

Energy Landscapes for Proteins: From Single Funnels to Multifunctional Systems

K. Röder,^{1,†} J. A. Joseph,^{1,†} Brooke E. Husic^{1,2} and D. J. Wales^{1,*}

1: Department of Chemistry, University of Cambridge, Lensfield Road, CB2 1EW, Cambridge, UK
2: Current affiliation: Chemistry Department, Stanford University, 333 Campus Drive, Stanford, CA 94305, USA

* Corresponding author: dw34@cam.ac.uk † These authors contributed equally.

Article type: Progress Report

Keywords: Energy landscapes, protein folding, multifunnel paradigm

In this report we advance the hypothesis that multifunctional systems may be associated with multifunnel potential and free energy landscapes, with particular focus on biomolecules. We compare systems that exhibit single, double, and multiple competing structures, and contrast multifunnel landscapes associated with misfolded amyloidogenic oligomers, which presumably do not arise as an evolutionary target. In this context, intrinsically disordered proteins could be considered intrinsically multifunctional molecules, associated with multifunnel landscapes. Potential energy landscape theory enables biomolecules to be treated in a common framework together with self-organising and multifunctional systems based on inorganic materials, atomic and molecular clusters, crystal polymorphs, and soft matter.

1 Introduction

The field of protein science underwent dramatic transformations with the groundbreaking experiments of Sanger^[1] and Kendrew,^[2] characterising the sequence of insulin and the three-dimensional structure of myoglobin, respectively. New questions were raised by *in vitro* studies on the refolding of ribonuclease.^[3] When treated with a chemical denaturant facilitating the cleavage of all disulphide bonds, the resulting denatured conformation contained no evidence of its native structure. However, when the denaturant was removed the protein spontaneously refolded. Anfinsen hypothesised that the ‘three-dimensional structure of a native protein in its normal physiological milieu is the one in which the Gibbs free energy of the whole system is its lowest,’ now known as the Thermodynamic Hypothesis.^[4–6] It follows from the Thermodynamic Hypothesis that the native conformation of the protein is determined entirely by its amino acid sequence in thermodynamic equilibrium.

Given that protein structure is governed by the amino acid sequence, Levinthal showed how improbable it would be for a protein to fold to its free energy minimum by sampling all possible amino acid conformations. He explained further that even if the dimensionality of the problem was reduced (e.g. by considering only the backbone and side-chain rotations), the time taken to fold by a random search would be unrealistically long.^[7] This dilemma, commonly known as Levinthal’s Paradox, has been extensively reviewed.^[8;9] While Levinthal was not able to reconcile theory and experiment, he was amongst the first to pioneer a number of important

ideas, including computational studies of protein folding^[10] and discussions on the existence of special folding pathways.^[11]

Levinthal’s thoughts on protein folding propelled subsequent discussions on the theory, and many of the models developed in the next three decades aimed to reconcile the apparent paradox, often based on the notion was that there must be some simplified mechanism explaining how proteins fold. In the present contribution we discuss how Levinthal’s Paradox fits into the broader context of self-organising, or structure-seeking systems, and show how this perspective can be extended to the more recent descriptions of double- and multifunnel systems.

The observation of ‘magic number’ atomic and molecular clusters, crystallisation, and formation of mesoscopic structures such as virus capsids, highlights the feasibility of events that appear to be statistically implausible in terms of random searches.^[12;13] The common principle that unifies our understanding of structure-seeking lies in the organisation of the underlying potential energy surface as a kinetic transition network with funnelling characteristics, where relaxation is effectively guided energetically downhill into a single structural ensemble. The ability to explore, sample, and calculate observable properties from this potential energy landscape provides the foundations for a theoretical framework that now extends beyond single funnels to double funnel^[14–17] (§7.2) and multifunnel (§7.3) systems with multifunctional capabilities.^[18;19] The evolution of these ideas is discussed below after we first summarise some of the experimental and computational approaches that have been developed to interrogate the properties of biomolecules. These sections are intended to provide some background for non-specialists, not a detailed review.

2 Overview of Experimental Techniques

2.1 Protein folding initiation methods

Protein folding is usually initiated from the denatured state. This process involves perturbing the prevailing conditions to produce a non-equilibrium ensemble, which can then relax to a new equilibrium state.

Rapid mixing-based methods, namely stopped-flow^[20–24] and quenched-flow,^[25–27] are historically the most common techniques employed to trigger protein folding. Generally, a denaturing agent is first used to unfold the protein. The protein-denaturant solution is then diluted by rapidly mixing a buffer that favours folding. Rapid-mixing techniques are appealing, since no chemical alterations need to be made to the protein under investigation. However, these methods suffer from limited time resolution due to the inherent dead times of the mixing apparatus, which can be on the order of milliseconds, compared to the fastest folding events that occur in nanoseconds or microseconds.

Alternatives, such as the continuous-flow technique,^[28–30] offer an improvement in time resolution, with dead times in the microsecond regime. Faster folding events, which occur in the early stages of folding, cannot be probed using these approaches. Nonetheless, rapid mixing-based methods have been instrumental in probing folding intermediates, particularly molten globules, as well as providing evidence in support of early secondary structure formation during folding.

Laser-induced temperature jumps are widely used to study protein folding events on the sub-millisecond time scale; initiated from cold- and heat-denatured states, as well as from folded states.^[31–41] For example, significant lowering of the temperature below physiological values leads to protein unfolding,^[42–44] and refolding may then be initiated by a rapid jump in temperature from the cold-denatured state. The observed kinetics are a combination of the forward and backward rates for this process,^[45] and can be used to determine folding time scales. Short laser pulses can excite the infrared (IR) vibrational modes of water, which

generally relax on the picosecond time scale, to produce a ultra-fast jump in temperature. The rapid temperature jump destabilises the denatured state and the protein subsequently refolds, allowing the study of fast folding events, such as secondary structure formation. Studies employing laser-induced temperature jump techniques were among the first to provide time-resolved structural dynamics for the helix-coil transition^[34] and β -hairpin formation.^[33]

Complementary to temperature-induced folding, rapid changes in pressure can also induce protein folding or unfolding.^[46–49] Generally, proteins denature under high pressures and may relax to the native state following a negative jump. Pressure perturbations can significantly alter the rate constant for folding; thus, reductions in the folding rate, via appropriate pressure jumps, can be employed to stabilise folding intermediates and characterise them.^[50;51] Microsecond timescales for folding are accessible,^[52] and observed mechanisms can be compared and contrasted to temperature-jump results.^[53]

Another means of initiating protein folding is via rapid electron transfer. This method is particularly useful in studies involving redox-active proteins, such as cytochrome *c*.^[54;55] Photo-induced ligand dissociation may also be utilised to trigger protein folding. Carbon monoxide (CO) is known to bind to the haem group of proteins such as myoglobin and cytochrome *c* and lead to unfolding. Rapid photolysis of the CO ligand causes the proteins to refold.^[56;57] Since the dissociation can occur on the sub-picosecond time scale, this technique is very useful in probing relatively fast folding events on a microsecond time scale.^[58] A more generally applicable approach is to engineer a photo-trigger into the protein, which can stabilise the unfolded state, and then be photo-cleaved irreversibly to initiate folding.^[59;60] Conversely, reversible folding is achieved using photo-switches that initiate folding via photo-induced isomerisation.^[61;62]

Mechanical force can also be employed to control protein folding. Atomic force microscopy^[63;64] and optical tweezers^[65;66] facilitate single-molecule protein folding studies, which are generally not possible using other techniques. The Jarzynski equality^[67–69] relates the free energy difference between two states and the irreversible work done on the system along a trajectory connecting them. Hummer and Szabo showed that this equality can be extended^[70;71] to a Liouville-type equation, e.g. describing systems evolving according to diffusion, Newtonian, Langevin, or Metropolis Monte Carlo dynamics, using the Feynman-Kac theorem. Based on their results, the free energy surface of a molecule can sometimes be reconstructed from single molecule manipulation experiments using atomic force microscopy or optical tweezers. For example, such calculations can be based on the cantilever positions in the experiment, and the resulting pulling curves. The theory can be applied to obtain kinetic data from experiments,^[72] intrinsic rates, activation free energies,^[73] and parameters defined with the framework of Kramers’ theory.^[74] Although the resolution is limited by Brownian motion, time resolution in the microsecond regime can be achieved.^[75]

While these methods probe the dynamics of protein folding by introducing non-equilibrium into the system, it is noteworthy that information about the folding process may also be collected under steady-state conditions, where a short equilibration time is introduced. In experiment, this is achieved through variations in temperature or concentration of chemical denaturants. In Differential Scanning Calorimetry (DSC),^[76;77] the heat capacity is recorded as a function of temperature and compared to a sample with no phase transition in the observed range. The differential between the observed heat capacities provides information about unfolding/folding events. DSC can be used on complex samples, such as blood plasma^[78] or cerebrospinal fluid,^[79] providing insight for *in vitro* behaviour of proteins. Measuring the conformational stability of proteins with respect to denaturing agents, such as urea or guanidine hydrochloride, yields solvent denaturation curves.^[80–82] Key applications of these measurements have been the study of contributions to the stability of folded proteins, especially the importance of hydrophobic interactions and hydrogen-bonds,^[83] and the interpretation of equilibrium-folding models.^[84]

2.2 Structural and kinetic probes

Pioneering work by Kendrew and colleagues in deciphering the structure of myoglobin via X-ray crystallography^[2] ushered in a new era of protein discovery. Although the spacial resolution of the initial X-ray crystal structure was low (approximately 6 Å), the details suggested that an intricate connection existed between protein structure and function. Subsequent X-ray structures for haemoglobin,^[85] lysozyme,^[86] and ribonuclease,^[87;88] among other proteins, further elucidated this connection. To date, X-ray crystallography has been the most extensively used technique to determine protein structure,^[89] accounting for over 80% of protein entries in the Protein Data Bank (PDB). The main challenge of this technique is in obtaining single crystals for X-ray diffraction. Membrane proteins, multidomain proteins that consist of flexible linkers, and intrinsically disordered proteins, are often difficult to crystallise, prohibiting characterisation by traditional X-ray diffraction. Furthermore, only a static representation of the protein is achieved, and the conformation adopted in solution may differ from the crystal structure. As an extension to this method, small angle X-ray scattering (SAXS), has been successfully applied to provide time-resolved structural data for proteins in solution, albeit at lower spacial resolution.^[90–97]

Cryo-electron microscopy (cryo-EM) uses the interactions of electrons with biomolecules to create images, as in transmission electron microscopy, but the system is kept at cryogenic temperatures.^[98] This technique provides atomic resolution for biomolecules close to their physiological environments,^[99;100] as the freezing happens rapidly, prohibiting rearrangements as observed in crystallisation.

Nuclear magnetic resonance (NMR) spectroscopy provides a wealth of information for protein folding,^[101–104] for example on the secondary structure, from backbone chemical shifts,^[105] and on exchange rates from line broadening^[106] and relaxation dispersion techniques^[107] employing multidimensional NMR.^[108] These procedures are commonly used to probe protein dynamics in the μ s to ms regime.^[109]

Scalar coupling constants and ^1H - ^1H cross peaks (COSY and NOESY) provide local distance constraints, and are particularly useful for structural characterisation of the folded state.^[103;110] NMR spectra for partially folded states suffer from poor dispersion of ^1H and ^{13}C resonances, and therefore fewer NMR restraints are available for structural characterisation.^[111] Two-dimensional ^1H - ^{15}N NMR correlation spectroscopy techniques, such as heteronuclear single quantum coherence spectroscopy (HSQC),^[112] provide reasonably dispersed spectra for these states, and ^{15}N relaxation data can be used to quantify backbone motion.^[111] Amide hydrogen-exchange techniques coupled with NMR are particularly useful for probing partially folded states.^[113]

The determination of backbone configurations (helical axis/backbone curvature) and the relative orientation of multi-domains in larger proteins may be investigated using residual dipolar coupling (RDC).^[114;115] RDCs have also been used to probe the structure and dynamics of unfolded^[116] and intrinsically disordered^[117] proteins.

In chemical exchange saturation transfer (CEST)^[118] experiments, different relaxation time scales in biomolecular systems are exploited. The solvent molecules in the bulk exhibit different relaxation behaviour due to their essentially homogeneous environments, while the solvent molecules interacting with the solute relax according to solvent-solute interactions. This difference allows the observation of exchange rates between occupied states,^[119] even for states with low occupancy. Analysis of the potential energy landscape for one particular example, namely the metabolite creatine, revealed a clear separation of configurations corresponding to proton transfer from zwitterionic creatine to water.^[120] Finally, for membrane proteins and amyloid fibrils, solid state NMR spectroscopy can be used.^[121]

Another useful technique for probing both equilibrium and kinetic aspects of protein folding (and unfolding) in solution is circular dichroism (CD),^[122–127] which is based on the difference

in absorption of right- and left-handed circularly polarised light by optically active molecules. The resolution of CD spectroscopy is low, but it is a fast method, which can be used in solution with variable pH, temperature and salinity. The method allows access to information about the secondary and tertiary structure,^[123;124] and can be used to monitor structural changes, providing insight into folding mechanisms and stability^[128] and folding intermediates.^[129] Furthermore, it can quantify structural changes in related proteins (e.g. mutants),^[130;131] and follow conformational transitions in proteins that are prone to misfolding and aggregation,^[132] such as amylin^[133–135] and β -amyloid.^[136–138]

Infrared (IR) spectroscopy is another well-established technique for investigating protein structure and dynamics,^[139;140] requiring moderate effort, no labelling, and yielding high temporal resolution (below 1 μ s).^[141–143] While complete structure prediction is challenging, techniques such as difference IR spectroscopy are very useful in detecting conformational changes, and are often employed to probe reaction sites (e.g. in enzyme-substrate interactions), folding intermediates, and protein flexibility in general.^[139;140] Two-dimensional infrared spectroscopy offers significant improvements for probing protein conformational dynamics on the sub-picosecond time scale, since vibrational modes are extremely sensitive.^[144] Along with Fourier transform infrared spectroscopy (FTIR),^[145] this technique has been instrumental in probing protein misfolding and aggregation, providing key insight into the evolution of various neurological disorders.^[146] As a complementary technique to IR spectroscopy, Raman spectroscopy also provides highly sensitive signatures for the secondary and tertiary structure of proteins.^[147;148]

Fluorescence techniques to interrogate protein folding^[40;149;150] either exploit native^[151;152] or non-native fluorescence.^[153] The quantum yields of fluorescent probes are extremely sensitive to the local environments and mobility, and fluorescence provides a high time resolution. In standard fluorescence studies, detailed structural characterisation is inhibited due to the inherent local nature of signals. Hence, these studies are most useful when performed in conjunction with other techniques, such as FTIR, CD and NMR.

Alternative approaches, namely Förster resonance energy transfer (FRET)^[154;155] and fluorescence correlation spectroscopy (FCS),^[156] provide more direct spacial information. In FRET studies, coupling effects between donor and acceptor fluorescent pairs are used to characterise protein folding events in terms of the corresponding distance.^[157] FRET has been used in interaction studies of proteins,^[158] in single-molecule protein folding studies,^[159–161] and even *in vivo*.^[162] FCS enables relaxation times to be computed.

3 Overview of Computer Simulations

Protein folding simulations can probe atomistic details of the folding process not amenable to most experimental techniques; in particular, high spacial (distance) and temporal (fastest motions) resolutions are achievable. The accuracy of protein simulations depends largely on the form and parameters of the energy functions used to represent the protein and the surrounding environment, and the sampling of conformational space. These factors are intrinsically linked to the available computing hardware, which can limit the length scales and duration of computer investigations.

3.1 Protein simulation techniques

The most basic protein simulations focus primarily on refining experimental structures, where model coordinates are derived mainly from X-ray crystallography studies.^[10;163–165] The earliest study employed coordinate fitting procedures (model building) to refine atomic coordinates.^[163] Subsequent refinement procedures sought to minimise the potential energy of the system with

respect to the Cartesian coordinates. In one treatment, Levitt and Lifson^[164] defined the potential energy as a function of the bond lengths, bond angles, dihedral angles, and non-bonded pairs, along with a constraint term, which ensured that the deviations of the atomic coordinates from experiment were kept to a minimum. Equilibrium bond lengths and angles were obtained from X-ray structures of small molecules, and torsional parameters were taken from the Ramachandran plot. These types of coordinate refinement procedures were used to optimise the geometries of single-domain globular proteins^[164–166] and to compute the conformational preferences of side-chains.^[167;168] Alternative energy minimisation procedures perturbed internal coordinates (dihedral angles) to search for lower energy structures.^[169]

These initial studies shed light on the complexity of the conformational space and revealed that, even in the vicinity of the native state, proteins exhibited significant conformational heterogeneity.^[170] To achieve better conformational sampling, Levitt and Warshel introduced a coarse-grained representation for proteins^[171] consisting of the C_α atom and the centroid of the side-chain. The key assumption was that a separation of time scales exists, permitting time-averaging of short-range motion without significantly affecting the folding process. They proposed that in the early stages of folding, long-range forces played a central role in restricting the conformational space by directing protein collapse.^[171]

Monte Carlo (MC) methods have also been extensively applied to proteins.^[172] Although these techniques do not probe protein dynamics directly, they can provide valuable information about the topography of the folding space. Exploration of conformational space with MC methods has been used extensively in conjunction with lattice models.^[173–175] Although simplified, lattice models of proteins played an important role in shaping the free energy landscape view of protein folding. In 1994, Hao and Scheraga^[176] suggested that folding may involve a first-order transition between the native state and the unfolded protein. In the same year, Socci and Onuchic^[177] used lattice models to classify polypeptides as either good- or non-folding sequences, where good folders exhibit a folding temperature that exceeds a hypothetical glass transition temperature.

