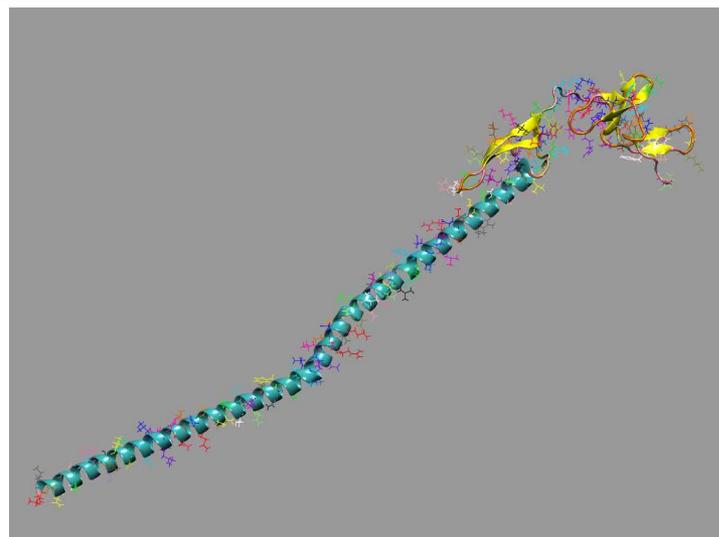
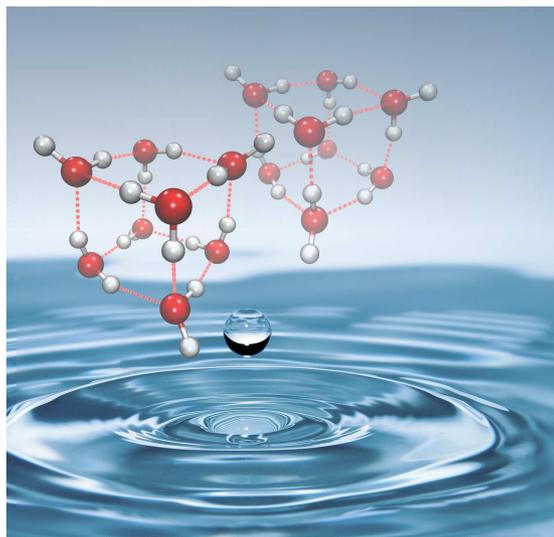


Energy landscapes: Structure, Dynamics, and Thermodynamics

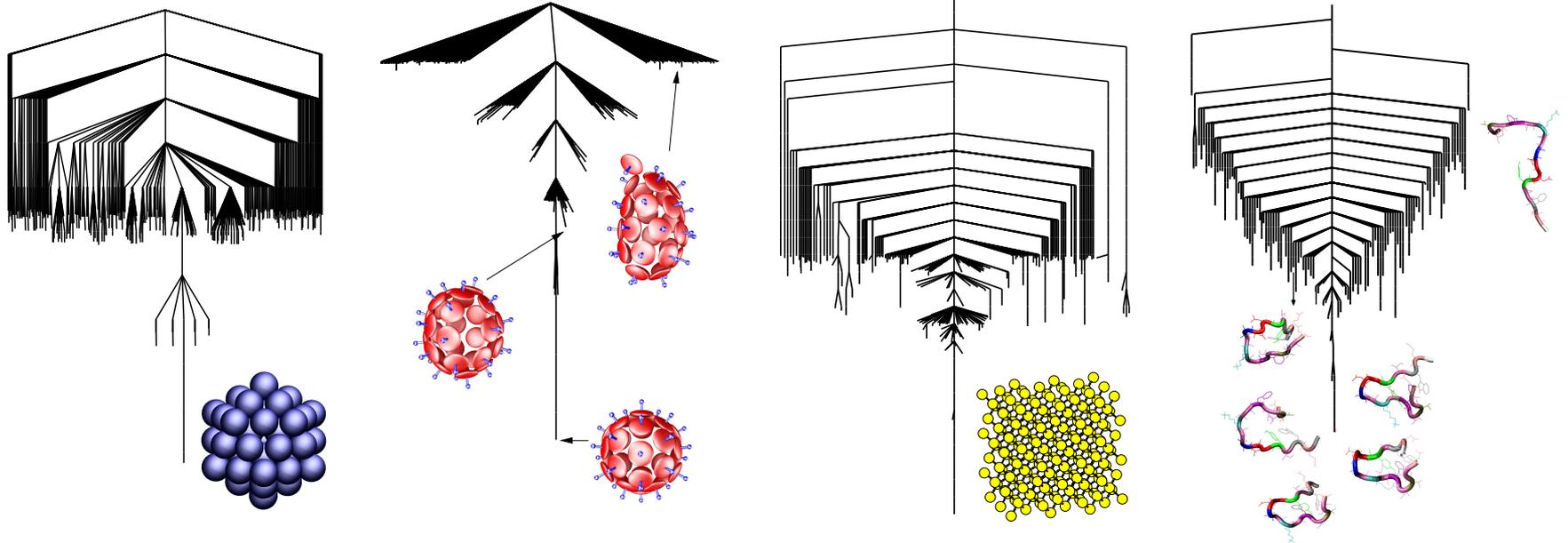
This conceptual and computational framework is based on **stationary points** (minima and transition states) of the **potential energy surface**.

- **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** and **pathways** can be located. **Larger** systems require appropriate **sampling**.

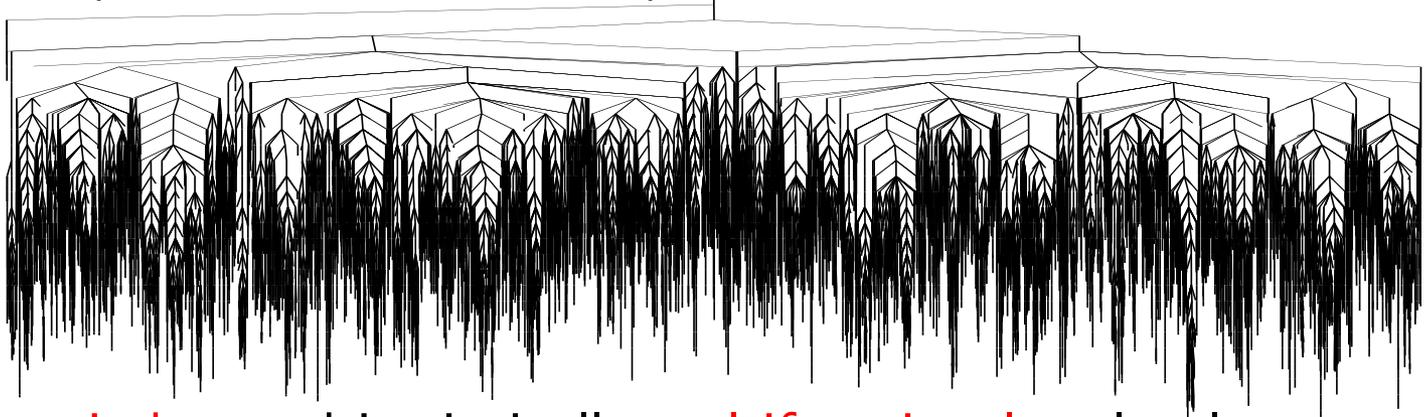


Self-Organisation is Encoded in Single Funnel Landscapes



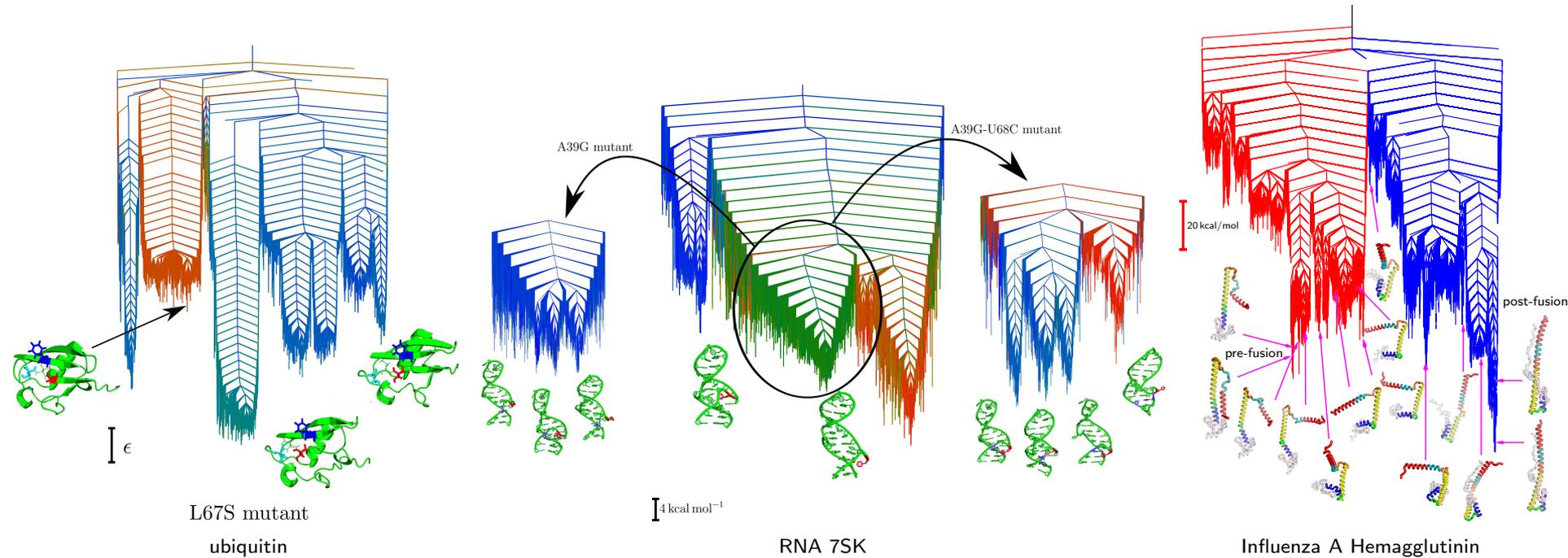
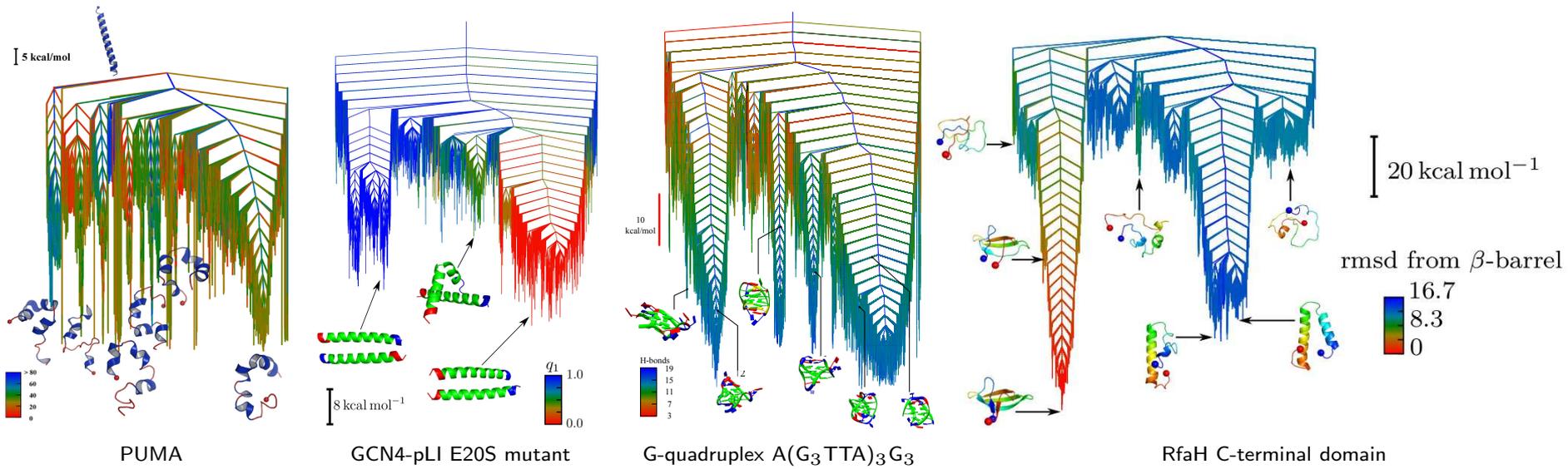
(Above) Energy landscapes for systems with **self-organising** properties.

(Below) A **glassy** landscape. (*Phil. Trans. Roy. Soc. A*, **363**, 357, 2005).

