Dynamical Signatures of Multifunnel Energy Landscapes

David J Wales*

ABSTRACT: Multifunctional systems, such as molecular switches, exhibit multifunnel energy landscapes associated with the alternative functional states. In this contribution the multifunnel organization is decoded from dynamical signatures in the first passage time distribution between reactants and products. Characteristic relaxation rates are revealed by analyzing the kinetics as a function of the observation time scale, which scans the underlying distribution. Extracting the corresponding dynamical signatures provides direct insight into the organization of the molecular energy landscape, which will facilitate a rational design of target functionality. Examples are illustrated for multifunnel landscapes in biomolecular systems and an atomic cluster.

The relaxation times measured in an experiment are determined by the structure of the underlying energy landscape, which encodes the pathways between products and reactants, and the associated barriers and rates. To interpret the corresponding properties, we need to know what is actually being measured and the time scale of the observation. This time scale determines which transformations are feasible and which states define the free energy minima that are averaged over in local equilibrium. Distinguishing different product and reactant states on different time scales has direct consequences for observations, ranging from resolution of tunnelling splittings in spectroscopy to thermodynamic measurements subject to broken ergodicity.

Interconversion rates report on the mean first passage time (MFPT) between reactant and product states. In this contribution we demonstrate how an analysis of the first passage time (FPT) distribution provides access to a wealth of additional information, including direct information on the organization of the underlying energy landscape. In particular, it enables us to resolve the signatures corresponding to distinct funnels, associated with local free energy minima and kinetically convergent pathways.

Single funnel landscapes correspond to good structure-seeking systems, such as magic number clusters and biomolecules that have evolved to fulfill a particular function. Multifunnel landscapes provide the opportunity to encode multiple functions, with the potential to act as molecular switches or bind alternative partners that stabilize particular structural motifs. The existence of multiple relaxation times could also be significant in terms of functionality. The dynamical signatures of landscapes capable of encoding multifunctional properties are therefore of great interest. Avoiding such kinetic traps is also a key principle in the design of efficient self-assembly.

Double funnel landscapes have previously been characterized for atomic clusters with competing low-energy morphologies. These clusters exhibit two characteristic slow relaxation time scales, one for direct relaxation to the global minimum and the slowest for paths that are diverted first to the competing funnel and must then escape from this kinetic trap. The present results show how the dynamical signatures of multiple funnels are connected to the organization of the landscape and how the observation time scale determines the outcome of corresponding measurements. Competing funnels that correspond to kinetic traps produce distinct peaks in the first passage time distribution on a logarithmic time scale, consistent with long time tails associated with protein folding and multieponential kinetics. Here we address the underlying kinetic transition network directly, to analyze how characteristics of the landscape, especially kinetic traps and the corresponding barriers, translate into features of the first passage time distribution. The theory applies to any network and any observation time scale where the kinetics are described by a linear master equation. This approach is complementary to schemes that focus on observables for single-molecule kinetics, where first passage time information has previously been exploited to make mechanistic inferences and detect.

Received: April 26, 2022
Accepted: June 16, 2022
Published: July 8, 2022
intermediate states. These previous applications to colloidal dynamics, folding kinetics, and transport through a pore provide some indication of the potential insight that might be achieved in future work. The present contribution reveals how multiple kinetic traps can be detected and visualized directly from an analysis of the first passage time distribution. It also demonstrates that corresponding features arise in the mean first passage time as the observation time scale of an experiment changes.

Analysis of the First Passage Time Distribution. We consider the first passage time distribution between product and reactant states, denoted $A$ and $B$, for a kinetic transition network (KTN). We assume that the network has already been created using some rare events methodology, which could be based on geometry optimization and coarse-graining into minima and transition states of the underlying energy landscape via discrete path sampling, or methods based on explicit dynamics. Having obtained a KTN, with nodes corresponding to local potential or free energy minima, and edges associated with forward and backward rates, the kinetics are extracted from a master equation formulation, as detailed in the Methods.

In the initial tests, the FPT distribution was calculated using both an analytical formulation and kinetic Monte Carlo (kMC) methods to validate the results and test convergence. The analytical formulation follows from a direct analysis of the rate equations via eigendecomposition. As shown in the Methods, the first passage time distribution $p(t)$ and the corresponding probability distribution for $y = \ln t$, $P(y)$, can be written as

$$p(t) = \sum \nu_i e^{-\nu_i z(t)} A_i \quad \text{and} \quad P(y) = \sum \nu_i e^{\nu_i y - \nu_i e^{y}} A_i \quad \text{(1)}$$

where $-\nu_i$ are the eigenvalues of the matrix defining the master equation dynamics, with $\nu_i > 0$, and the $A_i$ are amplitudes, which depend on the eigenvectors of the matrix and the initial condition defined by the reactants. Eigendecomposition results from routines DSYEVR and DGEEV were checked for consistency, as in previous work. Real negative eigenvalues are assured because the absorbing Markov chain derives from a Markov process that is reversible. The distribution for $\ln t$ is particularly useful because it separates the distinct time scales clearly, as illustrated below.

