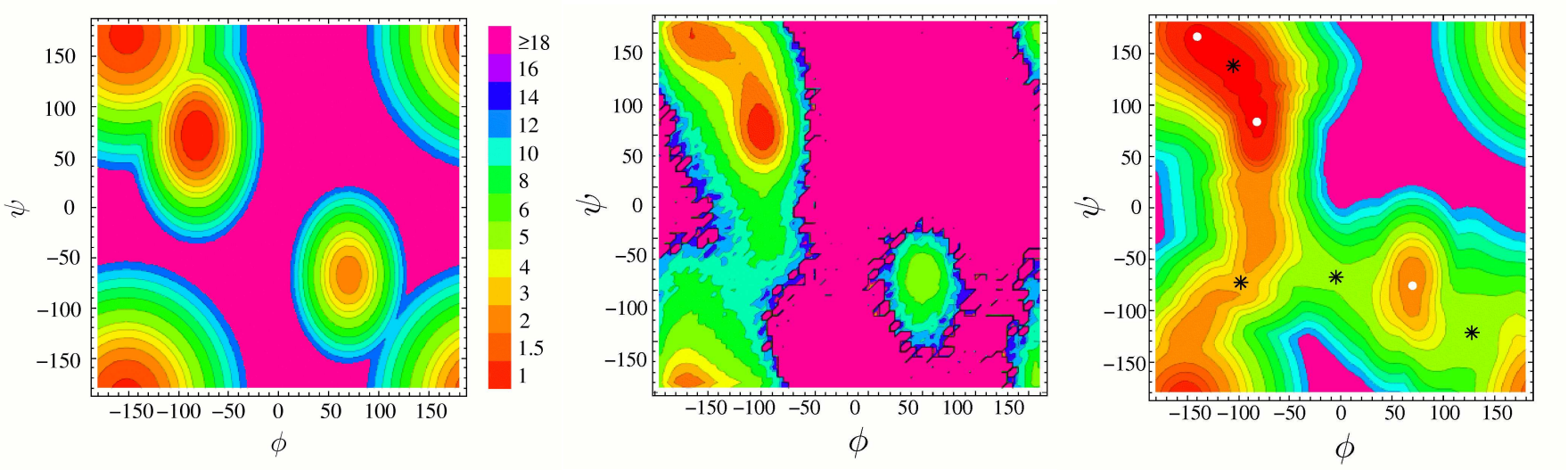
Energy Landscapes: Clusters, Glasses, and Biomolecules

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)
- Superposition approach for thermodynamics (*Chem. Phys. Lett.*, **466**, 105, 2008)
- Discrete path sampling for global kinetics (*Mol. Phys.*, **100**, 3285, 2002)

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



The Reaction Path Hamiltonian Superposition Approach

The total partition function as a function of order parameter a is constructed as a superposition of contributions from local minima, $Z_i(a,T)$, and configurations taken from the pathways that connect them, $Z_r^\dagger(a,T)$:

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

$$Z_r^\dagger(a,T) = \left(\frac{kT}{h}\right)^\kappa \frac{\delta_r \exp\left(-V_r^\dagger/kT\right)}{\left(\overline{\nu}_r^\dagger\right)^{\kappa-1} 2\pi kT\sqrt{A_r^\dagger}} \exp\left[-\frac{\left(a-a_r^\dagger\right)^2}{2kTA_r^\dagger}\right],$$

where $\overline{\nu}_i$ is the geometric mean of the normal mode frequencies, $\nu_{i,\gamma}$, V_i and a_i are the potential energy and order parameter for minimum i, $\kappa = 3N-6$, δ_r is a displacement, $\dagger$ labels transition states, and

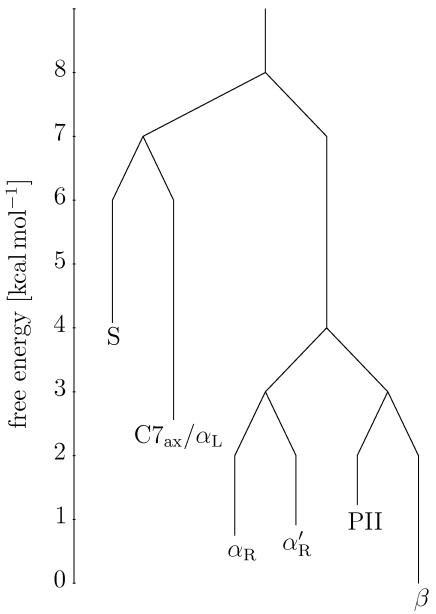
$$A_i = \sum_{\gamma=1}^{\kappa} \left[\left.\frac{\partial a(\mathbf{q}_i)}{\partial q_{i,\gamma}}\right|_{\mathbf{q}_i=\mathbf{0}} \frac{1}{2\pi\nu_{i,\gamma}}\right]^2.$$

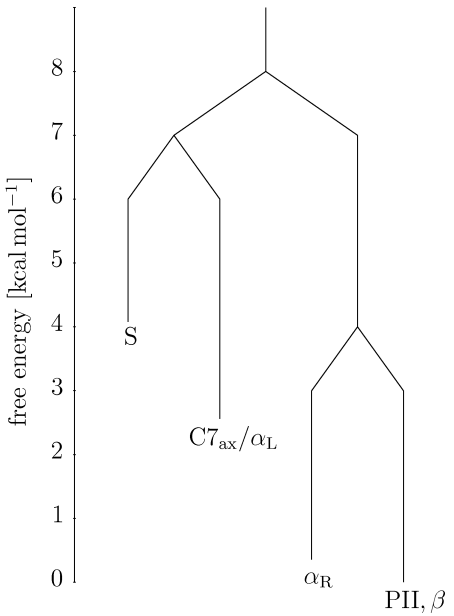
The method can be extended for projections onto additional order parameters.

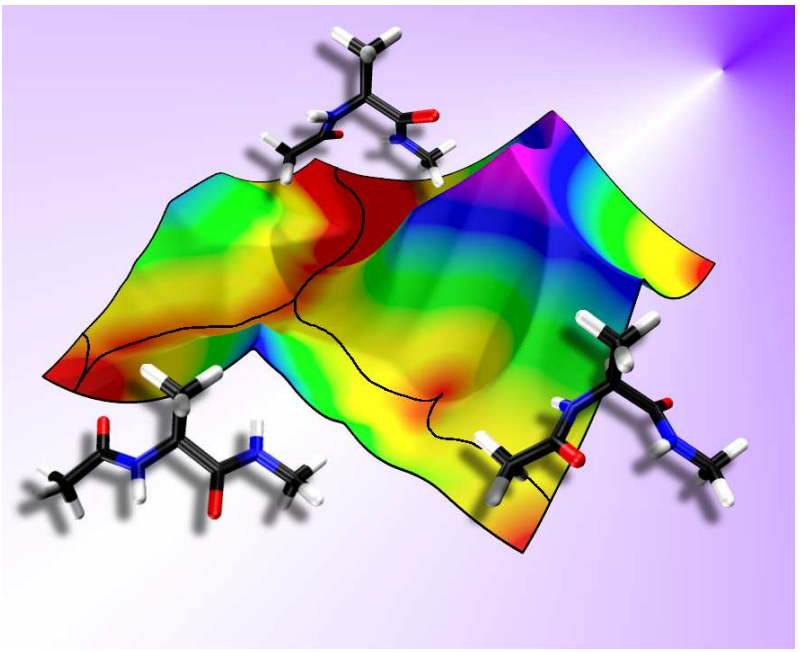
The 'filling in' problem for barrier regions in low-dimensional projections due to overlapping distributions can be avoided using disconnectivity graphs.

