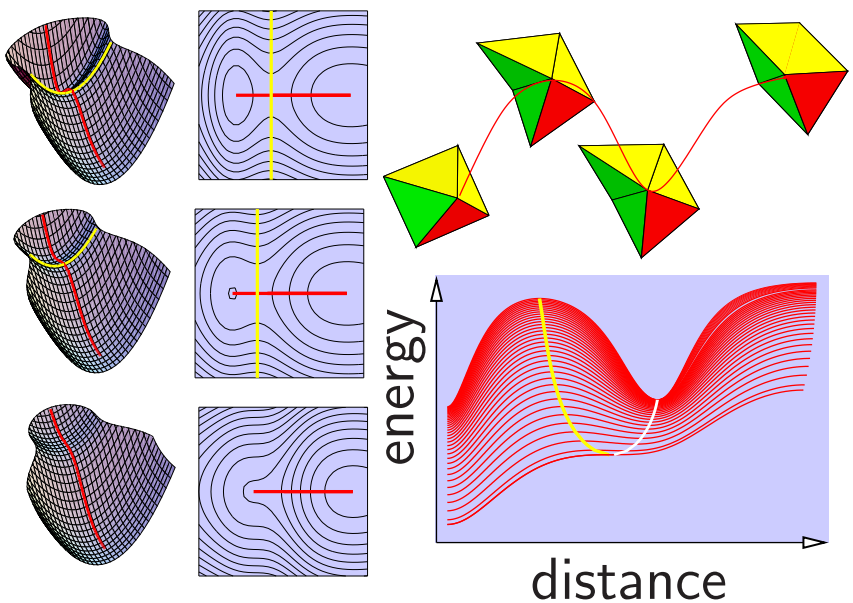
The splitting lemma reduces the dimensionality to the essential variables, while transversality guarantees that the resulting classifications are universal.

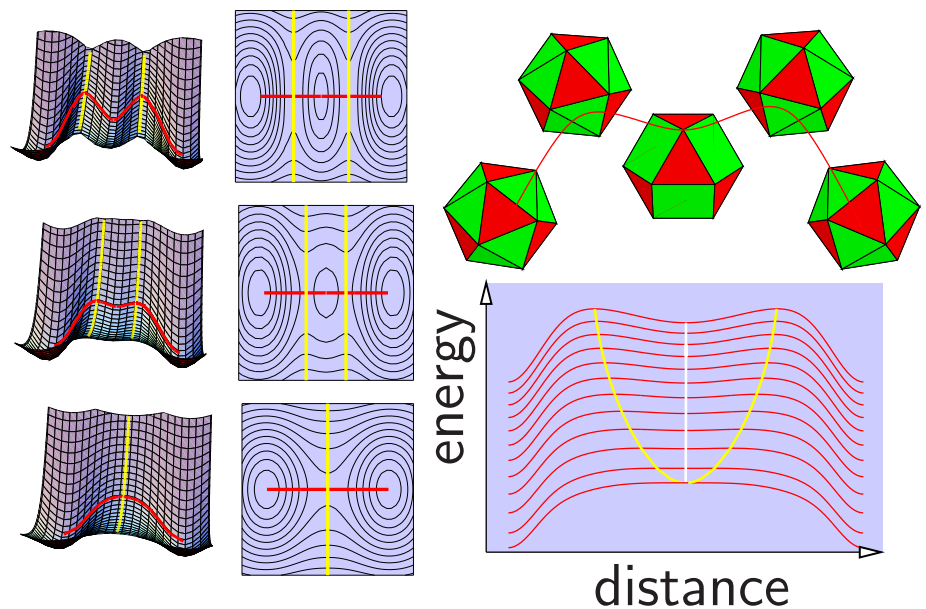
The simplest one-parameter catastrophes are the fold, $f(x) = \frac{1}{3}x^3 + ax$, and the symmetrical cusp, $f(x) = \frac{1}{4}x^4 + \frac{1}{2}ax^2$.