To examine the dependence of the MFPT on the observation time scale, we can integrate $p(t)$ up to a cutoff time $t_{\text{obs}}$, giving

$$\mathcal{T}(t_{\text{obs}}) = \sum_i \frac{A_i}{\nu_i z(t_{\text{obs}})} [1 - e^{-\nu_i z(t_{\text{obs}})} (1 + \nu_i t_{\text{obs}})] \quad \text{(2)}$$

where $z(t_{\text{obs}})$ is the normalization for the restricted distribution. Hence we might interpret this quantity in terms of transformations observed in an ensemble of molecules, where the experiment measures a signal corresponding to appearance of product or disappearance of reactant. We are therefore considering “feasible” kinetic processes, analogous to feasible operations that determine the appropriate molecular symmetry group and tunnelling splittings for nonrigid molecules.
When evaluated numerically the analytical results in eq 1 eventually suffer from loss of precision as the temperature decreases. Standard rejection-free kMC schemes also enable the FPT distribution to be extracted, but they become unfeasible at low temperatures because the number of steps increases rapidly as the temperature decreases. However, a formulation that samples the distribution of transition times for second-neighbor jumps (leapfrog kMC, Methods) is effective for the model landscapes used to test all the methodology. Results are first presented for one such model, which is representative of the test cases considered. Quantitative agreement was obtained between eigendecomposition, conventional kMC, and leapfrog kMC in the temperature ranges where alternative approaches were numerically tractable. The model is defined by a kinetic transition network, which includes all the information required to calculate the partition functions for the minima and transition states, using equilibrium statistical mechanics and unimolecular rate theory to extract thermodynamic properties and the rate constants between connected minima, as reviewed elsewhere. The corresponding landscape is designed to feature kinetic traps and an associated range of time scales, with local minima separated by low barriers in four branches of the disconnectivity graph shown in Figure 1. These minima exhibit multiple recrossings in kMC trajectories, highlighting the “flickering” problem in kMC simulations and increasing the escape time from the traps.

Figures 1 and 2 illustrate representative results at temperatures $T = 2$ and $T = 1$, below which the eigendecomposition becomes numerically unstable. Temperature is measured in the same units as the potential energy corresponding to Figure 1a. Four peaks are evident in the probability distributions $P(\ln t)$ and $h(\ln s)$ for the first passage time $t$ and the corresponding number of kMC steps $s$, which is also known as the dynamical activity. Hence these distributions report directly on the kinetic traps and the corresponding barriers that must be overcome to escape them. At the lower temperature the peaks are shifted to significantly higher values of $t$ and $s$, especially for the three slower features, which correspond to escape from the three deeper traps in the landscape (Figure 1). The peaks are also more clearly separated at $T = 1$. Hence there is a direct correspondence between the organization of the landscape and the probability distributions for the first passage time and the dynamical activity. We see quantitative agreement between the eigendecomposition and kMC results (where accessible) for these examples, which are representative of all the runs conducted at different temperatures for various model landscapes. The four peaks correspond to trajectories that reach $A$ via the four distinct branches on the landscape, starting from $B$.

The effect of cutting off the average for the mean first passage time at a maximum observation time $t_{\text{obs}}$ can be explored using eq 2. Selected results are shown in Figure 2. We see that $T_{AB}(t_{\text{obs}})$ exhibits four distinct steps corresponding to escape from the four kinetic traps. $T_{AB}(t_{\text{obs}})$ converges to the usual mean first passage time, $T_{AB}$, in the limit of long observation time scales, where

$$\lim_{t_{\text{obs}} \to \infty} T_{AB}(t_{\text{obs}}) = \sum_{i=1}^{l_{\text{max}}} \frac{A_i}{\nu_i} \sum_{j=1}^{l_{\text{max}}} A_j.$$

The four arrows in Figure 2b,d correspond to performing the sums in eq 3 up to $l_{\text{max}} = 1$, $l_{\text{max}} = 2$, and $l_{\text{max}} = 3$, where $\nu_{\text{max}}$ corresponds to the slowest relaxation. In the temperature range where the analytical results are reliable, the value obtained for

Figure 2. Probability distributions of $\ln s$ at temperatures (a) $T = 2$ and (c) $T = 1$ for the model landscape shown in Figure 1. For $T = 2$ two histograms are superimposed for kMC and leapfrog kMC results in orange and blue, respectively. (b, d) Associated MFPT $T_{AB}(t_{\text{obs}})$ as a function of the observation time scale cutoff $t_{\text{obs}}$ calculated from the analytical eigendecomposition formulation in eq 2. The red arrows mark the position of steps in $T_{AB}(t_{\text{obs}})$ predicted from eq 1, which correspond closely to the peaks in $P(\ln t)$ in Figure 1.
$T_{AB}$ agrees precisely with graph transformation,$^{37,46-49}$ which provides a numerically stable deterministic method for calculating this quantity.

Equation 2 shows that distinct features are expected in $T_{AB}(t_{obs})$ if the slow relaxation time scales defined by the values of $1/\nu$ are well-separated. This situation corresponds to kinetics where accurate fits might be obtained as a combination of exponential decay processes.$^{19,21,22}$ In contrast, for more complex systems, such as glasses, we would not expect to resolve such features.

**Results for Molecular Multifunnel Landscapes.** We now consider applications to multifunnel landscapes for an atomic cluster, a designed bistable peptide, and an RNA switch. Full details of how these databases were created are available in previous work; $^{16,50,51}$ here we show that an analysis of the first passage time distribution and the observation time scale provides new insight into the dynamics, with a direct connection to the underlying energy landscape.