The effect of regrouping for a barrier threshold of 3 kcal/mol is shown below for AMBER(ff03)/GB$^{\mathrm{OCB}}$ (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

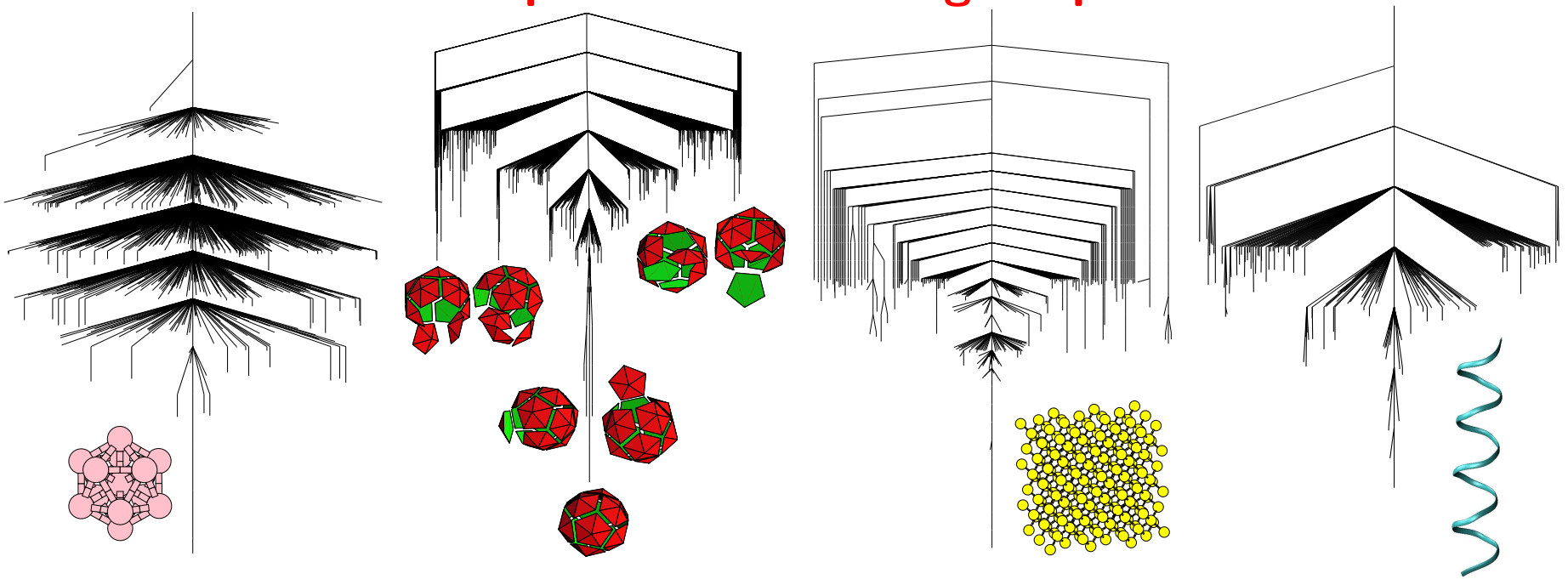
$$F^{\dagger}_{LJ}(T) = -kT \ln \sum_{l \leftarrow j} Z^{\dagger}_{lj}(T), \quad \text{and} \quad k_{LJ}(T) = \frac{kT}{h} e^{-\left[F^{\dagger}_{LJ}(T) - F_J(T)\right]/kT}.$$







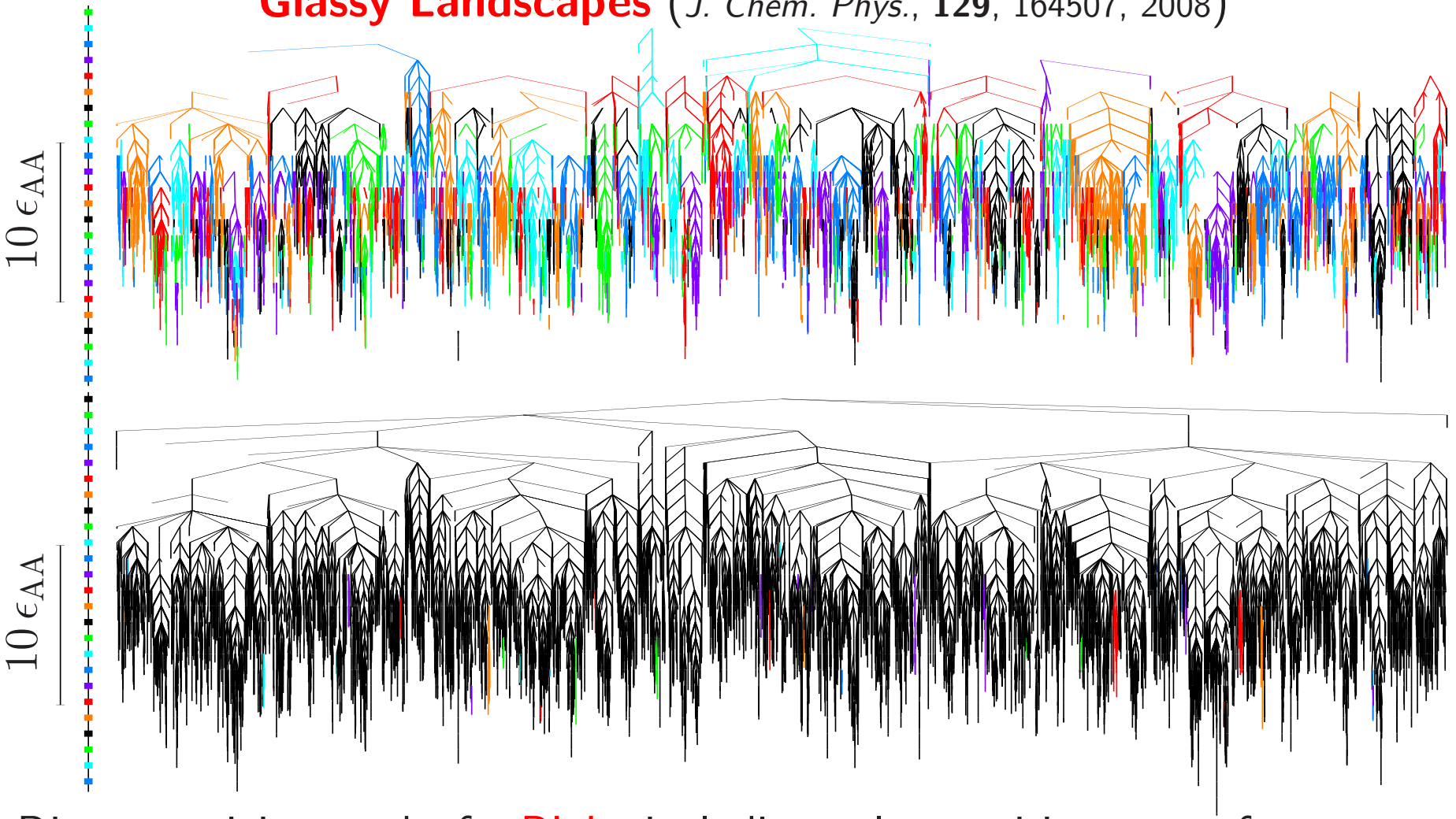
Landscapes with Funnelling Properties



The palm tree structure appears for a diverse range of landscapes, including the LJ_{13} cluster, icosahedral shells composed of pentagonal and hexagonal pyramids, crystalline (Stillinger-Weber) silicon, and the polyalanine ala_{16}.

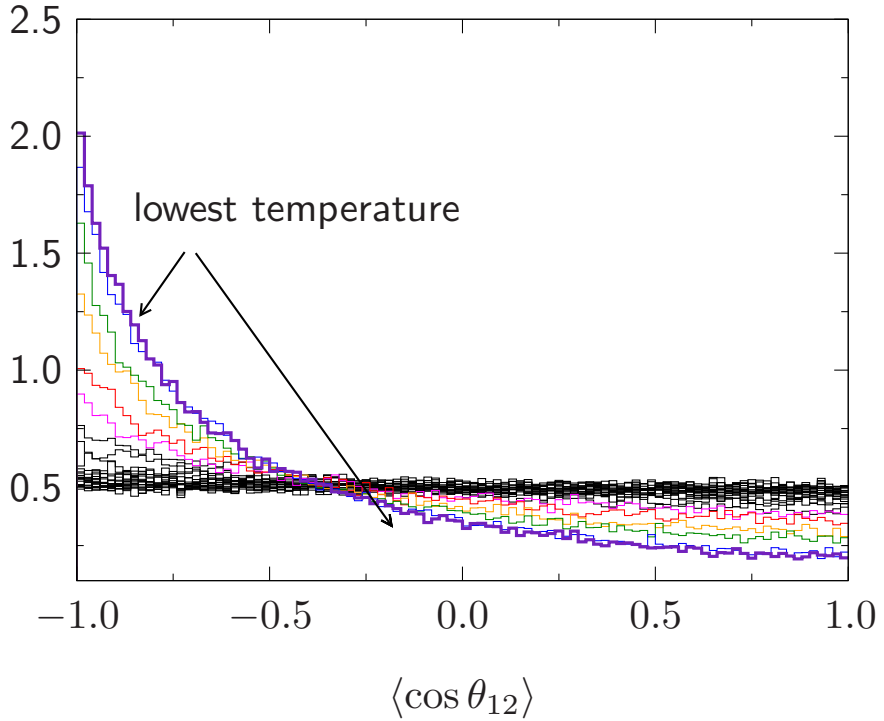
We associate this pattern with 'funnelling' properties, minimal frustration, large T_f/T_g, or hierarchical constraints. Such landscapes may guide the non-random searches that result in magic number clusters, crystallisation, self-assembly, and protein folding (*Phil. Trans. Roy. Soc. A*, **363**, 357, 2005).

