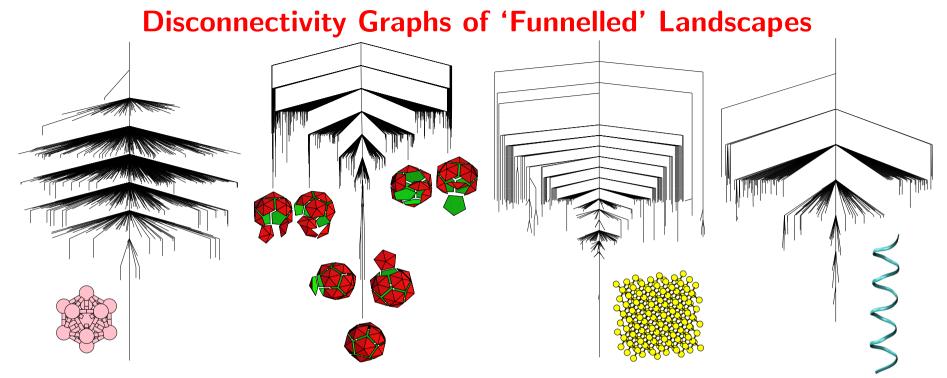
Energy Landscapes for Soft Matter

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

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

For m weakly coupled subsystems of N/m=n atoms each the number of local minima satisfies $f_{\min}(mn)=f_{\min}(n)^m$, so that $f_{\min}(N)=e^{\alpha N}$. For transition states we expect $f_{\rm ts}(mn)=mf_{\min}(n)^{m-1}f_{\rm ts}(n)$ so that $f_{\rm ts}(N)=Ne^{\alpha N}=Nf_{\min}(N)$. Two important consequences:

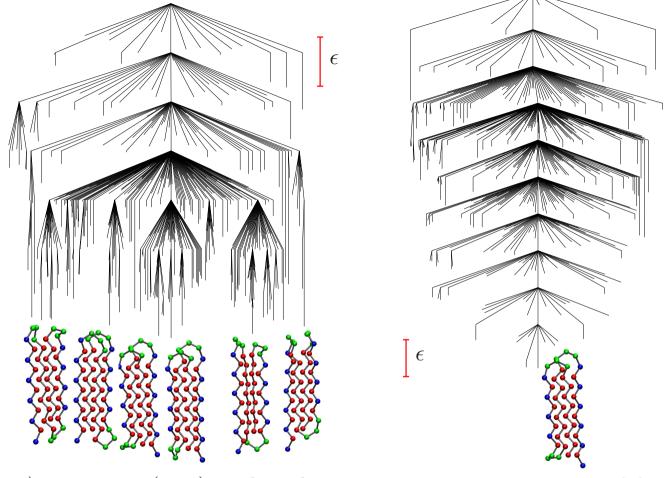
- Appropriate sampling schemes are required for larger systems,
- Low-dimensional projections of the landscape can only represent the connectivity faithfully for a few anharmonic degrees of freedom.



The nonrandom searches that result in magic number clusters, crystallisation, self-assembly, and protein folding are associated with a 'palm tree' organisation of the potential energy landscape (*Phil. Trans. Roy. Soc. A*, **363**, 357, 2005).

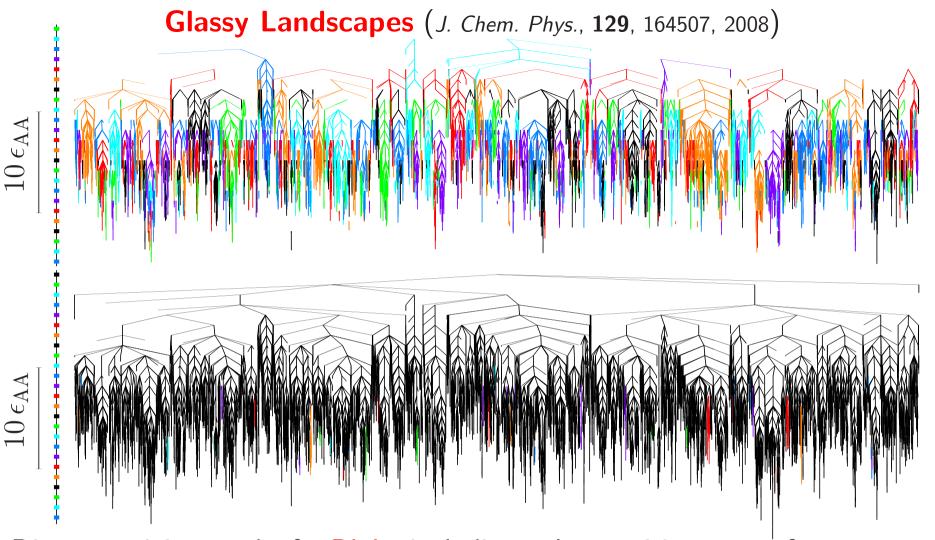
This 'funnelling' pattern has been verified for various structure-seeking systems, including the LJ_{13} cluster, icosahedral shells composed of pentagonal and hexagonal pyramids, crystalline (Stillinger-Weber) silicon, and the polyalanine ala₁₆. Large systems can exhibit relatively simple phenomenology.