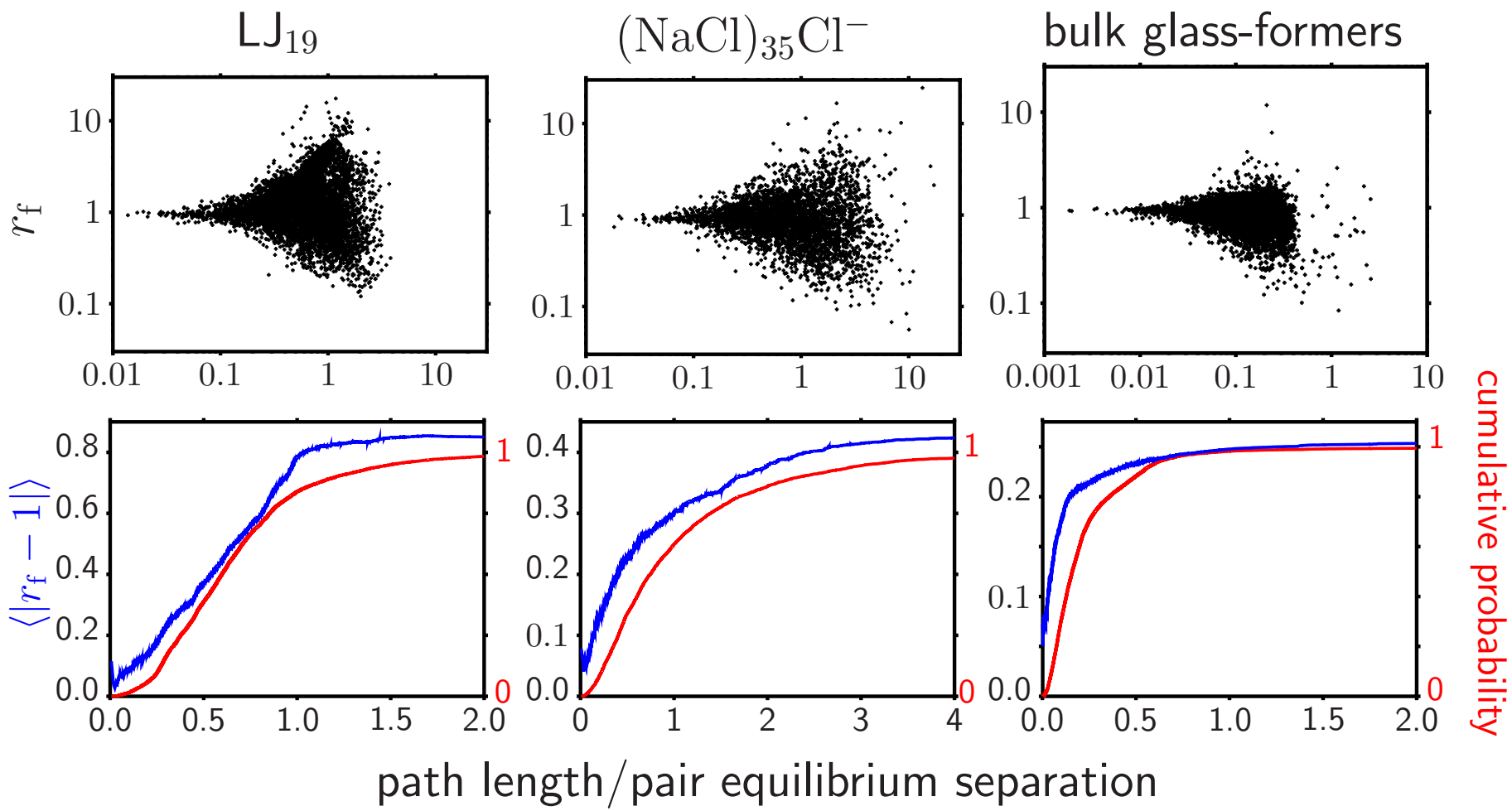
Geometries of the fold and cusp catastrophes.