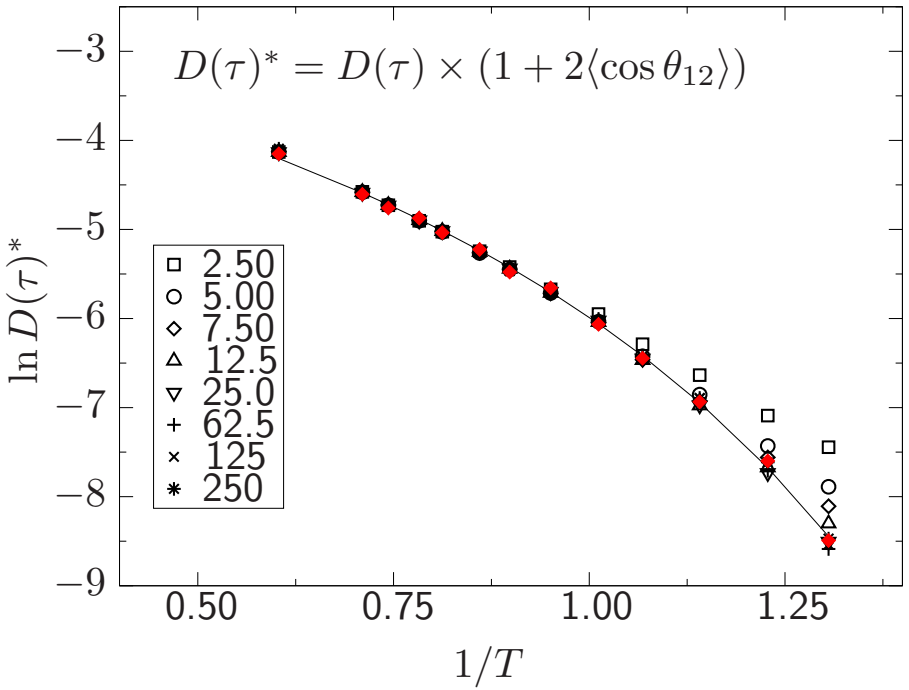
Glassy Landscapes (*J. Chem. Phys.*, **129**, 164507, 2008)



Disconnectivity graphs for BLJ_{60} 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.

'Fragile' Dynamics for the BLJ Solid (*Phys. Rev. B*, **74**, 134202, 2006)



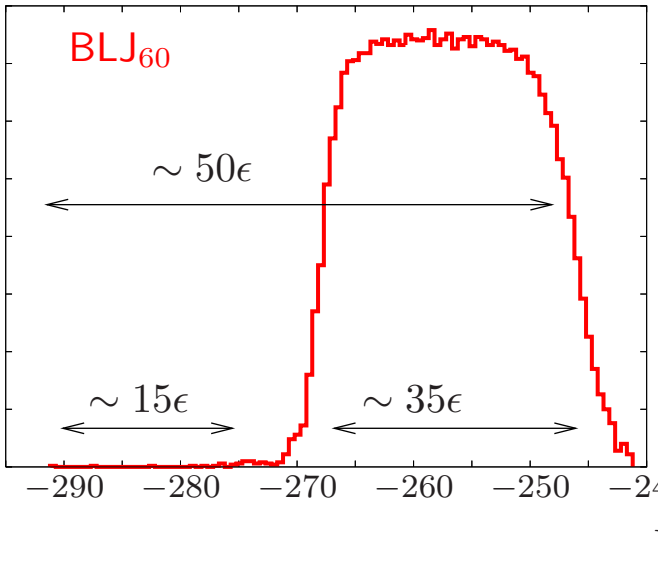


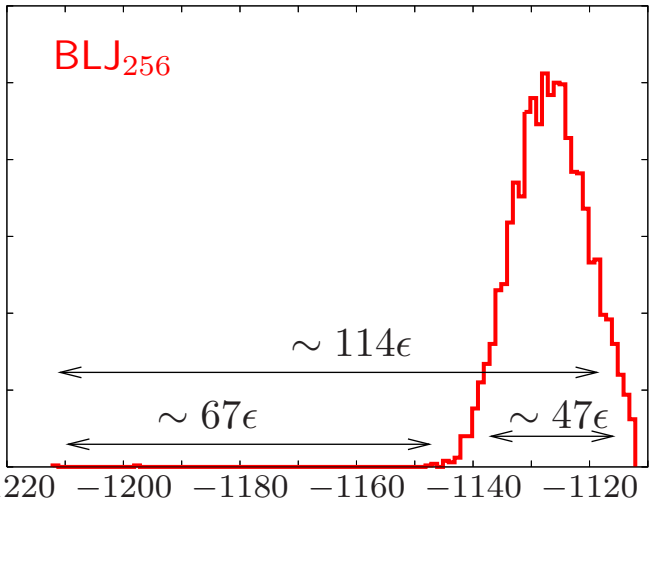
Diffusion constants calculated for increasing time windows, $D(\tau)$, show underlying Arrhenius behaviour for short time scales, converging to super-Arrhenius behaviour for intervals corresponding to local ergodicity.

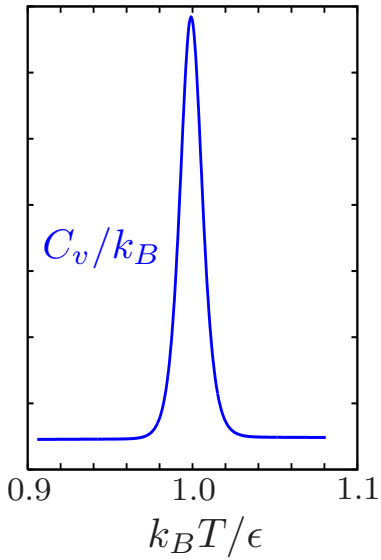
Displacements within successive time windows are negatively correlated on average (left). $D(\tau)$ can be corrected as $D(\tau)^* = D(\tau) \times (1 + 2\langle \cos\theta_{12} \rangle)$.

The dynamics are explicitly heterogeneous on non-ergodic time scales.

Thermodynamics of the BLJ Solid (*J. Chem. Phys.*, **127**, 044508, 2007)







Equilibrium thermodynamic properties such as $\Omega(E)$ and C_v (right, for BLJ_{320}) can be obtained from parallel tempering, despite the extensive energy gap in the probability distribution for local minima between the crystal and the amorphous states (left).

We were unable to converge Wang-Landau calculations because the probability of return to the crystal from the amorphous region is so low, even when a two-dimensional scheme based on $\Omega(E, Q_6)$ was used.

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

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

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

where $\mathbf{I}$ is the identity matrix, and $\widetilde{\mathbf{p}}$ 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: $\mathbf{R}_k \equiv \dfrac{\partial \mathbf{R}}{\partial p_k} = \dfrac{p_k \sin\theta}{\theta}\,\widetilde{\mathbf{p}}^2 + (1-\cos\theta)(\widetilde{\mathbf{p}}_k\widetilde{\mathbf{p}} + \widetilde{\mathbf{p}}\widetilde{\mathbf{p}}_k) + \dfrac{p_k\cos\theta}{\theta}\widetilde{\mathbf{p}} + \sin\theta\,\widetilde{\mathbf{p}}_k, \quad \text{with} \quad \widetilde{\mathbf{p}}_1 = \dfrac{1}{\theta^3}\begin{pmatrix} 0 & p_1p_3 & -p_1p_2 \\ -p_1p_3 & 0 & p_1^2-\theta^2 \\ p_1p_2 & \theta^2-p_1^2 & 0 \end{pmatrix}$

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

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

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

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

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

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

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

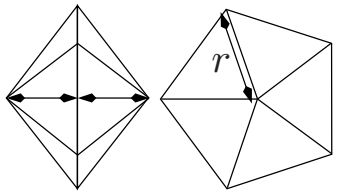
$$\frac{\partial^2 U_{ij}^{IJ}}{\partial r_k^I \partial r_l^J} = f_2(r_{ij})r_{ij,k}r_{ij,l}\epsilon_{IJ} + f_1(r_{ij})\delta_{kl}\epsilon_{IJ},$$