Frustrated or Funnelled? A Model Protein



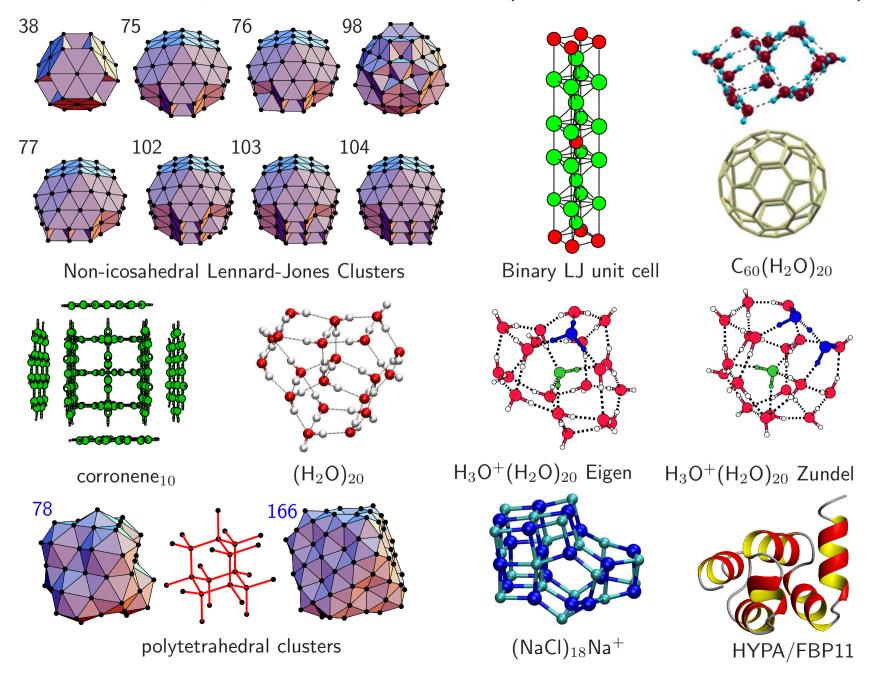
For $B_9N_3(LB)_4N_3B_9N_3(LB)_5L$ low-lying minima are separated by high barriers, where the beads are B=hydrophobic, L=hydrophilic, and N=neutral.

This 'frustration' affects the observed heat capacity and results in distinct relaxation time scales, which are eliminated in the corresponding Gō model.

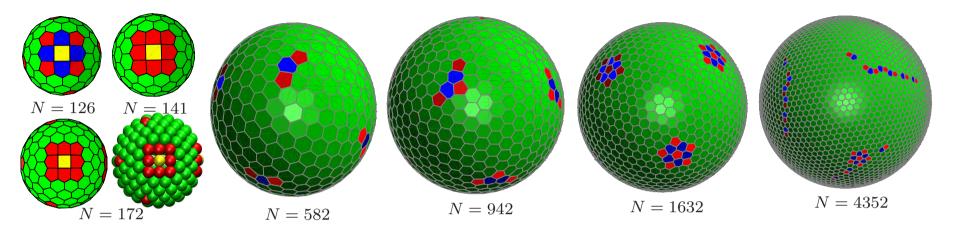


Disconnectivity graphs for BLJ₆₀ including only transition states for noncage-breaking (top) and cage-breaking (bottom) paths. Changes in colour indicate disjoint sets of minima. Cage-breaking transitions, defined by two nearest-neighbour changes, define a higher order metabasin structure.

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



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.

Angle-Axis Coordinates for Rigid Bodies (PCCP, 11, 1970, 2009)

Rodrigues' formula for the rotation matrix ${\bf R}$ corresponding to a rotation of magnitude ${m heta}=(p_1^2+p_2^2+p_3^2)^{1/2}$ around the axis defined by ${\bf p}$ is

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

where I is the identity matrix, and $\tilde{\mathbf{p}}$ is the skew-symmetric matrix

$$\widetilde{\mathbf{p}} = \frac{1}{\theta} \begin{pmatrix} 0 & -p_3 & p_2 \\ p_3 & 0 & -p_1 \\ -p_2 & p_1 & 0 \end{pmatrix}.$$

The product of $\widetilde{\mathbf{p}}$ and any vector \mathbf{v} returns the cross product: $\widetilde{\mathbf{p}}\mathbf{v} = \hat{\mathbf{p}} \times \mathbf{v}$.

All terms involving rigid-body angle-axis coordinates can be obtained by the action of the rotation matrix and its derivatives, whose forms are programmed in system-independent subroutines.

The angle-axis representation is free of singularities and constraints.