Atomistic molecular dynamics (MD) is probably the most commonly used technique to simulate proteins.^[178–180] Due to limited computing power, early MD work^[181–183] often probed motion in the neighbourhood of the native state, which was employed to refine X-ray/NMR structures. In 1981, Northrup *et al.* reported that B factors measured in X-ray crystallography studies showed good agreement with the mean-square fluctuations of atoms in MD simulations.^[183] The stability of hydrogen-bonds was also directly related to proton exchange rates in NMR studies, and the role of hydrogen-bonding in the folded state was investigated at high spatial resolution.^[182] Additionally, to achieve greater computing efficiency, NOE distances were used as constraints for structure prediction from extended states.^[184]

Advances over the past two decades in computational hardware, software, and the underlying theory have enabled MD simulation to access increasingly long timescales.^[185] The D. E. Shaw Research group has developed a custom supercomputer, Anton, which contains chips specifically designed for MD calculations.^[186] In 2010, Shaw and coauthors reported a 1 ms simulation of the 58-residue bovine pancreatic trypsin inhibitor (BPTI) protein in the folded state. For comparison, the first protein MD simulation was performed on the same system by McCammon, Gelin, and Karplus in 1977^[181] for a duration of 3.2 ps. Advances in distributed computing in platforms, such as Folding@home,^[187] IBM BlueGene,^[188] and GPU-GRID,^[189] have also extended accessible timescales for MD.

The hardware and software advances described above have generated protein folding datasets from atomistic MD with multiple folding and unfolding events.^[190–193] In particular, Lindorff-Larsen *et al.*^[193] reported MD simulations of twelve small fast-folding proteins with a variety of secondary structure combinations,^[194] and these datasets have been used for many subsequent analyses.^[195–198]

3.2 Enhanced sampling of protein conformational space

The barriers associated with folding can be high enough so that in a typical simulation the frequency at which the system acquires sufficient energy to cross them is low. Hence, folding can be regarded as a ‘rare event’, and standard simulations can be trapped in a local region of configuration space, a situation referred to as broken ergodicity. To achieve improved sampling numerous MC/MD based algorithms and protocols have been proposed.

A common way to address broken ergodicity is the introduction of a bias into the simulation, such as an elevated temperature. This approach drives unfolding and probes the corresponding variation in structure and dynamics. Similarly, pressure variations, low pH, and denaturants can be used to unfold proteins and to study the effects of solvent on protein denaturation.^[199;200] For example, MD simulations were able to probe, in atomistic detail, the disruption of hydrogen-bonds in proteins by water during denaturation.^[201] However, questions have been raised on the extent to which the principle of microscopic reversibility would hold under such non-equilibrium conditions, and on how much information about the reverse process (folding) can be inferred from unfolding.^[202] Furthermore, while elevated temperatures increase the likelihood of overcoming high energy barriers, they reduce the sampling in lower energy regions.

Replica exchange methods (REMs), specifically parallel tempering (PT) approaches,^[203;204] provide a powerful enhanced sampling framework. Here, multiple independent copies of the system are simulated at different temperatures via MC^[203] or MD sampling.^[204] Exchanges between replicas are attempted after some fixed interval (MC steps or MD time steps), or alternatively, the temperatures of adjacent copies may be swapped. The temperature distribution should be chosen to allow a uniform acceptance probability over the range of temperatures in the simulation, and a geometric progression has been found to fulfil this criterion.^[205] Further analysis based on the potential energy landscape perspective enables optimal temperature spacings to be estimated under general conditions.^[206]

The high temperature replicas are more likely to overcome energy barriers, providing a better global sampling, while the low temperature replicas provide more detailed local sampling. Due to exchange between replicas, the trajectories are discontinuous. As a result the system does not obey a master equation. Nonetheless, replica exchange MD (REMD),^[204] in particular, has been applied to study a wide range of protein systems, including small peptides in explicit/implicit solvent,^[207–209] amyloid-forming proteins,^[210;211] and chaperones.^[212] REMs can also take advantage of parallel computing and have also been adapted to run on distributed computing, as in multiplexed-REMD.^[213]

A more direct form of biasing involves umbrella sampling.^[214] Here a biasing potential (umbrella) is added and changes along a predefined coordinate, to direct sampling towards specific regions. The first protein studies employing umbrella sampling were published in the 1980s,^[215;216] for example probing slow conformational changes in proteins.^[215] Free energy surfaces based on the biased simulations (potential of mean force) can be derived using the weighted histogram analysis method (WHAM).^[217;218] This analysis may retain some artefacts of the bias if convergence is not reached for each histogram.^[219] Newer methods^[219–223] overcome this problem by including kinetic information from the simulation.

During the 1990s Brooks and colleagues^[224–228] presented several studies using an analogous biased-sampling technique, which they used to study peptides in solution. Other related approaches, which probe the free energy via biased potentials, include metadynamics,^[229–231] steered MD,^[232;233] and targeted MD.^[234;235] The latter two methods may be useful for studying protein-ligand binding (and unbinding) and the effects of mechanical force on protein structure, similar to AFM experiments. Alternatively, energy ranges may be used to focus the sampling, as in the multicanonical^[236;237] and Wang-Landau^[238] approaches.

Two key drawbacks are generally encountered in biased simulations. First, the choice of a

reaction coordinate or collective variables may introduce systematic errors.^[239] Furthermore, the bias leads to an alteration of the landscape itself. The elongation of amyloid fibrils provides an illustrative example. Bacci *et al.*^[240] showed that some features encountered in high temperature unfolding,^[241] replica exchange MD,^[242] and umbrella sampling^[243] were at least partly introduced by the methods used. Zheng *et al.*^[244] compared simulations with and without biasing for the elongation process and encountered significant differences. As discussed in §6, simulation methodology based on geometry optimisation to explore the potential energy landscape is largely complementary to the techniques discussed so far. Using this approach it was shown that the landscape for fibril elongation is multifunneled in character,^[245] which means that biasing schemes are likely to average over regions of configuration space that are actually separated by relatively high barriers. This problem is associated with low-dimensional projections of the landscape,^[246–254] and necessitates careful interpretation of the corresponding results. Despite this caveat, the studies using biased setups^[241–243] were successful in providing important data to improve our understanding of the elongation process.

The second set of methods that is commonly employed to study energy landscapes, alluded to above, is based on a master equation and is discussed in §6. The configuration space is partitioned into a number of distinct states, and reactive trajectories that connect them are analysed.

Sampling schemes that produce the energy density of states provide access to the free energy via the partition function. The last techniques that we will mention here are nested sampling, introduced by Skilling,^[255] and basin-sampling.^[256] Nested sampling has been applied to a variety of problems, including hard spheres,^[257] atomic clusters,^[258] phase diagrams,^[259–261] and even quantum partition functions.^[262] Superposition enhanced nested sampling (SENS)^[263] combines the efficiency of global optimisation techniques in the low energy region of the landscape with nested sampling for the high energy regions. The same philosophy underpins the basin-sampling approach, where a two-dimensional density of states is defined, and fitted to a model anharmonic form for bins in the potential energy distribution of local minima.^[256] Evaluating the efficiency of such methods for biomolecules could lead to further advances in tackling broken ergodicity.

4 Models of protein folding

Sequential and nucleation models were among the first suggestions for how proteins fold. The premiss of these theories was that since the volume of the configuration space is so large, there must be some initial event or unique sequence of events that leads to folding, thus reducing the subsequent number of possibilities. Levinthal considered the case for specific folding pathways, which might guide the protein from the denatured state to the folded state.^[11] In the Cooperative Sequential Model^[264] it was argued that, in addition to following a unique pathway, protein folding is initiated by a nucleation event. A similar model was proposed by Wetlaufer, the Nucleation Model,^[265] in which the rate of folding depends on the formation of an initial nucleus. The subsequent growth was suggested to occur either by adding neighbouring residues (fast kinetics) or distal residues (slow kinetics). In subsequent nucleation-based models, namely the Stageswise Mechanism^[266] and the Cluster Model,^[267] it was proposed that there exist multiple nucleation sites (so-called ‘centres of crystallisation’ or clusters) along the polypeptide chain, which eventually merge or collapse to produce the native structure.

Around the same time, Karplus and Weaver presented arguments for the Diffusion-Collision Model,^[268–270] where protein folding begins with the formation of transient segments of secondary structure, or microdomains. The authors suggested that if microdomains were made up of a small number of amino acids, the protein could efficiently sample all the available conformations for a given microdomain, thereby avoiding Levinthal’s paradox. Once formed,

microdomains were supposed to move diffusively and collide with each other. The rate-limiting step was attributed to the formation of microdomain intermediates, produced when collisions lead to coalescence. Unlike classical nucleation models, the Diffusion-Collision model provided recipes for extracting quantitative information about the folding process; for example, formulations for the folding rate were derived based on the physical properties of the microdomains.^[270]

In the Noninteracting Local Structure Model,^[271] a statistical mechanical approach to the protein folding problem was taken. A local structure was defined as a continuous segment of the polypeptide chain that adopted an equivalent conformation in the native structure. An important element of the model is that the interactions between local structures are assumed to be negligible. Additionally, the free energy of a given local structure is estimated from the atomic coordinates of the native configuration, thus providing a means of computing the partition function. This simplified model was used to demonstrate how protein folding might proceed by first forming local structures, which subsequently grow or merge (Growth-Merge Model)^[271] to yield the native structure.

Kim and Baldwin described the Framework Model^[272] for protein folding, as experimental evidence suggested the existence of folding intermediates that contained significant secondary structure. Consequently, the authors argued that during folding, hydrogen-bonded secondary structure is formed first, followed by tertiary interactions. Hence, in the Framework Model, it was suggested that stable secondary structure formed independently of tertiary structure.

Another proposed mechanism by Dill attributed protein folding to a driving force from the association of hydrophobic residues to avoid contact with the solvent, the Hydrophobic Collapse Model.^[273] The protein would undergo rapid collapse around the hydrophobic side-chains and then fold slowly, from the compact intermediate to the native state. In this scenario, the intermediate contains little secondary structure, in direct contrast to the Framework Model.^[272]

Finally, in the Nucleation-Condensation Model,^[274] arguments for two-state folding without folding intermediates were presented. Unlike earlier nucleation models, in which the nucleus was defined as a small incipient localised region, in this model the nucleus was assumed to be large and diffuse, and emerged in the transition state. Thus, the nucleus would correspond to the best-formed interactions in the transition state, and be stabilised by both local and long-range interactions. Moreover, in nucleation-condensation, secondary and tertiary interactions occur simultaneously, leading to a two-state folding process.

The preceding models, though not an exhaustive list, provide a summary of how ideas evolved in the field of protein folding after Anfinsen’s experiments. Although the initial impetus was to resolve Levinthal’s paradox, many of the later theories sought to account for more detailed experimental observations. However, to derive a theoretical framework capable of explaining self-organisation in general requires a more fundamental perspective, as we discuss below in §4.1 and §5. In principle, observable properties including structure, dynamics, and thermodynamics, are encoded in the underlying potential energy surface. Developing the potential energy landscape as both a conceptual and a computational tool has enabled us to understand how emergent properties depend upon the organisation of the landscape, and to explain how efficient relaxation to a single well-defined structure or structural ensemble can arise. In this context, naturally occurring biomolecules that have evolved to fulfil a single function take their place alongside systems such as ‘magic number’ clusters of atoms and molecules, which are selected by their special stability under evolution in specific conditions, such as molecular beams.^[12;13] The ability to visualise the structure of the landscape via disconnectivity graphs^[14;275] enabled such systems to be identified with single funnel motifs (§7.1). More recently, this paradigm has been extended to associate multifunctional systems with multifunnel landscapes.^[18;19] Once again, we find that nature has already exploited this design principle in evolved biomolecules.

4.1 Free energy landscapes

Initial considerations of free energy landscapes for proteins came from analogies suggested by glassy systems.^[56;276] To quantify this idea, Bryngelson and Wolynes formulated a Hamiltonian based on spin-glass models^[277] to describe energetic frustration,^[278] defined in terms of competing low-lying minima separated by large barriers compared to the relevant thermal energy. This approach led to the principle of minimal frustration for naturally occurring proteins, based on the intuition that folded structures have secondary and tertiary interactions satisfying a wide range of packing and bonding requirements and that favourable contacts are optimised in the native state. A similar idea was used by Gō in early models of protein folding,^[279] and a clear visualisation was later obtained by comparing the landscape for a model protein^[280;281] with the corresponding Gō model.^[282]

The organisation of the protein free energy landscape requires us to consider structural ensembles with similar configurations and energies. The folding process can then be seen as a progressive organisation of ensembles. The key principles that circumvent Levinthal’s paradox are that, under physiological conditions, (1) the free energy of the folded state is lower than the denatured state, (2) all states are not equally probable, and (3) the free energy landscape is inherently biased towards the native state.

Diverse conformations with comparable energies are accessible in the unfolded state, leading to a high internal entropy. Conversely, in the folded state inter- and intramolecular interactions are generally stabilising, and there are smaller fluctuations in structure, corresponding to reduced configurational entropy. Protein folding therefore corresponds to the molecular analogue of a first-order phase transition, as the balance of contributions to free energy changes with temperature, a scenario anticipated by Mirsky and Pauling.^[283]

Bryngelson and Wolynes suggested that the protein free energy landscape exhibits two funnels:^[278] one based on entropy, containing the unfolded states, the other favoured by enthalpy, containing the native fold. In the denatured state, where the conformational entropy is high, the free energy landscape is relatively flat. As the protein folds, formation of native contacts leads to a thermodynamic barrier to unfolding, which limits the number of accessible conformations and guides dynamical trajectories downhill, in the direction of the free energy gradient. As a result, the protein does not need to search the entire conformational space to locate the native state.

From the perspective of kinetics, Leopold *et al.* interpreted the minimally frustrated free energy landscape in terms of a set of kinetically convergent pathways, or ‘folding funnel’.^[284] This view provides a solution to the Levinthal paradox, and a physical basis for folding.^[201;285] It was suggested that funnelled free energy landscapes should be expected for naturally occurring proteins,^[286–288] as a natural consequence of evolution. Further work indicated that native sequences are robust with respect to point mutations,^[289] due to the existence of superfunnels in sequence space centred around stable sequence.^[290]

In contrast, a random polypeptide chain is not expected to have a well-defined folded state, because random heteropolymers contain both stabilising (local) and destabilising (non-local) contacts leading to energetic frustration.^[278] In natural proteins we expect strong interactions to be primarily native contacts, with few interactions opposing folding; hence there is minimal energetic frustration.^[278;291]

The free energy landscape may support many local minima, due to the interplay of entropic and enthalpic terms.^[286] The dynamics between various states will depend on the barriers and the overall funnelled organisation. The barrier heights in various parts of the landscape dictate whether intermediates accumulate during folding.^[292] Additionally, formation of non-native contacts (increased frustration) may lead to misfolded states, corresponding to deep wells on the landscape that act as kinetic traps, slowing down folding.^[284;285]

The notion of a free energy funnel is compatible with a variety of different mechanistic

scenarios. For example, a uniform attraction of hydrophobic residues may favour rapid collapse of the unfolded state into a compact globule (folding intermediate) that slowly rearranges into the native state. Alternatively, local interactions along the chain may facilitate transient secondary structure formation in the denatured state and eventual coalescence to the native fold. Each process will have associated energetic and kinetic barriers, and the gradient of the energy will depend on the relative stability of the denatured and folded states. However, the free energy landscape model can be adopted to interpret the various folding scenarios, and to account for the dynamics.^[293] Descriptions based on the notion of more specific pathways^[294;295] are also compatible with the organisation of the free energy landscape as a folding funnel if the pathway is recognised as an ensemble.^[12;296–298] Here the ensemble would include analogous steps through configurations with minor structural variations, such as rotation of irrelevant side-chains, compatible with the essential mechanistic features.

Using Φ -value analysis,^[299;300] Fersht showed that to describe folding accurately, the folding transition state needs to be identified.^[301] He explained that experiment and theory yield complementary results if information about the transition state is used.^[302] Subsequently, many studies have shown that the free energy landscape perspective for biomolecular folding agrees well with experimental observations.^[303–306] These efforts include analysis of the free energy landscape using atomic force microscopy^[307;308] and fluorescence spectroscopy.^[309]

The free energy landscape perspective can be extended to treat a wide variety of problems in molecular and condensed matter science, beyond the applications to glassy systems and proteins introduced above. Folding of nucleic acids such as RNA,^[310] and misfolding associated with disease,^[311] provide important examples. Free energy landscapes for RNA folding have been presented,^[312–314] and some recent results based on potential energy landscapes are mentioned in the following sections.

5 Potential Energy Landscapes

The free energy landscapes considered above in §4.1, and indeed all the observable properties of molecular and condensed matter systems, are encoded at a fundamental level by the potential energy landscape (PEL). It is the potential energy as a function of atomic coordinates that defines structure, dynamics, and thermodynamics, in both classical and quantum mechanical descriptions (within the Born-Oppenheimer approximation). The remainder of this report will focus on this potential energy landscape perspective, which has been described in detail and reviewed in a variety of articles.^[12;13;253;315;316] Key aspects that arise in our discussion of self-organisation and multifunctional systems will be revisited below.