The 38-atom cluster LJ$_{38}$ bound by the Lennard-Jones potential provides a well-understood example of a double-funnel molecule,$^{13,15,16}$ with low-lying structures based on close-packed (A) and icosahedral (B) structures (Figure 3). The same database and product/reactant definitions were used here as in previous work.$^{17}$ Two time scales are clearly visible in the probability distributions for ln $t$ and $T(t_{obs})$ for relaxation to either A or B when the initial occupation probability is localized in a high-lying minimum. These results are for a reduced temperature of $k_BT/\epsilon = 1.25$, just above the equilibrium temperature between the two morphologies, at the limit where eigendecomposition is numerically stable. For relaxation to A, the B region acts as a kinetic trap, and vice versa, producing the peaks in $P(\ln t)$ around $\ln t = 22$ and $\ln t = 20$. These longer time-scale features correspond to switching between the two competing morphologies. The peaks around $\ln t = 7$ and $\ln t = 5$ are for direct relaxation to A and B, respectively.

The precise features observed in $P(\ln t)$ depend on the initial distribution. For all three examples considered here the reactants were selected to highlight distinct relaxation time scales. Different experimental setups could probe the dependence of relaxation times on the preparation of the initial conditions, which could provide further insight into the organization of the landscape.

Figure 4 shows results for the 18-residue peptide DP5, which was designed to exhibit competing $\alpha$-helical and $\beta$-hairpin structures.$^{15}$ The interconversion mechanism is of interest for the insight it may provide into pathways to amyloid formation. A subset of the database was selected that supports low-lying hairpin and helical morphologies and the key interconversion pathway between them. Relaxation to both these structures from a higher energy minimum at 300 K reveals two peaks in $P(\ln t)$ corresponding to direct and indirect relaxation via the $\alpha$ and $\beta$ regions. In $T(t_{obs})$ exhibits two steps in each case.

Figure 5 shows the results for a nine base pair RNA duplex that features a noncanonical adenine–adenine base pair. NMR results show that this system exhibits switching between major and minor forms, and previous simulations indicate that the dominant kinetic pathways feature stacked intermediates.$^{50,54}$ Another funnel containing structures with the AS base flipped

![Figure 3](https://doi.org/10.1021/jacs.jpclett.2c01258)
out was found to be favorable at higher temperature, where the balance between enthalpy and entropy shifts. One of these minima with A5 flipped out was selected as the starting point, and the figure shows results for relaxation to three low-lying minima in the major and minor forms at 300 K. Three distinct time scales are discernible for the minor form target, with steps in \( \ln(t_{\text{obs}}) \) translating directly into peaks for the \( P(\ln t) \) distribution. The peaks corresponding to the two slower processes overlap, producing a shoulder in the distribution around \( \ln t = -6 \).

**DISCUSSION**

The mean first passage time between reactants and products is the usual focus of attention for an analysis of rates in systems that exhibit conventional single-exponential kinetics. However, for systems that exhibit multiple relaxation time scales the first passage time distribution encodes information that reports directly on the structure of the underlying energy landscape. Such behavior is expected for designed or evolved molecules that can act as switches, where the equilibrium between competing structures may be shifted by external factors. In particular, the ability of experiments to probe a wider range of time scales at higher resolution has the potential to provide new insight into such phenomena and probe the multifunnel/multifunctional hypothesis. Evolved biomolecules that perform different cellular functions provide a testing ground; for example, intrinsically disordered proteins may correspond to intrinsically multifunctional systems, where function is encoded in a multifunnel landscape.

To connect the features of the first passage time distribution with experiment we must take account of what is actually being measured; the question of what to calculate logically precedes the problems associated with how to do the calculation. The first issue is the definition of reactant and product states, which may combine structures with analogous experimental signatures that are separated by insurmountable barriers. These structures must be lumped together in the analysis of kinetics; otherwise, additional slow relaxation processes will interfere with the interpretation of the data. The accessible experimental time scale is also key. If our calculations can probe processes corresponding to kinetic traps that are not observable, then including these transformations in the analysis will produce calculated rates that are too slow. We should therefore focus on feasible processes, which produce observable experimental signatures in the same way that the molecular symmetry group is defined for nonrigid molecules in terms of observable tunnelling splittings. The results also suggest that caution may be required when rates are calculated by inverting mean first passage times, which may be a particular issue for multiscale or coarse-grained models.

The feasible molecular rearrangements also determine the appropriate free energy minima, since the concept of equilibrium is determined by the observation time scale and, hence, the interplay with kinetics. In this contribution, these issues have been addressed directly, by analyzing the first
passage time. The time scales associated with escape from kinetic traps in the energy landscape can produce clear signatures in the first passage time distribution and in the mean first passage time calculated as a function of the observation time scale. These features are resolvable for processes that correspond to well-separated time scales and become more distinct at a lower temperature. The effective time scale could also be tuned by changing experimental conditions, such as temperature or the strength of an applied field. The techniques developed here will enable detailed information about the organization of the energy landscape to be extracted and, hence, facilitate the design of multifunctional materials.

**METHODS**

**First Passage Time Distributions from Eigendecomposition.** The first passage time probability distribution \( p(t) \) can be obtained by considering the products as an absorbing state, since the dynamics are unchanged up to the point of absorption. \(^{35,37}\) Considering \( \mathcal{A} \leftarrow \mathcal{B} \) for specificity, and denoting the other intervening states as \( I \), the appropriate master equation for the occupation probabilities \( P_I(t) \) for states in \( I \cup \mathcal{B} \) is

\[
\begin{bmatrix}
\dot{P}_I(t) \\
\dot{P}_B(t)
\end{bmatrix} = \begin{bmatrix}
K_{II} - D_I & K_{IB} \\
K_{BI} & K_{BB} - D_B
\end{bmatrix} \begin{bmatrix}
P_I(t) \\
P_B(t)
\end{bmatrix}
\]

\[
\equiv MP_{IUB}(t).
\] (4)

Here \( K_{XY} \) is the matrix containing the individual rate constants between connected states, with dimension \( N_X \times N_Y \). \( D_X \) is a diagonal matrix of dimension \( N_X \times N_X \) with elements corresponding to the total escape rates from each state.