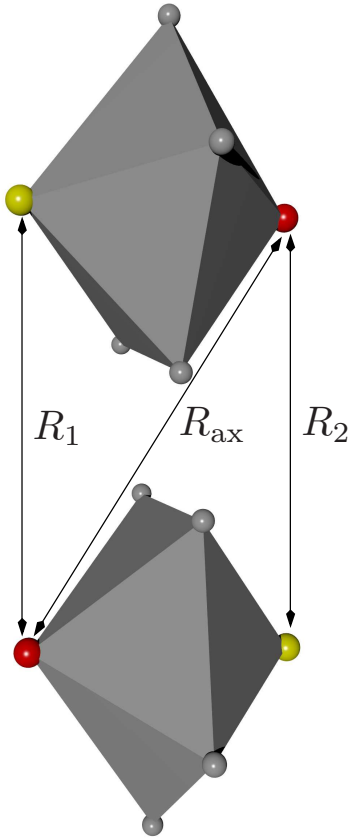
$$\frac{\partial^2 U_{ij}^{IJ}}{\partial p_k^I \partial p_l^J} = f_2(r_{ij})(\mathbf{r}_{ij}\cdot\mathbf{R}_k^I\mathbf{r}_i^0)(\mathbf{r}_{ij}\cdot\mathbf{R}_l^I\mathbf{r}_i^0)\delta_{IJ} - f_2(r_{ij})(\mathbf{r}_{ij}\cdot\mathbf{R}_k^I\mathbf{r}_i^0)(\mathbf{r}_{ij}\cdot\mathbf{R}_l^J\mathbf{r}_j^0)(1-\delta_{IJ})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

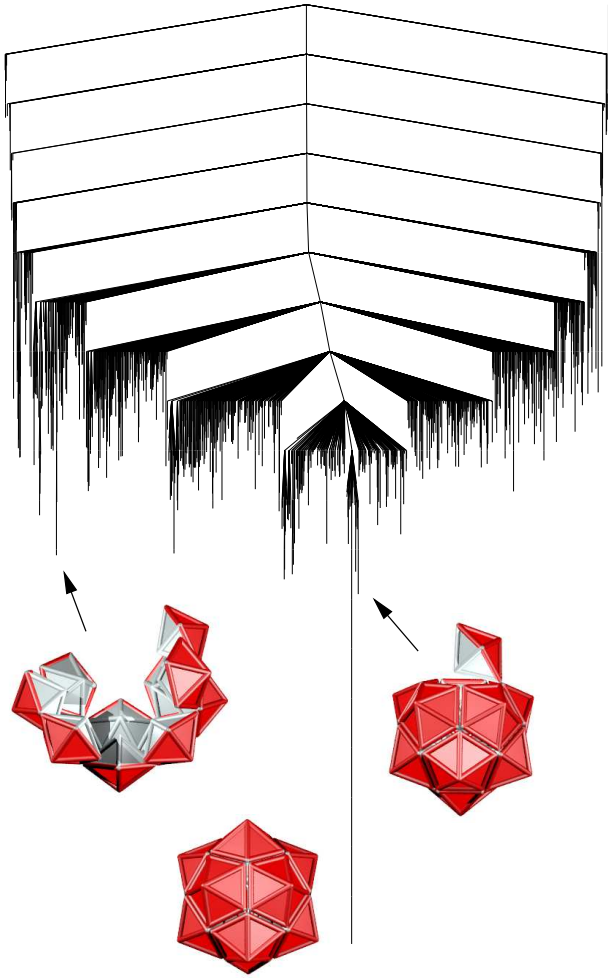
$$\frac{\partial^2 U_{ij}^{IJ}}{\partial r_k^I \partial p_l^J} = f_2(r_{ij})(\mathbf{r}_{ij}\cdot\mathbf{R}_l^I\mathbf{r}_i^0)r_{ij,k}\delta_{IJ} - f_2(r_{ij})(\mathbf{r}_{ij}\cdot\mathbf{R}_l^J\mathbf{r}_j^0)r_{ij,k}(1-\delta_{IJ}) + f_1(r_{ij})[\mathbf{R}_k^I\mathbf{r}_i^0]_l\delta_{IJ} - f_1(r_{ij})[\mathbf{R}_l^J\mathbf{r}_j^0]_l(1-\delta_{IJ}).$$

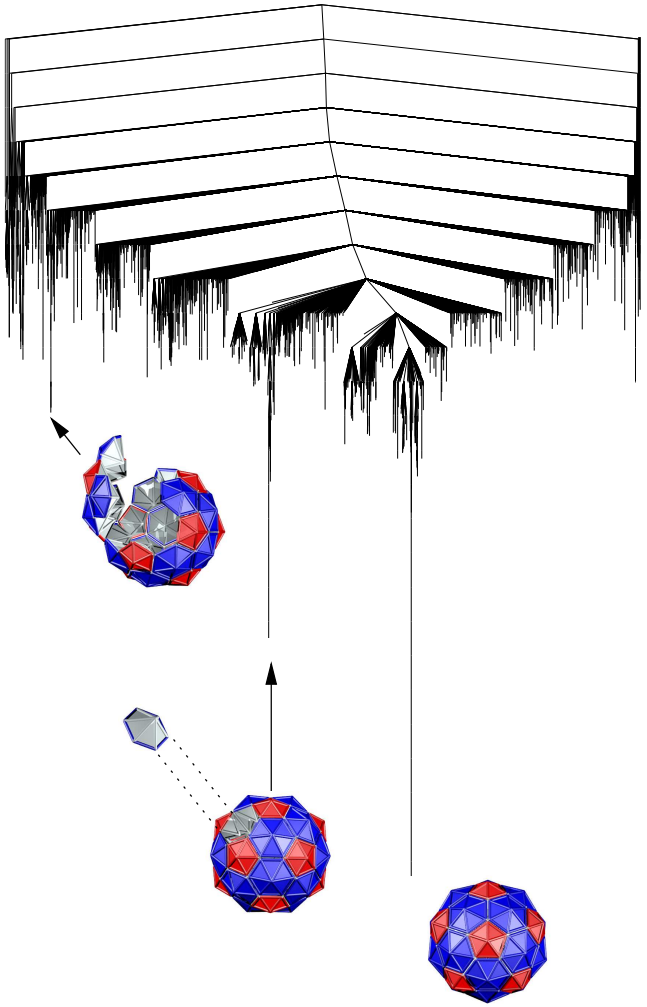
where $f_1(r_{ij}) = f'_{ij}(r_{ij})/r_{ij}$, $f_2(r_{ij}) = f'_1(r_{ij})/r_{ij}$, $\epsilon_{IJ} = 1$ for $I = J$ and $\epsilon_{IJ} = -1$ for $I \neq J$, and δ_{IJ} is the Kronecker delta.