The potential energy surface is directly involved in theory and calculations of molecular properties through gradients, defining forces and dynamics, and the energy density of states or the Boltzmann factor that appears in microcanonical and canonical partition functions. Focusing attention on stationary points of the PEL, where the gradient vanishes, provides a coarse-graining that provides a convenient framework in which to develop both conceptual and computational tools. Characterising local minima, and the transition states that connect them via steepest-descent pathways, provides tools for structure prediction (global optimisation), enhanced sampling of thermodynamic properties, and rare event dynamics. Here we adopt the geometrical definition of a transition state, as a stationary point with precisely one negative Hessian eigenvalue.^[317] In this framework, sampling leads to a database of local minima and transition states, which can be combined into a kinetic transition network,^[251;253;318;319] as discussed in §6. Thermodynamic and kinetic properties are extracted from the database using the tools of statistical mechanics and unimolecular rate theory for the individual rate constants associated with each transition state.^[320;321] Most of the computational effort is invested in the geometry optimisation procedures, especially the location of transition states

(details can be found elsewhere^[12;13;253;315;316]). In brief, we employ custom LBFGS^[322] (limited memory Broyden,^[323] Fletcher,^[324] Goldfarb,^[325] Shanno^[326]) routines for minimisation, a doubly-nudged^[327] elastic band^[328;329] (DNEB) algorithm to locate transition state candidates, and hybrid eigenvector-following^[330–333] to refine these candidates accurately.

Coarse-graining the landscape into stationary points provides a natural way to resolve the thermodynamic and kinetic properties into well-defined components. In particular, the global partition function, $Z(T)$, can be decomposed into contributions from the catchment basins of local minima, α . This is the superposition approach,^[12;334–338] where we write

$$Z(T) = \sum_{\alpha} Z_{\alpha}(T), \quad (1)$$

and $Z_{\alpha}(T)$ includes a statistical weighting to account for permutation-inversion isomers. The equilibrium occupation probability for structure α at temperature T is then $p_{\alpha}^{\text{eq}}(T) = Z_{\alpha}(T)/Z(T)$, which we use in §7. This approach can be formally exact, but is usually implemented using a harmonic normal mode approximation for the vibrational densities of states.^[12] In this case $Z(T)$ includes landscape anharmonicity, which arises from the sum over different minima, but not well anharmonicity for the individual contributions. Approximate treatments of well anharmonicity and quantum effects are available.^[335;339–342]

The component partition functions can also be used to define free energies for the local minima, which can be regrouped to construct free energy states corresponding to the temporal resolution of the experimental setup of interest.^[343] We elaborate on the dependence of free energy upon observation time scale^[344] in §7, below. Local free energies can also be used in global optimisation procedures for structure prediction, to account directly for local temperature dependent effects.^[345;346]

The following sections describe how the potential energy landscape approach, formulated in terms of local minima and transition states, can be analysed as a kinetic transition network (§6) and visualised to diagnose how the landscape encodes function. This framework could be viewed as a new formulation of the structure-function paradigm, where the structure now refers to the organisation of the potential energy landscape, rather than the molecular structure itself.

6 Kinetic Transition Networks and the Master Equation Approach

To analyse the kinetics associated with a kinetic transition network we generally assume that the local transition rates depend only on the current state, not the history of a trajectory. This memoryless assumption corresponds to Markovian dynamics, and the resulting set of linear kinetic equations for the flow of probability is known as a master equation. The master equation approach has a long history in the field of chemical dynamics,^[347;348] and provides foundations for the methodology discussed in §6.1 and §6.2 below. We first consider networks constructed from the computational potential energy landscape perspective described above in §5, which is based on geometry optimisation. Coarse-graining the landscape into local minima and the transition states that connect them provides a natural decomposition into states,^[12;349–351] which can be regrouped into sets of local minima with associated free energies to account for local equilibration on short time scales.^[343] Regrouping may also be necessary for comparison with experiment, depending on resolution and how separate states are distinguished operationally. Representing the dynamics in terms of discrete jumps between states also entails approximations, which can be analysed in the context of a quasi-stationary distribution.^[352]

The master equation has an analytical solution, but this direct approach encounters problems of efficiency and precision when the number of states is large, or the relaxation time scale

of interest is slow, even when regrouping is employed.^[353–355] Instead, phenomenological rates can be extracted using kinetic Monte Carlo methods.^[356–358] However, the most efficient approach is based on deterministic schemes to systematically eliminate states and renormalise the waiting times and branching probabilities of the remaining nodes in the network.^[359;360] For all the rates extracted from static networks obtained by discrete path sampling (§6.1) reported in the Examples (§8), deterministic graph transformation procedures were employed.

6.1 Discrete Path Sampling

The potential energy landscape is a high-dimensional surface, but it can be characterised by the stationary points that it supports, in particular local minima and transition states. One possibility for exploring an energy landscape is therefore given by characterising the local minima of interest and the transition states between them. Discrete path sampling (DPS)^[361;362] provides a systematic framework to create and expand a database of local minima and transition states that captures the essential properties of the potential energy landscape. A discrete path is defined as a connected series of local minima and transition states that describe one way of changing the conformation of a chemical system from a starting configuration, A , to a final configuration, B , via a set of intervening minima, I . Each transition state connects two local minima, which are located by following approximate steepest-descent pathways in the direction of the eigenvector associated with the imaginary frequency normal mode of the transition state.

The principal objective of discrete path sampling is to create a kinetic transition network (KTN),^[251;363] from which it is possible to extract reactive pathways between different conformational states and the corresponding rate constants. From discrete path sampling, structural, mechanistic, thermodynamic and kinetic data can be obtained from simulation within one theoretical framework using statistical mechanics and unimolecular rate theory,^[320;321] without projection onto order parameters or the use of reaction coordinates. As the location of stationary points is based on geometry optimisation the sampling is largely independent of the height of energy barriers and broken ergodicity can be overcome.

6.2 Methods Based on Explicit Dynamics

Another family of methods to treat rare events employs explicit dynamics, and many of these schemes correspond to some form of transition path sampling (TPS).^[239;364] Initially Pratt^[365] suggested using importance sampling to study rare event paths. Dellago *et al.* then developed the original implementation of TPS by importance sampling of the pathway ensemble for dynamic trajectories that connect stable states.^[239] The sampling is usually based on shooting moves,^[246;366] where a new path is found by creating a new trajectory shooting off a sampled path. This methodology, which has been developed in various versions, can provide decorrelated paths representing the transition path ensemble. A more efficient method based on the same ideas is transition interface sampling (TIS).^[367–369] Instead of sampling the path ensemble for the full distance between reactants and products, a number of intermediate transition hypersurfaces are defined, and the sampling aims to identify the fluxes through these surfaces, shortening the computational time required.

A related approach is provided by sampling weighted path ensembles.^[370–372] Here, multiple trajectories are simulated, and if unexplored regions of configuration space are encountered, trajectories can spawn correctly weighted daughter trajectories. The method can treat equilibrium^[373] and non-equilibrium processes,^[371] and has recently been employed in protein-peptide binding studies^[374] and for processes involving multiple pathways.^[375] The forward flux sampling scheme can also treat rare events away from equilibrium.^[376;377] Applications have been presented for processes ranging from crystallisation^[378] to genetic toggle switches.^[379]

Milestoning^[380–383] again aims to sample transition paths. Here, a low-dimensional representation of the reaction is divided into multiple short trajectories between milestones. Each trajectory only samples a small part of the landscape and therefore large energy barriers can be broken down into smaller contributions, which are easier to treat. From the short trajectories the probability of transitions between the milestones can be calculated, leading to reactive fluxes from the stationary solution. Recently, milestoning has been applied to the anthrax channel, showing that it can be used for large systems.^[384]

The string method^[385–387] is related to TPS and milestoning, and also attempts to find the reactive paths. A reactive path may be restricted to a small space around an initial guess, and its lower bound can be optimised to find the path within the subspace with the highest reactive flux. This approach reduces the problem to finding the geodesic, and corresponds to minimising the action along a parametrised path. The restrictions may be broadened adaptively to allow for better sampling. In this formulation the string method is equivalent to adaptive milestoning, maximising the global flux. The string method has been applied to biomolecules,^[387–389] where collective variables have been employed to increase efficiency. Another application involves analysis of Markov state models (discussed below),^[390] or use of the string path to obtain the optimal position of the milestones for milestoning.^[382;391;392]

6.3 Markov state models

Construction of kinetic transition networks from MD or MC simulations is often referred to as Markov state modelling. Instead of running one long simulation, the goal here is to build statistically robust models from numerous independent simulations. One way to connect independent simulations is to model the system dynamics as Markovian; in other words, the future dynamics of the system are a function only of the current description in configuration or phase space, and not the history of the trajectory.^[393–397] In 1983, Zwanzig suggested the use of a kinetic master equation to describe classical dynamics using a short memory approximation.^[393] This approach requires the projection of a system onto nonoverlapping cells spanning all of phase space (i.e. onto characteristic functions). Zwanzig noted that if the cells are Markovian, there is a ‘remarkably simple’ way to estimate transition rates between those cells.

Later, Zwanzig and coauthors used this approach to comment on Levinthal’s paradox using an idealised approach.^[398] In this study, protein states were represented by a list of binary flags, representing whether each residue is in a correct or incorrect position. By biasing against incorrect local positions, a much smaller search space is required than Levinthal’s paradox assumes. In this initial study, the folded state, with all the residues correct, was treated as a sink, and folding was considered irreversible. A later study extended this analysis to treat reversible folding.^[399]

Dividing phase space into cells such that transitions between them are approximately Markovian provides another route to a linear master equation.^[394–397] To create such a representation from explicit MD simulation, trajectories are first decomposed into states, and a transition probability matrix is then built for the coarse-grained representation. The transition matrix is derived based on the observed transitions between the states, which are assumed to be history-independent (i.e. Markovian). Hence, the resulting MSM encodes both the thermodynamics and dynamics of the folding process.^[400–403] This approach is well suited to exploit distributed computing platforms, such as Folding@home,^[187] but can also be applied to a single long trajectory. However, since MSM analyses do not require that simulations are performed as a single long trajectory, greater simulation times can be obtained from runs performed in parallel.

Results based on MSMs have been interpreted in terms of a hub-like character of the native state, where the folded protein is generally accessible via multiple paths from the heterogeneous unfolded state.^[196;404] This observation seems to be consistent with the idea that the single

funnel potential energy landscapes associated with structure-seeking systems (see §7.1 below) may exhibit small-world properties,^[343;405;406] where a randomly selected minimum is usually connected to the global minimum via a relatively small number of transition states.^[407] Such characteristics have been associated with various biomolecules.^[318;343;408;409] It was also shown that non-native states play a crucial role in slowing down the folding rate, as previously reported in experimental studies.^[404] Voelz *et al.*^[191] presented folding pathways for a millisecond-folder modelled atomistically with implicit solvent. The estimated folding time of 1 ms was in good agreement with the experimental value of 1.5 ms. Later, folding simulations of an 80-residue protein were conducted at the atomic level with explicit solvent representation.^[192] This study further highlighted the power of the MSM technique in exploiting large amounts of MD data to yield thermodynamic and dynamic insights on time and length scales much longer than the conventional nanosecond to millisecond simulation limit.

In principle it is also possible to analyse the energy landscape by considering the generator of the dynamics.^[397;410–416] The eigenvalues of the generator can be grouped by their absolute value, and the smallest set of eigenvalues corresponds to the slowest processes, which are usually the focus of attention. Considering only the lowest M eigenvalues it is possible to reduce the system to a Markov chain with M states.^[417] However, the computation of the eigenvalues can be numerically challenging. A case study for an atomic cluster used an existing database created by discrete path sampling to test this approach^[417] and associated transition path theory to describe the underlying Markov chains.^[410]

7 Organisation of the Landscape: Emergent Functionality

A faithful representation of the potential energy landscape is complicated by the high dimensionality. In addition, the number of stationary points increases exponentially with system size,^[334;418;419] and the number of transition states per minimum is expected to increase linearly.^[420] This increase in connectivity cannot be effectively represented by surfaces in three dimensions.^[316] A common alternative approach is to project the landscape onto lower dimensions, using order parameters or collective variables, but such methods will inevitably lose information and average out properties of the landscape.^[246–254]

Kunz and Berry^[349;350] suggested representing the landscape using monotonic sequences, defined as a series of minima and connecting transition states, where the energy of the minima decreases. A set of sequences that lead to a particular minimum constitute a monotonic sequence basin, which should be related to the idea of a folding funnel^[284] discussed in §4.1.

In the present contribution we visualise the potential or free energy landscape using dis-connectivity graphs, following Becker and Karplus.^[275] To construct a dis-connectivity graph, a fixed energy interval, ΔE , is chosen, and a superbasis analysis is conducted at a discrete series of energies separated by ΔE . A superbasis analysis at energy E classes the local minima into disjoint sets. Two minima are in the same superbasis if there is a discrete path (i.e. a series of local minima and intervening transition states) on which no point has an energy higher than E . This procedure is illustrated in **Figure 1**. This construction is similar to ‘energy lid’ and ‘energy threshold’ procedures;^[421–424] calculating transition states directly reveals the precise organisation of the landscape very efficiently. Each superbasis is represented by a node at the corresponding energy. Nodes between neighbouring energy levels are connected if they are the same superbasis or superbases that have merged at the higher energy. The arrangement of nodes on the horizontal axis is arbitrary, and is usually chosen to provide the clearest representation, where lines do not cross. However, the nodes can also encode additional information, for example using colours based on structural features, or variable line widths and positions.^[425]

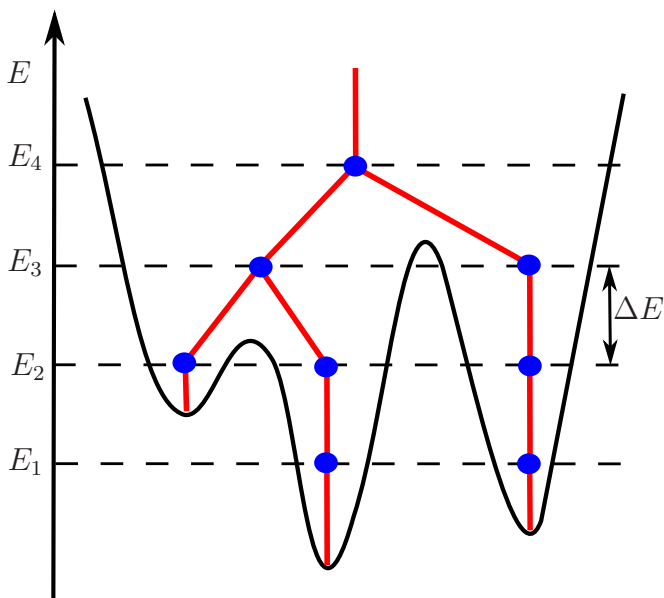


Figure 1: Construction of a disconnectivity graph. The dashed lines indicate the discrete series of energies used for the superbasin analysis. The blue dots represent the nodes for the basins at each of the energies, and the red lines are the disconnectivity graph.

Disconnectivity graphs provide a powerful way to think about potential and free energy landscapes, which immediately provided new insight when applications to existing databases were presented.^[14] We emphasise that our calculations of thermodynamic properties and rates employ densities of states calculated for the individual minima and transition states, as outlined in §5. Usually we use normal mode analysis and a harmonic approximation to characterise the corresponding partition functions and minimum-to-minimum rate constants. However, once a database has been constructed, post-processing with anharmonic and quantum mechanical descriptions can be used,^[335;339–342] together with unimolecular rate theory beyond the simplest transition state theory approximation,^[320;321]

Free energy disconnectivity graphs can be calculated from the local partition functions,^[353;426] and regrouped into states that are expected to equilibrate rapidly compared to the rate-determining step of interest,^[343] if such a time scale separation exists. Regrouping schemes can account for experimental details, by lumping states that are not distinguished experimentally. The free energy disconnectivity graph can therefore be a function of the observation time scale.^[344] To quantify this connection we can associate the average lifetime τ with the reciprocal rate constant, $1/k$, assuming first order kinetics, and employ the Eyring-Polanyi formulation of the rate constant^[427–429] in terms of the free energy of activation, ΔG^\ddagger , to write

$$k = \frac{k_B T}{h} e^{-\Delta G^\ddagger / k_B T}, \quad (2)$$

where k_B is the Boltzmann constant, h is Planck’s constant, and T is the temperature. For a given observation time scale we then recursively regroup free energy minima separated by barriers less than the corresponding value of G^\ddagger . A simple example is provided for illustration in **Figure 2** for alanine dipeptide.^[344] Here we distinguish enantiomers but not permutational isomers, leading to two sets of three minima in the potential energy disconnectivity graph, separated by a high barrier (Figure 2a). For an observation time longer than $\tau = 4.7 \times 10^{-12}$ s, which corresponds to a regrouping threshold of 2.0 kcal/mol, the lowest two minima in each of the two groups can interconvert, leading to the structure in Figure 2b. The three minima in each of the two main branches merge when the observation time exceeds $\tau = 9.8 \times 10^{-7}$ s (Figure 2c) for a regrouping threshold of 9.3 kcal/mol. The remaining barrier is so large that the

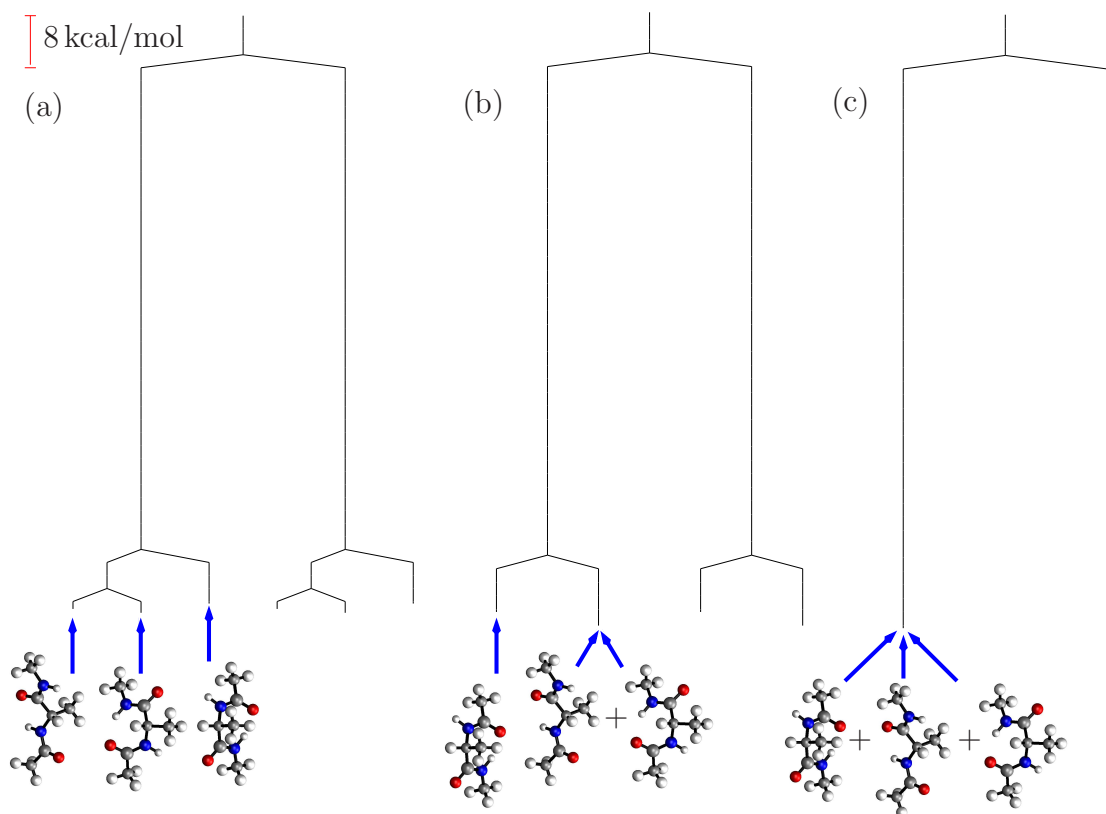


Figure 2: Disconnectivity graphs for alanine dipeptide when permutational isomers are lumped together.^[344] (a) Potential energy graph, (b) free energy graph for regrouping threshold 2.0 kcal/mol ($\tau = 4.7 \times 10^{-12}$ s), and (c) free energy graph for regrouping threshold 9.3 kcal/mol ($\tau = 9.8 \times 10^{-7}$ s). In each panel the two groups separated by the high barrier are enantiomers.

corresponding enantiomers would only become indistinguishable for observation times beyond around $\tau = 5.7 \times 10^{48}$ s (not shown).