**Figure 5.** (a) Disconnectivity graph \(^{13,36}\) for a subset of the database \(^{39}\) for a nine base pair RNA duplex containing 29 minima. The FPT distributions correspond to relaxation from the minimum with adenine A5 flipped out, shown in full at the top left of the graph, to three target minima. Two of these structures correspond to the major form (red and blue) with A14 stacked between A6 and A15, and the third target is the minor form (green), where A5 is stacked between A6 and A15. The corresponding structures are illustrated for magnifications of the central base region to highlight the alternative base stacking. The two distinct components of the major funnel, highlighted in red and blue, differ by a 180° rotation of the A5 purine ring about the bond to the ribose ring. (right) FPT diagnostics calculated at 300 K from eigendecomposition; all times are in seconds. (b) \( \ln(T_{\text{obs}}) \) exhibits two clear steps for relaxation to the two major form minima corresponding to (c) bimodal probability distributions \( P(\ln t) \). In \( \ln(T_{\text{obs}}) \) for the minor form has an additional feature around \( \ln(T_{\text{obs}}/s) = -6 \), which produces a shoulder in the \( P(\ln t) \) distribution. (inset) Magnification of the region around \( \ln t/s = -20 \), revealing additional small peaks in \( P(\ln t) \).
in $X$, so that $D_X = \sum_j K_{ij}$. When the $A$ states are absorbing the escape rates from $A$ are zero, so $K_{X:A} = 0$, but transitions to $A$ appear in $D_Y$ and $D_B$. The formal solution to eq 4 is $P_{fUB}(t) = \exp(Mt)P_{fUB}(0)$. Equating the probability flux out of $I \cup B$ with the probability that the first passage time for $A \leftrightarrow B$ lies between $t$ and $t + dt$ implies that

$$p(t)dt = -(1_f P_I + 1_B P_B)dt \equiv -1_I P_I P_{fUB} dt$$

(5)

for initial distribution $P_I(0) = 0$, where $1_f$ is a row vector of ones with dimension $|I_f|$. Hence

$$p(t) = -1_I P_I M \exp(Mt) P_{fUB}(0)$$

(6)

which we now write in terms of the eigendecomposition of $M$

$$M = -\sum_i \nu_i w_i^R \otimes w_i^T$$

(7)

where $\otimes$ is the is the diadic (outer) product, the eigenvalues $-\nu_i < 0$ are negative for a connected network, and $w_i^R$ and $w_i^T$ are the left (row) and right (column) eigenvectors of $M$, with $w_i^R w_i^T = \delta_{ij}$. The probability distribution for the first passage time is therefore

$$p(t) = \sum_i \nu_i e^{-\nu_i t} 1_{fUB} (w_i^R \otimes w_i^T) P_{fUB}(0) \equiv \sum_i \nu_i e^{-\nu_i t} A_i$$

(8)

The moments of this distribution are the expectation values of $t^n$

$$\langle t^n \rangle = \sum_i \frac{n!}{\nu_i^{n+1}} 1_{fUB} (w_i^R \otimes w_i^T) P_{fUB}(0)$$

(9)

with MFPT $T_{AB}$ and variance $\mathcal{V}_{AB}$ corresponding to $\langle t \rangle$ and $\langle t^2 \rangle$, respectively.

In the present analysis we used two extensions of these results. To highlight the dynamical signature of a multifunnel landscape we need the probability distribution of $y = \ln t$, denoted $P(y)$, with

$$P(y) = p(t)/(dy/dt) = p(e^y) / (1/t) = e^y p(e^y)$$

$$= \sum_i \nu_i e^{y - \nu_i \exp(y)} 1_{fUB} (w_i^R \otimes w_i^T) P_{fUB}(0)$$

$$\equiv \sum_i \nu_i e^{y - \nu_i \exp(y)} A_i$$

(10)

We can also calculate the moments corresponding to experiments with a maximum observation time scale $t_{obs}$ by integrating up to this limit instead of infinity to obtain

$$\langle t^n \rangle_{obs} = \sum_i 1_{fUB} (w_i^R \otimes w_i^T) P_{fUB}(0) \frac{\Gamma(n + 1) - \Gamma(n + 1, \nu_i t_{obs})}{\nu_i z(t_{obs})}$$

(11)

with normalization

$$z(t_{obs}) = \int_0^{t_{obs}} p(t) dt = \sum_i \left(1 - e^{-\nu_i t_{obs}}\right) 1_{fUB} (w_i^R \otimes w_i^T) P_{fUB}(0)$$

(12)

In particular

$$T_{AB}(t_{obs}) = \sum_i 1_{fUB} (w_i^R \otimes w_i^T) P_{fUB}(0) \frac{\nu_i z(t_{obs})}{\nu_i z(t_{obs})} \left(1 - e^{-\nu_i t_{obs}}(1 + \nu_i t_{obs})\right)$$

$$\equiv \sum_i \frac{A_i}{\nu_i z(t_{obs})} \left(1 - e^{-\nu_i t_{obs}}(1 + \nu_i t_{obs})\right)$$

(13)

$T_{AB}(t_{obs})$ is the MFPT averaged over events that occur on the accessible observation time scale.