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





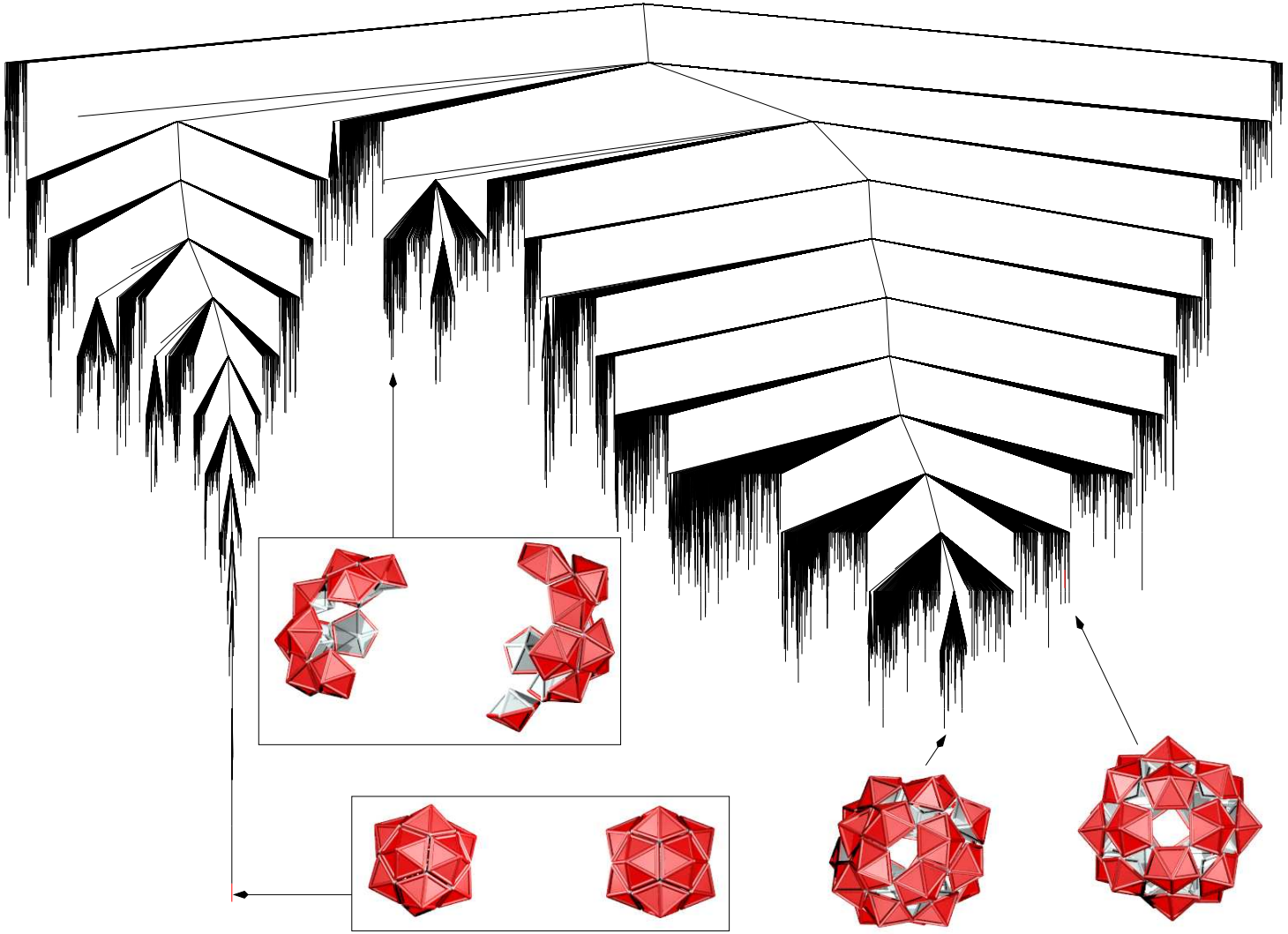




Palm tree disconnectivity graphs with I_h global minima are found for $T = 1$ and $T = 3$ shells constructed from pentagonal and hexagonal pyramids.

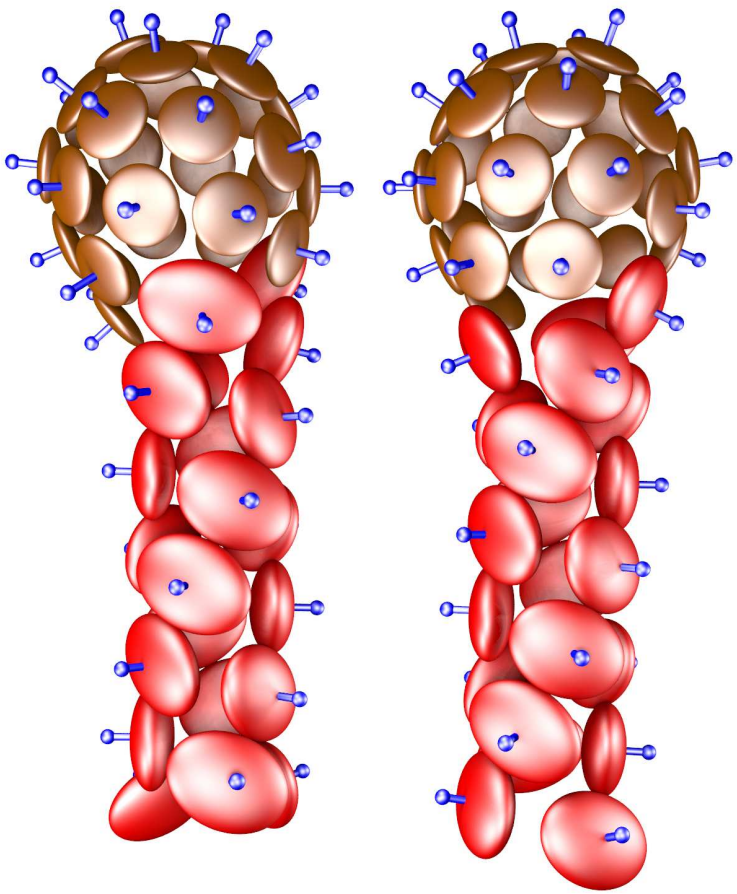
Landscapes of this form are associated with good structure-seekers.

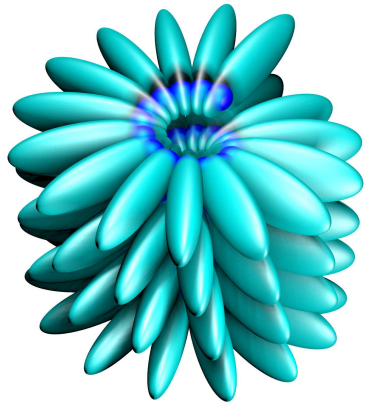
24 Pentagonal Pyramids



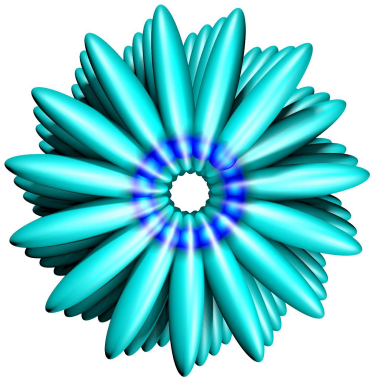
For the same parameters two $T = 1$ icosahedra are similar in energy to a single shell based on a snub cube. *Polyoma virus* capsid protein VP_1 forms a left-handed snub cube from alkaline solution in the absence of the genome.

Modelling Mesoscopic Structures





side

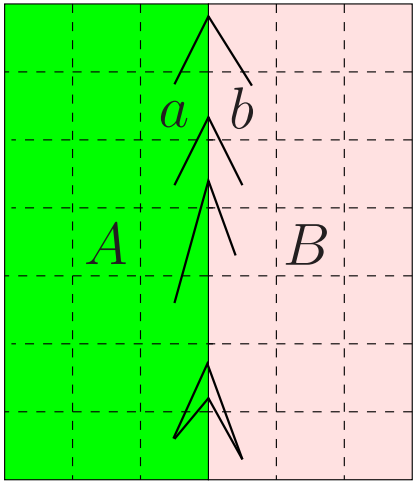


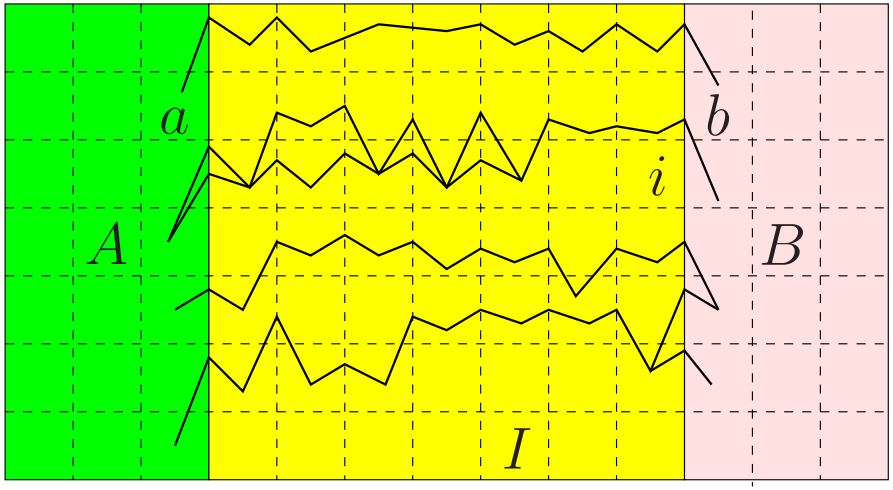
top

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.

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





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

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, $p_b^{\rm eq}$:

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

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.