7.1 Self-Organisation is Encoded in Single Funnel Potential Energy Landscapes

Applications of disconnectivity graphs to a diverse range of molecular and condensed matter systems provided new insight into how characteristic observables are encoded in the organisation of the underlying potential energy landscape.^[12;14;17] Self-organising or structure-seeking^[351] systems, with efficient relaxation to a well-defined structure or structural ensemble, are associated with ‘palm tree’ single funnel potential energy landscapes.^[14] This motif assures that there is a single free energy minimum, which is kinetically accessible over a wide range of temperature or total energy. Examples are illustrated in **Figure 3** for two atomic clusters, providing a direct realisation of a minimally frustrated potential energy landscape that supports the kinetic pathways associated with funnelling characteristics in earlier work.^[278;285;433;434] This is the structure that we expect for biomolecules that have evolved to deliver a single function, and ‘magic number’ clusters that emerge in molecular beams. It is important to recognise the common features of these apparently very different systems, since fundamental insights and advances in theory and simulation may be directly transferable. The initial application of ideas from spin-glass theory to biomolecules provides an early example.^[278]

Systems with a single funnel PEL are likely to be particularly remarkable, especially when they result from a global potential energy minimum with high symmetry, which is often the

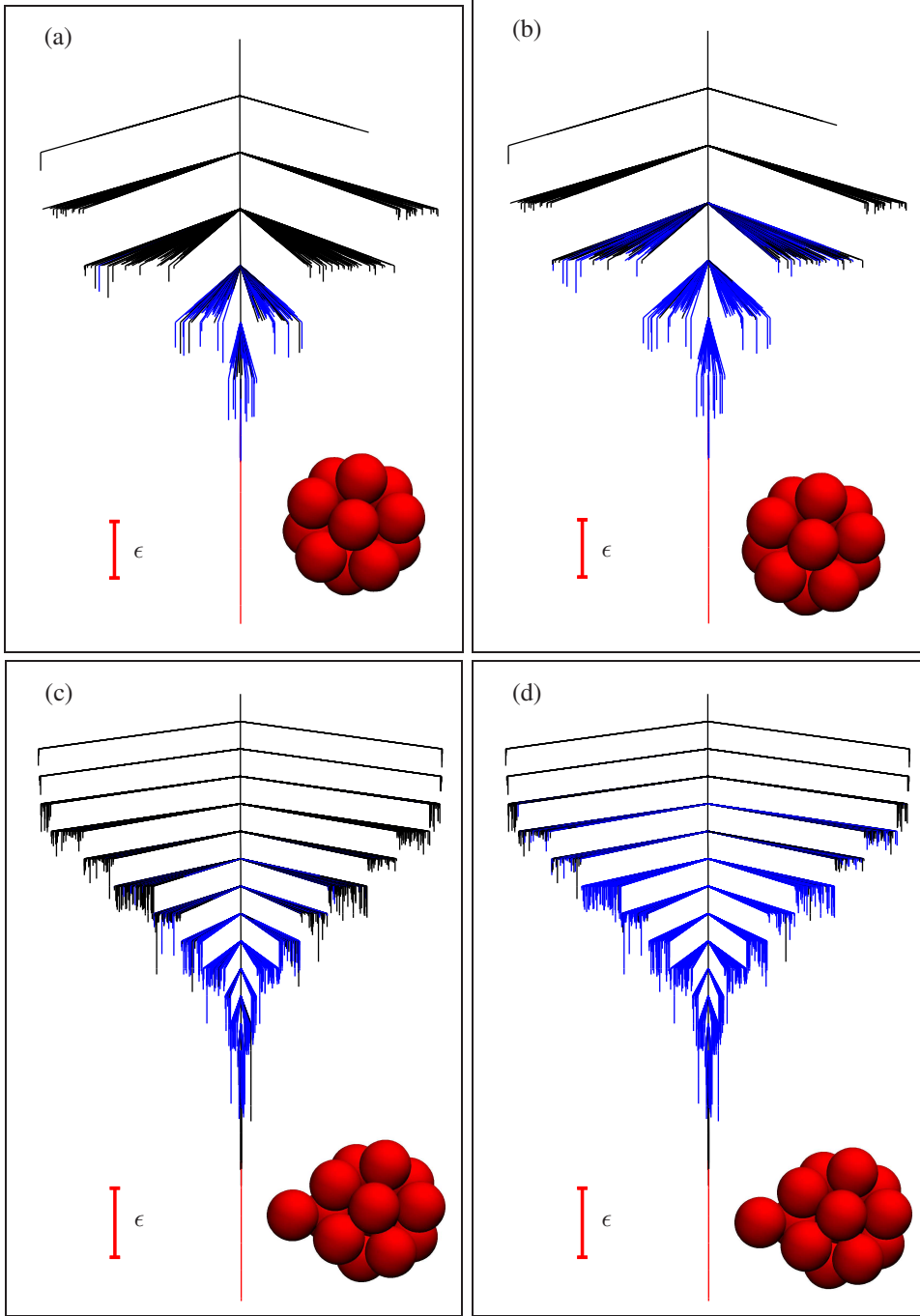


Figure 3: Disconnectivity graphs for 13- and 14-atom atomic clusters bound by the Lennard-Jones potential^[430] (denoted LJ_{13} and LJ_{14}) with branches coloured according to the occupation probability gradients for local minima α , $\partial p_{\alpha}^{\text{eq}}(T)/\partial T$ (see §4.1), as a function of temperature at the melting point.^[431] ϵ is the binding energy for a pair of atoms, and an illustration of the global minimum is included for each graph. Both the Mackay icosahedron^[432] global minimum for LJ_{13} and the capped icosahedron for LJ_{14} have single funnel potential energy landscapes, corresponding to the ‘palm tree’ motif associated with efficient relaxation to the global minimum.^[14] Panels (a) and (b) are for LJ_{13} where minima contributing (a) 90% and (b) 99% to the calculated heat capacity are coloured blue and red for positive and negative $\partial p_{\alpha}^{\text{eq}}(T)/\partial T$ at $k_B T/\epsilon = 0.28$. Analogous results are shown for LJ_{14} in panels (c) and (d). For LJ_{13} anharmonic vibrational densities of states were employed in the heat capacity calculations, while for LJ_{14} a simpler harmonic normal mode approximation was used.^[431]

case.^[12;435;436] These landscapes are exceptional, but a key theme of the present contribution is to highlight how this framework can be generalised to multifunctional systems. We first consider the simplest extension for double-funnel PEL, where frustration exists in the form of two competing morphologies, separated by a high barrier. This organisation provides the possibility for the system to switch between alternative functions, perhaps tuned by application of an external field, or interaction with alternative binding partners. Indeed, a double funnel PEL may actually be a necessary condition for encoding such functionality.

7.2 Competing Structures: Characteristics of Double-Funnel Landscapes

Double-funnel potential energy landscapes have been analysed extensively for atomic clusters. The most interesting systems exhibit a low temperature heat capacity peak representing the finite system analogue of a first-order phase transition.^[12;16;338;437–444] At low temperature the global free energy minimum corresponds to the global potential energy minimum, but at higher temperature the greater entropy of the alternative low-lying morphology shifts the equilibrium population to the other funnel. Hence a change in temperature represents the most straightforward means of switching the structure, and therefore potential functionality.

A quantitative analysis of contributions to the heat capacity, C_V , can be developed within the superposition framework for both harmonic and anharmonic formulations.^[431] The total heat capacity can be expressed in terms of occupation probability derivatives, $\partial p_\alpha^{\text{eq}}(T)/\partial T$ (§4.1), for local minima α with respect to temperature, and the magnitude of the contributions correlates with the probability flux, enabling us to identify the key states without defining a structural order parameter. The sign of the derivative defines the low- and high-temperature states involved in the transition that produces the heat capacity feature, while the magnitude of the derivative determines the relative importance of the individual contributions to C_V . Colouring branches in the corresponding disconnectivity graph according to the sign of the probability flux, and including only the states that make the largest contributions up to a fixed fraction (for example, 90% or 99%) of the total, provides fundamental insight into how the transition is encoded in the underlying landscape. This colouring scheme is used for the atomic clusters in Figure 3, where there is a single high temperature heat capacity feature corresponding to the finite system analogue of melting. Here we see that the melting transition corresponds to probability flow from the global minimum (low energy, low entropy) to a much larger number of high energy minima, in clear contrast to the low temperature transitions described for CGN4-pLI in §8.

Double-funnel systems, where the global potential energy minimum is associated with lower entropy and a relatively small fraction of configuration space, provide useful benchmark problems for global optimisation and structure prediction.^[437;445–447] Calculating equilibrium thermodynamic properties for landscapes associated with broken ergodicity is also challenging.^[12;16;338;437–444] Still harder is the computation of transition rates between the competing morphologies, where atomic clusters once again provide benchmarks that enable rare events methods to be tested and developed.^[16;361;362;448–450] All of this methodology, for structure prediction, enhanced thermodynamic sampling, and rare events dynamics, can be applied to biomolecules that support multifunnel landscapes.

A particularly clear double funnel structure occurs for the bacterial regulatory protein RfaH,^[451;452] which exhibits an extreme case of fold-switching. The C-terminal domain (CTD) of RfaH transitions from an α -helical hairpin structure to a five-stranded β -barrel scaffold. In the α -helical state, the CTD is known to mask the RNA polymerase binding site on the N-terminal domain (NTD) of RfaH, thereby regulating transcription.^[453;454] However, in the presence of a specific DNA target (the so-called *ops* site),^[452;455;456] interactions between the

transcribing RNAP and RfaH lead to domain separation, and the CTD refolds to the β -sheet conformer.^[454] This conformer then binds to ribosomal protein S10, ultimately assisting in protein translation.^[454] Hence, RfaH-CTD has two distinct folds with two specific functions, which are encoded by two prominent funnels in the potential energy landscape, as shown in Figure 4.^[457]

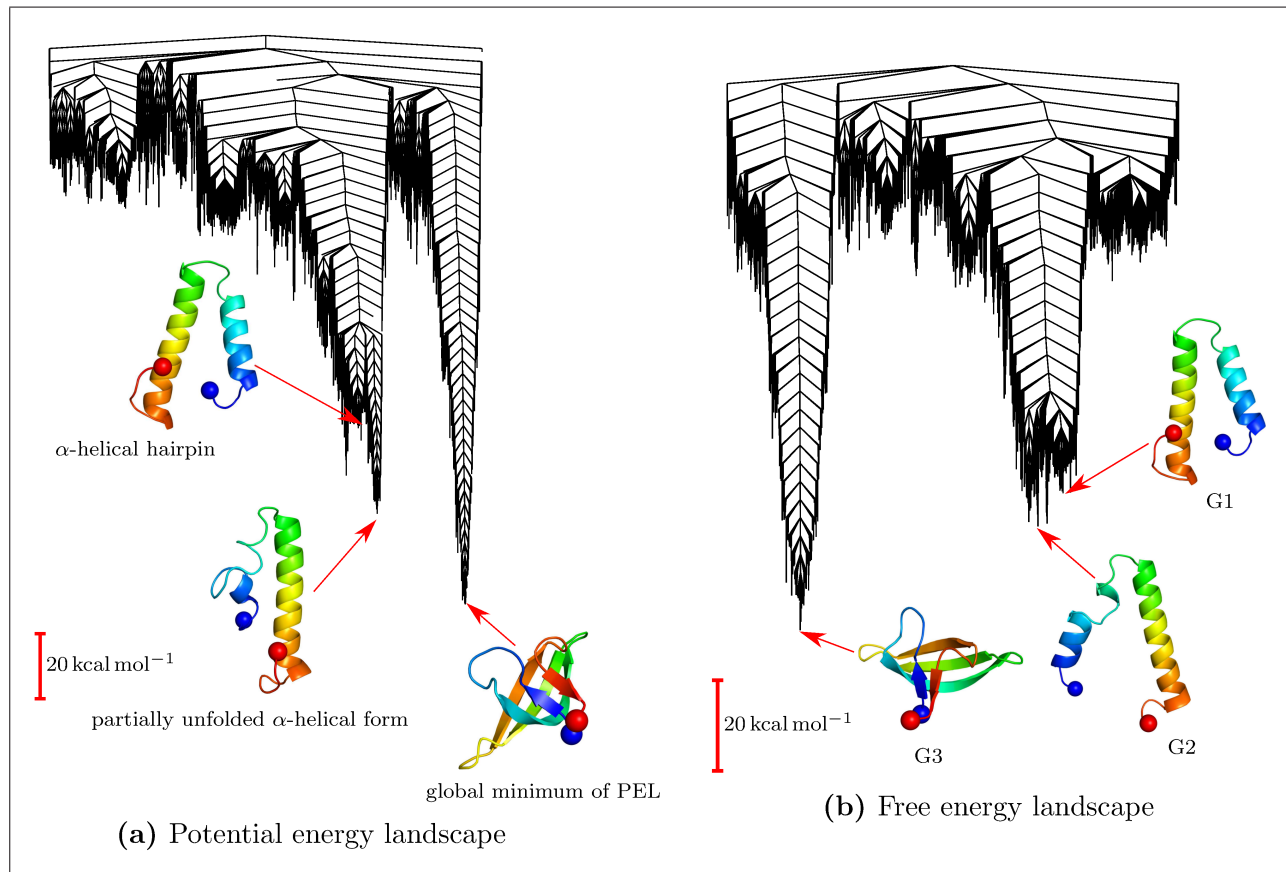


Figure 4: (a) Potential energy and (b) free energy disconnectivity graphs for the isolated RfaH-CTD. The lowest energy α -helical conformer (partially unfolded), the α -helical conformer with maximum helical content (hairpin), and the β -barrel scaffold (i.e. the global minimum of the PEL) are all illustrated on the graph.^[457] For panel (b) the free energies were computed at 310 K with minima and transition states regrouped^[343] based on an energy threshold of 5 kcal mol⁻¹. Selected free energy groups (G1–G3) are shown.

7.3 Multifunnel Landscapes for Biomolecules

The importance of proteins that exhibit heterogeneous structural ensembles in the absence of specific binding partners is becoming increasingly apparent; a cross-section of recent activity has recently appeared.^[458] These systems are commonly referred to as intrinsically disordered proteins, and in many cases they are associated with the ability to perform different functions in response to alternative stimuli. Hence, we expect these molecules to exhibit multifunnel potential energy landscapes, with incipient structure and properties that are realised under specific conditions, which might be environmental, or linked to association with a membrane or a particular ligand. A corresponding multifunnel organisation of the potential energy landscape could encode these incipient characteristics.

A detailed description of the potential energy landscape for such a protein was presented

for the p53 upregulated modulator of apoptosis (PUMA) peptide.^[18] It revealed a number of distinct funnels separated by energy barriers that are large compared with the available thermal energy at physiological temperatures. Furthermore, the most heavily populated structures at equilibrium are not those observed in bound complexes, suggesting an induced fit binding mechanism. While this observation may seem to violate the principle of minimal frustration, it can be argued that induced fit means that the peptide is optimised within a certain set of environmental conditions and the binding partner is one of these factors. Hence the landscape becomes minimally frustrated in the presence of a binding partner, rather than in isolation. This point of view further implies that upon binding the landscape changes to provide a bias towards a single structural ensemble, i.e. the bound state. Consequently, we still may consider the functionality and sequence of PUMA to be evolved for specific functions. A number of examples are now known to exhibit similar behaviour.^[458–460]

The evolution of PUMA is likely to contrast with peptide sequences that are prone to aggregation, such as aFFE and GNNQQNY. These fragments of amyloid-forming proteins are interesting because they retain aggregation behaviour at small sizes. However, the disorder in these systems is not a consequence of evolution towards biological functionality, so we do not expect them to exhibit the same characteristics as IDPs. Given enough monomers aggregation arises because there is no competing structural ensemble on the landscape.