First Passage Time Distributions from Leapfrog Kinetic Monte Carlo. The analytical formulations break down at low temperatures because of numerical precision problems. The low-temperature regime is often the focus of interest, where the key transformations correspond to rare events. To access the FPT distribution we have previously employed kinetic path sampling (kPS), which uses the graph transformation (GT) approach. The graph transformation component enables kPS to be used in the metastable regime, so long as the subnetworks describe the metastable sets of nodes appropriately. This framework can provide first passage time distributions at temperatures where conventional kMC procedures are unfeasible due to flickering between nodes and excessively long path lengths. Graph transformation (GT) can be used to calculate the mean first passage time reliably at low temperature and provides a reference value for comparison with averages over the FPT distribution.

The FPT distributions for the model landscapes considered as test cases in the present work can be obtained efficiently using an extension of the leapfrog kMC scheme, introduced in earlier work to mitigate the flickering problem. In this method we employ renormalized branching probabilities and waiting times for the current minimum in a kMC simulation, to allow steps to second neighbors or directly connected absorbing states, if there are any.

Using the expectation value for the waiting time over all the implicit transitions conserves the overall MFPT between reactants and products, as in the GT approach. For the current state $i$, with neighbors $j \neq A$, we consider steps to any neighbor of $i$ in the absorbing set $A$ and to all second neighbors of $i$ that are directly connected to $j \neq A$, excluding $i$ itself. These events can be combined with any number of $i \leftrightarrow j$ recrossings in any order. The relative probability of $n$ total recrossings with $n_j$ for neighbor $j$ and $\sum_{j \neq A} n_j = n$ is

$$n! \prod_{j \neq A} (P_{ij})^{n_j} / \prod_{j \neq A} n_j!$$

(14)

which is a term in the multinomial expansion $(\sum_{j \neq A} P_{ij})^n$. Here $P_{ij}$ is the branching probability of moving to connected minimum $i$ from minimum $j$. The probability of $n$ total recrossings is

$$\left(\sum_{j \neq A} P_{ij}\right)^n / \left(\sum_{m=0}^{\infty} \left(\sum_{j \neq A} P_{ij}\right)^m\right) = \tilde{\beta}^n (1 - \tilde{\beta})$$

(15)

with $\tilde{\beta} = \sum_{j \neq A} P_{ij}$. The probability of stepping to any second neighbor $k \neq i$ with any number of recrossings is
$\sum_{k\neq i} \sum_{\mathcal{A} \not=i} P_{kij}^f / (1 - \beta_i)$, and the probability of stepping to any neighbor of $i$ in $\mathcal{A}$ is $\sum_{j \not=i} P_{ij} / (1 - \beta_i)$, which together sum to one. The path probabilities are conserved if we renormalize $P_{kij}^f = \sum_{\mathcal{A} \not=i} P_{kij} / (1 - \beta_i)$ for $k \neq i$ and $P_{ij}^f = P_{ij} / (1 - \beta_i)$ for $i \in \mathcal{A}$, and the MFPT to products is conserved by setting $\tau_i^f = (\tau_i + \sum_{\mathcal{A} \not=i} \tau_{kij}^f) / (1 - \beta_i)$, where $\tau_i$ is the expected waiting time for a transition out of minimum $i$.

**Sampling Distributions with Leapfrog Kinetic Monte Carlo.** To calculate the full FPT in the present analysis, rather than just the mean value that was the focus of interest in the original development, we sample the number of kMC steps and associated waiting time for each leapfrog move from the appropriate probability distribution. Repeated kMC trajectories then sample the FPT distribution, which generally appeared well-converged for 100,000 repeats using different random number sequences. The results presented are based on sets of 200,000 leapfrog kMC runs, which only required a few seconds of computer time in these tests. In contrast, conventional kMC simulations and eigendecomposition cannot access the FPT at lower temperature. Some of the leapfrog trajectories considered in the results correspond to over $10^{15}$ conventional kMC steps.

To sample the number of recrossings, we note that the probability of $N$ recrossings or fewer is $(1 - \beta_i) \sum_{n=1}^{N} \beta_i^n$. Generating a random number $x$ from a uniform distribution in the range $[0, 1]$ and solving $x = 1 - \beta_i^{n+1}$ produces $N = \ln(1 - x)/\ln \beta_i - 1$ as the sampled number of recrossings, each of which involves two kMC steps. Sampling the associated waiting time is more involved, because the waiting time associated with each recrossing depends on the neighboring node $j$.

The recrossing problem corresponds to sampling $m$ random variables, which may be collected in a column vector $(X_1, X_2, \ldots, X_m)^T$, distributed according to a multinomial distribution with terms $N! \prod_{a=1}^{m} \nu_{a}^{n_a} / \prod_{a=1}^{m} \nu_{a}!$, where $\sum_{a=1}^{m} \nu_{a} = 1$ and $\sum_{a=1}^{m} n_a = \sum_{a=1}^{m} X_a = N$. For transitions out of minimum $i$ this formulation corresponds to probabilities $\pi_j = P_{ij} / \sum_{\mathcal{A} \not=i} P_{ij}$ for large $N$ the multinomial distribution can be represented by a multivariate normal distribution, which can be sampled $^{37}$ by generating $m - 1$ independent random variables identically distributed as $\mathcal{N}(0, 1)$, with

$$
\begin{bmatrix}
X_1 \\
X_2 \\
\vdots \\
X_m 
\end{bmatrix} = \sqrt{N} \text{diag}(\mathbf{u}) Q \begin{bmatrix}
\sigma_1 \\
\sigma_2 \\
\vdots \\
\sigma_m 
\end{bmatrix} + N \begin{bmatrix}
z_1 \\
z_2 \\
\vdots \\
z_m 
\end{bmatrix}
$$

(16)

where $\mathbf{u}$ is the vector with components $u_a = \sqrt{\nu_a}$, $z_a$ is a random normal variable distributed as $\mathcal{N}(0, 1)$, and $Q$ is an orthogonal matrix with final column $\mathbf{u}$. $\text{diag}(\mathbf{u})$ is the diagonal matrix with elements $u_a$. Choosing $Q = \mathbf{L}_m - 2v \otimes v^T$ can be shown to provide an appropriate formulation, $^{37}$ for $v_a = (\delta_a - u_a)/\sqrt{2(1 - u_a)}$.