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:

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

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)

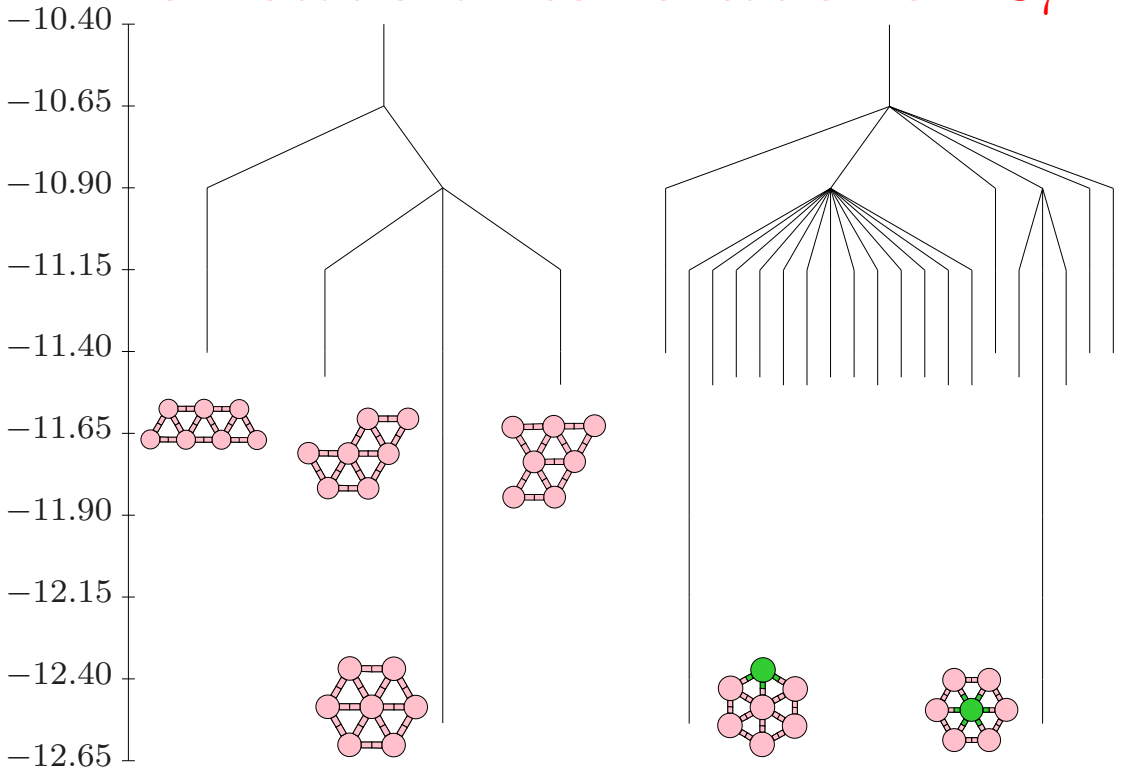
The graph transformation procedure is non-stochastic and non-iterative. Minima, x, are progressively removed, while the branching probabilities and waiting times in adjacent minima, $\beta \in \Gamma$, are renormalised:

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

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

kT/K	$\Delta F_{\mathrm{barrier}}$	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	

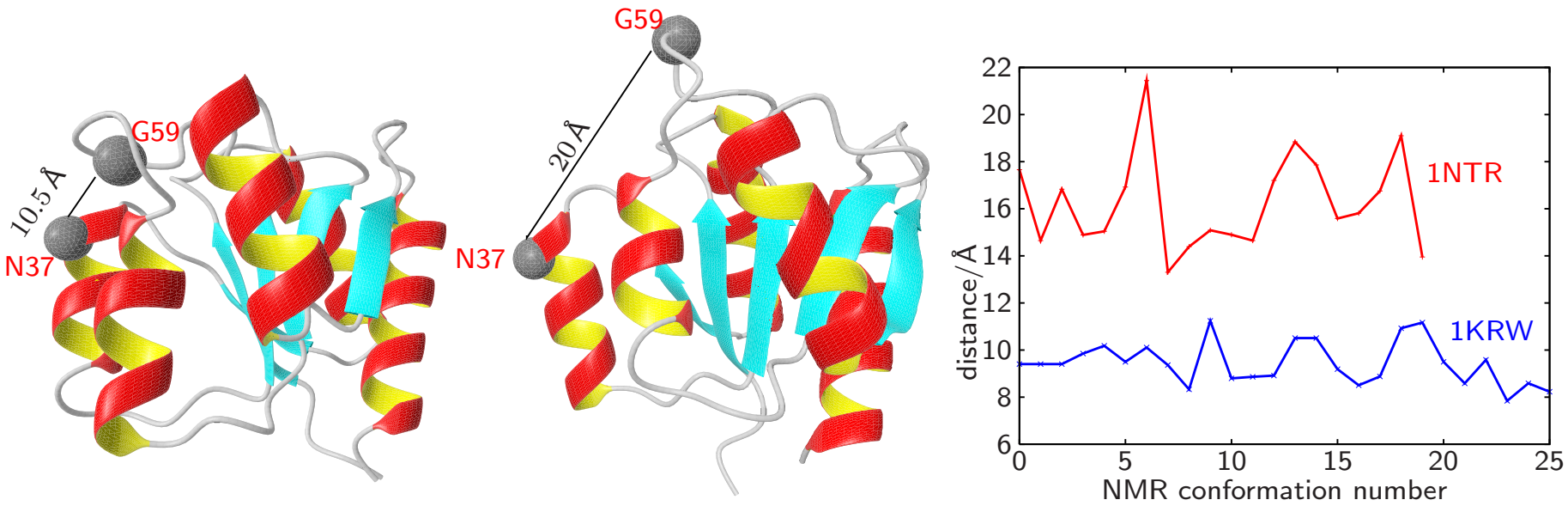
Permutational Isomerisation of LJ_7^{2D}



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.

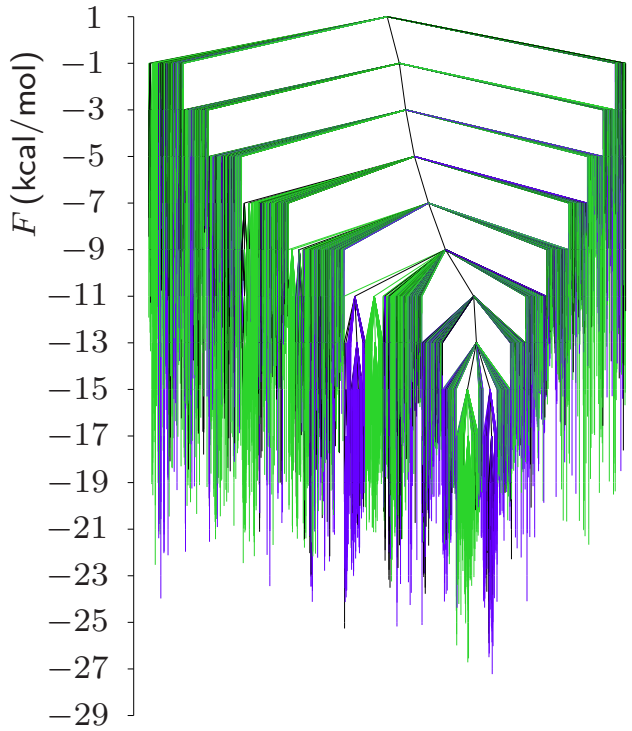
The fastest ten paths contribute about 74% of the total rate constant at $kT/\epsilon = 0.05$. Various combinations of diamond-square-diamond rearrangements make significant contributions.

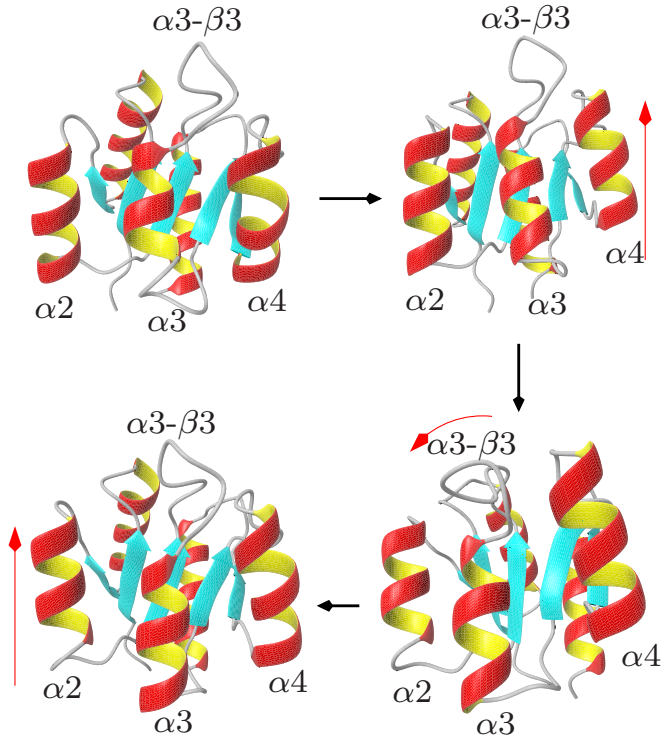
Conformational Change in NtrC (*J. Phys. Chem. B*, **112**, 2456, 2008)



Nitrogen regulatory protein C (NtrC) belongs to the bacterial signal transduction pathway. Phosphorylation at aspartate-54 plays a key role in activating the protein, and is associated with a relatively large conformational change.

The separation of residues N37 and G59 (left) distinguishes NMR conformations for the open (inactive) and closed (active) forms (right). After phosphorylation, equilibrium involves more than 99% of the active form, suggesting a two-state, allosteric behaviour, rather than induced fit.