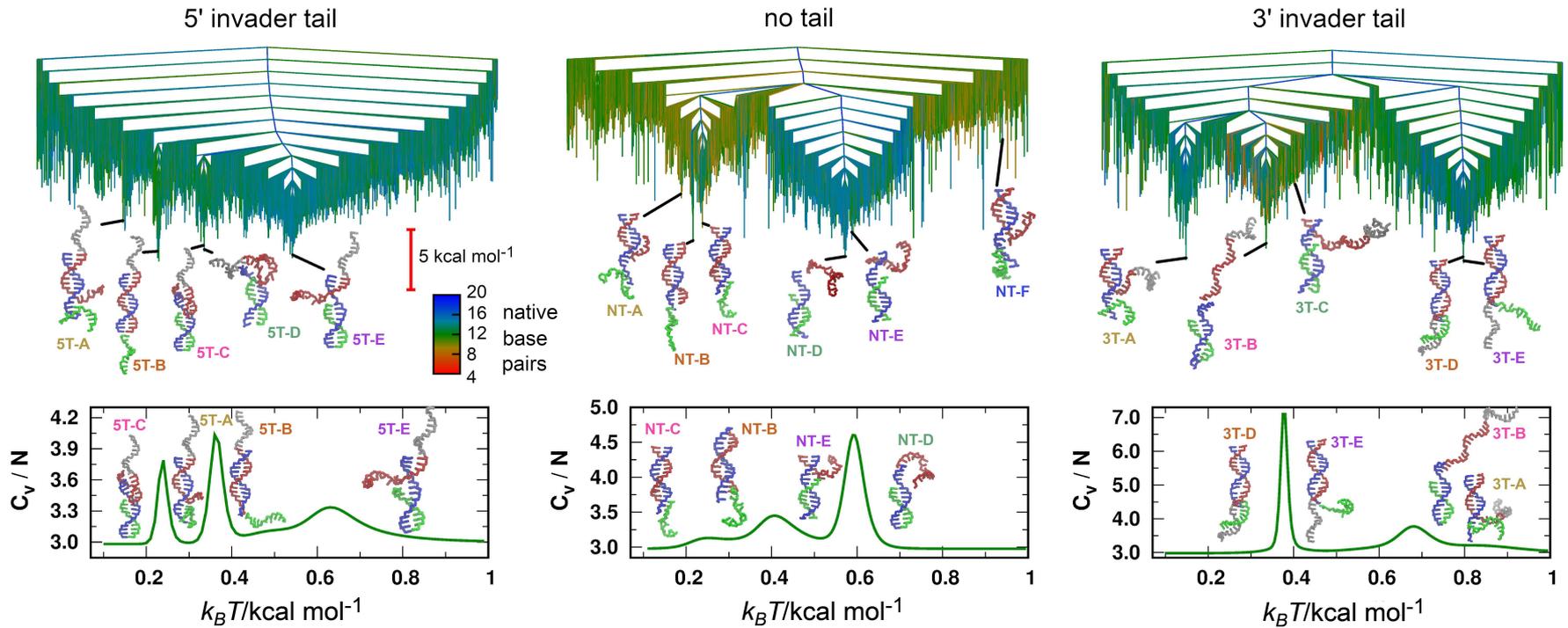


Molecular **switches** and intrinsically **multifunctional** molecules correspond to **multifunnel** landscapes (*Adv. Theory and Simulations*, **2**, 1800175, 2019).

Multifunnel Landscapes Can Encode Multifunctional Systems



Dynamic DNA Nanotechnology: Toehold Strand Displacement



Chemical reaction networks can be programmed using DNA strand displacement reactions. An **invading** strand displaces the **incumbent** strand from **substrate** via a sticky toehold domain.

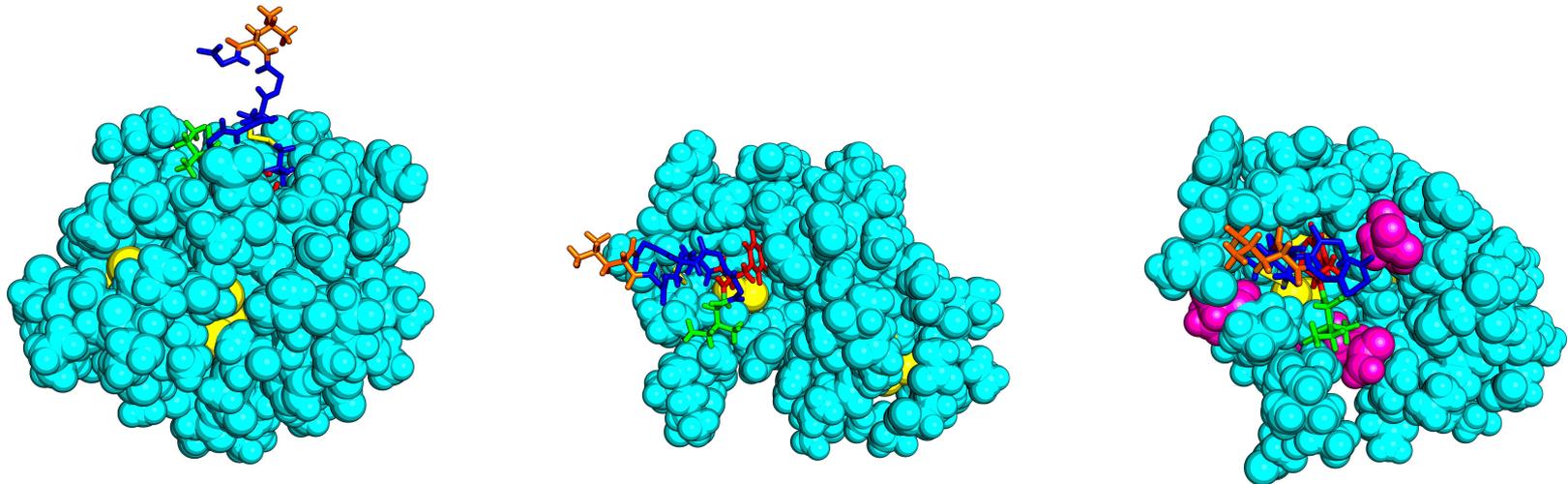
5' and 3' tails on the invader accelerate and retard the displacement, respectively, allowing kinetic modulation of a **catalytic circuit** for signal **amplification** or **pulse** generation.

Mutational Basin-Hopping (*J. Phys. Chem. Lett.*, **9**, 6169, 2018)

This generalised basin-hopping approach uses **biminima** for a peptide or nucleic acid **sequence** to optimise a target **property**.

A **binding energy** measure was optimised for **nonapeptide** hormone **vasopressin** and **carrier** protein **neurophysin II**, which also binds **oxytocin**.

MBH runs were started from the **crystal structure** of neurophysin II/oxytocin:

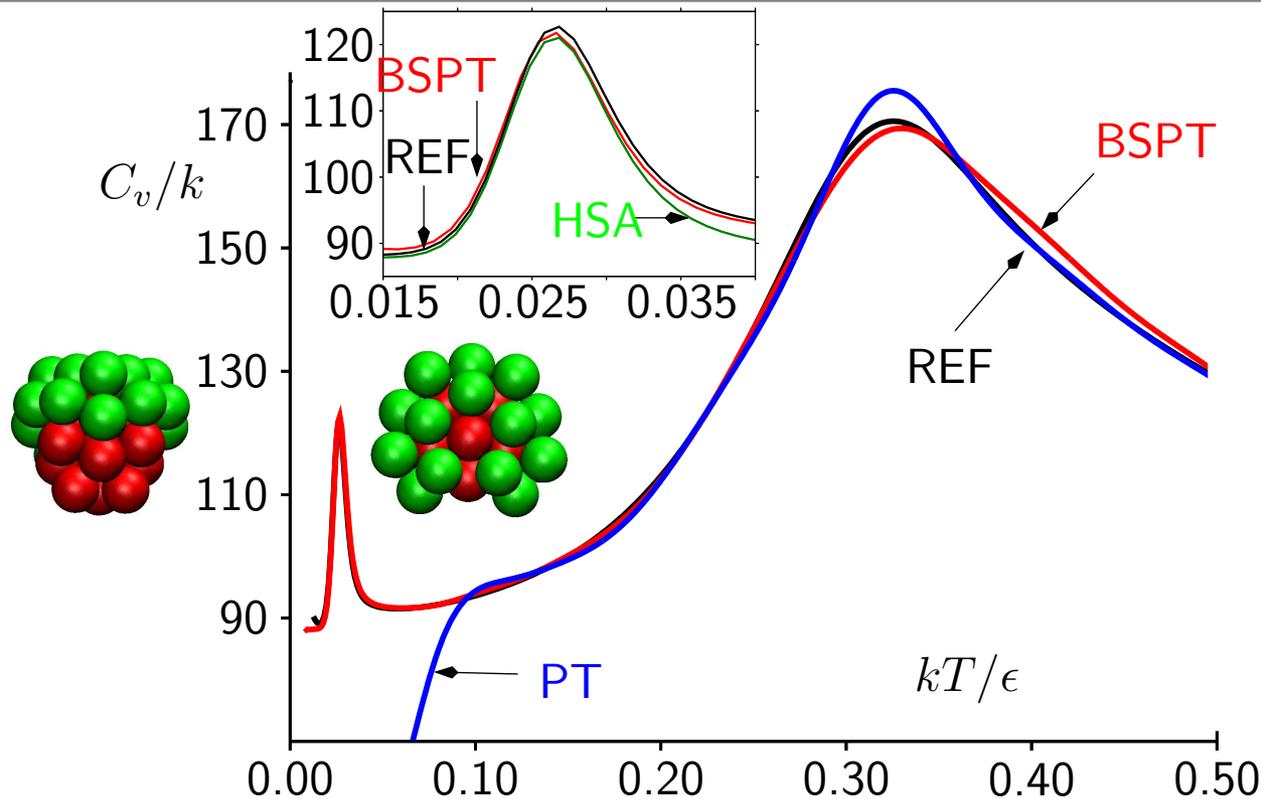


Vasopressin is located after overall **I3F** and **L8R** mutations.

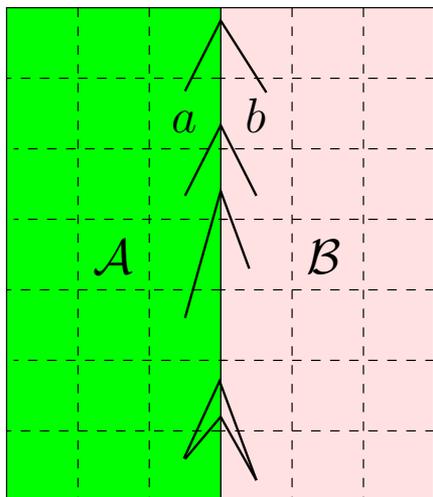
Mutations of key **sites** in the **carrier** (right) improved binding to **oxytocin**.

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

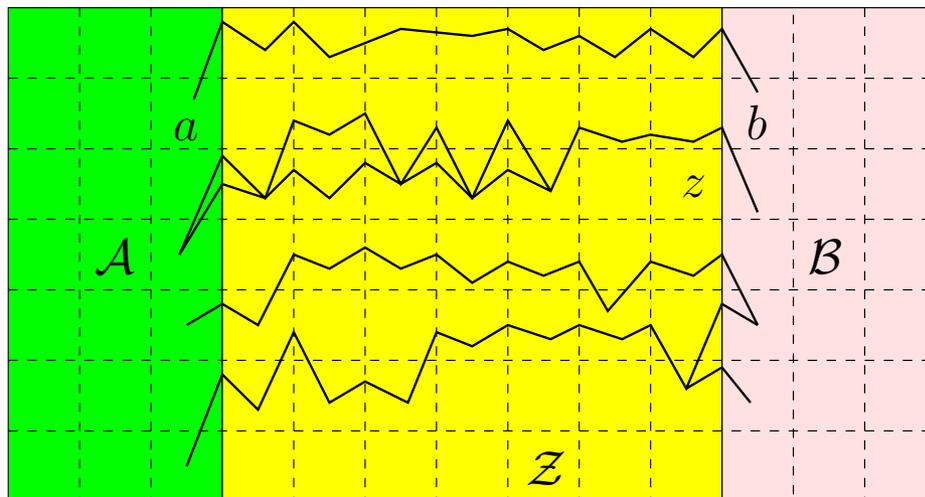
Broken ergodicity is treated by basin-hopping and the configuration space corresponding to high temperature is sampled by parallel tempering. A two-dimensional anharmonic form is used to combine the density of states. Accurate thermodynamics were obtained for the solid-solid phase transition in LJ_{31} in 21.8 minutes compared to 110.5 hours for parallel tempering.