1st derivatives: $\mathbf{R}_k \equiv \frac{\partial \mathbf{R}}{\partial p_k} = \frac{p_k \sin \theta}{\theta} \, \tilde{\mathbf{p}}^2 + (1 - \cos \theta) (\tilde{\mathbf{p}}_k \tilde{\mathbf{p}} + \tilde{\mathbf{p}} \tilde{\mathbf{p}}_k) + \frac{p_k \cos \theta}{\theta} \tilde{\mathbf{p}} + \sin \theta \, \tilde{\mathbf{p}}_k, \quad \text{with} \quad \tilde{\mathbf{p}}_1 = \frac{1}{\theta^3} \begin{pmatrix} 0 & p_1 p_3 & -p_1 p_2 \\ -p_1 p_3 & 0 & p_1^2 - \theta^2 \\ p_1 p_2 & \theta^2 - p_1^2 & 0 \end{pmatrix}$

2nd derivatives:
$$\mathbf{R}_{kk} \equiv \frac{\partial^2 \mathbf{R}}{\partial p_k^2} = \frac{2p_k \sin \theta}{\theta} (\widetilde{\mathbf{p}}_k \widetilde{\mathbf{p}} + \widetilde{\mathbf{p}} \widetilde{\mathbf{p}}_k) + \left(\frac{p_k^2 \cos \theta}{\theta^2} - \frac{p_k^2 \sin \theta}{\theta^3} + \frac{\sin \theta}{\theta} \right) \widetilde{\mathbf{p}}^2$$
$$+ (1 - \cos \theta) (2\widetilde{\mathbf{p}}_k^2 + \widetilde{\mathbf{p}}_{kk} \widetilde{\mathbf{p}} + \widetilde{\mathbf{p}} \widetilde{\mathbf{p}}_{kk}) + (-\frac{p_k^2 \sin \theta}{\theta^2} - \frac{p_k^2 \cos \theta}{\theta^3} + \frac{\cos \theta}{\theta}) \widetilde{\mathbf{p}} + \frac{2p_k \cos \theta}{\theta} \widetilde{\mathbf{p}}_k + \sin \theta \widetilde{\mathbf{p}}_{kk},$$

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

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

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 $U_{i,j}^{IJ}$ the potential energy is

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

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

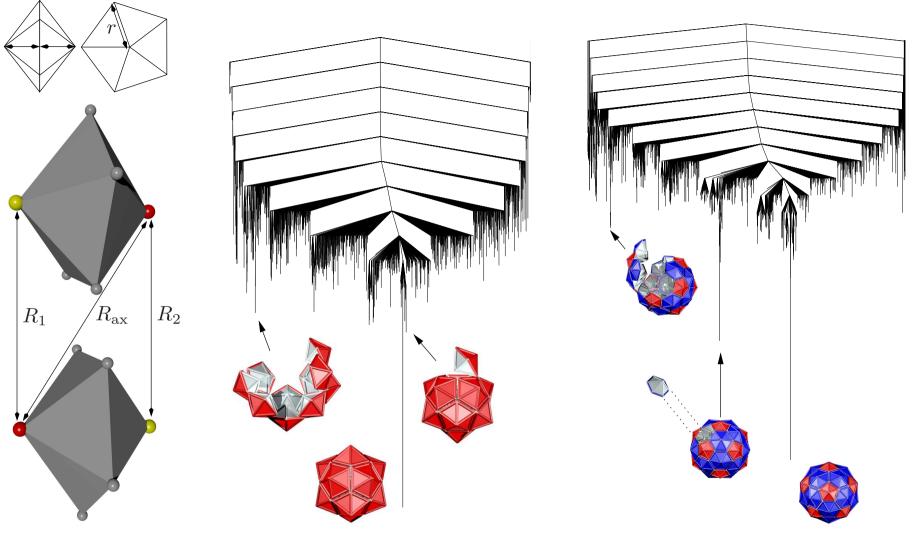
$$\frac{\partial^2 U_{ij}^{IJ}}{\partial r^I \partial r^J} = f_2(r_{ij}) r_{ij,k} r_{ij,l} \epsilon_{IJ} + f_1(r_{ij}) \delta_{kl} \epsilon_{IJ},$$

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

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

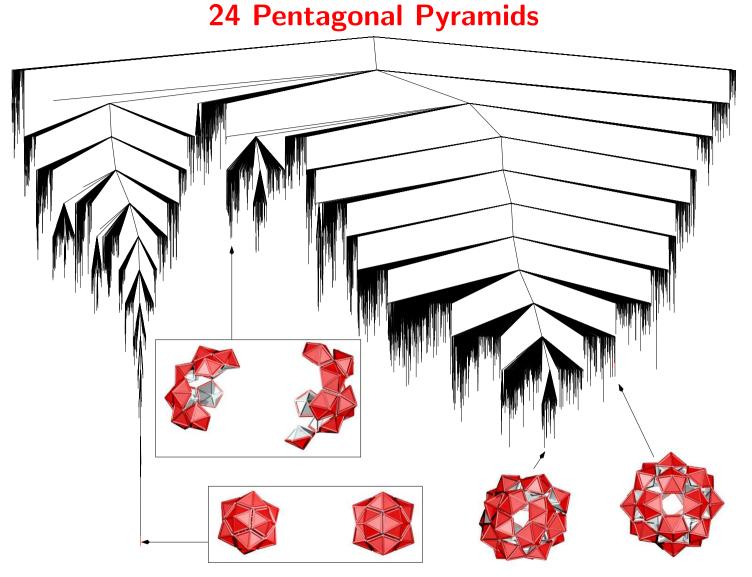
where $f_1(r_{ij}) = f'_{ij}(r_{ij})/r_{ij}$, $f_2(r_{ij}) = f'_1(r_{ij})/r_{ij}$, $\epsilon_{IJ} = 1$ for I = J and $\epsilon_{IJ} = -1$ for $I \neq J$, and δ_{IJ} is the Kronecker delta.