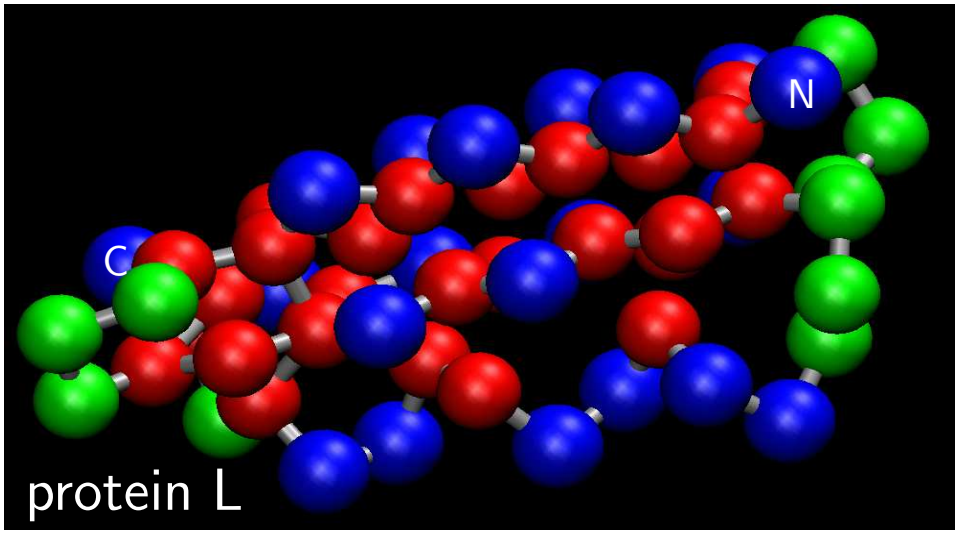


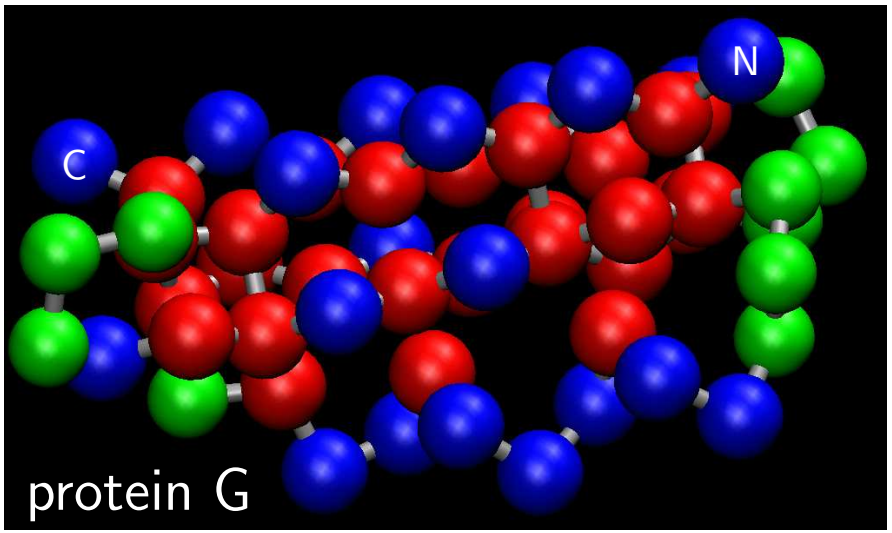
For the lowest two sets of active (blue) and inactive (green) structures the calculated transition rate is about 200 s^{-1} (CHARMM19/EEF1).

α4 shifts up relative to α2 and α3 before the α3-β3 loop closes.

Ring stacking of Y94 and Y101 stabilises the closed conformation. Charged residues R3 and K46 become exposed to solvent and the unhinging of α2 and α5 causes the surface area to increase in the closed form.

BLN Models of Protein L and Protein G



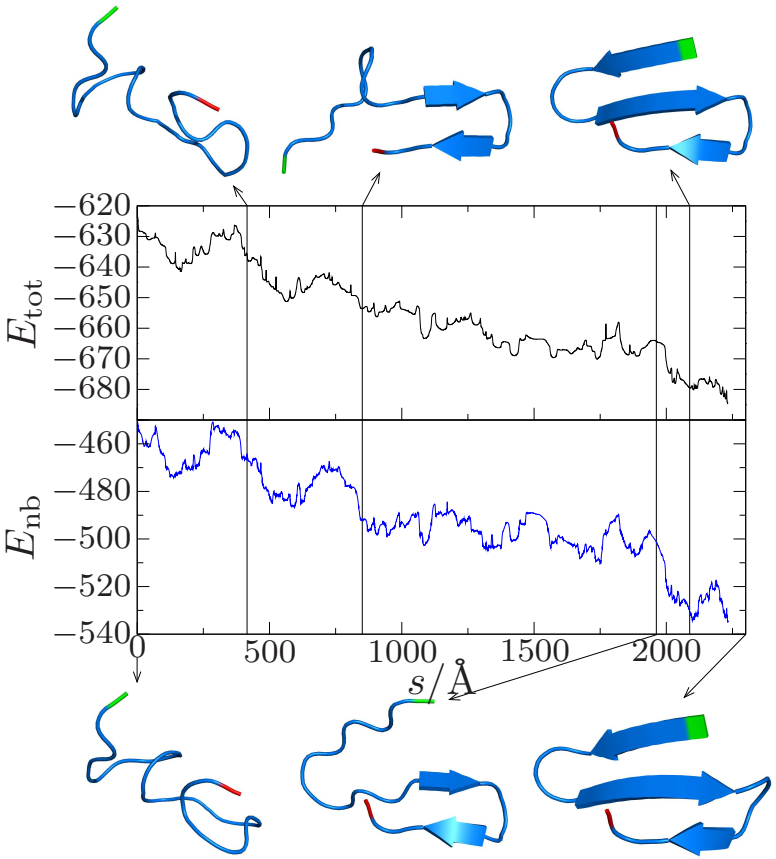


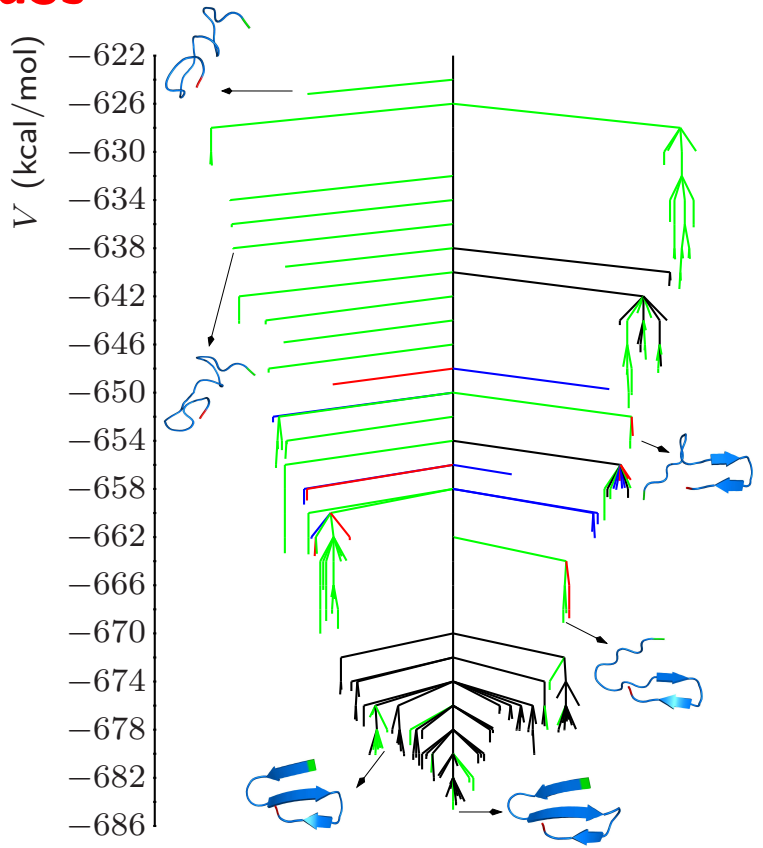
The global minima for BLN sequences designed for proteins L and G are easily located by basin-hopping. Both have a central α-helix packed against a four-stranded β-sheet composed of two β-hairpins.

Protein L forms the N-terminal hairpin 1 first, followed by the C-terminal hairpin 2, but the order is reversed for protein G, which may also exhibit an early intermediate (Head-Gordon et al.), in agreement with experiment.

Initial DPS results for the folding rate constant in protein L agree with previous Langevin dynamics to within an order of magnitude.