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



no intervening minima

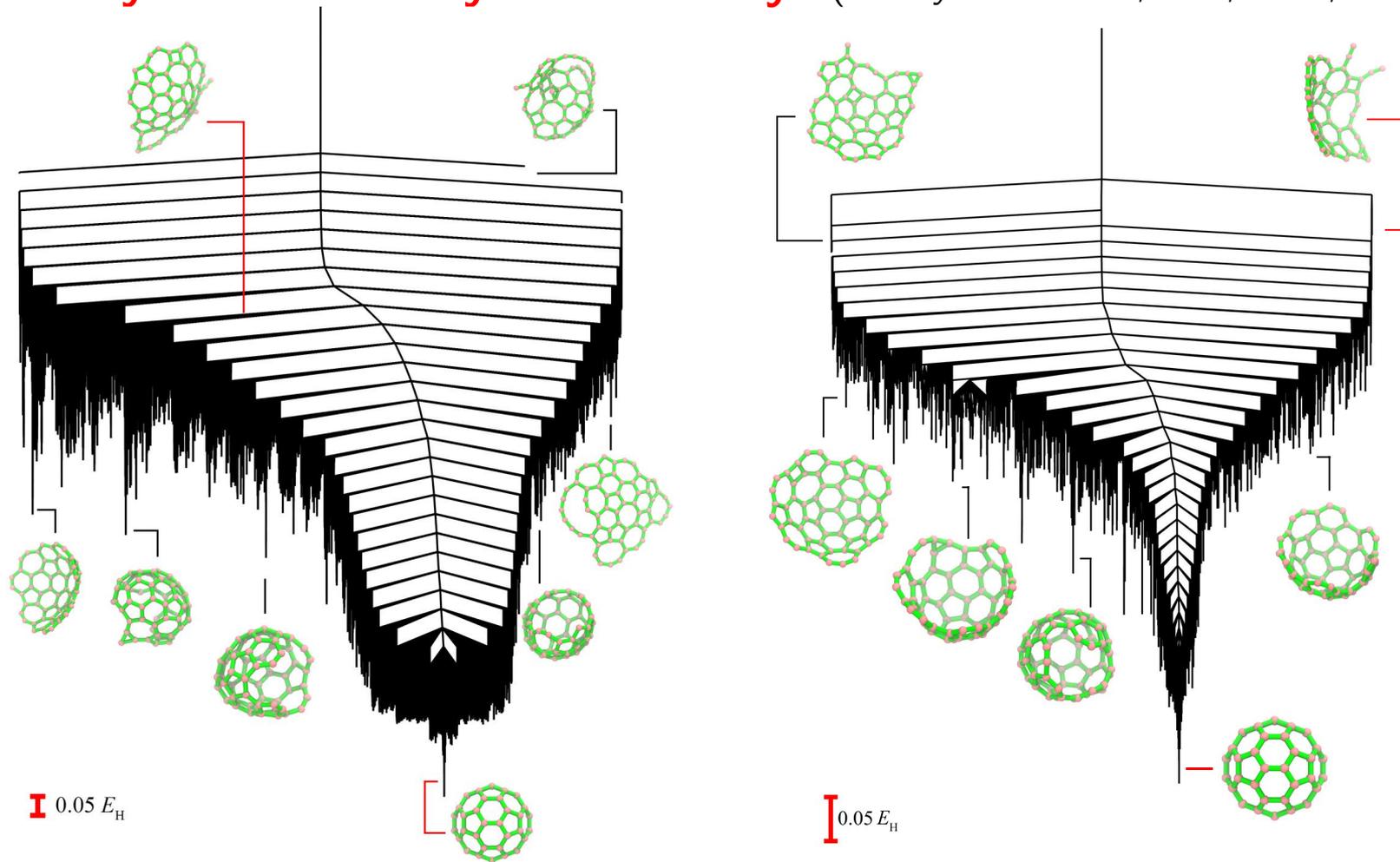


products intervening minima reactants

Networks of local minima and transition states are sampled using **geometry optimisation**. No **reaction coordinate** and no **projection** error.

The mean first passage time from $\mathcal{S} = \mathcal{Z} \cup \mathcal{B}$ is $\mathcal{T}_{AS} = \tau_S \mathbf{G}_S \mathbf{P}_S(0)$, with **waiting** times $\tau_s = 1 / \sum_{\alpha} K_{\alpha s}$, **fundamental** matrix $\mathbf{G}_S = [\mathbb{I}_S - \mathbf{B}_{SS}]^{-1}$, and initial occupation **probabilities** $P_s(0)$. \mathbb{I}_S is the **identity** matrix and $B_{s's} = K_{s's} \tau_s$ is a **branching** probability for rate constant $K_{s's}$.

Buckybowl to Buckyball Pathways (J. Phys. Chem. A, 126, 2342, 2022)



We compared two quantum-based potentials: **SCC-DFTB** and **GFN2-xTB**.

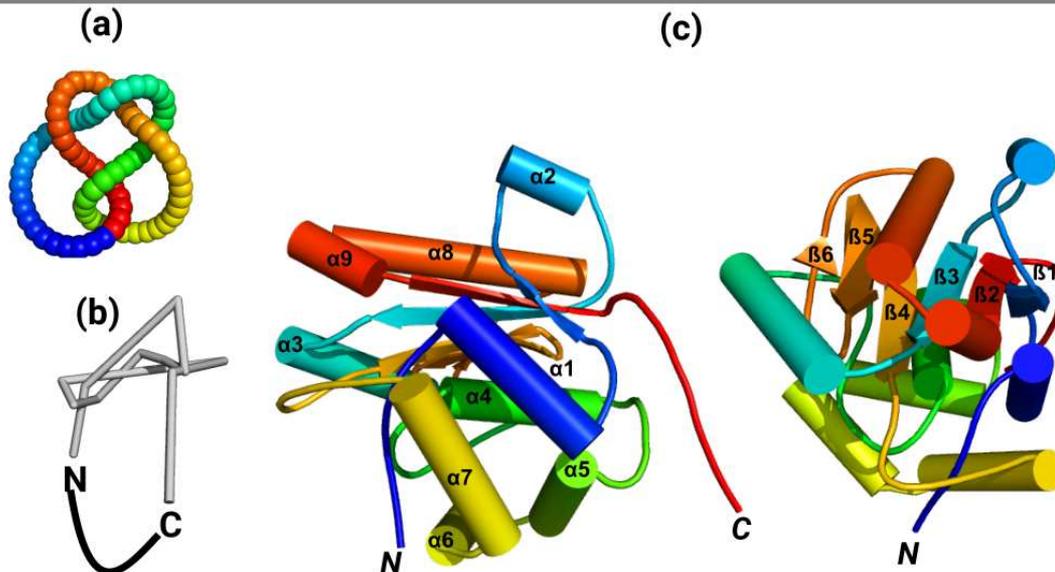
The **highest** energy structures are **buckybowls** with **dangling** chains.

The **fastest** paths between them each have around **50** transition states.

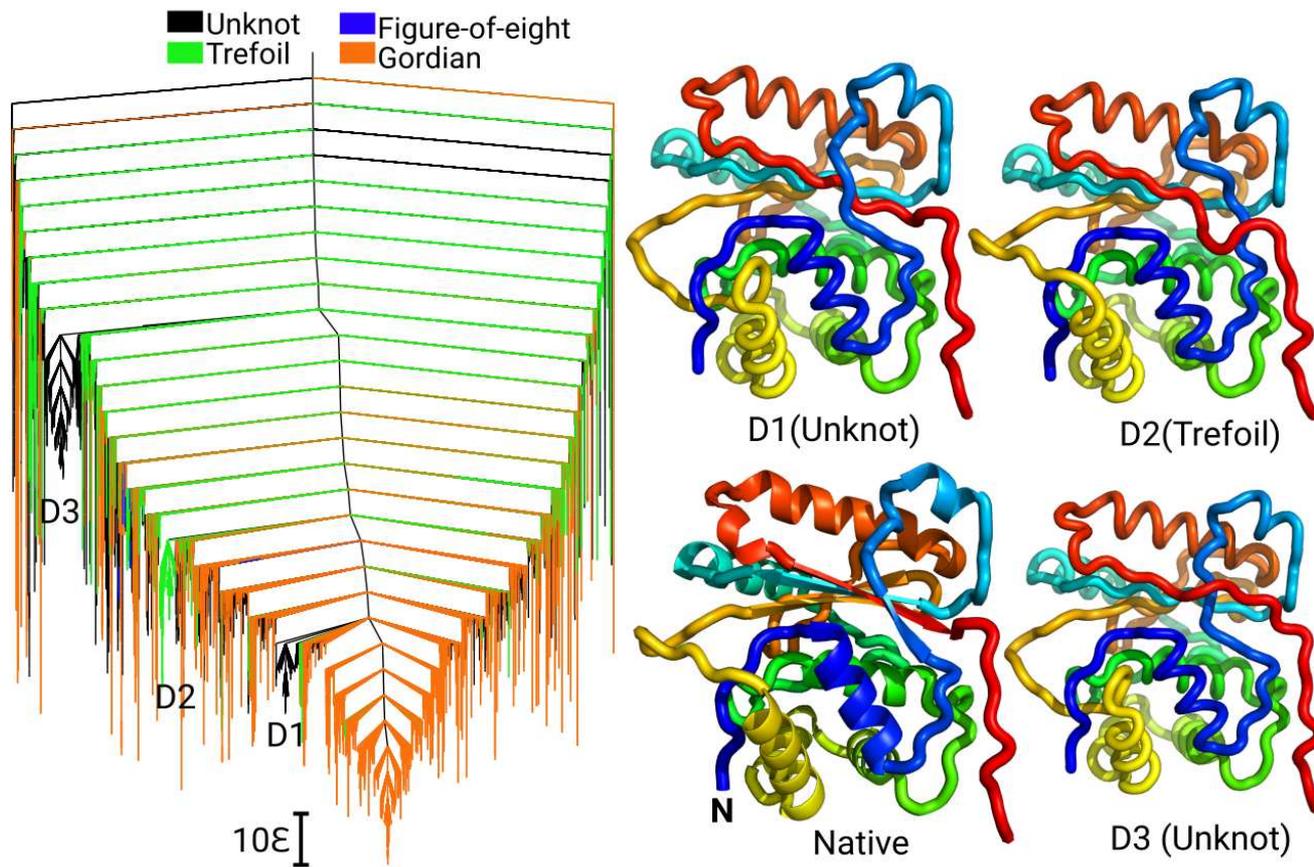
Cutting a Gordian Knot

Ubiquitin C-terminal hydrolase isoenzyme L1 is a deubiquitinating enzyme linked to **Parkinson's** and **Alzheimer's**. The N- and C-termini have 5 and 10 residues, while **216** residues form a shallow **Gordian** knot.

The **initial** path required **quasi-continuous interpolation**, penalising **internal minima** for distances between discrete images (*JCTC*, **8**, 5020, 2012).



- (a) The ideal 5_2 or **Gordian** knot with a closed loop coloured in the N- to C-direction from **blue** to **red**. (b) Sketch of **backbone** and **loop-closure**. (c) **Secondary** structure of UCH-L1 with nine α -**helices** and six β -**sheets**.



The landscape features **Gordian**, **figure-of-eight**, and **trefoil** knot configurations, with regions corresponding to a **jammed** trefoil knot (D2) and to **unknotted** structures (D1 and D3).

Several distinct classes of **folding** pathway exist, including either **trefoil** or **figure-of-eight** intermediates, or a **direct** conversion from the unknot.

Decoding Chemical Kinetics (*JPCL*, 13, 6349, 2022)

The **first passage time** distribution $p(t)$ can be written as

$$p(t) = \sum_{\ell} \lambda_{\ell} e^{-\lambda_{\ell} t} A_{\ell}, \quad \text{and for } y = \ln t, \quad \mathcal{P}(y) = \sum_{\ell} \lambda_{\ell} e^{y - \lambda_{\ell} \exp(y)} A_{\ell},$$

where $-\lambda_{\ell} < 0$ are the **eigenvalues** of the matrix defining the **master equation** dynamics for **absorbing** products, and the A_{ℓ} are **amplitudes**, which depend on the **eigenvectors** and the **initial condition**.