Similar observations apply for amyloid- β ,^[245] where we identify multifunnel landscape, in agreement with experimental studies^[461] and expectations that amyloid-forming peptides are disordered.^[462;463] However, in contrast to KFFE, there are clear separations between distinct structural ensembles for amyloid- β dimers (see **Figure 5**). This feature persists for larger assemblies,^[245] and is associated with kinetically insurmountable energy barriers, likely limiting the possibility of equilibration in experiment. The occurrence of disorder in these systems has been a focus of research, largely driven by the implication of aggregation in a number of diseases. Interpretation of these landscapes in terms of the principle of minimal frustration is an interesting open question. One possibility is that evolution of these sequences to avoid aggregation behaviour has not yet progressed very far, perhaps because amyloid diseases are often associated with ageing, and the rapid increase in life expectancy is a relatively recent phenomenon. Alternatively, these proteins may perhaps have evolved in constrained environments, and away from these conditions small perturbations lead to a dramatic deviation from the usual functionality.

For small peptides intrinsic disorder may not produce numerous competing funnels in the landscape. For example, the KFFE monomer and the hormones oxytocin and vasopressin have all been characterised as intrinsically disordered, but their energy landscapes are not multifunneled. These peptides are simply too small to accommodate a large number of competing interactions, or to have a large number of strong interactions that would define a particularly stable, native fold in the first place. The observed structural flexibility is therefore a result of relatively low stabilisation of the global minimum compared to denatured structures, rather than due to the competition of a large number of different folds.

A multifunnel landscape has also been characterised for a four-fold telomere repeat in DNA, which forms a variety of G-quadruplex structures.^[464] These repeat base sequences play a key role in preventing damage to the termini of chromosomes, with quadruplex formation decreasing the activity of the telomerase enzyme, which normally maintains the length of the repeat sequences. A combinatorial number of arrangements can be constructed for the G tetrads, some of which are physically realisable. The complexity of the resulting multifunnel landscape makes low-dimensional projections problematical, and leads to long time scales for interconversion of alternative morphologies.^[464;465]

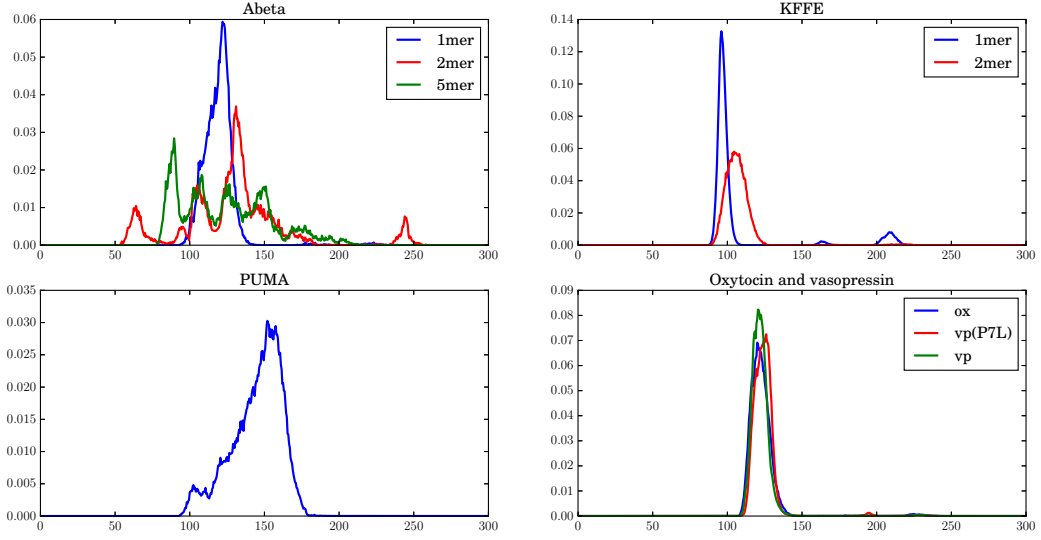


Figure 5: Distribution of the potential energy for local potential energy minima of $A\beta_{42}$ monomers, dimers and pentamers, KFFE monomers and dimers, PUMA and oxytocin and vasopressin. The difference between IDPs and small peptides like KFFE and oxytocin is clearly visible. Furthermore, the separation of structural ensembles in larger assemblies for $A\beta_{42}$, which relates to high energy barriers, is apparent.^[245]

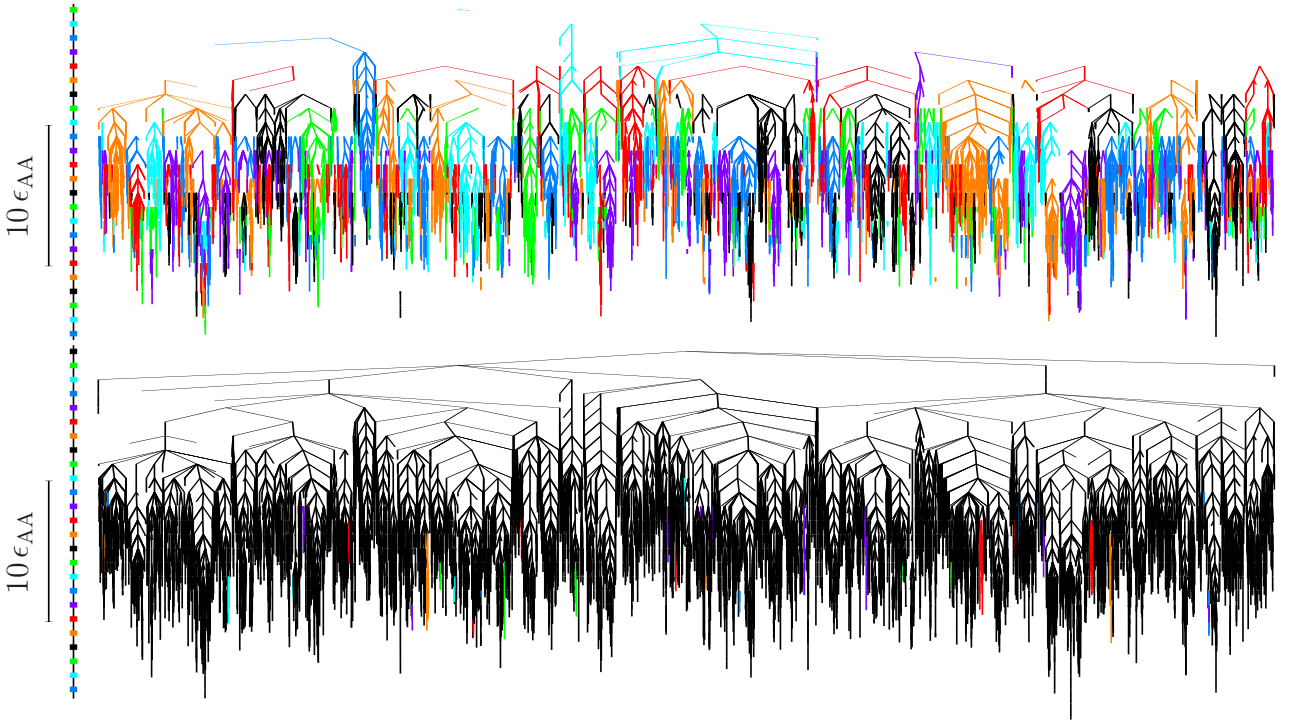


Figure 6: Disconnectivity graph for a binary atomic structural glass-former where transition states corresponding to cage-breaking (top) and non-cage-breaking rearrangements (bottom) are removed to analyse the organisation (adapted with permission).^[466] Groups of minima that retain mutual connectivity are coloured according to the energy at which they are disconnected from the rest of the graph, defined in the colour scale on the vertical axis. The model is a periodically repeated supercell containing sixty atoms of two types, A and B. ϵ_{AA} is the unit of energy, and corresponds to the pair well depth for two A atoms.^[466]

Structural glasses are probably the most extreme example of multifunnel landscapes, and visualisation for a variety of systems has produced direct insight into the glass transition.^[466–469] We reproduce an example here to highlight the complexity that results from competition between a combinatorial number of low-lying amorphous minima, providing a distinct contrast with the landscapes visualised for biomolecules. As the available thermal energy decreases, the system becomes localised in a diminishing region of configuration space on any given finite observation time scale. The local properties within any one of these regions are essentially the same, since each one comprises a similar set of amorphous local minima (**Figure 6**). The region that the system relaxes into will depend on the starting conditions and the cooling rate. This scenario is basically consistent with the model that underlies random first order transition theory.^[470;471]

Figure 6 reveals both the multifunnel character and additional hierarchical organisation, which becomes clear when the pathways corresponding to individual transition states are classified in terms of cage-breaking rearrangements for individual atoms. Here, the cage refers to the set of nearest-neighbours surrounding a given atom, so that cage-breaking events are necessary for atomic diffusion. Removing the corresponding transition states (top panel of Figure 6) and colouring the disjoint sets of minima that can interconvert without a cage-break, reveals additional hierarchical organisation.^[466] The top panel is fragmented into regions that have been identified with ‘metabasins’, while the graph in the lower panel remains highly connected, indicating that most minima can still interconvert if cage-breaking pathways are retained and non-cage-breaking pathways are removed.

Although the origin of the glass transition seems clear from the organisation of the potential energy surface, the associated observable properties are much more subtle, because they depend upon the progressive decrease of accessible configuration space as a function of the observation time scale and cooling schedule. The precise energy and length scales that define the landscape are also system dependent, although the characteristic hierarchical organisation is analogous for the model glass formers that have been visualised so far.^[466–469]

8 Further Examples

The suggestion that multifunctional biomolecules are expected to exhibit multifunnel energy landscapes has a number of consequences, and opens up new research avenues. For example, a multifunnel scenario introduces significant energy barriers on the energy landscape, which may lead to problems of broken ergodicity for simulations, and require us to consider the effect of time scale resolution in experiment. With respect to the principle of minimal frustration, the question arises of how evolution has optimised multifunctional biomolecules. There is also the issue of how perturbations such as ligand binding, solvation conditions, or external fields, may shift the stability or occupation of alternative structural ensembles.

The interplay of sequence alterations, i.e. point mutations, with evolutionary optimisation, and the balance of stability between morphologies is particularly interesting. Nelson and Onuchic^[289] showed that the principle of minimal frustration stabilises energy landscapes with respect to sequence mutations for single funnel landscapes. This result is based on the fact that sequences have evolved to optimise native contacts to support the native state, leading to a deep funnel. Their results indicate that if a polypeptide still folds after a sequence mutation, the adopted native fold will closely resemble the original native structure, and the energy landscape will only exhibit subtle differences. Selection pressures influencing the stability, and hence the overall molecular functionality, should therefore affect the evolution of sequences.^[287;288]

A study by Tina *et al.*^[472] used a coarse-grained model to test the effects of mutations in more detail. Between a third and a half of the mutations were found to change the folding time, but preserve the native state. However, some mutations actually led to misfolding, changing

the energy landscape more significantly. This result was confirmed using amino acid propensities.^[473] Considering the sequence space and comparing the relative stability with respect to mutations, it was shown that there are ‘superfunnels’ centred around very robust sequences.^[290] Further work showed that the changes in interactions are non-linear, and particularly large shifts upon mutation are observed when hydrophilic residues are added in a hydrophobic region, disturbing the secondary structure.^[474] In general, these results imply that stability with respect to mutations is only marginal, and is shifted by more disruptive mutations.^[475;476] More recently, it was shown that mutations change the populations of different structural ensembles and folding pathways,^[477] modulate protein-ligand association,^[478] and lead to shifts in the free energy landscape of biomolecules.^[479;480]

We have recently analysed how sequence mutations change multifunnel energy landscapes.^[19;481] Across a number of examples, including proteins and RNA, the principle of minimal frustration holds in a more generalised way. In multifunctional systems the energy landscape was found to exhibit an extended minimal frustration by exhibiting the minimum number of funnels necessary to support the distinct structural ensembles required for functionality. In this section we discuss two examples in more detail, namely the coiled-coil GCN4-pLI^[482] and ubiquitin.^[483]

Both systems are the focus of active contemporary research, and hence a wealth of experimental data is available. The coiled-coil is a common motif in biological systems,^[484] with the leucine zipper of the yeast transcription factor GCN4^[485] and the derived family of coiled-coils^[486] of particular interest. Depending on the solvent and the exact sequence, multiple oligomer sizes for these coiled-coils have been observed,^[487;488] and further competition exists between parallel and antiparallel alignments of the α -helices. This competition is likely to be encoded in a multifunnel energy landscape,^[489] and FRET experiments^[490] have been used to observe the balance between two such states for the Rop dimer, another coiled-coil, confirming this picture, following theoretical predictions of a multifunneled scenario.^[491] In the GCN4-pLI tetramer^[486] a single point mutation leads to the crystallisation of two different configurations, while in the wild type only one is observed.^[492]

The ubiquitination of proteins is one of the two main post-translational protein modifications,^[493] making ubiquitin a versatile, important and, appropriately, a ubiquitous component of biological systems.^[494] A large number of similar crystal structures have been reported for ubiquitin (Ub configuration).^[495] While some dynamic regions of ubiquitin are known, for example the four C-terminal residues, and the $\beta 1/\beta 2$ -loop,^[496–498] hydrogen-exchange experiments have been reproduced using a single-funnel picture.^[499] In phosphorylated ubiquitin, a significant structural change was observed, revealing a second conformation with a slipped C-terminal region moving the $\beta 5$ -strand and changing the Ser65-loop in crystal structures^[500] and NMR experiments.^[501] Recently, Gladkova *et al.*^[502] reported that an analogous retracted conformation (Ub-CR) exists in unphosphorylated ubiquitin, with important implications for the Ser65 phosphorylation, which is linked to mitophagy, the process of isolation and clearance of damaged parts of the mitochondria.^[503;504]

In both systems we have found that the energy landscape for the native sequence exhibits two funnels (see **Figure 8 and 7**, panel (a)). For the coiled-coil these funnels correspond to the parallel and antiparallel alignments, and for ubiquitin they are the canonical and the retracted configurations. In the coiled-coil example the higher energy funnel is inaccessible at physiological temperatures,^[482] but for ubiquitin it has a small but significant occupation probability.^[502] Modelling ubiquitin in terms of single funnel therefore can therefore describe some properties well;^[499] however, as experimental resolution increases and our understanding of complex biological process increases, the multifunnel scenario becomes more important.

In both cases single point mutations alter the experimentally observed behaviour,^[492;502] which can be explained by a detailed study of the pathways between the funnels (see for example Figure 8 b, c and d).^[482;483] More generally we have found that sequence mutations

introduce a number of changes in the energy landscape: the existing funnels shift in width and depth, the relative energies and barriers between the corresponding ensembles change, and ultimately new funnels may appear while others may disappear.

The multifunnel character of the landscape, apparent from disconnectivity graphs, may produce a number of characteristic observable features, as discussed for double-funnel landscapes in §7.2. One example is the analysis of the heat capacity for the coiled-coil (see Figure 7 c and d), where a clear difference between the heat capacity curves is observed between native and mutant sequence,^[482] associated with thermal accessibility of both competing states in the mutant. Furthermore, the distribution of the minima in potential energy reveals clear and distinct structural ensembles for the different funnels (see Figures 7e and 8e). From the structure-function paradigm, these different ensembles are expected to have complementary functions. The distribution of transition states in potential energy tracks the distribution for the minima, which can be viewed partly as a consequence of the Hammond postulate^[505] or the Bell-Evans-Polanyi principle.^[428;506] These trends in turn can be understood from catastrophe theory.^[12;507] The separation of structural ensembles contrasts markedly with results for intrinsically disordered peptides, where there are no distinct features in the potential energy distribution of minima corresponding to the multiple funnels in a disconnectivity graph (see Figure 5).

Another consequence of point mutations is increased frustration, which correlates with greater structural variation.^[19] As a consequence, alterations of naturally occurring sequences to obtain new functionality, as well as *de novo* sequence designs, will most likely exhibit frustrated landscapes, leading to a number of competing structural ensembles and functions. The design process should therefore aim to reduce the frustration, while at the same time improving the functionality. The mutational basin-hopping approach has recently been developed to realise these conditions.^[508]

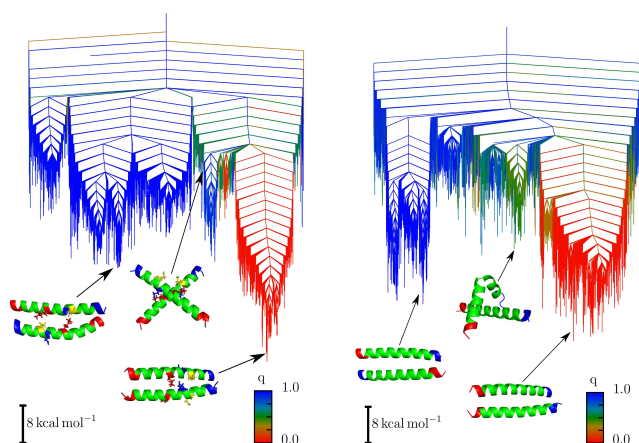
9 Conclusions

In this report we have attempted to summarise recent developments in our understanding of potential energy landscapes, focusing on biomolecules, and providing context in terms of earlier models and alternative experimental, theoretical, and computational approaches. The association of distinct functions with alternative structural ensembles and corresponding features of the landscape provides a common framework for explaining emergent properties and designing new targets, which should be applicable throughout molecular science. The examples we have highlighted include double-funnel and multiple-funnel landscapes, and we have compared molecules where multifunctional capabilities have evolved with systems where disorder probably has no functional role.