Self-Assembly of Icosahedral Shells (PCCP, 11, 2098-2104, 2009)



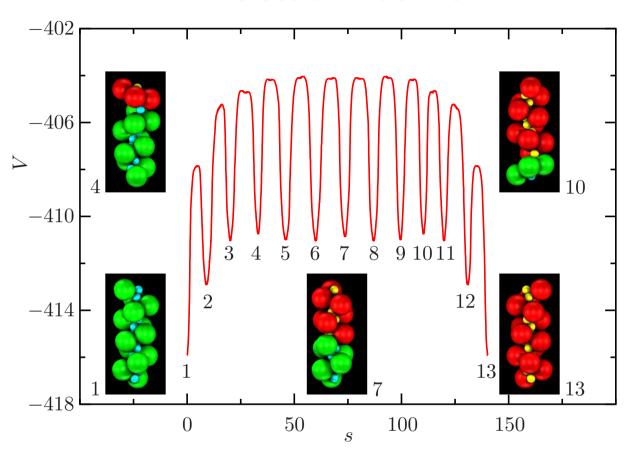
Palm tree disconnectivity graphs with I_h global minima are found for T=1 and T=3 shells constructed from pentagonal and hexagonal pyramids.

Landscapes of this form are associated with good structure-seekers.



For the same parameters two T=1 icosahedra are similar in energy to a single shell based on a snub cube. *Polyoma virus* capsid protein VP_1 forms a left-handed snub cube from alkaline solution in the absence of the genome.

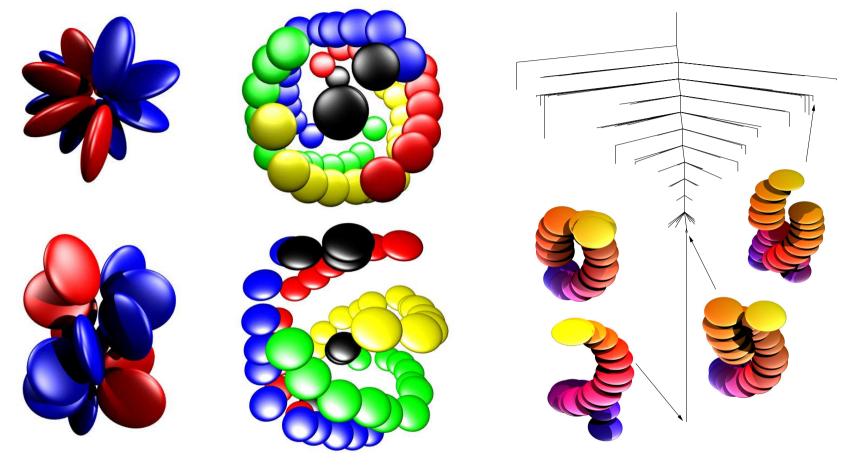
A Molecular Machine



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.

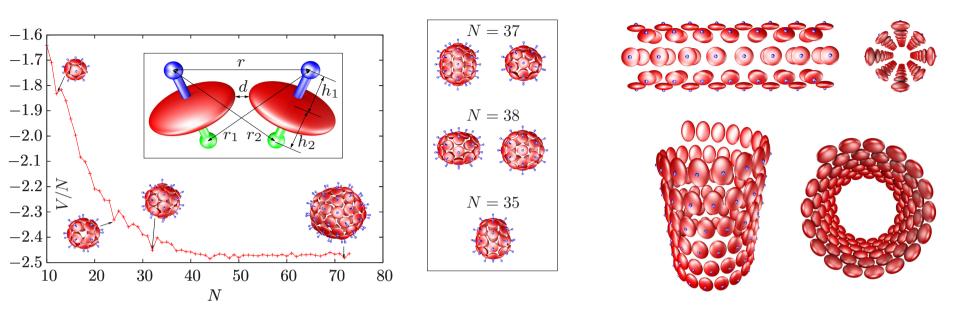
Clusters of Ellipsoids (Phys. Rev. Lett., 99, 086106, 2007)



Clusters of discoids bound by the Paramonov-Yaliraki potential exhibit helical global minima when the dimer has a shifted stacked configuration.

The corresponding energy landscapes generally have single funnel topologies for both single and multiple strand helices.