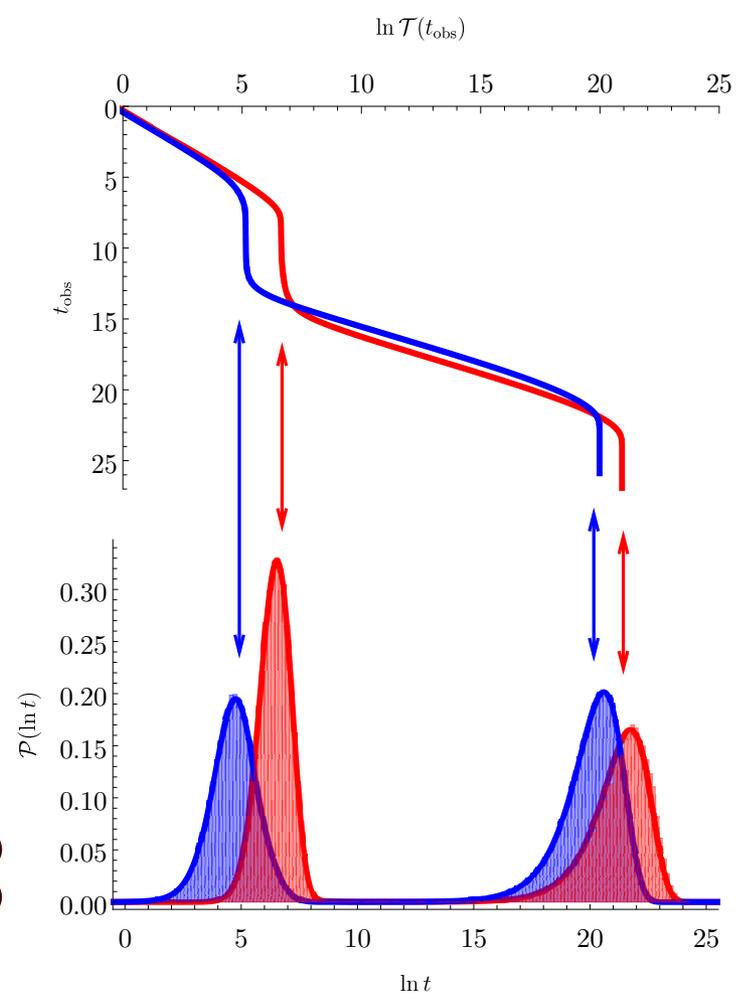
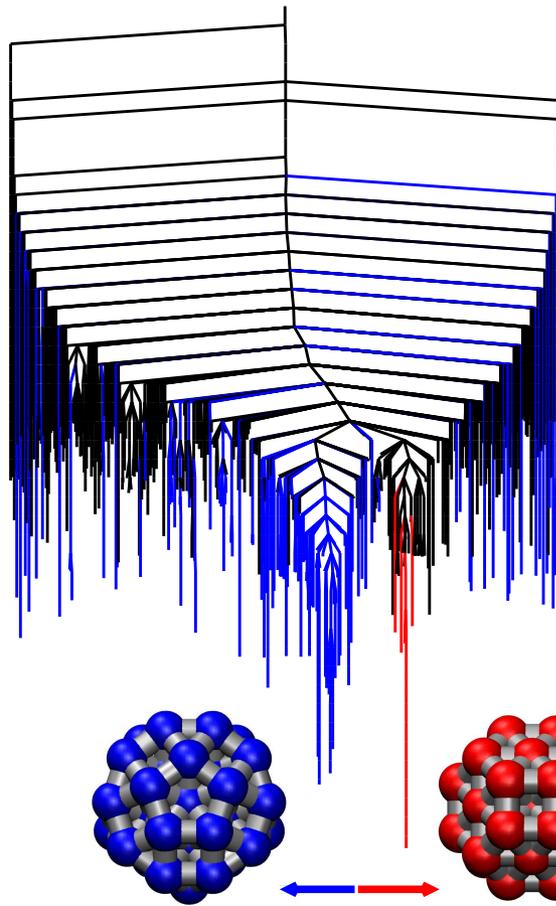
The **mean** first passage time can be defined for **observation** timescale t_{obs} :

$$\mathcal{T}(t_{\text{obs}}) = \sum_{\ell} \frac{A_{\ell}}{\lambda_{\ell} z(t_{\text{obs}})} \left[1 - e^{-\lambda_{\ell} t_{\text{obs}}} (1 + \lambda_{\ell} t_{\text{obs}}) \right],$$

where $z(t_{\text{obs}})$ is the **normalisation** for the restricted distribution.

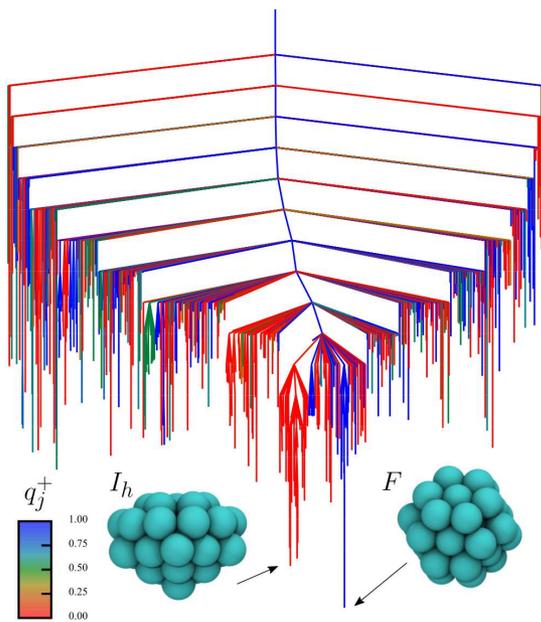
$$\lim_{t_{\text{obs}} \rightarrow \infty} \mathcal{T}(t_{\text{obs}}) = \sum_{\ell=1}^{\ell_{\text{max}}} \frac{A_{\ell}}{\lambda_{\ell}} / \sum_{\ell=1}^{\ell_{\text{max}}} A_{\ell}.$$

$\mathcal{T}(t_{\text{obs}})$ exhibits **steps** corresponding to **escape** from kinetic **traps** where we sum up to ℓ_{max} , $\ell_{\text{max}} - 1$, $\ell_{\text{max}} - 2 \dots$, with ℓ_{max} the **slowest** relaxation.

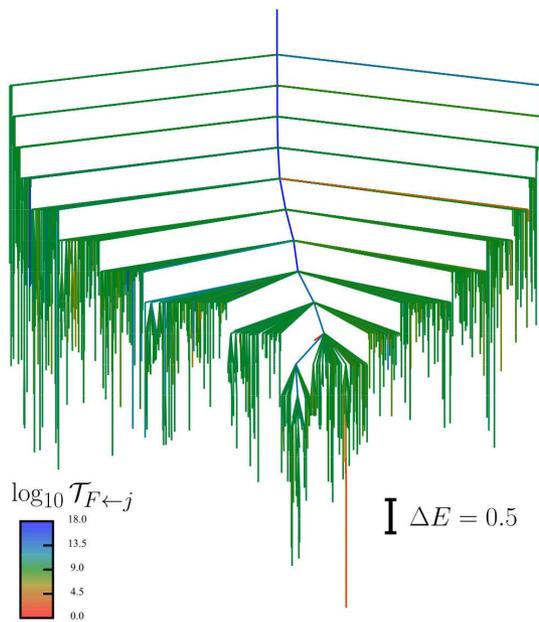


In the **double-funnel** LJ_{38} cluster **two** timescales appear in $\mathcal{P}(\ln t)$ and $\mathcal{T}(t_{\text{obs}})$ for **relaxation** from a high energy minimum to the competing **close-packed** and **icosahedral** structures. The **longer** time scale corresponds to **switching** between morphologies in each case.

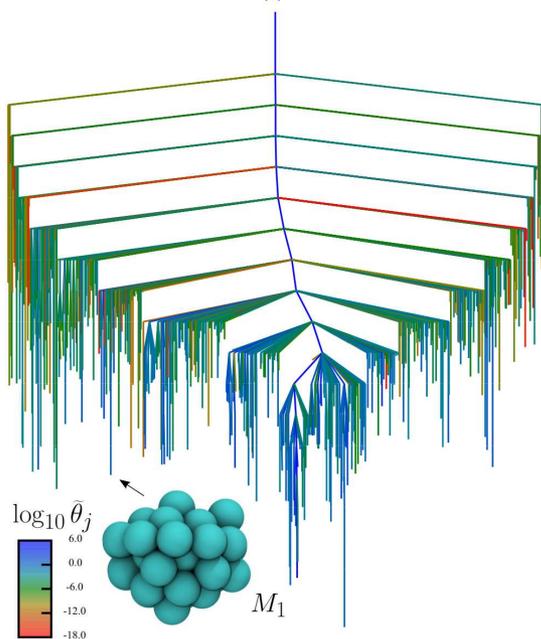
Dynamical Properties for LJ₃₈ (Phys. Rev. E, 104, 015301, 2021)



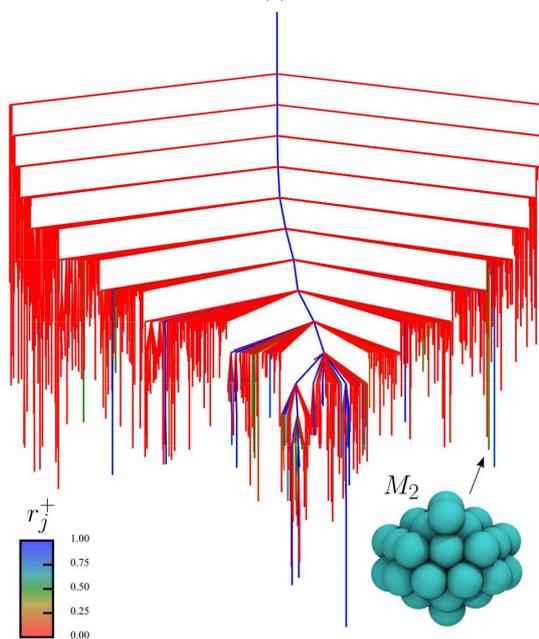
(a)



(b)



(c)



(d)

Coloured graphs:

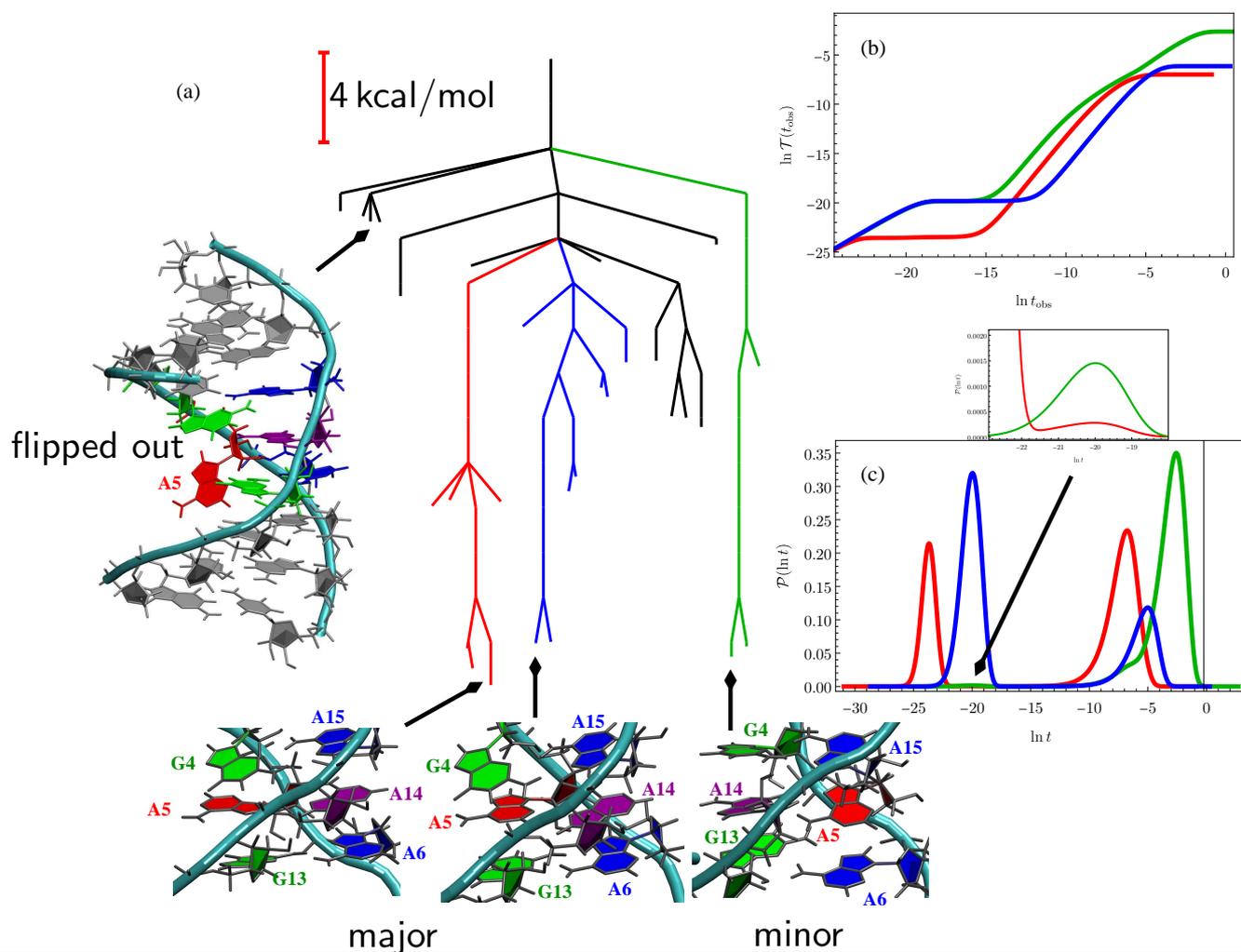
(a) **Committor** probability, q_j^+ .