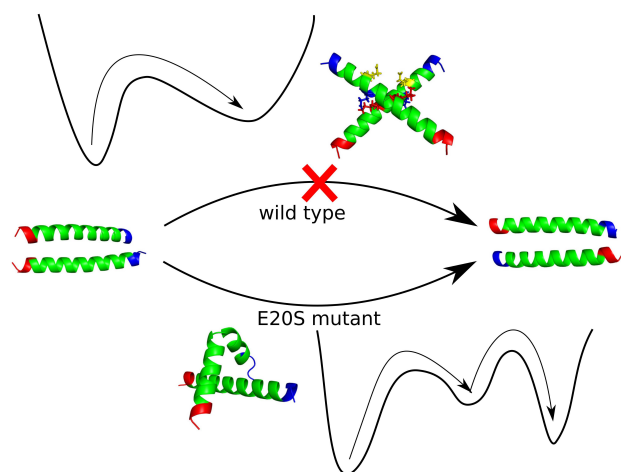
Continuing improvements in theory and computational methods should provide access to larger and more elaborate systems in future work. The current state-of-the-art can provide the foundations to advance our understanding and design of functionality on new scales of complexity.

References

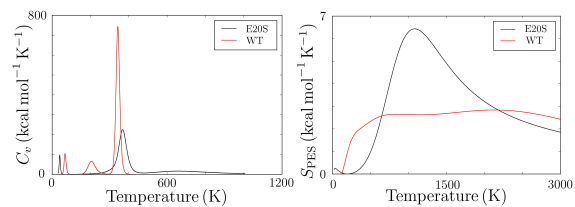
- [1] F. Sanger, *Biochem. J.* **45**, 563 (1949).
- [2] J. C. Kendrew, G. Bodo, H. M. Dintzis, R. G. Parrish, H. Wyckoff and D. C. Phillips, *Nature* **181**, 662 (1958).
- [3] E. Haber and C. B. Anfinsen, *J. Biol. Chem.* **237**, 1839 (1962).



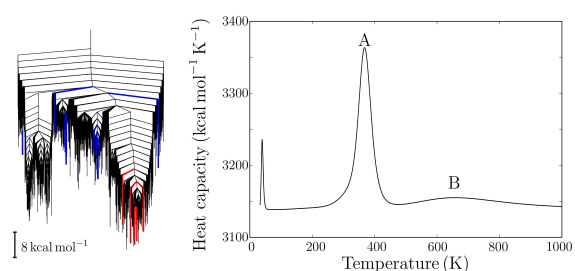
(a) Native sequence and E20S mutant PE landscapes



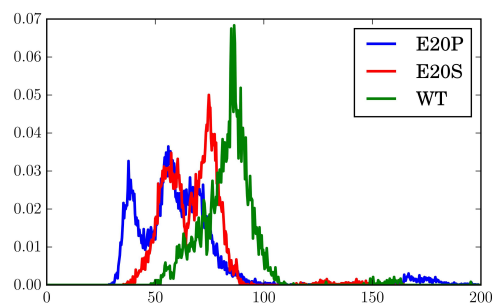
(b) Schematic of the transition mechanism



(c) Heat capacity and landscape entropy

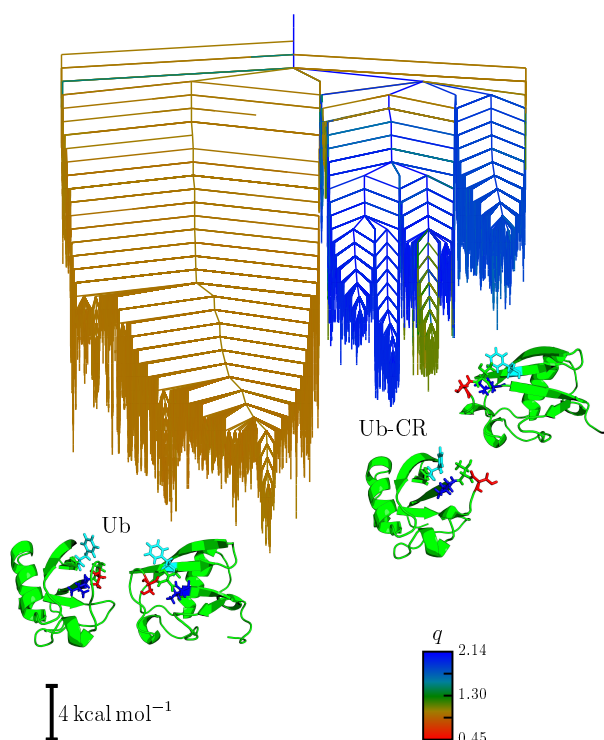


(d) Heat capacity analysis

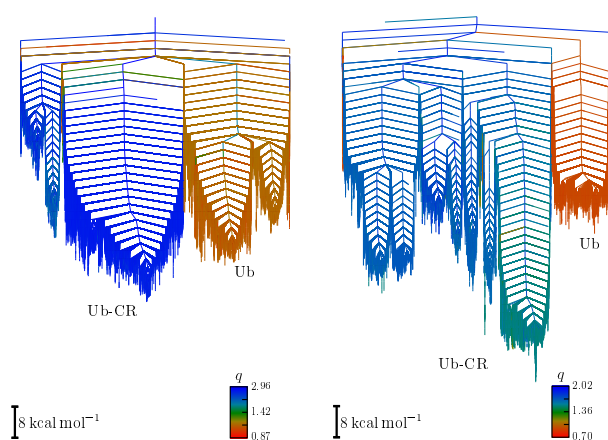


(e) Distribution of potential energy for the minima

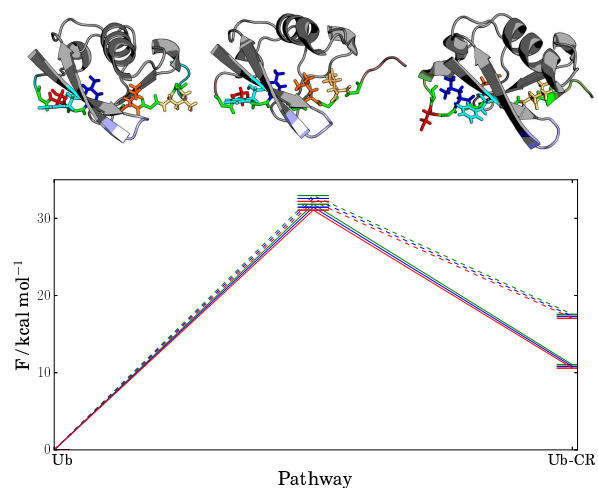
Figure 7: Case study for the multifunnel energy landscape for GCN4-pLI. a) The energy landscapes for the native sequence and the E20S mutant show multifunnel character, though the observed behaviour^[492] differs based on the relative energies between them (b). The difference between the two landscapes is apparent in the heat capacity for both cases (c and d). The distribution of potential energy for the minima show a number of overlapping, but distinct ensembles for the wild type and the mutants.^[482] Adapted in parts under the terms of the CC-BY licence.^[482] Copyright 2017, ACS.



(a) Wild type ubiquitin



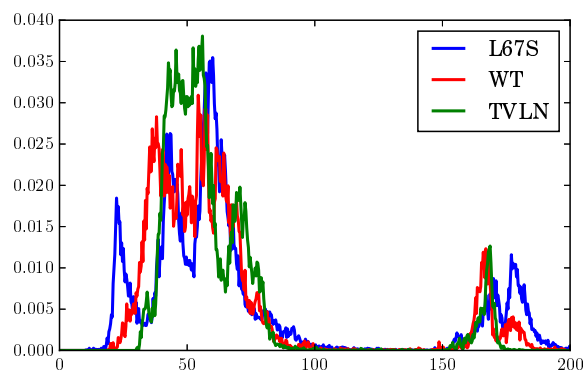
(b) T66V+L67N (left) and L67S mutant (right)



(c) Mechanism and free energy barriers in the wild type

Sequence	T (°C)	ΔG^{exp}	ΔG^{calc}
Wild type	45	19.3	20.6
T66V+L67N	45	15.8	13.0
F4A	25	19.1	18.9
	45	17.4	17.7

(d) Free energy barriers in kcal mol⁻¹



(e) Distribution of potential energy of minima

Figure 8: Case study for the multifunnel energy landscape for ubiquitin. a) The energy landscape for ubiquitin shows two funnels, accommodating the canonical (Ub) and retracted fold (Ub-CR). b) The multifunnel character becomes more pronounced in mutants (here T66V+L67N and L67S) due to an increased frustration and changes in funnel depth, width and relative positions are apparent. The double funnel character in the wildtype can be related to the mechanism of retraction (c) and simulation matches experimental observations^[502] well (d). The distribution of potential energy for the minima show a number of overlapping, but distinct ensembles for the wild type and the mutants. More information can be found in previously published work.^[483] In parts adapted with permission.^[483] Copyright 2018, ACS.

- [4] C. J. Epstein, R. F. Goldberger and C. B. Anfinsen, in *Cold Spring Harb. Symp. Quant. Biol.*, vol. 28, pp. 439–449. Cold Spring Harbor Laboratory Press (1963).
- [5] C. B. Anfinsen, *Biochem. J.* **128**, 737 (1972).
- [6] C. B. Anfinsen, *Science* **181**, 223 (1973).
- [7] C. Levinthal, in *Mossbauer Spectroscopy in Biological Systems, Proceedings of a Meeting held at Allerton House, Monticello, Illinois*, edited by P. Debrunner, J. Tsibris and E. Munck, p. 22, Urbana (1969), University of Illinois Press.
- [8] M. Karplus, *Fold. Des.* **2**, S69 (1997).
- [9] C. M. Dobson and M. Karplus, *Curr. Opin. Struc. Biol.* **9**, 92 (1999).
- [10] C. Levinthal, *Sci. Am.* **214**, 42 (1966).
- [11] C. Levinthal, *J. Chim. Phys. Pcb.* **65**, 44 (1968).
- [12] D. J. Wales, *Energy Landscapes*, Cambridge University Press, Cambridge (2003).
- [13] D. J. Wales, *Phil. Trans. Roy. Soc. A* **363**, 357 (2005).
- [14] D. J. Wales, M. A. Miller and T. R. Walsh, *Nature* **394**, 758 (1998).
- [15] M. A. Miller, J. P. K. Doye and D. J. Wales, *J. Chem. Phys.* **110**, 328 (1999).
- [16] J. P. K. Doye, M. A. Miller and D. J. Wales, *J. Chem. Phys.* **110**, 6896 (1999).
- [17] D. J. Wales and T. V. Bogdan, *J. Phys. Chem. B* **110**, 20765 (2006).
- [18] Y. Chebaro, A. J. Ballard, D. Chakraborty and D. J. Wales, *Sci. Rep.* **5**, 10386 (2015).
- [19] K. Röder and D. J. Wales, *J. Phys. Chem. B* p. in press (2018).
- [20] A. M. Labhardt, *Proc. Natl. Acad. Sci. USA* **81**, 7674 (1984).
- [21] M. Ikeguchi, K. Kuwajima, M. Mitani and S. Sugai, *Biochemistry* **25**, 6965 (1986).
- [22] R. I. Gilmanashin and O. B. Ptitsyn, *FEBS Lett.* **223**, 327 (1987).
- [23] K. Kuwajima, H. Yamaya, S. Miwa, S. Sugai and T. Nagamura, *FEBS Lett.* **221**, 115 (1987).
- [24] K. Kuwajima, A. Sakuraoka, S. Fueki, M. Yoneyama and S. Sugai, *Biochemistry* **27**, 7419 (1988).
- [25] H. Roder and K. Wüthrich, *Proteins: Struct., Funct., Bioinf.* **1**, 34 (1986).
- [26] H. Roder, G. A. Elöve and S. W. Englander, *Nature* **335**, 700 (1988).
- [27] J. B. Udgaonkar and R. L. Baldwin, *Nature* **335**, 694 (1988).
- [28] S. Takahashi, S.-R. Yeh, T. K. Das, C.-K. Chan, D. S. Gottfried and D. L. Rousseau, *Nat. Struct. Mol. Biol.* **4**, 44 (1997).
- [29] C.-K. Chan, Y. Hu, S. Takahashi, D. L. Rousseau, W. A. Eaton and J. Hofrichter, *Proc. Natl. Acad. Sci. USA* **94**, 1779 (1997).

- [30] M. C. R. Shastry, S. D. Luck and H. Roder, *Biophys. J.* **74**, 2714 (1998).
- [31] C. M. Phillips, Y. Mizutani and R. M. Hochstrasser, *Proc. Natl. Acad. Sci. USA* **92**, 7292 (1995).
- [32] S. Williams, T. P. Causgrove, R. Gilmanshin, K. S. Fang, R. H. Callender, W. H. Woodruff and R. B. Dyer, *Biochemistry* **35**, 691 (1996).
- [33] V. Muñoz, P. A. Thompson, J. Hofrichter and W. A. Eaton, *Nature* **390**, 196 (1997).
- [34] P. A. Thompson, W. A. Eaton and J. Hofrichter, *Biochemistry* **36**, 9200 (1997).
- [35] R. Gilmanshin, S. Williams, R. H. Callender, W. H. Woodruff and R. B. Dyer, *Proc. Natl. Acad. Sci. USA* **94**, 3709 (1997).
- [36] U. Mayor, C. M. Johnson, V. Daggett and A. R. Fersht, *Proc. Natl. Acad. Sci. USA* **97**, 13518 (2000).
- [37] C. D. Snow, L. Qiu, D. Du, F. Gai, S. J. Hagen and V. S. Pande, *Proc. Natl. Acad. Sci. USA* **101**, 4077 (2004).
- [38] J. Kubelka, T. K. Chiu, D. R. Davies, W. A. Eaton and J. Hofrichter, *J. Mol. Biol.* **359**, 546 (2006).
- [39] J. Chen, D. L. Rempel and M. L. Gross, *J. Am. Chem. Soc.* **132**, 15502 (2010).
- [40] C. M. Davis, S. Xiao, D. P. Raleigh and R. B. Dyer, *J. Am. Chem. Soc.* **134**, 14476 (2012).
- [41] X. Zhang and A. Tokmakoff, *Biophys. J.* **114**, 209a (2018).
- [42] P. L. Privalov, *Crit. Rev. Biochem. Mol. Biol.* **25**, 281 (1990).
- [43] J. Zhang, X. Peng, A. Jonas and J. Jonas, *Biochemistry* **34**, 8631 (1995).
- [44] J. Sabelko, J. Ervin and M. Gruebele, *J. Phys. Chem. B* **102**, 1806 (1998).
- [45] R. H. Callender, R. B. Dyer, R. Gilmanshin and W. H. Woodruff, *Annu. Rev. Phys. Chem.* **49**, 173 (1998).
- [46] K. M. Pryse, T. G. Bruckman, B. W. Maxfield and E. L. Elson, *Biochemistry* **31**, 5127 (1992).
- [47] G. Desai, G. Panick, M. Zein, R. Winter and C. A. Royer, *J. Mol. Biol.* **288**, 461 (1999).
- [48] J. Woenckhaus, R. Köhling, P. Thiyagarajan, K. C. Littrell, S. Seifert, C. A. Royer and R. Winter, *Biophys. J.* **80**, 1518 (2001).
- [49] T. R. Alderson, C. Charlier, D. A. Torchia, P. Anfinrud and A. Bax, *J. Am. Chem. Soc.* **139**, 11036 (2017).
- [50] X. Peng, J. Jonas and J. L. Silva, *Proc. Natl. Acad. Sci. USA* **90**, 1776 (1993).
- [51] M. W. Lassalle, H. Yamada and K. Akasaka, *J. Mol. Biol.* **298**, 293 (2000).
- [52] C. Dumont, T. Emilsson and M. Gruebele, *Nat. Methods* **6**, 515 (2009).