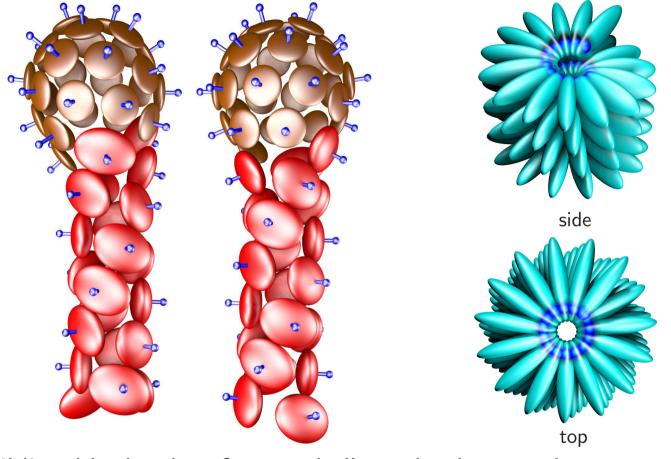
Emergent Behaviour from Simple Building Blocks



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

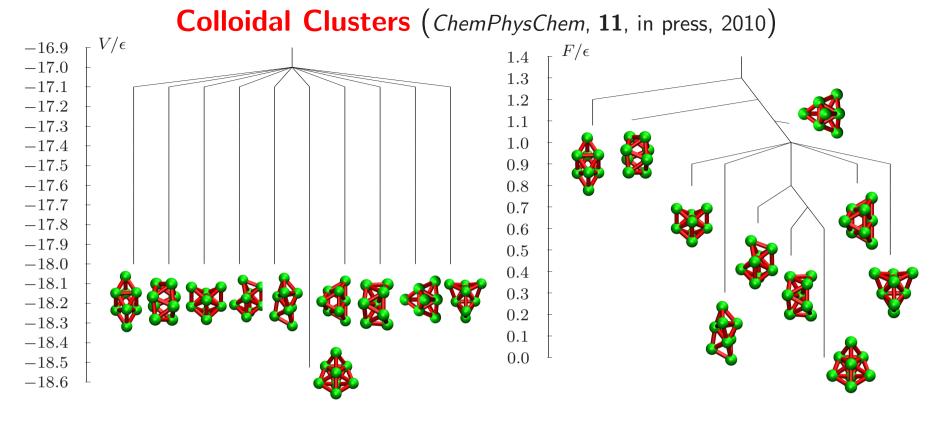
Global minima for the oblate ellipsoids include icosahedra for N=12, 32 and 72 (T=1,3 and 7), the snub cube observed for polyoma virus capsids at N=24, and conical, biaxial, prolate, and oblate shells at other sizes.

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



Mixing 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 structures of colloidal clusters formed from polystyrene microspheres by depletion interactions have recently been resolved using optical microscopy.

For these short-ranged interactions the isomer populations are quantitatively reproduced by the partition functions $Z_{\alpha}=e^{-\beta V_{\alpha}}/o_{\alpha}(\beta h\overline{\nu}_{\alpha})^{3N-6}=e^{-F_{\alpha}/k_BT}$, calculated for a Morse potential with range parameter $\rho=30$. The order of the point group, o_{α} , plays a key role, as shown for 8-particle clusters, above.

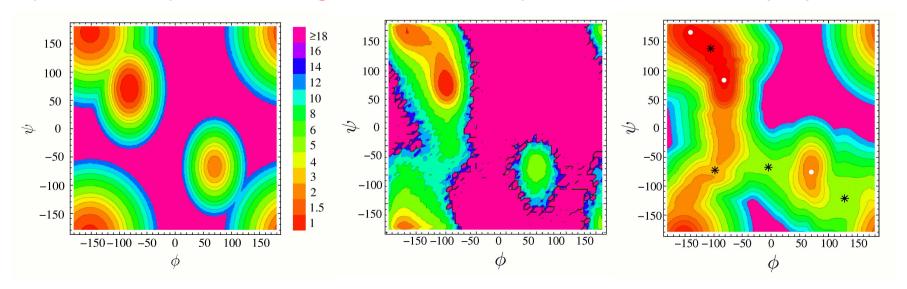
A Modified Superposition Approach (Chem. Phys. Lett., 466, 105, 2008)

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

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

where A_i is a weighted sum of order parameter derivatives.

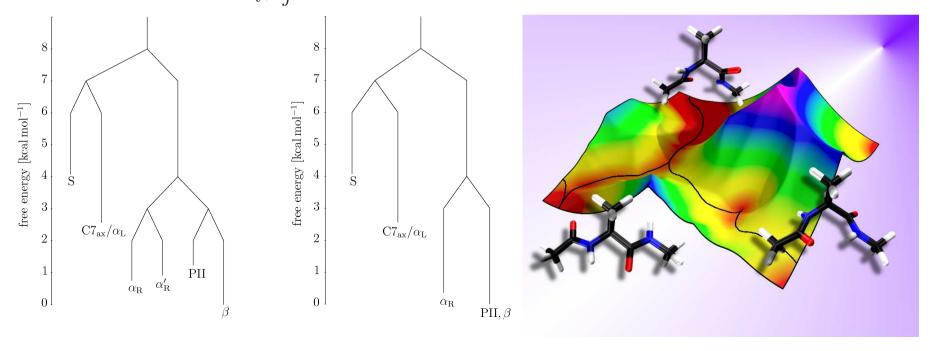
Free energy surfaces for alanine dipeptide (CHARMM22/vacuum) from superposition, replica exchange, and reaction path Hamiltonian superposition:



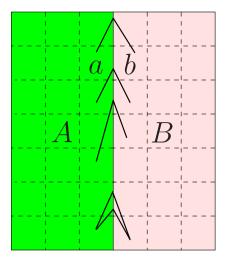
The 'filling in' problem for barrier regions in low-dimensional projections due to overlapping distributions can be avoided using disconnectivity graphs.