(b) **MFPTs** $\mathcal{T}_{F \leftarrow j}$.

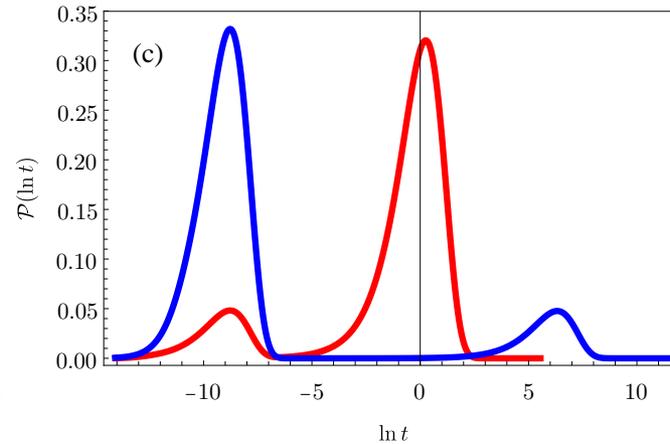
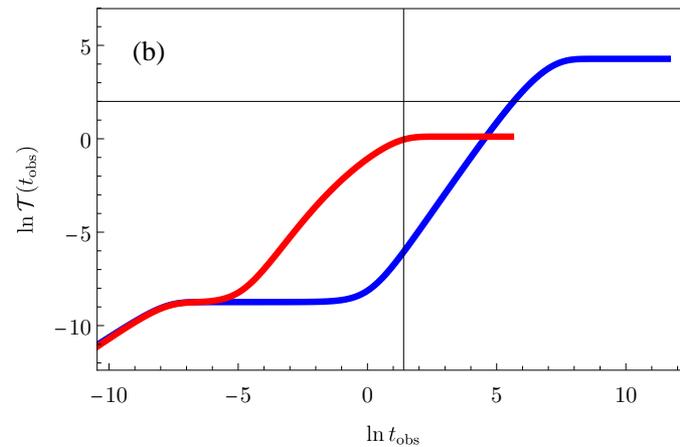
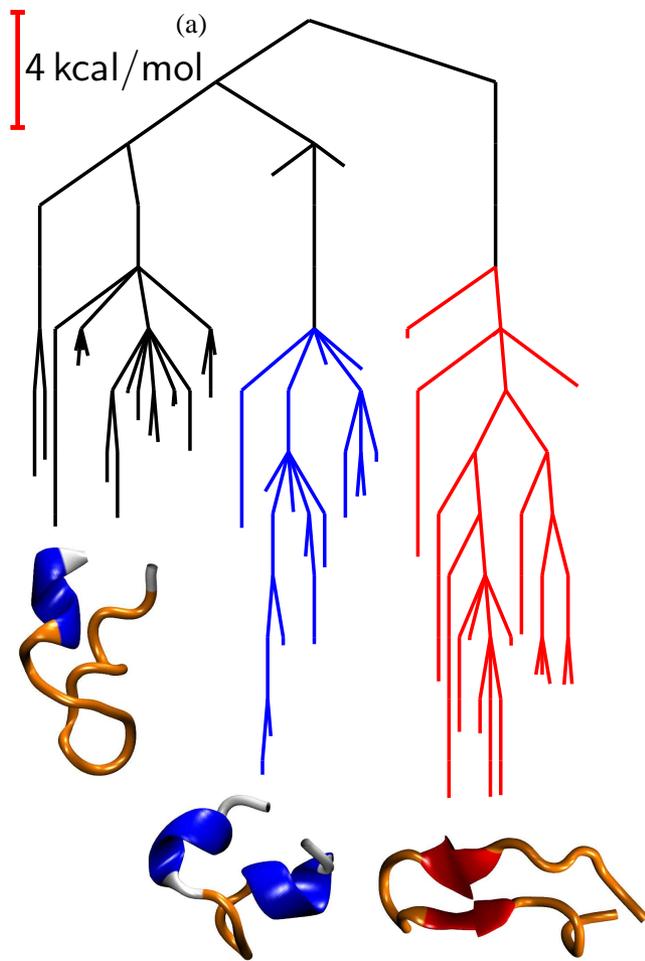
(c) Expected numbers of **node visits** along $F \leftarrow I_h$ reactive paths, $\tilde{\theta}_j$.

(d) **Reactive** visitation probabilities r_j^+ .

Nodes M_1 and M_2 both have high **visitation** probabilities, but only M_2 is a dynamical **bottleneck** with a **committor** probability close to 0.5.

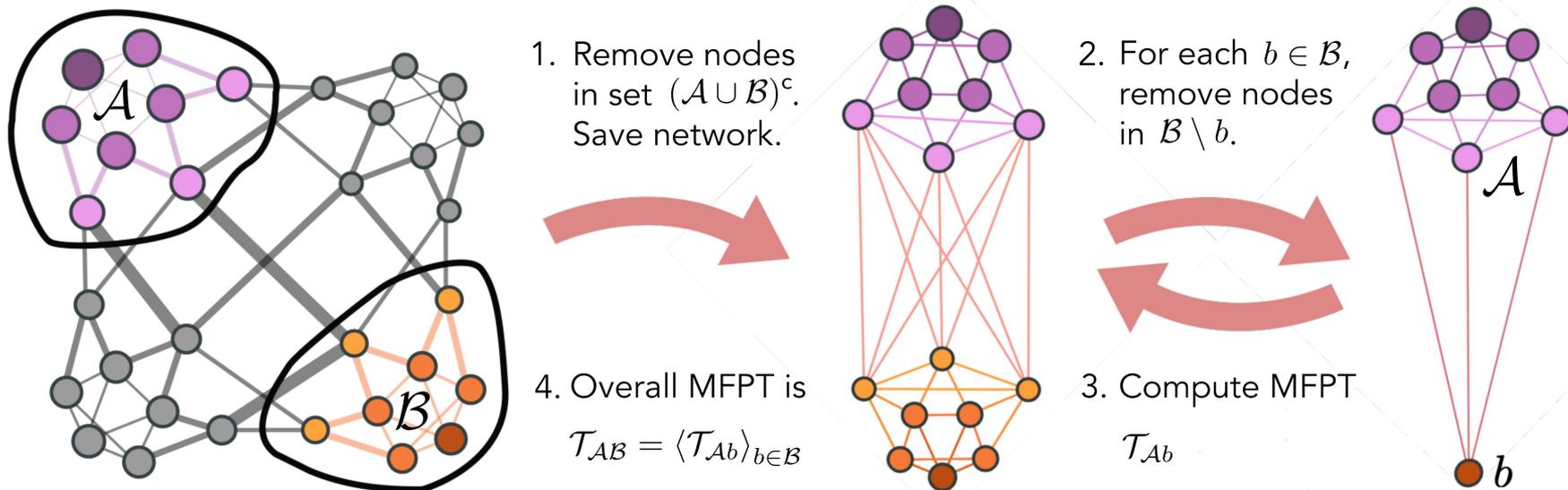


The landscape for a nine base pair RNA duplex with a non-canonical adenine-adenine contact exhibits major and minor forms with A14 or A5 stacked between A6 and A15. Three distinct time scales are discernible for relaxation to the minor form, with corresponding steps in $\ln \mathcal{T}(t_{\text{obs}})$.



Designed 18-residue peptide **DP5** exhibits competing α -helical and β -hairpin structures. The **interconversion** mechanism may provide insight into **amyloid** formation. $\mathcal{P}(\ln t)$ exhibits two peaks for **direct** and **indirect** relaxation, with corresponding steps in $\ln \mathcal{T}(t_{\text{obs}})$.

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



Minima, z , are progressively **removed**, and the branching probabilities and waiting times in adjacent minima, β , are **renormalised**, maintaining precision for highly **metastable** systems using $1 - B_{zz} = \sum_{\epsilon \neq z} B_{\epsilon z}$:

$$B'_{\gamma\beta} = B_{\gamma\beta} + B_{\gamma z} B_{z\beta} \sum_{m=0}^{\infty} B_{zz}^m = B_{\gamma\beta} + \frac{B_{\gamma z} B_{z\beta}}{1 - B_{zz}}, \quad \tau'_{\beta} = \tau_{\beta} + \frac{B_{z\beta} \tau_z}{1 - B_{zz}}.$$

The **MFPT** from every **reactant** state to the set of **product** states is **conserved**, with an execution time **independent** of temperature.

Partial Graph Transformation (*JCP*, **153**, 244108, 2020; *Phil. Trans. A*, in press)

The full **FPT** distribution gives more detailed insight than the **MFPT**, revealing **multiple** distinct relaxation time scales.

The most **interesting** networks are likely to be the most **ill-conditioned**.

The FPT distribution may be **approximately** conserved for **partial** graph transformation, eliminating **selected** nodes \mathcal{Z} and **retaining** nodes $\mathcal{S}^{\mathcal{Z}}$.

The **renormalised** branching matrix $\mathbf{B}_{\mathcal{S}^{\mathcal{Z}}\mathcal{S}^{\mathcal{Z}}}^{\mathcal{Z}}$ and waiting times $\tau_{\mathcal{S}^{\mathcal{Z}}}^{\mathcal{Z}}$ define a rate matrix $\mathbf{Q}_{\mathcal{S}^{\mathcal{Z}}}^{\mathcal{Z}} = [\mathbf{B}_{\mathcal{S}^{\mathcal{Z}}\mathcal{S}^{\mathcal{Z}}}^{\mathcal{Z}} - \mathbb{I}_{\mathcal{S}^{\mathcal{Z}}}] \mathbf{D}_{\mathcal{S}^{\mathcal{Z}}}^{\mathcal{Z}}$, where $[\mathbf{D}_{\mathcal{S}^{\mathcal{Z}}}^{\mathcal{Z}}]_{ij} = \delta_{ij} / [\tau_{\mathcal{S}^{\mathcal{Z}}}^{\mathcal{Z}}]_i$.

We require the new **steady state** distribution $\pi_{\mathcal{S}^{\mathcal{Z}}}^{\mathcal{Z}}$ to satisfy the **balance** condition $\mathbf{B}_{\mathcal{S}^{\mathcal{Z}}\mathcal{S}^{\mathcal{Z}}}^{\mathcal{Z}} \mathbf{D}_{\mathcal{S}^{\mathcal{Z}}}^{\mathcal{Z}} \pi_{\mathcal{S}^{\mathcal{Z}}}^{\mathcal{Z}} = \mathbf{D}_{\mathcal{S}^{\mathcal{Z}}}^{\mathcal{Z}} \pi_{\mathcal{S}^{\mathcal{Z}}}^{\mathcal{Z}}$, producing $[\mathbf{D}_{\mathcal{S}^{\mathcal{Z}}}^{\mathcal{Z}} \pi_{\mathcal{S}^{\mathcal{Z}}}^{\mathcal{Z}}]_i = [\mathbf{D}\pi]_i$, so

$$[\pi_{\mathcal{S}^{\mathcal{Z}}}^{\mathcal{Z}}]_i / [\tau_{\mathcal{S}^{\mathcal{Z}}}^{\mathcal{Z}}]_i = \pi_i / \tau_i \quad \text{or} \quad \pi_i^{\mathcal{Z}} / \pi_i = \tau_i^{\mathcal{Z}} / \tau_i \quad \text{for} \quad i \in \mathcal{S}^{\mathcal{Z}}.$$