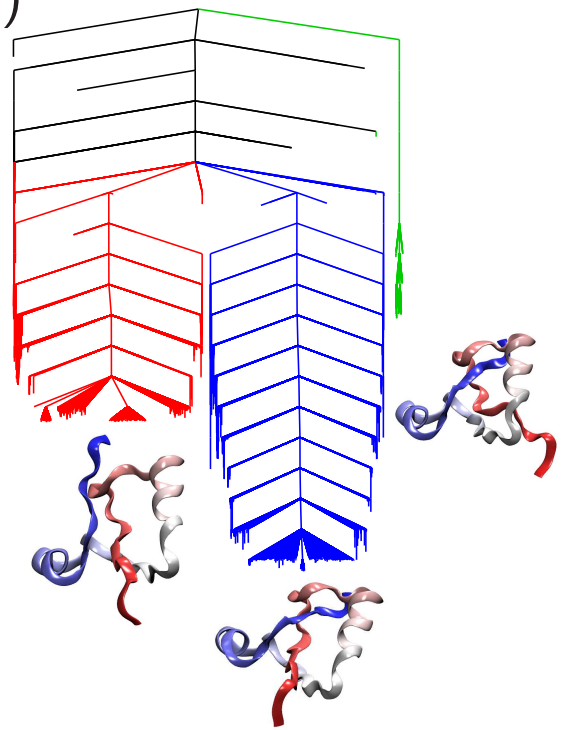




For systems with a fixed potential we effectively have a snap-shot of parameter space. On average, r_{f} remains close to unity for many pathways in both model clusters and bulk, providing an explanation for Hammond's postulate.

A Knotted Protein (*PLoS Comput. Biol.*, **6**, e1000835, 2010)

Quasi-Continuous Interpolation (*JCTC*, **8**, 5020, 2012)

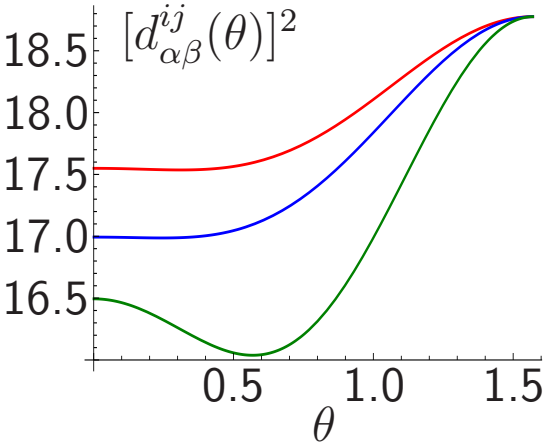


The tRNA methyltransferase protein 1UAM contains a deep trefoil knot (right).

The folding pathway exhibits two slipknot-type steps for a truncated (residues 78–135) Gō model using an associated memory Hamiltonian and initial QCI.

The QCI potential preserves the covalent bonding framework, with short-range repulsion between unconstrained atoms. An internal minimum for atoms α and β between images i and j occurs at

$$[d^{ij}_{\alpha\beta}(\theta)]^2 = \frac{\left|\mathbf{r}^i_\alpha - \mathbf{r}^i_\beta\right|^2 \left|\mathbf{r}^j_\alpha - \mathbf{r}^j_\beta\right|^2 - \left[\left(\mathbf{r}^i_\alpha - \mathbf{r}^i_\beta\right)\cdot\left(\mathbf{r}^j_\alpha - \mathbf{r}^j_\beta\right)\right]^2}{\left|\mathbf{r}^i_\alpha - \mathbf{r}^i_\beta - \mathbf{r}^j_\alpha + \mathbf{r}^j_\beta\right|^2}.$$



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