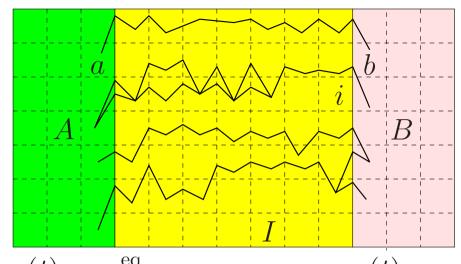
The effect of regrouping for a barrier threshold of $3 \, \text{kcal/mol}$ is shown below for AMBER(ff03)/GB^{OCB} (left) and compared with the CHARMM22/vacuum surface (right). Free energy of group J: $F_J(T) = -kT \ln \sum_{j \in J} Z_j(T)$ with

$$F_{LJ}^{\dagger}(T) = -kT \ln \sum_{l \leftarrow i} Z_{lj}^{\dagger}(T), \quad \text{and} \quad k_{LJ}(T) = \frac{kT}{h} e^{-\left[F_{LJ}^{\dagger}(T) - F_J(T)\right]/kT}.$$



Discrete Path Sampling (*Mol. Phys.*, **100**, 3285, 2002).





$$\text{no intervening minima} \quad \frac{p_a(t)}{p_{a'}(t)} = \frac{p_a^{\text{eq}}}{p_{a'}^{\text{eq}}} \quad \dot{p}_i(t) = 0 \quad \quad \frac{p_b(t)}{p_{b'}(t)} = \frac{p_b^{\text{eq}}}{p_{b'}^{\text{eq}}}$$

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_h^{eq} :

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

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.

Discrete path sampling is a framework for growing databases of stationary points that are relevant to global kinetics (Int. Rev. Phys. Chem., 25, 237, 2006).

A hierarchy of expressions can be obtained for the rate constants:

$$k_{AB}^{SS} = \frac{1}{p_B^{eq}} \sum_{b \in B} \frac{C_b^A p_b^{eq}}{\tau_b}, \quad k_{AB}^{NSS} = \frac{1}{p_B^{eq}} \sum_{b \in B} \frac{C_b^A p_b^{eq}}{t_b}, \quad k_{AB} = \frac{1}{p_B^{eq}} \sum_{b \in B} \frac{p_b^{eq}}{T_{Ab}}.$$

 τ_b , t_b and \mathcal{T}_{Ab} are the mean waiting times for a transition from b to an adjacent minimum, to any member of $A \cup B$, and to the A set, with $\tau_b \leq t_b \leq \mathcal{T}_{Ab}$.

 k_{AB} is formally exact within a Markov assumption for transitions between the states, which can be regrouped. Additional approximations come from incomplete sampling, and the densities of states and transition state theory used to describe the local thermodynamics and kinetics.

Calculating k_{AB} using diagonalisation, successive overrelaxation (SOR), or kinetic Monte Carlo (KMC) can become unfeasible for large databases.

Kinetic Analysis by Graph Transformation (JCP, 124, 234110, 2006)

The graph transformation procedure is non-stochastic and non-iterative. Minima, x, are progressively removed, while the branching probabilities and waiting times in adjacent minima, $\beta \in \Gamma$, 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:

$\overline{kT/K}$	$\Delta F_{ m barrier}$	$N_{ m min}$	$N_{ m ts}$	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	

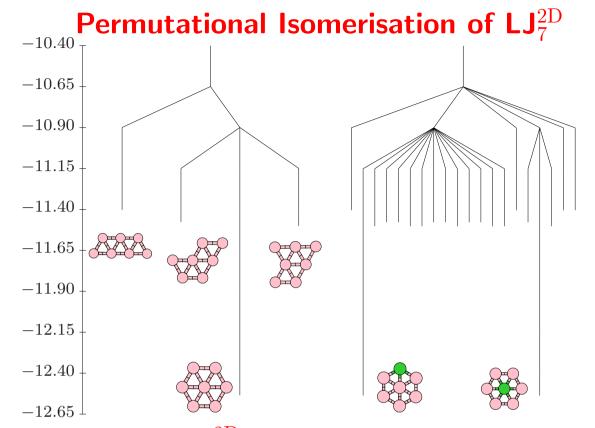
Finding Stationary Points

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

Transition states: single-ended searches use hybrid eigenvector-following (*Phys. Rev. B*, **59**, 3969, 1999; *Chem. Phys. Lett.*, **341**, 185, 2001), double-ended searches use the doubly-nudged elastic band approach (*J. Chem. Phys.*, **120**, 2082, 2004).