The **renormalised** equilibrium occupation probabilities and branching probabilities can be used to define effective **free energies** for the retained minima, $f_s(T)$, and the transition states that connect them, $f_{ss'}^\dagger(T)$:

$$f_s(T) = -k_B T \ln[\boldsymbol{\pi}_{\mathcal{S}^Z}]_s,$$

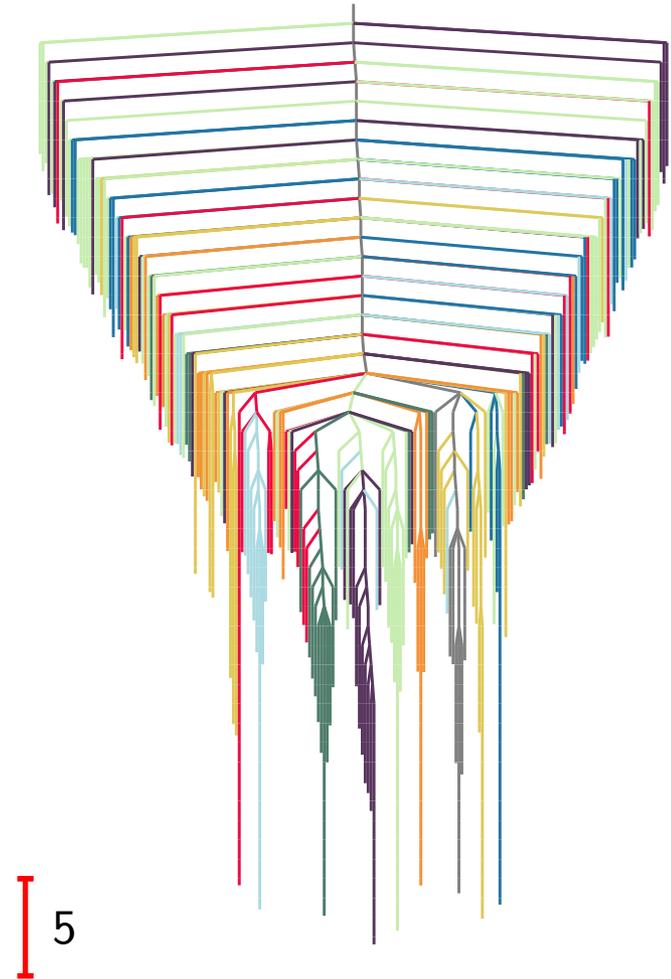
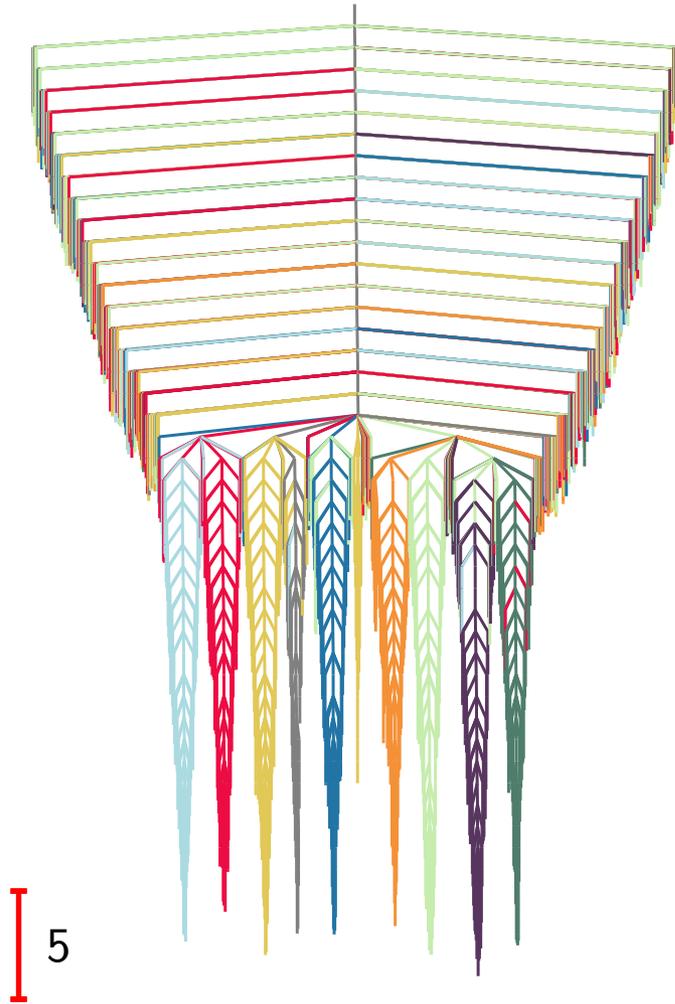
$$\begin{aligned} f_{ss'}^\dagger(T) &= f_{s'}(T) - k_B T \ln[\mathbf{K}_{\mathcal{S}^Z \mathcal{S}^Z}^Z]_{ss'} + k_B T \ln(k_B T/h), \\ &= f_s(T) - k_B T \ln[\mathbf{K}_{\mathcal{S}^Z \mathcal{S}^Z}^Z]_{s's} + k_B T \ln(k_B T/h), \end{aligned}$$

where the **rate constant** $[\mathbf{K}_{\mathcal{S}^Z \mathcal{S}^Z}^Z]_{s's}$ in \mathcal{S}^Z is $[\mathbf{B}_{\mathcal{S}^Z \mathcal{S}^Z}^Z]_{s's} / [\boldsymbol{\tau}_{\mathcal{S}^Z}]_s$.

The free energy of the **transition state** is defined to **reproduce** $[\mathbf{K}_{\mathcal{S}^Z \mathcal{S}^Z}^Z]_{s's}$:

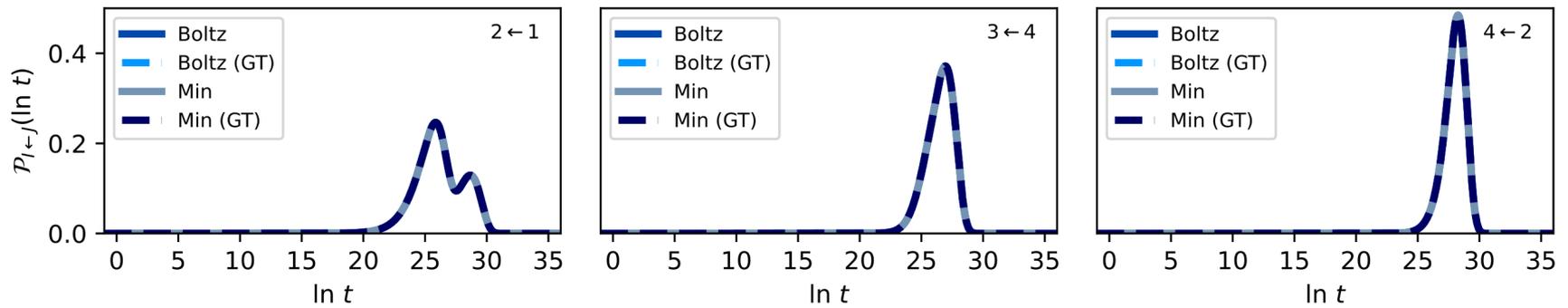
$$[\mathbf{K}_{\mathcal{S}^Z \mathcal{S}^Z}^Z]_{s's} = \frac{k_B T}{h} \exp \left[-\frac{\left(f_{ss'}^\dagger(T) - f_s(T) \right)}{k_B T} \right].$$

Hence we obtain a **disconnectivity graph** for the reduced network \mathcal{S}^Z .

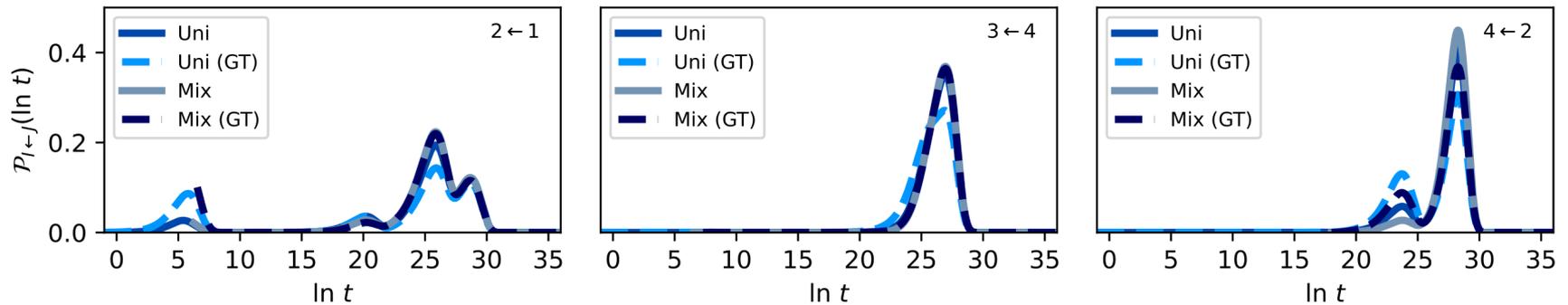


Disconnectivity graphs for (left) a nine-community model network of **994** minima, and (right) a **reduced** network of **215** states obtained from partial GT. The colour scheme highlights the different **communities**.

a)



b)



FPT distributions for transitions between different **communities** in the nine-community model at $T = 1$ for the **full** and graph transformed (**GT**) networks.

The system is **initialised** in different local distributions within the starting communities.

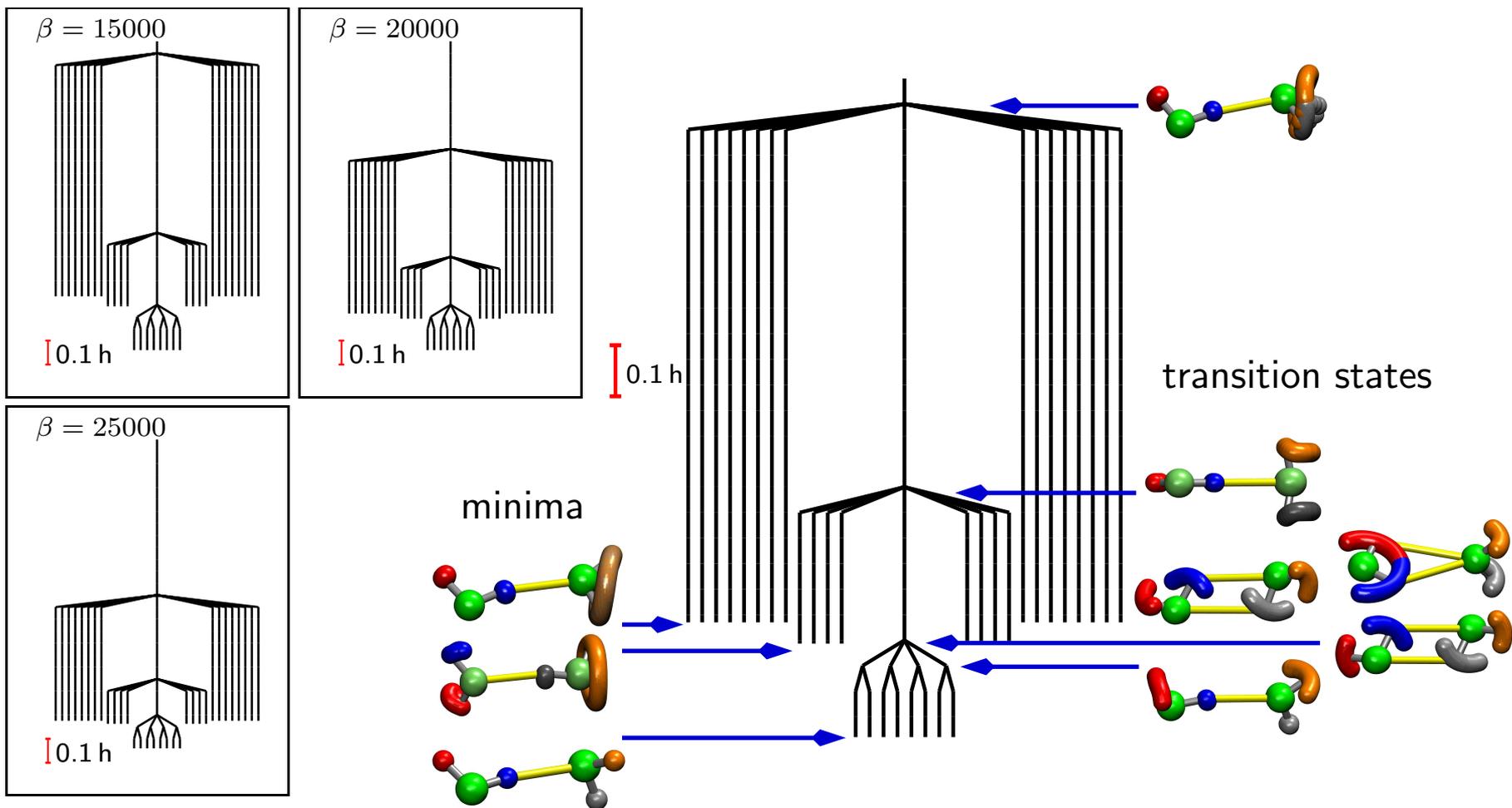
Boltz: probability is the local **Boltzmann** distribution within the starting community,

Min: probability is **localised** in the global free energy minimum of the starting community,