- [53] A. J. Wirth, Y. Liu, M. B. Prigozhin, K. Schulten and M. Gruebele, *J. Am. Chem. Soc.* **137**, 7152 (2015).
- [54] T. Pascher, J. P. Chesick, J. R. Winkler and H. B. Gray, *Science* **271**, 1558 (1996).
- [55] E. Chen, P. Wittung-Stafshede and D. S. Kliger, *J. Am. Chem. Soc.* **121**, 3811 (1999).
- [56] A. Ansari, J. Berendzen, S. F. Bowne, H. Frauenfelder, I. E. Iben, T. B. Sauke, E. Shyamsunder and R. D. Young, *Proc. Natl. Acad. Sci. USA* **82**, 5000 (1985).
- [57] C. M. Jones, E. R. Henry, Y. Hu, C.-K. Chan, S. D. Luck, A. Bhuyan, H. Roder, J. Hofrichter and W. A. Eaton, *Proc. Natl. Acad. Sci. USA* **90**, 11860 (1993).
- [58] M. Gruebele, in *Protein Folding, Misfolding and Aggregation: Classical Themes and Novel Approaches*, edited by V. Muñoz, pp. 106–138. Royal Society of Chemistry (2008).
- [59] K. C. Hansen, R. S. Rock, R. W. Larsen and S. I. Chan, *J. Am. Chem. Soc.* **122**, 11567 (2000).
- [60] L. Redecke, S. Binder, M. I. Y. Elmallah, R. Broadbent, C. Tilkorn, B. Schulz, P. May, A. Goos, A. Eich, M. Rübhausen and C. Betzel, *Free Radic. Biol. Med.* **46**, 1353 (2009).
- [61] T. E. Schrader, W. J. Schreier, T. Cordes, F. O. Koller, G. Babitzki, R. Denschlag, C. Renner, M. Löweneck, S.-L. Dong, L. Moroder, , P. Tavan and W. Zinth, *Proc. Natl. Acad. Sci. USA* **104**, 15729 (2007).
- [62] A. A. Beharry and G. A. Woolley, *Chem. Soc. Rev.* **40**, 4422 (2011).
- [63] M. Rief, M. Gautel, F. Oesterhelt, J. M. Fernandez and H. E. Gaub, *Science* **276**, 1109 (1997).
- [64] J. M. Fernandez and H. Li, *Science* **303**, 1674 (2004).
- [65] M. S. Z. Kellermayer, S. B. Smith, H. L. Granzier and C. Bustamante, *Science* **276**, 1112 (1997).
- [66] E. A. Shank, C. Cecconi, J. W. Dill, S. Marqusee and C. Bustamante, *Nature* **465**, 637 (2010).
- [67] C. Jarzynski, *Phys. Rev. Lett.* **78**, 2690 (1997).
- [68] C. Jarzynski, *Phys. Rev. E* **56**, 5018 (1997).
- [69] C. Jarzynski, *C. R. Physique* **8**, 495 (2007).
- [70] G. Hummer and A. Szabo, *Proc. Natl. Acad. Sci. USA* **98**, 3658 (2001).
- [71] G. Hummer and A. Szabo, *Proc. Natl. Acad. Sci. USA* **107**, 21441 (2010).
- [72] O. K. Dudko, J. Mathé, A. Szabo, A. Meller and G. Hummer, *Biophys. J.* **92**, 4188 (2007).
- [73] O. Dudko, G. Hummer and A. Szabo, *Phys. Rev. Lett.* **96**, 108101 (2006).
- [74] O. K. Dudko, G. Hummer and A. Szabo, *Proc. Natl. Acad. Sci. USA* **105**, 15755 (2008).
- [75] G. Zoldak, J. Stigler, B. Pelz, H. B. Li and M. Rief, *Proc. Natl. Acad. Sci. USA* **110**, 18156 (2013).

- [76] W. M. Jackson and J. F. Brandts, *Biochemistry* **9**, 2294 ().
- [77] C. M. Johnson, *Arch. Biochem. Biophys.* **531**, 100 (2013).
- [78] N. C. Garbett, M. J. J., A. B. Jenson and J. B. Chaires, *Biophys. J.* **94**, 1377 (2008).
- [79] A. A. Chagovetz, R. L. Jensen, L. Recht, M. Glantz and A. M. Chaovetz, *J. Neuro-Oncol.* **105**, 499 (2011).
- [80] C. Tanford, *J. Am. Chem. Soc.* **86**, 2050 ().
- [81] R. F. Greene and C. N. Pace, *J. Biol. Chem.* **249**, 5388 (1974).
- [82] C. N. Pace, *Methods Enzymol.* **259**, 538 (1995).
- [83] C. N. Pace, B. A. Shirley, M. McNutt and K. Gajiwala, *FASEB J.* **10**, 75 (1996).
- [84] J. Walters, S. L. Milam and A. C. Clark, *Methods Enzymol.* **455**, 1 (2009).
- [85] M. F. Perutz, M. G. Rossmann, A. F. Cullis, H. Muirhead, G. Will and A. C. T. North, *Nature* **185**, 416 (1960).
- [86] C. C. F. Blake, D. F. Koenig, G. A. Mair, A. C. T. North, D. C. Phillips and V. R. Sarma, *Nature* **206**, 757 (1965).
- [87] G. Kartha, J. Bello and D. Harker, *Nature* **213**, 862 (1967).
- [88] H. W. Wyckoff, K. D. Hardman, N. M. Allewell, T. Inagami, L. N. Johnson and F. M. Richards, *J. Biol. Chem.* **242**, 3984 (1967).
- [89] Y. Shi, *Cell* **159**, 995 (2014).
- [90] D. J. Segel, A. Bachmann, J. Hofrichter, K. O. Hodgson, S. Doniach and T. Kiefhaber, *J. Mol. Biol.* **288**, 489 (1999).
- [91] K. W. Plaxco, I. S. Millett, D. J. Segel, S. Doniach and D. Baker, *Nat. Struct. Mol. Biol.* **6**, 554 (1999).
- [92] L. Pollack, M. W. Tate, A. C. Finnefrock, C. Kalidas, S. Trotter, N. C. Darnton, L. Lurio, R. H. Austin, C. A. Batt, S. M. Gruner and S. G. J. Mochrie, *Phys. Rev. Lett.* **86**, 4962 (2001).
- [93] D. I. Svergun and M. H. Koch, *Curr. Opin. Struct. Biol.* **12**, 654 (2002).
- [94] D. I. Svergun and M. H. J. Koch, *Rep. Prog. Phys.* **66**, 1735 (2003).
- [95] J. Lipfert and S. Doniach, *Annu. Rev. Bioph. Biom.* **36**, 307 (2007).
- [96] H. D. T. Mertens and D. I. Svergun, *J. Struct. Biol.* **172**, 128 (2010).
- [97] A. G. Kikhney and D. I. Svergun, *FEBS Lett.* **589**, 2570 (2015).
- [98] M. Adrian, J. Dubochet, J. Lepault and A. W. McDowell, *Nature* **308**, 32 (1984).
- [99] R. Henderson, J. Baldwin, T. Ceska, F. Zemlin, E. Beckmann and K. Downing, *J. Mol. Biol.* **213**, 899 (1990).
- [100] E. Callaway, *Nature* **525**, 172 (2015).

- [101] K. Wuthrich, *NMR of proteins and nucleic acids*, Wiley (1986).
- [102] K. Wüthrich, *Nat. Struct. Mol. Biol.* **8**, 923 (2001).
- [103] H. J. Dyson and P. E. Wright, *Chem. Rev.* **104**, 3607 (2004).
- [104] A. E. Smith, Z. Zhang, G. J. Pielak and C. Li, *Curr. Opin. Struct. Biol.* **30**, 7 (2015).
- [105] D. S. Wishart and B. D. Sykes, *Methods Enzymol.* **239**, 363 (1994).
- [106] J. K. Myers and T. G. Oas, *Annu. Rev. Biochem.* **71**, 783 (2002).
- [107] P. Neudecker, P. Lundström and L. E. Kay, *Biophys. J.* **96**, 2045 (2009).
- [108] K. H. Gardner and L. E. Kay, *Annu. Rev. Biophys. Biomol. Struct.* **27**, 357 (1998).
- [109] A. Mittermaier and L. E. Kay, *Science* **312**, 224 (2006).
- [110] D. Marion and K. Wüthrich, *Biochem. Biophys. Res. Commun.* **113**, 967 (1983).
- [111] D. R. Shortle, *Curr. Opin. Struct. Biol.* **6**, 24 (1996).
- [112] G. Bodenhausen and D. J. Ruben, *Chem. Phys. Lett.* **69**, 185 (1980).
- [113] S. W. Englander and L. Mayne, *Annu. Rev. Biophys. Biomol. Struct.* **21**, 243 (1992).
- [114] N. Tjandra and A. Bax, *Science* **278**, 1111 (1997).
- [115] J.-C. Hus, D. Marion and M. Blackledge, *J. Am. Chem. Soc.* **123**, 1541 (2001).
- [116] R. Mohana-Borges, N. K. Goto, G. J. A. Kroon, H. J. Dyson and P. E. Wright, *J. Mol. Biol.* **340**, 1131 (2004).
- [117] M. R. Jensen, R. W. H. Ruigrok and M. Blackledge, *Curr. Opin. Struct. Biol.* **23**, 426 (2013).
- [118] K. M. Ward, A. H. Aletras and R. S. Balaban, *J. Magn. Reson.* **143**, 79 (2000).
- [119] P. Vallurupalli, G. Bouvignies and L. E. Kay, *J. Am. Soc. Chem.* **134**, 8148 (2012).
- [120] O. Ivchenko, C. S. Whittleston, J. M. Carr, P. Imhof, S. Goerke, P. Bachert and D. J. Wales, *J. Phys. Chem. B* **118**, 1969 (2014).
- [121] R. Tycko, *Annu. Rev. Phys. Chem.* **62**, 279 (2011).
- [122] N. J. Greenfield and G. D. Fasman, *Biochemistry* **8**, 4108 (1969).
- [123] S. M. Kelly and N. C. Price, *Biochim. Biophys. Acta, Protein Struct. Mol. Enzymol.* **1338**, 161 (1997).
- [124] N. J. Greenfield, *TrAC, Trends Anal. Chem.* **18**, 236 (1999).
- [125] S. M. Kelly, T. J. Jess and N. C. Price, *Biochim. Biophys. Acta, Proteins Proteomics* **1751**, 119 (2005).
- [126] N. J. Greenfield, *Nat. Protoc.* **1**, 2876 (2007).
- [127] L. Whitmore and B. A. Wallace, *Biopolymers* **89**, 392 (2008).

- [128] N. J. Greenfield, *Nat. Protoc.* **1**, 2876 (2006).
- [129] S. W. Provencher and J. Gloeckner, *Biochemistry* **20**, 33 (1981).
- [130] S. Vuilleumier, J. Sancho, R. Loewenthal and A. R. Fersht, *Biochemistry* **32**, 10303 (1993).
- [131] N. Sreerama, M. C. Manning, M. E. Powers, J.-X. Zhang, D. P. Goldenberg and R. W. Woody, *Biochemistry* **38**, 10814 (1999).
- [132] S. Benjwal, S. Verma, K.-H. Röhm and O. Gursky, *Protein Sci.* **15**, 635 (2006).
- [133] L. R. McLean and A. Balasubramaniam, *Biochim. Biophys. Acta, Protein Struct. Mol. Enzymol.* **1122**, 317 (1992).
- [134] C. Goldsbury, K. Goldie, J. Pellaud, J. Seelig, P. Frey, S. A. Müller, J. Kistler, G. J. S. Cooper and U. Aebi, *J. Struct. Biol.* **130**, 352 (2000).
- [135] D. Kourouski, R. K. Dukor, X. Lu, L. A. Nafie and I. K. Lednev, *Biophys. J.* **103**, 522 (2012).
- [136] C. J. Barrow, A. Yasuda, P. T. M. Kenny and M. G. Zagorski, *J. Mol. Biol.* **225**, 1075 (1992).
- [137] E. Terzi, G. Hoelzemann and J. Seelig, *Biochemistry* **33**, 1345 (1994).
- [138] P. K. Mandal and J. W. Pettegrew, *Neurochem. Res.* **29**, 2267 (2004).
- [139] R. B. Dyer, F. Gai, W. H. Woodruff, R. Gilmanishin and R. H. Callender, *Acc. Chem. Res.* **31**, 709 (1998).
- [140] A. Barth, *Biochim. Biophys. Acta, Bioenerg.* **1767**, 1073 (2007).
- [141] W. K. Surewicz, H. H. Mantsch and D. Chapman, *Biochemistry* **32**, 389 (1993).
- [142] J. Kong and S. Yu, *Acta Biochim. Biophys. Sin.* **39**, 549 (2007).
- [143] H. Yang, S. Yang, J. Kong, A. Dong and S. Yu, *Nat. Protoc.* **10**, 382 (2015).
- [144] A. Ghosh, J. S. Ostrander and M. T. Zanni, *Chem. Rev.* **117**, 10726 (2017).
- [145] J. Seo, W. Hoffmann, S. Warnke, X. Huang, S. Gewinner, W. Schöllkopf, M. T. Bowers, G. von Helden and K. Pagel, *Nat. Chem.* **9**, 39 (2017).
- [146] S. J. Roeters, A. Iyer, G. Pletikapić, V. Kogan, V. Subramaniam and S. Woutersen, *Sci. Rep.* **7**, 41051 (2017).
- [147] G. J. Thomas Jr, *Annu. Rev. Biophys. Biomol. Struct.* **28**, 1 (1999).
- [148] R. Tuma, *J. Raman Spectrosc.* **36**, 307 (2005).
- [149] J. M. Beechem and L. Brand, *Annu. Rev. Biochem.* **54**, 43 (1985).
- [150] C. A. Royer, *Chem. Rev.* **106**, 1769 (2006).
- [151] S. Khorasanizadeh, I. D. Peters and H. Roder, *Nat. Struct. Mol. Biol.* **3**, 193 (1996).

- [152] L. Qiu, S. A. Pabit, A. E. Roitberg and S. J. Hagen, *J. Am. Chem. Soc.* **124**, 12952 (2002).
- [153] A. Hawe, M. Sutter and W. Jiskoot, *Pharm. Res.* **25**, 1487 (2008).
- [154] T. Förster, *Ann. Phys.* **437**, 55 (1948).
- [155] R. M. Clegg, *Curr. Opin. Biotechnol.* **6**, 103 (1995).
- [156] E. Haustein and P. Schwille, *Methods* **29**, 153 (2003).
- [157] T. Förster, in *Modern Quantum Chemistry. Istanbul Lectures. Part III: Action of Light and Organic Crystals*, vol. 3, pp. 93–137. Academic Press (1965).
- [158] B. A. Pollok and R. Heim, *Trends Cell Biol.* **9**, 57 (1999).
- [159] X. Michalet, S. Weiss and M. Jäger, *Chem. Rev.* **106**, 1785 (2006).
- [160] B. Schuler and W. A. Eaton, *Curr. Opin. Struct. Biol.* **18**, 16 (2008).
- [161] H. S. Chung, K. McHale, J. M. Louis and W. A. Eaton, *Science* **335**, 981 (2012).
- [162] R. B. Sekar and A. Periasamy, *J. Cell Biol.* **160**, 629 (2003).
- [163] R. Diamond, *Acta Crystallogr.* **21**, 253 (1966).
- [164] M. Levitt and S. Lifson, *J. Mol. Biol.* **46**, 269 (1969).
- [165] M. Levitt, *J. Mol. Biol.* **82**, 393 (1974).
- [166] B. R. Gelin and M. Karplus, *Proc. Natl. Acad. Sci. USA* **74**, 801 (1977).
- [167] B. R. Gelin and M. Karplus, *Proc. Natl. Acad. Sci. USA* **72**, 2002 (1975).
- [168] R. Hetzel, K. Wüthrich, J. Deisenhofer and R. Huber, *Biophys. Struct. Mech.* **2**, 159 (1976).
- [169] K. D. Gibson and H. A. Scheraga, *Proc. Natl. Acad. Sci. USA* **63**, 9 (1969).
- [170] G. Némethy and H. A. Scheraga, *Q. Rev. Biophys.* **10**, 239 (1977).
- [171] M. Levitt and A. Warshel, *Nature* **253**, 694 (1975).
- [172] U. H. E. Hansmann and Y. Okamoto, *Curr. Opin. Struct. Biol.* **9**, 177 (1999).
- [173] J. Skolnick and A. Kolinski, *Science* **250**, 1121 (1990).
- [174] E. Shakhnovich, G. Farztdinov, A. M. Gutin and M. Karplus, *Phys. Rev. Lett.* **67**, 1665 (1991).
- [175] A. Kolinski and J. Skolnick, *Proteins: Struct., Funct., Bioinf.* **18**, 338 (1994).
- [176] M.-H. Hao and H. A. Scheraga, *J. Phys. Chem.* **98**, 4940 (1994).
- [177] N. D. Socci and J. N. Onuchic, *J. Chem. Phys.* **101**, 1519 (1994).
- [178] M. Karplus and J. A. McCammon, *Nat. Struct. Mol. Biol.* **9**, 646 (2002).

- [179] T. Hansson, C. Oostenbrink and W. F. van Gunsteren, *Curr. Opin. Struct. Biol.* **12**, 190 (2002).
- [180] M. Karplus and J. Kuriyan, *Proc. Natl. Acad. Sci. USA* **102**, 6679 (2005).
- [181] J. A. McCammon, B. R. Gelin and M. Karplus, *Nature* **267**, 585 (1977).
- [182] M. Levitt, *Nature* **294**, 379 (1981).
- [183] S. H. Northrup, M. R. Pear, J. D. Morgan, J. A. McCammon and M. Karplus, *J. Mol. Biol.* **153**, 1087 (1981).
- [184] M. Karplus and G. A. Petsko, *Nature* **347**, 631 (1990).
- [185] R. O. Dror, R. M. Dirks, J. Grossman, H. Xu and D. E. Shaw, *Annu. Rev. Biophys.* **41**, 429 (2012).
- [186] D. E. Shaw, R. O. Dror, J. K. Salmon, J. Grossman, K. M. Mackenzie, J. A. Bank, C. Young, M. M. Deneroff, B. Batson, K. J. Bowers, E. Chow, M. Eastwood, D. Ierardi, J. Klepeis, J. Kuskin, R. Larson, K. Lindorff-Larsen, P. Maragakis, M. Moraes, S. Piana, Y. Shan and B. Towles, in *Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis*, p. 39 (2009).
- [187] M. Shirts and V. S. Pande, *Science* **290**, 1903 (2000).
- [188] F. Allen, G. Almasi, W. Andreoni, D. Beece, B. J. Berne, A. Bright, J. Brunheroto, C. Cascaval, J. Castanos, P. Coteus, P. Crumley, A. Curioni, M. Denneau, W. Donath, M. Eleftheriou, B. Flitch, B. Fleischer, C. J. Georgiou, R. Germain, M. Giampapa, D. Gresh, M. Gupta, R. Haring, H. Ho, P. Hochschild, S. Hummel, T. Jonas, D. Lieber, G. Martyna, K. Maturu, J. Moreira, D. Newns, M. Newton, R. Philhower, T. Picunko, J. Pitera, M. Pitman, R. Rand, A. Royyuru, V. Salapura, A. Sanomiya, R. Shah, Y. Sham, S. Singh, M. Snir, F. Suits, R. Swetz, W. C. Swope, N. Vishnumurthy, T. J. C. Ward, H. Warren and R. Zhou, *IBM Syst. J.* **40**, 310 (2001).
- [189] I. Buch, M. J. Harvey, T. Giorgino, D. P. Anderson and G. De Fabritiis, *J. Chem. Inf. Model.* **50**, 397 (2010).
- [190] D. E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R. O. Dror, M. P. Eastwood, J. A. Bank, J. M. Jumper, J. K. Salmon, Y. Shan and W. Wriggers, *Science* **330**, 341 (2010).
- [191] V. A. Voelz, G. R. Bowman, K. Beauchamp and V. S. Pande, *J. Am. Chem. Soc.* **132**, 1526 (2010).
- [192] G. R. Bowman, V. A. Voelz and V. S. Pande, *J. Am. Chem. Soc.* **133**, 664 (2011).
- [193] K. Lindorff-Larsen, S. Piana, R. O. Dror and D. E. Shaw, *Science* **334**, 517 (2011).
- [194] J. Kubelka, J. Hofrichter and W. A. Eaton, *Curr. Opin. Struct. Biol.* **14**, 76 (2004).
- [195] K. A. Beauchamp, R. McGibbon, Y.-S. Lin and V. S. Pande, *Proc. Natl. Acad. Sci.* **109**, 17807 (2012).
- [196] A. Dickson and C. L. Brooks III, *J. Am. Chem. Soc.* **135**, 4729 (2013).
- [197] J. K. Weber, R. L. Jack and V. S. Pande, *J. Am. Chem. Soc.* **135**, 5501 (2013).