Folding of Beta3s

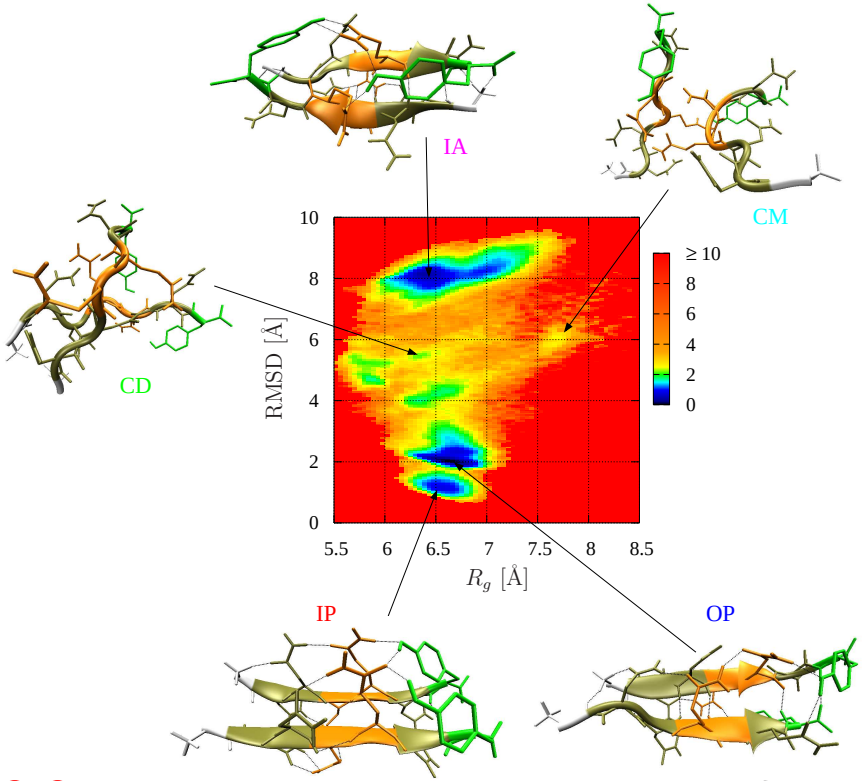


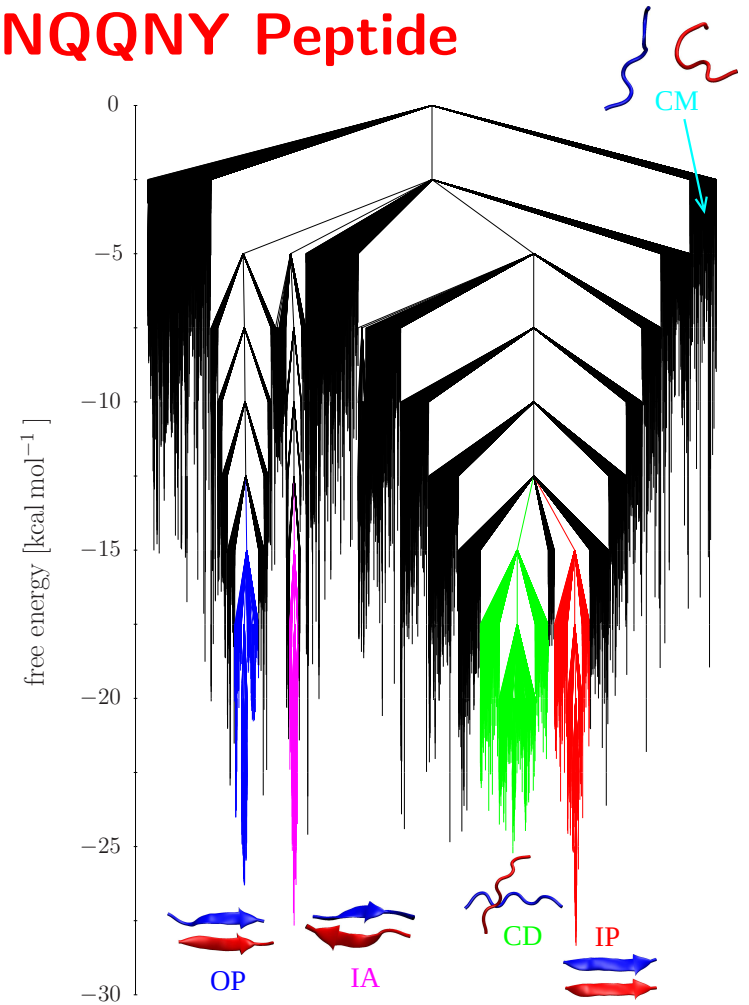


Beta3s is a designed 20-residue peptide with a three-stranded antiparallel β-sheet. Folding with CHARMM19/EEF1 involves early formation of the C-terminal hairpin followed by docking of the N-terminal strand.

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).

Aggregation of the GNNQQNY Peptide

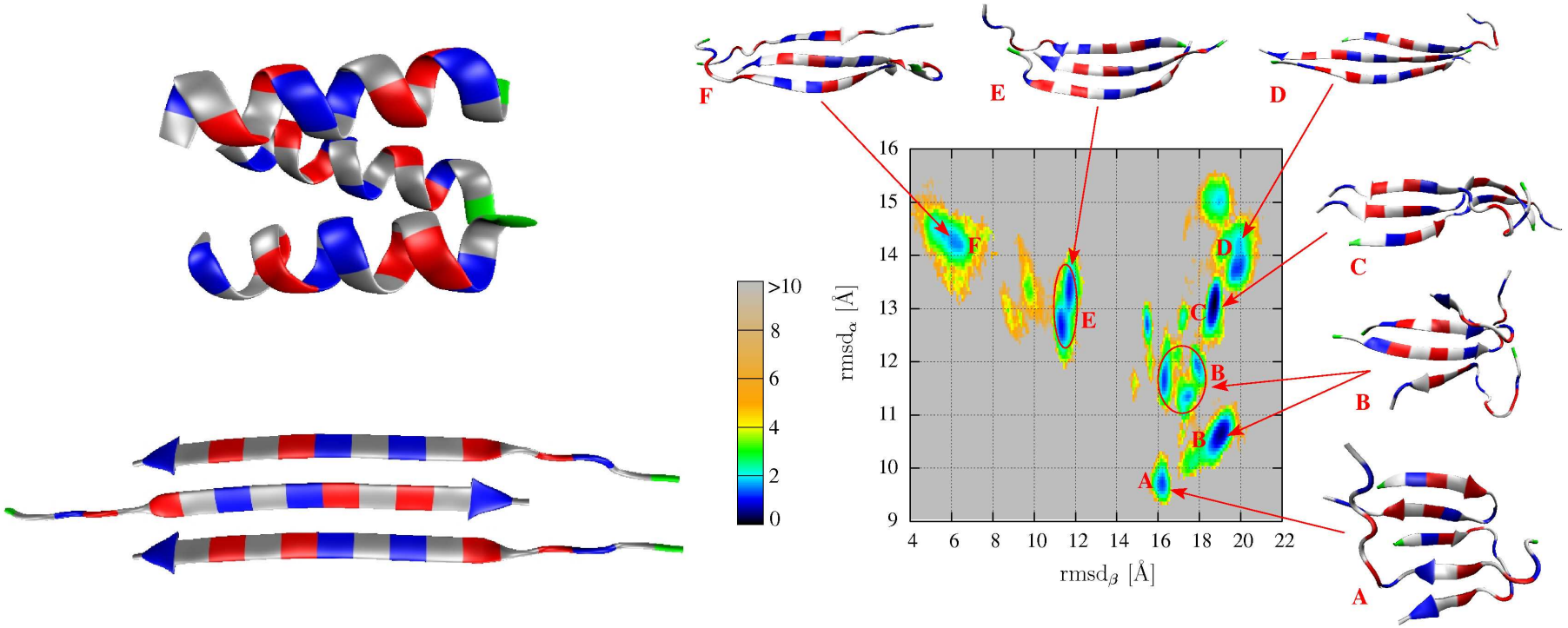




GNNQQNY is a polar heptapeptide from the N-terminal prion-determining region of the 685 residue yeast prion protein Sup35. Dimer free energy minima are in-register parallel, IP, off-register parallel, OP, and antiparallel, IA, sheets.

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

Amyloid Formation in ccβ (*J. Phys. Chem. B*, **112**, 9998, 2008)



The designed peptide ccβ adopts a trimeric coiled-coil structure, but transforms to give amyloid fibrils on raising the temperature to 310 K.

CHARMM19/EEF1 free energy surface includes the β-sheet structure deduced from experiment. Paths between the α-helical trimer and β-sheet structure involve over 1000 transition states.