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.

- 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

PUMA
GCN4-pLI E20S mutant
G-quadruplex A(G3 TTA)3 G3
RfaH C-terminal domain

L67S mutant
ubiquitin
RNA 7SK
Influenza A Hemagglutinin

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.

This generalised basin-hopping approach uses bminima 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.
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 \((\textit{Mol. Phys.}, 100, 3285, 2002; 102, 891, 2004)\).

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 \(S = \mathcal{Z} \cup \mathcal{B}\) is \(T_{AS} = \tau_S G_S P_S(0)\), with waiting times \(\tau_S = 1/\sum \alpha K_{\alpha S}\), fundamental matrix \(G_S = [I_S - B_{SS}]^{-1}\), and initial occupation probabilities \(P_S(0)\). \(I_S\) is the identity matrix and \(B_{s's} = K_{s's} \tau_S\) is a branching probability for rate constant \(K_{s's}\).
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.
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 $\alpha$-helices and six $\beta$-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.
The first passage time distribution $p(t)$ can be written as

$$p(t) = \sum_\ell \lambda_\ell e^{-\lambda_\ell t} A_\ell,$$

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

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

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

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

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

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

$\mathcal{T}(t_{\text{obs}})$ exhibits steps corresponding to escape from kinetic traps where we sum up to $\ell_{\text{max}}$, $\ell_{\text{max}} - 1$, $\ell_{\text{max}} - 2$, ... , with $\ell_{\text{max}}$ 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.
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 $\ln T(t_{\text{obs}})$. 
Designed 18-residue peptide DP5 exhibits competing $\alpha$-helical and $\beta$-hairpin structures. The interconversion mechanism may provide insight into amyloid formation. $P(\ln t)$ exhibits two peaks for direct and indirect relaxation, with corresponding steps in $\ln \mathcal{T}(t_{\text{obs}})$. 
Minima, $z$, are progressively removed, and the branching probabilities and waiting times in adjacent minima, $\beta$, 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.
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 $Z$ and retaining nodes $S^Z$.

The renormalised branching matrix $B_{S^ZS^Z}^Z$ and waiting times $\tau_{S^Z}^Z$ define a rate matrix $Q_{S^Z}^Z = [B_{S^ZS^Z}^Z - \mathbb{I}_{S^Z}]D_{S^Z}^Z$, where $[D_{S^Z}^Z]_{ij} = \delta_{ij}/[\tau_{S^Z}^Z]_i$.

We require the new steady state distribution $\pi_{S^Z}^Z$ to satisfy the balance condition $B_{S^ZS^Z}^ZD_{S^Z}^Z\pi_{S^Z}^Z = D_{S^Z}^Z\pi_{S^Z}^Z$, producing $[D_{S^Z}^Z\pi_{S^Z}^Z]_i = [D\pi]_i$, so $[\pi_{S^Z}^Z]_i/[\tau_{S^Z}^Z]_i = \pi_i/\tau_i$ or $\pi_i/\pi_i = \tau_i^Z/\tau_i$ for $i \in S^Z$. 
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, $f^\dagger_{ss'}(T)$:

\[
f_s(T) = -k_B T \ln[\pi_{S^Z}]_s,
\]
\[
f^\dagger_{ss'}(T) = f_{s'}(T) - k_B T \ln[\mathbf{K}_{S^Z S^Z}]_{ss'} + k_B T \ln (k_B T/h),
\]

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

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

\[
[K_{S^Z S^Z}]_{s's} = \frac{k_B T}{h} \exp \left[ -\frac{(f^\dagger_{ss'}(T) - f_s(T))}{k_B T} \right].
\]

Hence we obtain a disconnectivity graph for the reduced network $S^Z$. 
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.
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 $\tau_{\text{mix}}$. 
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).
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, $\theta_i$, plus entangling layers consisting of CNOT two-qubit gates.

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

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

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

Compact, accurate wavefunctions result from generalised basin-hopping for the $t$ and the discrete space of ordered fermionic operators, $\mu$.
SCF Landscapes for Benzene

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.)
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 $\mu_k$ is

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

*S* is the set of valid assignment matrices $r$, 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.

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.
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 $K = 6$).