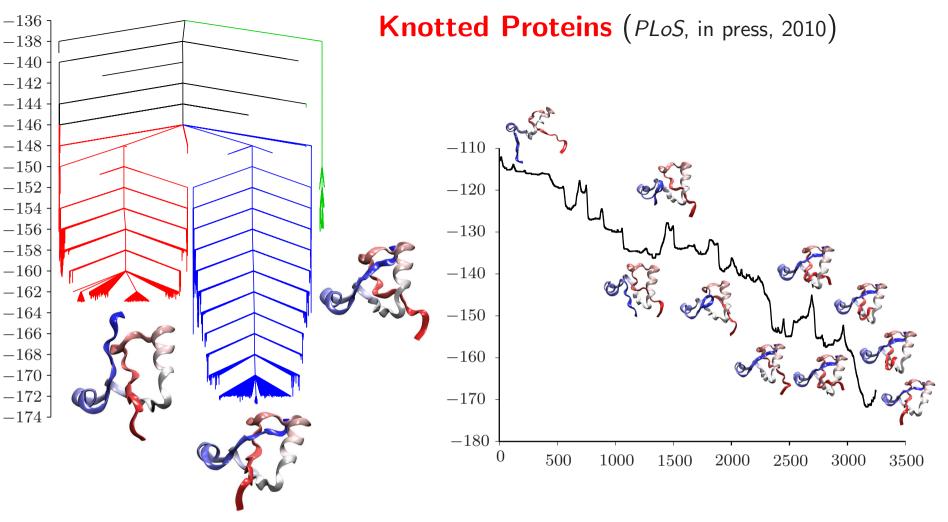
These methods can be combined with electronic structure calculations.

- Hydrocarbon dissociation on Pt $\{110\}$ (1×2) (J. Chem. Phys., 126, 044710, 2007). For ethane, low barriers (0.3 to 0.4 eV) are found for initial formation of ethene and ethylidene, medium barriers (0.7 to 1.1 eV) are found for dehydrogenation of C_2H_4 , and higher barriers for further dehydrogenation.
- Ammonia synthesis and dissociation on Fe{211}: the Haber-Bosch process. We predict that atomic nitrogen can be hydrogenated above around 340 K, with ammonia being evolved at temperatures above 570-670 K.



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.

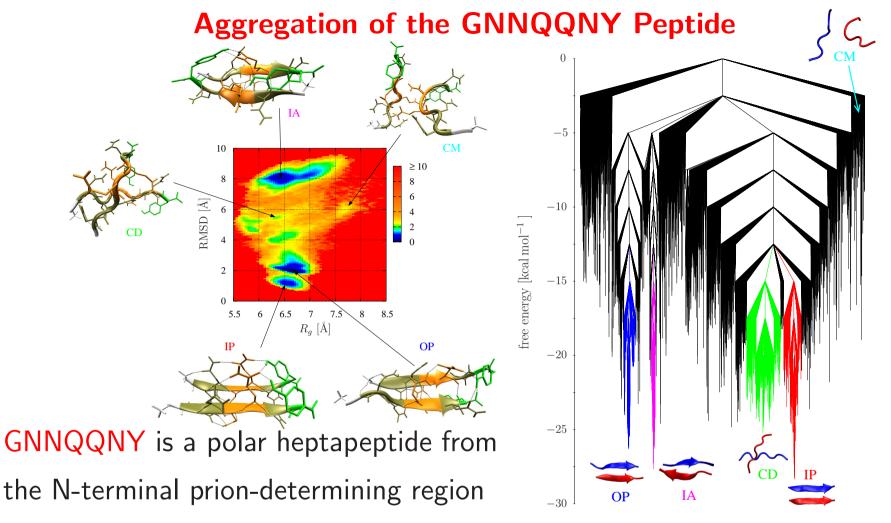
The fastest ten paths contribute about 74% of the total rate constant at $kT/\epsilon=0.05$. Various combinations of diamond-square-diamond rearrangements make significant contributions.



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

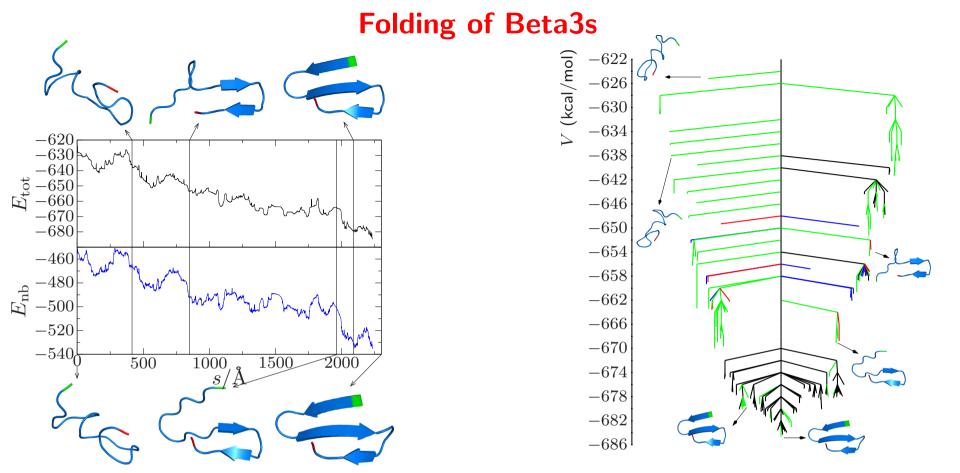
The folding pathway has two slipknot-type steps for a truncated (residues 78–135) Gō model representation using an associated memory Hamiltonian.