Uni: probability is **uniform** over all states in the community,

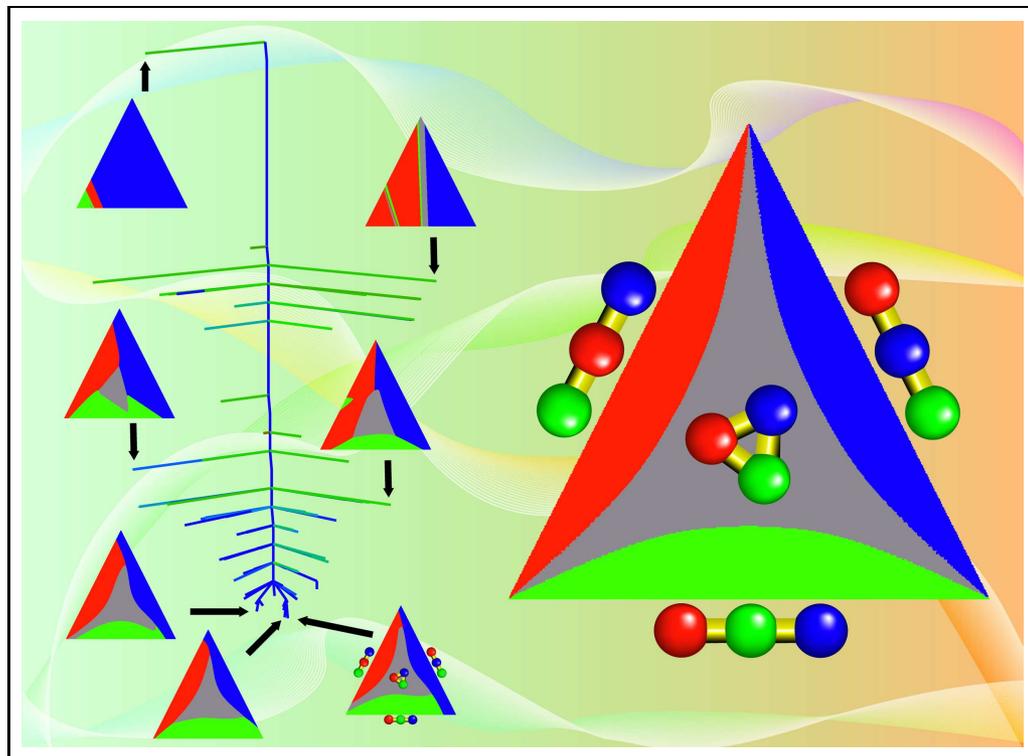
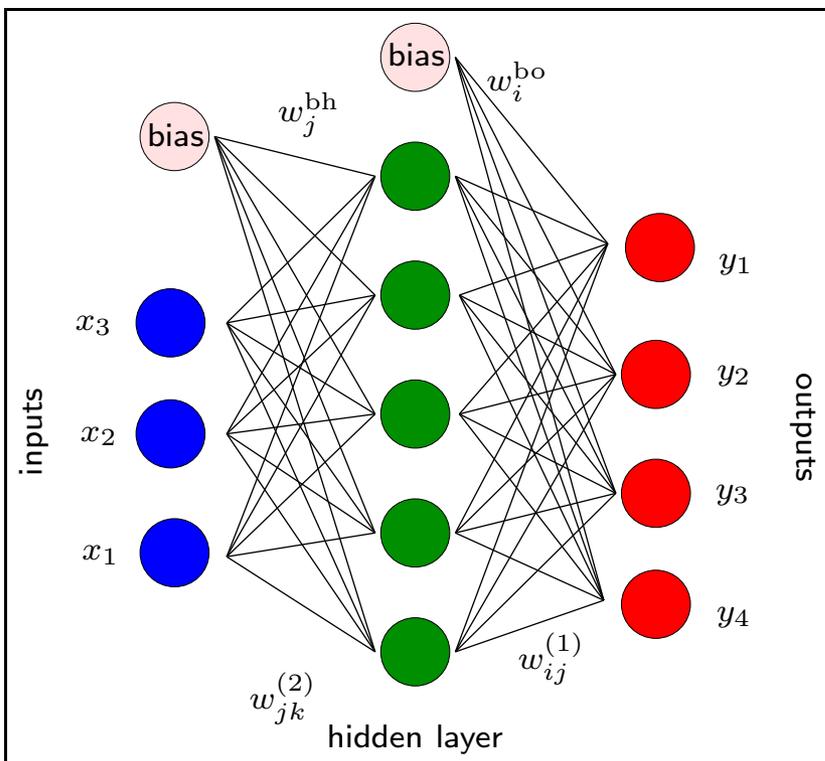
Mix: probability is initially **uniform**, then conditioned to remain within the starting community for a **mixing** time τ_{mix} .

Path Integral Energy Landscapes for Water Dimer (*JCTC*, 15, 33, 2019)



The ring polymer **landscape** of $(\text{H}_2\text{O})_2$ includes **classical** and **delocalised** minima and transition states for the **MBPOL** potential (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.

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

Energy Landscapes for Quantum Circuits

The **variational quantum eigensolver** algorithm provides a practical approach for contemporary **noisy intermediate-scale quantum** devices.

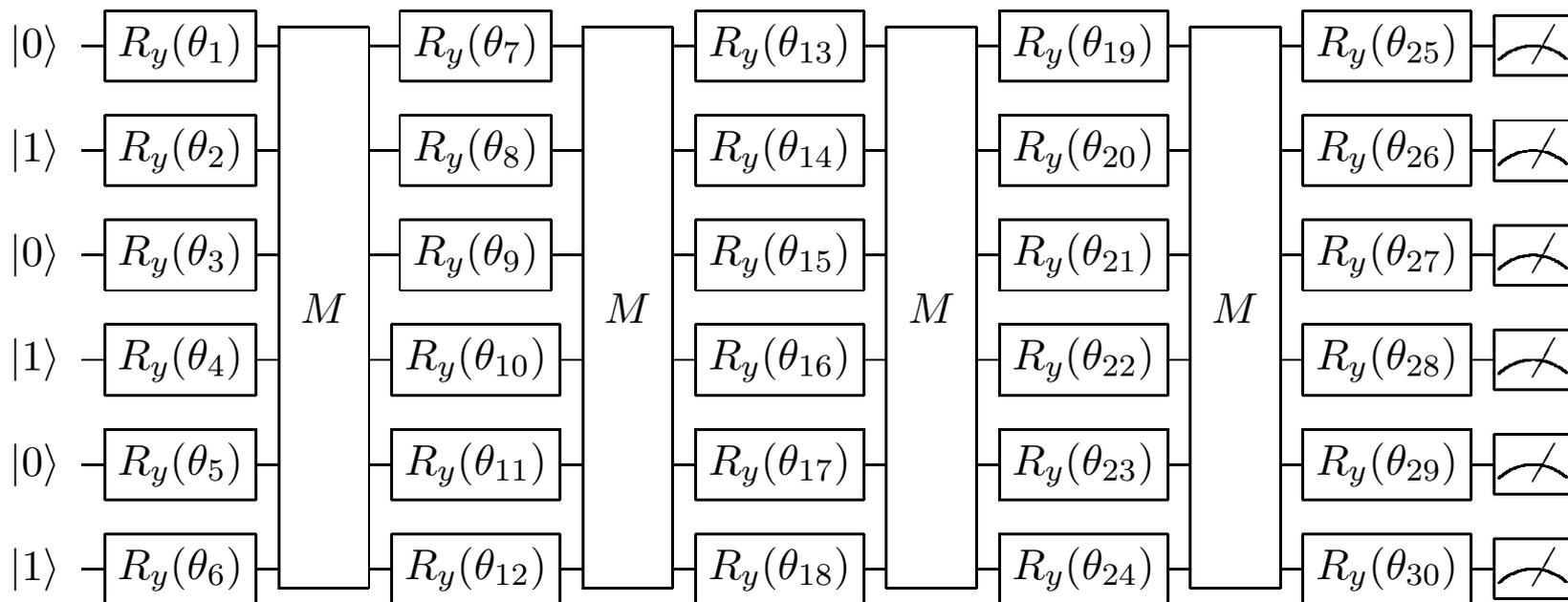
The **electronic** Hamiltonian \hat{H}_e is mapped onto a **qubit** Hamiltonian \hat{H}_q , which would be measured on a quantum computer using a suitable **circuit**.

We have investigated **hardware-efficient** variational circuits with L layers of R_y one-qubit **rotation** gates for each qubit and associated rotational parameters, θ_i , plus **entangling** layers consisting of **CNOT** two-qubit gates.

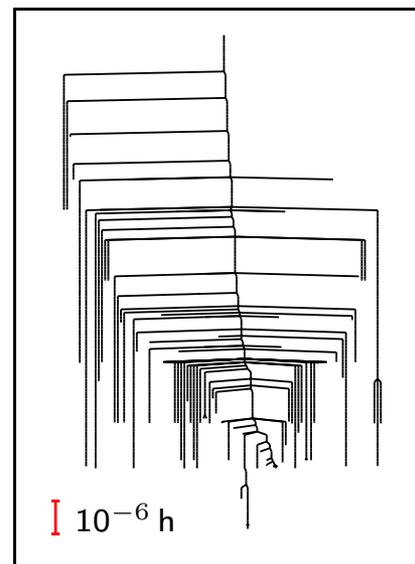
An **initial** Hartree-Fock state **evolves** as $|\psi(\boldsymbol{\theta})\rangle = \hat{U}(\boldsymbol{\theta}) |\psi(0)\rangle$ for unitary

$$\hat{U}(\boldsymbol{\theta}) = \begin{cases} \bigotimes_{i=1}^N R_y(\theta_i), & L = 1 \\ \prod_{j=1}^{L-1} \left[\bigotimes_{k=Nj+1}^{N(j+1)} R_y(\theta_k) \bigotimes_{i=1}^{N-1} \text{CNOT}(q_i, q_{i+1}) \right] \bigotimes_{i=1}^N R_y(\theta_i), & L > 1 \end{cases}$$

Six qubit hardware-efficient circuit for BeH_2 with a minimal basis:



This circuit supports various **local** minima (right).
Fixing selected angles **simplifies** the landscape, while maintaining the **accuracy** of the global minimum.
Hence analysis of $E(\theta)$ landscapes provides **design** principles for **optimising** circuits.

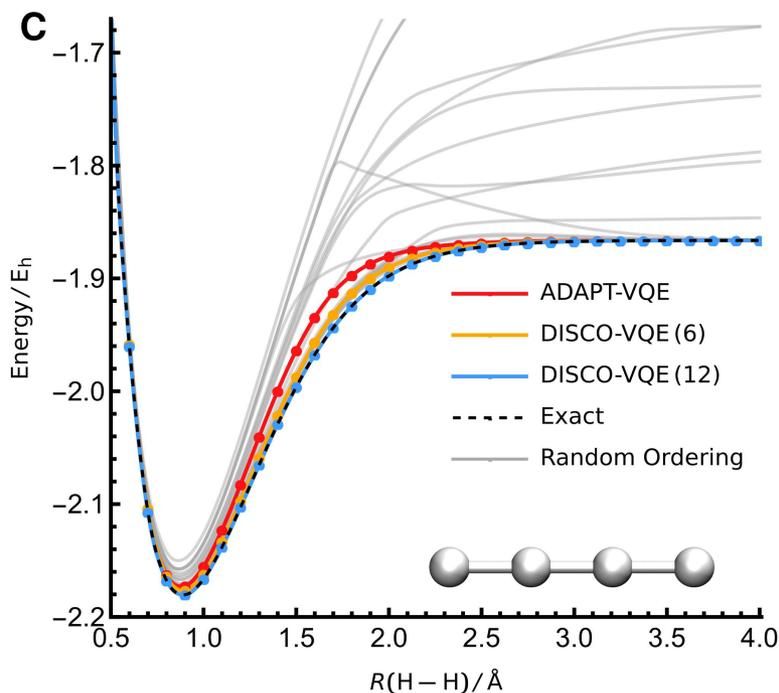
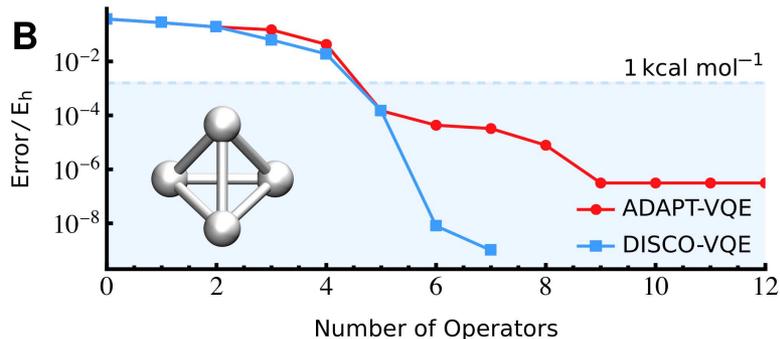
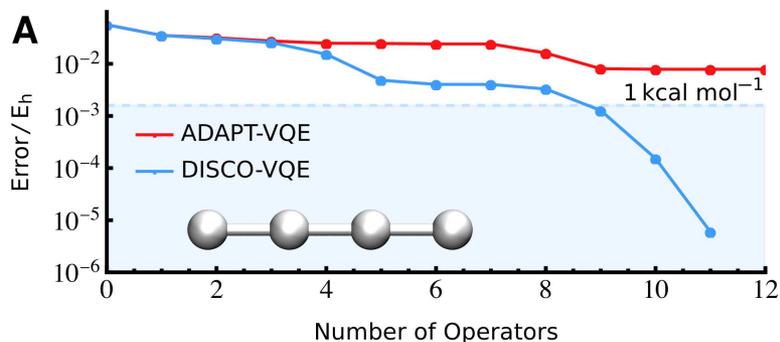