The multinormal sampling provides an approximate set of $n_i$ values and associated waiting times $n_i (\tau_i + \tau_j)$ for each leapfrog transition, which is accurate when $\sum_{\mathcal{A} \not=i} n_i = N$ is large. Multinormal sampling was therefore applied for minima with $\beta_i > 0.95$. Alternatively, good agreement with eigendecomposition and standard kMC can also be achieved using the mean waiting time for a single recrossing step $\tau_i^f$, defined above, multiplied by the number of recrossings sampled from the appropriate distribution. The contribution of the exit steps and waiting times to a second neighbor or adjacent absorbing state are added to the recrossing contribution. For these exit steps the waiting time was sampled from a Poisson escape distribution using $-\tau_i \ln x$ for a random number $x$ drawn from the uniform distribution $[0, 1]$ for each minimum $\sigma_j$ as for the standard kMC runs. Nodes with only a single possible recrossing were treated separately, since we only need to sample the total number of recrosings for one edge of the network in that case, and the multinormal distribution is not required.

Leapfrog moves were used when $\beta_i$ exceeded a given threshold; the results are not sensitive to this threshold, and all the values tested in the range from 0.1 to 0.9 worked very well. To avoid loss of precision when $\beta_i$ approaches 1 we use the identity $1 - \sum_{\mathcal{A} \not=i} P_{ij}^f = \sum_{\mathcal{A} \not=i} P_{ij} + \sum_{\mathcal{A} \not=i} \sum_{k \not=i} P_{kij}^f$, as in previous work, $^{56}$ where $j$ and $k$ run over first neighbors of $i$ and first neighbors of $j$, respectively, excluding return to $i$.

I am grateful to a referee for noting the existence of an earlier alternative leapfrog KMC scheme, $^{58,59}$ which could also have been used to check the eigendecomposition calculations. For more complex landscapes the kps formulation would probably be needed, $^{32,56}$ and as a further check kPS simulations were run for the model landscape in Figure 1 using the DISCOTRESS package. $^{60}$ The FPT distributions obtained from these independent runs agree very well with the results obtained using eigendecomposition and the leapfrog scheme described above.

■ AUTHOR INFORMATION

**Corresponding Author**

David J Wales – Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, U.K.; orcid.org/0000-0002-3533-6645; Email: dw34@cam.ac.uk

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpcl.2c01258

**Notes**

All the calculations were performed with the open-source program PATHSAMPLE, available for download from www-wales.ch.cam.ac.uk. The data that support the findings of this study will also be available for download.

The author declares no competing financial interest.

**Biography**

David J. Wales received his BA degree from Cambridge University in 1985, PhD in 1988, and ScD in 2004. He was a Lindemann Trust Fellow in 1989, a Research Fellow at Downing College Cambridge in 1990, a Lloyd’s of London Tercentenary Fellow in 1991, and a Royal Society University Research Fellow from 1991 to 1998. In 1998 he was appointed to a Lectureship in Cambridge and is now Professor of Chemical Physics and Chair of the Theory group. He is also an International Chair at the Interdisciplinary Institute for Artificial Intelligence at 3iA Cote d’Azur. He was awarded the Cambridge University Norrish Prize for Chemistry and the Gonville and Caius College Schulchard Plate in 1985, the Meldola Medal and Prize in 1992 and the Tilden Prize in 2015, both by the Royal Society of Chemistry. He was a Baker Lecturer at Cornell University in 2005, the Inaugural Henry Frank Lecturer at the University of Pittsburgh in 2007, Distinguished Lecturer at the National Institute of Standards and Technology, United States, in 2018, and was awarded a Visiting Miller
Professorship at the University of California, Berkeley, for 2020. He was elected a Fellow of the Royal Society in 2016. In 2020 he became the inaugural recipient of the ICReDD Award from Hokkaido University and received a Humboldt Research Prize from the Alexander von Humboldt Foundation. His research primarily involves the exploration of energy landscapes, with applications to chemical biology, spectroscopy, machine learning, clusters, glasses, solids, and surfaces. Group home page, including the Cambridge Landscape Database and software downloads: www-wales.ch.cam.ac.uk.

**Acknowledgments**

D.J.W. gratefully acknowledges Dr. D. Chakraborty, for providing some of the original databases for the biomolecular examples, and financial support from the Engineering and Physical Sciences Research Council under EP/N035003/1. The author is also grateful to two referees for their helpful comments on the original manuscript. Institutional rights retention statement: for the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission.