The estimated rate constant is between 0.04 and $0.4 \,\mathrm{s}^{-1}$.



of the 685 residue yeast prion protein Sup35. Dimer free energy minima are in-register parallel, IP, off-register parallel, OP, and antiparallel, IA, sheets.

Dimer formation rates are estimated as milliseconds to seconds. Time scale for interconversion between dimers ranges from hours to days at 298 K.



Beta3s is a designed 20-residue peptide with a three-stranded antiparallel β -sheet. Folding with CHARMM19/EEF1 involves early formation of the C-terminal hairpin followed by docking of the N-terminal strand.

Mean first passage time is 300 ns at 298 K, consistent with other calculations and the experimental upper bound of 4000 ns (*J. Phys. Chem. B*, **112**, 8760, 2008).

A Connection Between Dynamics and Thermodynamics

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

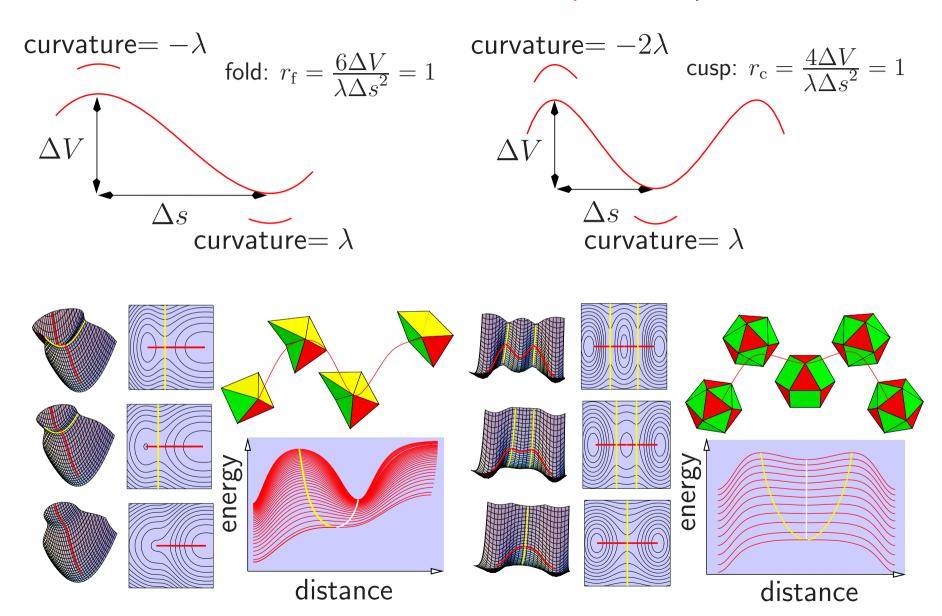
The organisation of families of PES's as a function of parameters in the potential is determined by the stationary points that possess additional zero Hessian eigenvalues, known as non-Morse points.

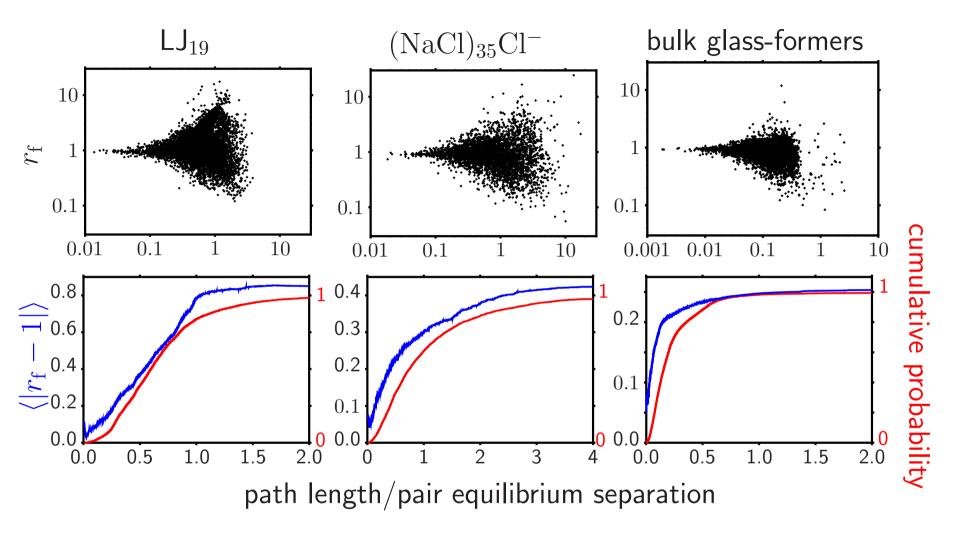
Catastrophe theory provides a local representation of the PES around non-Morse points as a function of both atomic coordinates and parameters.

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, $r_{\rm f}$ remains close to unity for many pathways in both model clusters and bulk, providing an explanation for Hammond's postulate.