Global Optimization for a Unitary Product States Ansatz

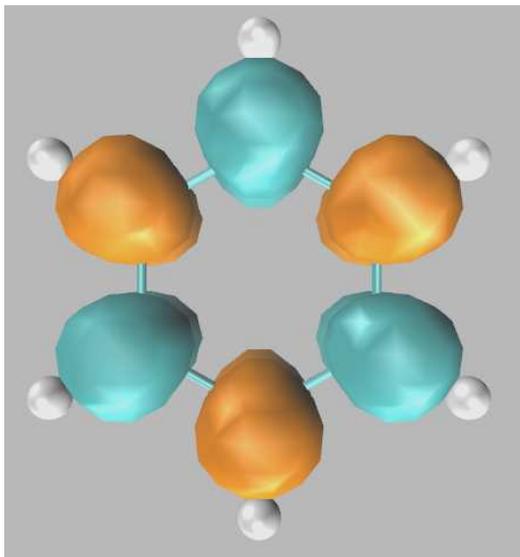
The **ansatz** employs anti-Hermitian **second-quantised** operators $\hat{\kappa}_{\mu_i}$, readily encoded on **quantum hardware**, with **continuous** amplitudes t and energy

$$E(\mathbf{t}, \boldsymbol{\mu}) = \langle \Phi_0 | \left(\prod_{i=M}^1 e^{-t_i \hat{\kappa}_{\mu_i}} \right) \hat{H} \left(\prod_{j=1}^M e^{t_j \hat{\kappa}_{\mu_j}} \right) | \Phi_0 \rangle .$$

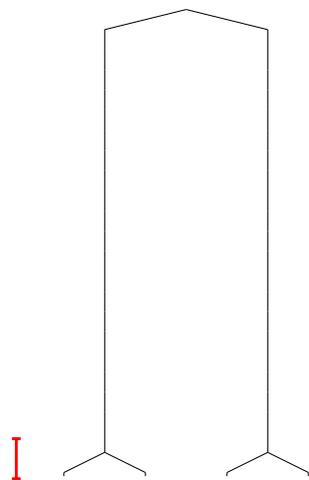
Compact, accurate wavefunctions result from **generalised basin-hopping** for the t and the **discrete** space of **ordered** fermionic operators, $\boldsymbol{\mu}$.



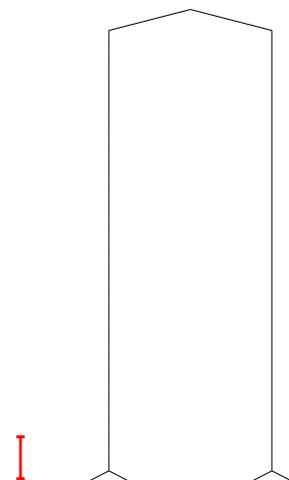
SCF Landscapes for Benzene



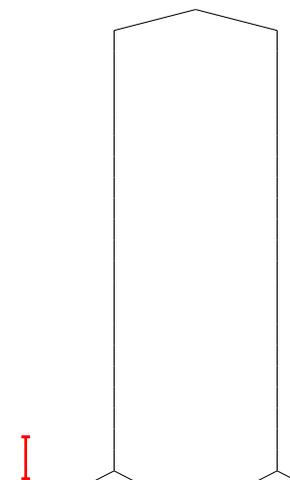
3-21G HF spin densities



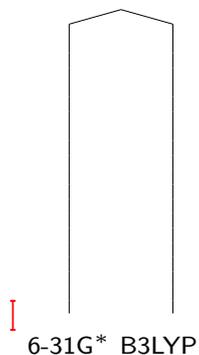
$$3\text{-}21\text{G HF } \langle \hat{S}^2 \rangle = 0.50$$



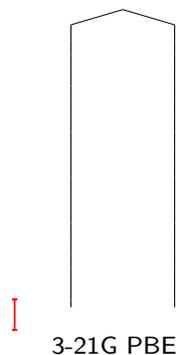
$$6\text{-}31\text{G}^* \text{ HF } \langle \hat{S}^2 \rangle = 0.49$$



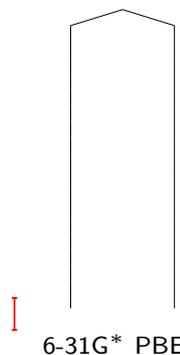
$$\text{cc-pVDZ HF } \langle \hat{S}^2 \rangle = 0.45$$



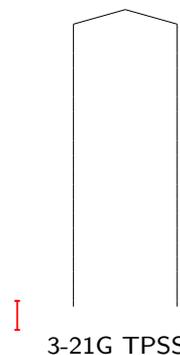
6-31G* B3LYP



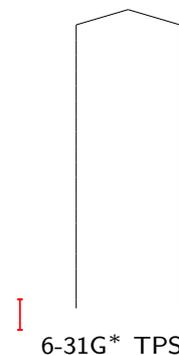
3-21G PBE



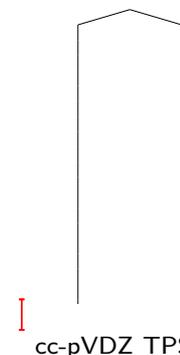
6-31G* PBE



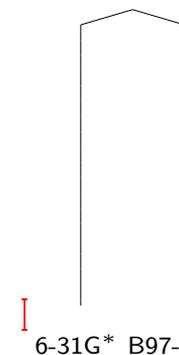
3-21G TPSS



6-31G* TPSS



cc-pVDZ TPSS



6-31G* B97-D3

Benzene exhibits symmetry breaking for the HF landscapes; for the DFT functionals $\langle \hat{S}^2 \rangle = 0$. (The scale bars are all 0.02 h.)

Landscapes for K-Means Clustering (*J. Chem. Phys.*, **156**, 054109, 2022)

The K -means algorithm exhibits **multiple** local minima for different **clustering** solutions. Locating the **global minimum** can be challenging.

The cost function for N data points and K clusters at positions μ_k is

$$J(\mu) = \min_{j \in S} \left(\sum_{i=1}^N \sum_{k=1}^K r_{ik}^{(j)} \|\mathbf{x}_i - \mu_k\|^2 \right).$$

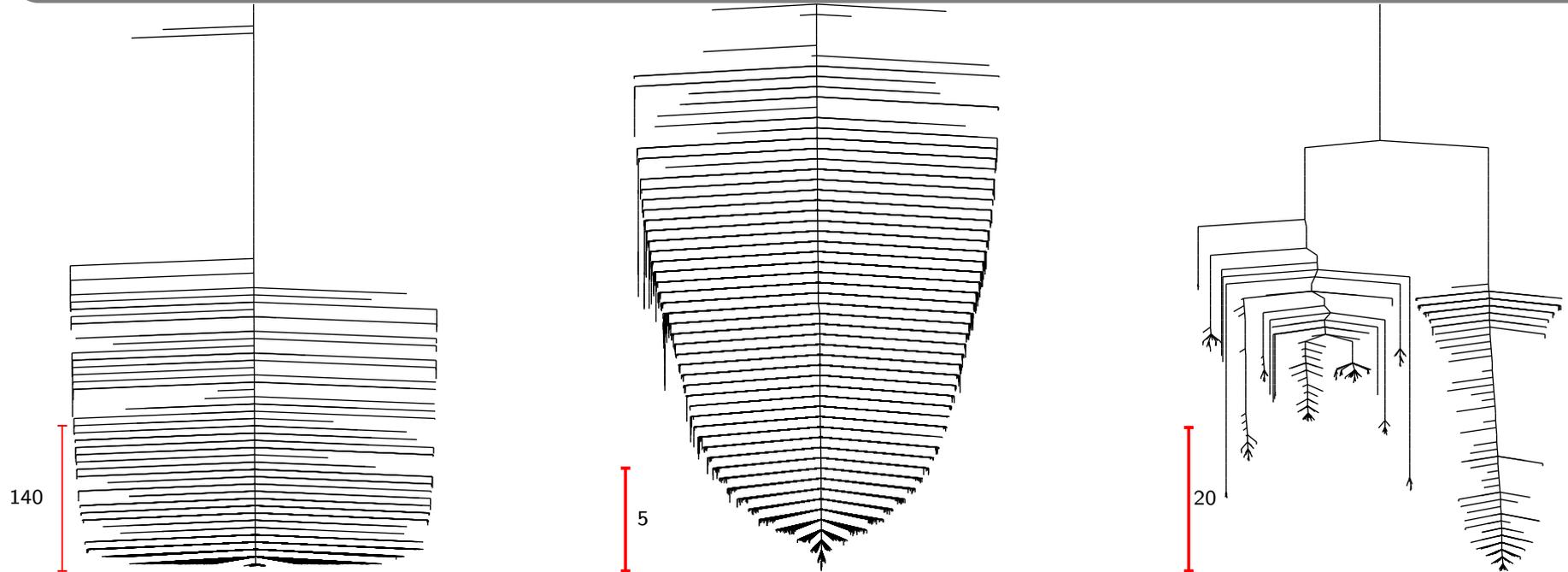
S is the set of valid **assignment** matrices \mathbf{r} , where each of the data points, with N_f features, is assigned to one of the clusters.

Local **minima** were identified using **iterative** assignment and **minimisation** (Lloyd's algorithm), and augmented by minima from **pathways**.

Derivatives of $J(\mu)$ are **undefined** at changes in cluster assignment. **Transition states** were identified using a **penalty** function to find the **minimum** barrier height between different assignments on quadratic **intersection** seams, a method developed to analyse **conical intersections**.

An important **application** is to **gene expression** data, which report on the cellular **environment**, where clustering can identify new **cancer** subtypes.

Accurate **classification** provides **specific** guidance for **treatment**.

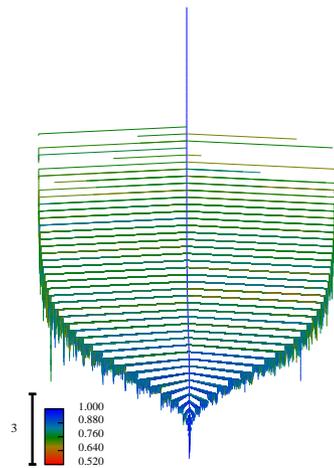


(a) $K = 2$ full database

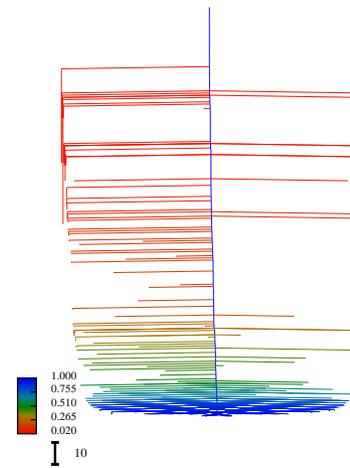
(b) $K = 2$ magnification

(c) $K = 3$

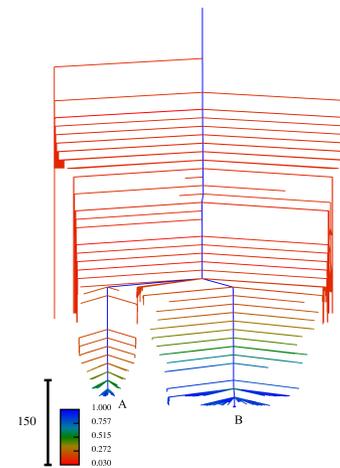
The K -means landscapes above are for a **bone marrow** dataset from paediatric acute lymphoblastic leukemia patients, based on **mRNA** analysis with 2526 genes. The **248** samples separate into **two** classes.



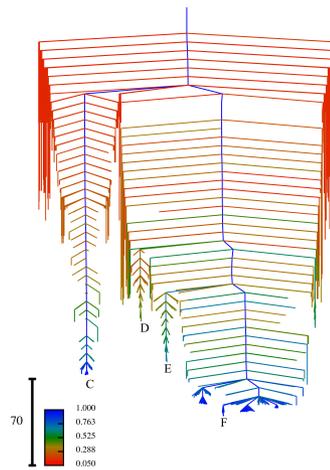
$K = 2$



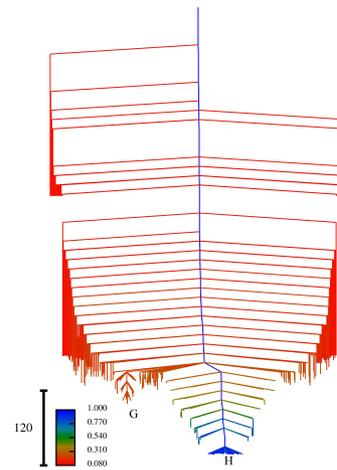
$K = 3$



$K = 4$



$K = 5$



$K = 6$

The **Yeoh** dataset has **two** different clinical assignments.

We find that the landscape has a largely **single** funnel structure when K matches either of the **true** number of classes ($K = 2$ or $K = 6$).