**References**

1. Longuet-Higgins, H. C. The symmetry groups of non-rigid molecules. *Mol. Phys.* 1963, 6, 445−460.
2. Ma, S. K. *Statistical mechanics: World Scientific; Singapore, 1985.*
3. Sibani, P.; Schön, J. C.; Salamon, P.; Andersson, J. Emergent hierarchical structures in complex-system dynamics. *Europhys. Lett.* 1993, 22, 479−485.
4. Baba, A.; Komatsuzaki, T. Construction of effective free energy landscape from single-molecule time series. *Proc. Natl. Acad. Sci. USA* 2007, 104, 19297−19302.
5. Wales, D. J.; Salamon, P. Observation time scale, free-energy landscapes, and molecular symmetry. *Proc. Natl. Acad. Sci. USA* 2014, 111, 617−622.
6. Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. Funneled pathways, and the energy landscape of protein folding: a synthesis. *Proteins: Struct., Funct. and Gen.* 1995, 21, 167−195.
7. Onuchic, J. N.; Wolynes, P. G.; Luthey-Schulten, Z.; Socci, N. D. Toward an outline of the topography of a realistic protein-folding funnel. *Proc. Natl. Acad. Sci. USA* 1995, 92, 3626−3630.
8. Leopold, P. E.; Montal, M.; Onuchic, J. N. Protein folding funnels: a kinetic approach to the sequence-structure relationship. *Proc. Natl. Acad. Sci. USA* 1992, 89, 8721−8725.
9. Chebaro, Y.; Balland, A. J.; Chakraborty, D.; Wales, D. J. Intrinsically disordered energy landscapes. *Sci. Rep.* 2015, 5, 10386.
10. Joseph, J. A.; Röder, K.; Chakraborty, D.; Mantell, R. G.; Wales, D. J. Exploring biomolecular energy landscapes. *Chem. Commun.* 2017, 53, 6974−6988.
11. Röder, K.; Joseph, J. A.; Husic, B. E.; Wales, D. J. Energy landscapes for proteins: from single funnels to multifunctional systems. *Adv. Theory Simulations* 2019, 2, 1800175.
12. Grant, J.; Jack, R. L.; Whitealam, S. Analyzing mechanisms and microscopic reversibility of self-assembly. *J. Chem. Phys.* 2011, 135, 214505.
13. Wales, D. J.; Miller, M. A.; Walsh, T. R. Archetypal energy landscapes. *Nature* 1998, 394, 758−760.
14. Miller, M. A.; Doye, J. P. K.; Wales, D. J. Structural relaxation in morse clusters: Energy landscapes. *J. Chem. Phys.* 1999, 110, 328−334.
15. Doye, J. P. K.; Miller, M. A.; Wales, D. J. The double-funnel energy landscape of the 38-atom Lennard-Jones cluster. *J. Chem. Phys.* 1999, 110, 6896−6906.
16. Wales, D. J.; Bogdan, T. V. Potential energy and free energy landscapes. *J. Phys. Chem. B* 2006, 110, 20765−20776.
17. Swinburne, T. D.; Kannan, D.; Sharpe, D. J.; Wales, D. J. Rare events and first passage time statistics from the energy landscape. *J. Chem. Phys.* 2020, 153, 134115.
18. Dutta, R.; Pollak, E. What can we learn from transition path time distributions for protein folding and unfolding? *Phys. Chem. Chem. Phys.* 2021, 23, 23787−23795.
19. Sabelko, J.; Ervin, J.; Gruebele, M. Observation of strange kinetics in protein folding. *Proc. Natl. Acad. Sci. U.S.A.* 1999, 96, 6031−6036.
20. Yang, W. Y.; Gruebele, M. Detection-dependent kinetics as a probe of folding landscape microstructure. *J. Am. Chem. Soc.* 2004, 126, 7758−7759.
21. Flynn, S. W.; Zhao, H. C.; Green, J. R. Measuring disorder in irreversible decay processes. *J. Chem. Phys.* 2014, 141, 104107.
22. Nichols, J. W.; Flynn, S. W.; Green, J. R. Order and disorder in irreversible decay processes. *J. Chem. Phys.* 2015, 142, 064113.
23. Cao, J.; Silbey, R. J. Generic schemes for single-molecule kinetics. 1: self-consistent pathway solutions for renewal processes. *J. Phys. Chem. B* 2008, 112, 12867−12880.
24. Li, X.; Kolomeisky, A. B. Mechanisms and topology determination of complex chemical and biological network systems from first-passage theoretical approach. *J. Chem. Phys.* 2013, 139, 144106.
25. Thorneycroft, A. L.; Gladrow, J.; Qing, Y.; Rico-Pasto, M.; Ritort, F.; Bayley, H.; Kolomeisky, A. B.; Keyser, U. F. Direct detection of molecular intermediates from first-passage times. *Science Advances* 2020, 6, 0.eaaz4642.
26. Noé, F.; Fischer, S. Transition networks for modeling the kinetics of conformational change in macromolecules. *Curr. Opin. Struct. Biol.* 2008, 18, 154−162.
27. Prada-Gracia, D.; Gómez-Gardenes, J.; Echenique, P.; Falo, F. Exploring the free energy landscape: from dynamics to networks and back. *PLoS Comput. Biol.* 2009, 5, No. e1000415.
28. Wales, D. J. Energy landscapes: some new horizons. *Curr. Opin. Struct. Biol.* 2010, 20, 3−10.
29. Wales, D. J. Discrete path sampling. *Mol. Phys.* 2002, 100, 3285−3306.
30. Wales, D. J. Some further applications of discrete path sampling to cluster isomerization. *Mol. Phys.* 2004, 102, 891−908.
31. Schütte, C.; Fischer, A.; Huisings, W.; Deufhard, P. A direct approach to conformational dynamics based on hybrid Monte Carlo. *J. Comput. Phys.* 1999, 151, 146−168.
32. Swee, W. C.; Pitera, J. W.; Suits, F. Describing protein folding kinetics by molecular dynamics simulations. 1. Theory. *J. Phys. Chem. B* 2004, 108, 6571−6581.
33. Chodera, J. D.; Singhal, N.; Pande, V. S.; Dill, K. A.; Swee, W. C. Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics. *J. Chem. Phys.* 2007, 126, 155101.
34. Husic, B. E.; Pande, V. S. Markov state models: from an art to a science. *J. Am. Chem. Soc.* 2018, 140, 2386−2396.
35. Swinburne, T. D.; Perez, D. Self-optimized construction of transition rate matrices from accelerated atomistic simulations with Bayesian uncertainty quantification. *Phys. Rev. Materials* 2018, 2, 053802.
36. Becker, O. M.; Karplus, M. The topology of multidimensional potential energy surfaces. Theory and application to peptide structure and kinetics. *J. Chem. Phys.* 1997, 106, 1495−1517.
37. Swinburne, T. D.; Wales, D. J. Defining, calculating, and converging observables of a kinetic transition network. *J. Chem. Theory Comput.* 2020, 16, 2661−2679.
38. Bortz, A. B.; Kalos, M. H.; Lebowitz, J. L. A new algorithm for Monte Carlo simulation of Ising spin systems. *J. Comput. Phys.* 1975, 17, 10−18.
39. Voter, A. F. Classically exact overlayer dynamics: Diffusion of rhodium clusters on Rh(100). *Phys. Rev. B* 1986, 34, 6819−6829.
40. Fichthon, K. A.; Weinberg, W. H. Theoretical foundations of dynamical Monte Carlo simulations. *J. Chem. Phys.* 1991, 95, 1090−1096.
(41) Wales, D. J. Exploring energy landscapes. Ann. Rev. Phys. Chem. 2018, 69, 401−425.
(42) Athènes, M.; Bulatov, V. V. Path factorization approach to stochastic simulations. Phys. Rev. Lett. 2014, 113, 230601.
(43) Sharpe, D. J.; Wales, D. J. Efficient and exact sampling of transition path ensembles on Markovian networks. J. Chem. Phys. 2020, 153, 024121.
(44) Lecomte, V.; Appert-Rolland, C.; van Wijland, F. Thermodynamic formalism for systems with markov dynamics. J. Stat. Phys. 2007, 127, S1−106.
(45) Newcomb, L. B.; Alaghemandi, M.; Green, J. R. Nonequilibrium phase coexistence and criticality near the second explosion limit of hydrogen combustion. J. Chem. Phys. 2017, 147, 034108.
(46) Trygubenko, S. A.; Wales, D. J. Graph transformation method for calculating waiting times in Markov chains. J. Chem. Phys. 2006, 124, 234110.
(47) Wales, D. J. Calculating rate constants and committor probabilities for transition networks by graph transformation. J. Chem. Phys. 2009, 130, 204111.
(48) Stevenson, J. D.; Wales, D. J. Communication: Analysing kinetic transition networks for rare events. J. Chem. Phys. 2014, 141, 041104.
(49) MacKay, R. S.; Robinson, J. D. Aggregation of Markov flows I: theory. Philos. Trans. R. Soc. A 2018, 376, 20170232.
(50) Chakraborty, D.; Wales, D. J. Dynamics of an adenine-adenine RNA conformational switch from discrete path sampling. J. Chem. Phys. 2019, 150, 125101.
(51) Chakraborty, D.; Chebaro, Y.; Wales, D. J. A multifunnel energy landscape encodes the competing α-helix and β-hairpin conformations for a designed peptide. Phys. Chem. Chem. Phys. 2020, 22, 1359−1370.
(52) Araki, M.; Tamura, A. Transformation of an alpha-helix peptide into a beta-hairpin induced by addition of a fragment results in creation of a coexisting state. Proteins 2007, 66, 860−868.
(53) Chen, G.; Kennedy, S. D.; Qiao, J.; Krugh, T. R.; Turner, D. H. An alternating sheared AA pair and elements of stability for a single sheared purine-purine pair flanked by sheared GA pairs in RNA. Biochemistry 2006, 45, 6889−6903.
(54) Van Nostrand, K. P.; Kennedy, S. D.; Turner, D. H.; Mathews, D. H. Molecular mechanics investigation of an adenine–adenine non-canonical pair conformational change. J. Chem. Theory Comput. 2011, 7, 3779−3792.
(55) Athènes, M.; Kaur, S.; Adjanor, G.; Vanacker, T.; Jourdan, T. Elastodiffusion and cluster mobilities using kinetic Monte Carlo simulations: Fast first-passage algorithms for reversible diffusion processes. Phys. Rev. Materials 2019, 3, 103802.
(56) Trygubenko, S. A.; Wales, D. J. Kinetic analysis of discrete path sampling stationary point databases. Mol. Phys. 2006, 104, 1497−1507.
(57) Georgii, H.-O. Stochastics: Introduction to Probability and Statistics; Walter de Gruyter: Berlin, Germany, 2007.
(58) Athènes, M.; Bellon, P.; Martin, G. Identification of novel diffusion cycles in B2 ordered phases by Monte Carlo simulation. Philosophical Magazine A 1997, 76, 565−585.
(59) Athènes, M.; Bellon, P.; Martin, G. Effects of atomic mobilities on phase separation kinetics: a Monte–Carlo study. Acta Mater. 2000, 48, 2675−2688.
(60) Sharpe, D. J. DISCOTRESS: Discrete State Continuous Time Rare Event Simulation Suite. https://github.com/danieljsharpe/DISCOTRESS, 2020.