- [198] B. E. Husic, R. T. McGibbon, M. M. Sultan and V. S. Pande, *J. Chem. Phys.* **145**, 194103 (2016).
- [199] V. Daggett and M. Levitt, *Curr. Opin. Struct. Biol.* **4**, 291 (1994).
- [200] A. Caflisch and M. Karplus, *Proc. Natl. Acad. Sci. USA* **91**, 1746 (1994).
- [201] M. Karplus and A. Šali, *Curr. Opin. Struct. Biol.* **5**, 58 (1995).
- [202] J.-E. Shea and C. L. Brooks III, *Annu. Rev. Phys. Chem.* **52**, 499 (2001).
- [203] U. H. E. Hansmann, *Chem. Phys. Lett.* **281**, 140 (1997).
- [204] Y. Sugita and Y. Okamoto, *Chem. Phys. Lett.* **314**, 141 (1999).
- [205] D. A. Kofke, *J. Chem. Phys.* **117**, 6911 (2002).
- [206] A. J. Ballard and D. J. Wales, *J. Chem. Theory Comput.* **10**, 5599 (2014).
- [207] K. Y. Sanbonmatsu and A. E. Garcia, *Proteins: Struct., Funct., Bioinf.* **46**, 225 (2002).
- [208] A. K. Felts, Y. Harano, E. Gallicchio and R. M. Levy, *Proteins: Struct., Funct., Bioinf.* **56**, 310 (2004).
- [209] P. H. Nguyen, G. Stock, E. Mittag, C. Hu and M. S. Li, *Proteins: Struct., Funct., Bioinf.* **61**, 795 (2005).
- [210] M. Cecchini, F. Rao, M. Seeber and A. Caflisch, *J. Chem. Phys.* **121**, 10748 (2004).
- [211] A. Baumketner and J.-E. Shea, *J. Mol. Biol.* **366**, 275 (2007).
- [212] S. Patel, E. Vierling and F. Tama, *Biophys. J.* **106**, 2644 (2014).
- [213] Y. M. Rhee and V. S. Pande, *Biophys. J.* **84**, 775 (2003).
- [214] G. Torrie and J. Valleau, *J. Comput. Phys.* **23**, 187 (1977).
- [215] S. H. Northrup, M. R. Pear, C.-Y. Lee, J. A. McCammon and M. Karplus, *Proc. Natl. Acad. Sci. USA* **79**, 4035 (1982).
- [216] A. Warshel, *Proc. Natl. Acad. Sci. USA* **81**, 444 (1984).
- [217] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen and P. A. Kollman, *J. Comput. Chem.* **13**, 1011 (1992).
- [218] C. Bartels, *Chem. Phys. Lett.* **331**, 446 (2000).
- [219] E. Rosta and G. Hummer, *J. Chem. Theory Comput.* **11**, 276 (2015).
- [220] A. S. Mey, H. Wu and F. Noé, *Phys. Rev. X* **4**, 041018 (2014).
- [221] H. Wu and F. Noé, *SIAM Multiscale Model. Simul.* **12**, 25 (2014).
- [222] H. Wu, A. S. J. S. Mey, E. Rosta and F. Noé, *J. Chem. Phys.* **141**, 214106 (2014).
- [223] L. S. Stelzl, A. Kells, E. Rosta and G. Hummer, *J. Chem. Theory Comput.* **13**, 6328 (2017).
- [224] D. J. Tobias, J. E. Mertz and C. L. Brooks III, *Biochemistry* **30**, 6054 (1991).

- [225] D. J. Tobias and C. L. Brooks III, *Biochemistry* **30**, 6059 (1991).
- [226] D. J. Tobias, S. F. Sneddon and C. L. Brooks III, *J. Mol. Biol.* **227**, 1244 (1992).
- [227] C. L. Brooks III and L. Nilsson, *J. Am. Chem. Soc.* **115**, 11034 (1993).
- [228] W. S. Young and C. L. Brooks III, *J. Mol. Biol.* **259**, 560 (1996).
- [229] A. Laio and M. Parrinello, *Proc. Natl. Acad. Sci. USA* **99**, 12562 (2002).
- [230] A. Laio and F. L. Gervasio, *Reports Prog. Phys.* **71**, 126601 (2008).
- [231] F. Baftizadeh, P. Cossio, F. Pietrucci and A. Laio, *Curr. Phys. Chem.* **2**, 79 (2012).
- [232] M. Balsera, S. Stepaniants, S. Izrailev, Y. Oono and K. Schulten, *Biophys. J.* **73**, 1281 (1997).
- [233] J. R. Gullingsrud, R. Braun and K. Schulten, *J. Comput. Phys.* **151**, 190 (1999).
- [234] J. Schlitter, M. Engels and P. Krüger, *J. Mol. Graph.* **12**, 84 (1994).
- [235] J. Ma and M. Karplus, *Proc. Natl. Acad. Sci. USA* **94**, 11905 (1997).
- [236] B. A. Berg and T. Neuhaus, *Phys. Lett. B* **267**, 249 (1991).
- [237] N. Nakajima, H. Nakamura and A. Kidera, *J. Phys. Chem. B* **101**, 817 (1997).
- [238] F. Wang and D. P. Landau, *Phys. Rev. Lett.* **86**, 2050 (2001).
- [239] C. Dellago, P. G. Bolhuis, F. S. Csajka and D. Chandler, *J. Chem. Phys.* **108**, 1964 (1998).
- [240] M. Bacci, J. Vymětal, M. Mihajlovic, A. Caffisch and A. Vitalis, *J. Chem. Theory Comput.* **13**, 5117 (2017).
- [241] N.-V. Buchete and G. Hummer, *Biophys. J.* **92**, 3032 (2007).
- [242] M. Han and U. H. E. Hansmann, *J. Chem. Phys.* **135**, 065101 (2011).
- [243] T. Gurry and C. M. Stultz, *Biochemistry* **53**, 6981 (2014).
- [244] W. Zheng, M.-Y. Tsai and P. G. Wolynes, *J. Am. Chem. Soc.* **139**, 16666 (2017).
- [245] K. Röder and D. J. Wales, *J. Am. Chem. Soc.* **140**, 4018 (2018).
- [246] P. G. Bolhuis, D. Chandler, C. Dellago and P. L. Geissler, *Annu. Rev. Phys. Chem.* **53**, 291 (2002).
- [247] S. V. Krivov and M. Karplus, *Proc. Nat. Acad. Sci. USA* **101**, 14766 (2004).
- [248] S. V. Krivov and M. Karplus, *J. Phys. Chem. B* **110**, 12689 (2006).
- [249] S. Muff and A. Caffisch, *Proteins: Struct., Func. and Bioinf.* **70**, 1185 (2008).
- [250] S. V. Krivov and M. Karplus, *Proc. Nat. Acad. Sci. USA* **105**, 13841 (2008).
- [251] F. Noé and S. Fischer, *Curr. Opin. Struc. Biol.* **18**, 154 (2008).
- [252] B. M. Dickson, D. E. Makarov and G. Henkelman, *J. Chem. Phys.* **131**, 074108 (2009).

- [253] D. J. Wales, *Curr. Op. Struct. Biol.* **20**, 3 (2010).
- [254] D. J. Wales, *J. Chem. Phys.* **142**, 130901 (2015).
- [255] J. Skilling, in *AIP Conference Proceedings*, vol. 735, pp. 395–405. AIP (2004).
- [256] D. J. Wales, *Chem. Phys. Lett.* **584**, 1 (2013).
- [257] L. B. Pártay, A. P. Bartók and G. Csányi, *Phys. Rev. E* **89**, 022302 (2014).
- [258] L. B. Pártay, A. P. Bartók and G. Csányi, *J. Phys. Chem. B* **114**, 10502 (2010).
- [259] L. B. Pártay, *Comp. Mater. Sci.* **149**, 153 (2018).
- [260] N. Bernstein, K. M. Salerno, L. B. Pártay and G. Csányi, *Phys Rev.* **96** (2017).
- [261] L. B. Pártay, C. Ortner, A. P. Bartok, C. J. Pickard and G. Csányi, *Phys. Chem. Chem. Phys.* **19**, 19369 (2017).
- [262] B. Szekeres, L. B. Pártay and E. Mátyus, *J. Chem. Theory Comput.* **14**, 4353 (2018).
- [263] S. Martiniani, J. D. Stevenson, D. J. Wales and D. Frenkel, *Phys. Rev. X* **4**, 031034 (2014).
- [264] T. Y. Tsong, R. L. Baldwin, P. McPhie and E. L. Elson, *J. Mol. Biol.* **63**, 453 (1972).
- [265] D. B. Wetlaufer, *Proc. Natl. Acad. Sci. USA* **70**, 697 (1973).
- [266] O. B. Ptitsyn and A. A. Rashin, *Dokl. Akad. Nauk SSSR* **213**, 473 (1973).
- [267] M. I. Kanehisa and T. Y. Tsong, *J. Mol. Biol.* **124**, 177 (1978).
- [268] M. Karplus and D. L. Weaver, *Nature* **260**, 404 (1976).
- [269] M. Karplus and D. L. Weaver, *Biopolymers* **18**, 1421 (1979).
- [270] M. Karplus and D. L. Weaver, *Protein Sci.* **3**, 650 (1994).
- [271] N. Go and H. Abe, *Biopolymers* **20**, 991 (1981).
- [272] P. S. Kim and R. L. Baldwin, *Annu. Rev. Biochem.* **51**, 459 (1982).
- [273] K. A. Dill, *Biochemistry* **24**, 1501 (1985).
- [274] A. R. Fersht, *Curr. Opin. Struct. Biol.* **7**, 3 (1997).
- [275] O. M. Becker and M. Karplus, *J. Chem. Phys.* **106**, 1495 (1998).
- [276] D. L. Stein, *Proc. Natl. Acad. Sci. USA* **82**, 3670 (1985).
- [277] S. Kirkpatrick, C. D. Gelatt and M. P. Vecchi, *Science* **220**, 671 (1983).
- [278] J. D. Bryngelson and P. G. Wolynes, *Proc. Natl. Acad. Sci. USA* **84**, 7524 (1987).
- [279] N. Gō, *Annu. Rev. Biophys. Bioeng.* **12**, 183 (1983).
- [280] J. D. Honeycutt and D. Thirumalai, *Proc. Natl. Acad. Sci. USA* **87**, 3526 (1990).
- [281] J. D. Honeycutt and D. Thirumalai, *Biopolymers* **32**, 695 (1992).

- [282] M. A. Miller and D. J. Wales, *J. Chem. Phys.* **111**, 6610 (1999).
- [283] A. E. Mirsky and L. Pauling, *Proc. Natl. Acad. Sci. USA* **22**, 439 (1936).
- [284] P. E. Leopold, M. Montal and J. N. Onuchic, *Proc. Natl. Acad. Sci. USA* **89**, 8721 (1992).
- [285] J. D. Bryngelson, J. N. Onuchic, N. D. Socci and P. G. Wolynes, *Proteins* **21**, 167 (1995).
- [286] J. N. Onuchic, Z. Luthey-Schulten and P. G. Wolynes, *Annu. Rev. Phys. Chem.* **48**, 545 (1997).
- [287] F. Morcos, N. P. Schafer, R. R. Cheng, J. N. Onuchic and P. G. Wolynes, *Proc. Natl. Acad. Sci. USA* **111**, 12408 (2014).
- [288] P. G. Wolynes, *Biochimie* **119**, 218 (2015).
- [289] E. D. Nelson and J. N. Onuchic, *Proc. Natl. Acad. Sci. USA* **95**, 10682 (1998).
- [290] E. Bornberg-Bauer and H. S. Chan, *Proc. Natl. Acad. Sci. USA* **96**, 10689 (1999).
- [291] J. D. Bryngelson and P. G. Wolynes, *J. Phys. Chem.* **93**, 6902 (1989).
- [292] C. M. Dobson, A. Šali and M. Karplus, *Angew. Chem., Int. Ed.* **37**, 868 (1998).
- [293] J. N. Onuchic and P. G. Wolynes, *Curr. Opin. Struct. Biol.* **14**, 70 (2004).
- [294] Y. Bai, T. R. Sosnick, L. Mayne and S. W. Englander, *Science* **269**, 192 (1995).
- [295] S. W. Englander and L. Mayne, *Proc. Natl. Acad. Sci. USA* **114**, 8253 (2017).
- [296] P. Wolynes, J. Onuchic and D. Thirumalai, *Science* **267**, 1619 (1995).
- [297] P. G. Wolynes, *Philos. T. Roy. Soc. A* **363**, 453 (2005).
- [298] W. A. Eaton and P. G. Wolynes, *Proc. Natl. Acad. Sci. USA* **114**, E9759 (2017).
- [299] A. R. Fersht, *FEBS Lett.* **325**, 5 (1993).
- [300] A. R. Fersht, *Biochem. Soc. Trans.* **22**, 267 (1994).
- [301] A. R. Fersht, *Curr. Opin. Struct. Biol.* **5**, 79 (1995).
- [302] V. Daggett, A. Li, L. S. Itzhaki, D. E. Otzen and A. R. Fersht, *J. Mol. Biol.* **257**, 430 (1996).
- [303] P. G. Wolynes, Z. Luthey-Schulten and J. N. Onuchic, *Chem. Biol.* **3**, 425 (1996).
- [304] A. R. Dinner, A. Šali, L. J. Smith, C. M. Dobson and M. Karplus, *Trends Biochem. Sci.* **25**, 331 (2000).
- [305] C. D. Snow, H. Nguyen, V. S. Pande and M. Gruebele, *Nature* **420**, 102 (2002).
- [306] P. G. Wolynes, W. A. Eaton and A. R. Fersht, *Proc. Natl. Acad. Sci. USA* **109**, 17770 (2012).
- [307] G. Žoldák and M. Rief, *Curr. Opin. Struct. Biol.* **23**, 48 (2013).
- [308] E. J. Guinn, B. Jagannathan and S. Marqusee, *Nat. Commun.* **6**, 6861 (2015).

- [309] B. Schuler, E. A. Lipman and W. A. Eaton, *Nature* **419**, 743 (2002).
- [310] T. R. Cech, in *The RNA World*, edited by R. F. Gesteland and J. F. Atkins, pp. 239–269, New York (1993), Cold Spring Harbor Lab. Press.
- [311] C. M. Dobson, *Philos. Trans. Roy. Soc. London Series B-Biological Sci.* **356**, 133 (2001).
- [312] D. Thirumalai and S. A. Woodson, *Acc. Chem. Res.* **29**, 433 (1996).
- [313] D. Thirumalai, *Proc. Natl. Acad. Sci. USA* **95**, 11506 (1998).
- [314] S. J. Chen and K. A. Dill, *Proc. Natl. Acad. Sci. USA* **97**, 646 (2000).
- [315] J. A. Joseph, K. Röder, D. Chakraborty, R. G. Mantell and D. J. Wales, *Chem. Commun.* **53**, 6974 (2017).
- [316] D. J. Wales, *Ann. Rev. Phys. Chem.* **69**, 401 (2018).
- [317] J. N. Murrell and K. J. Laidler, *Trans. Faraday. Soc.* **64**, 371 (1968).
- [318] F. Rao and A. Caffisch, *J. Mol. Biol.* **342**, 299 (2004).
- [319] D. Prada-Gracia, J. Gómez-Gardenes, P. Echenique and F. Fernando, *PLoS Comput. Biol.* **5**, e1000415 (2009).
- [320] W. Forst, *Theory of Unimolecular Reactions*, Academic Press, New York (1973).
- [321] K. J. Laidler, *Chemical Kinetics*, Harper & Row, New York (1987).
- [322] J. Nocedal, *Math. Comput.* **35**, 773 (1980).
- [323] C. G. Broyden, *J. Inst. Math. Appl.* **6**, 76 (1970).
- [324] R. Fletcher, *Comput. J.* **13**, 317 (1970).
- [325] D. Goldfarb, *Math. Comput.* **24**, 23 (1970).
- [326] D. F. Shanno, *Math. Comput.* **24**, 647 (1970).
- [327] S. A. Trygubenko and D. J. Wales, *J. Chem. Phys.* **120**, 2082 (2004).
- [328] G. Henkelman and H. Jónsson, *J. Chem. Phys.* **113**, 9978 (2000).
- [329] G. Henkelman, B. P. Uberuaga and H. Jónsson, *J. Chem. Phys.* **113**, 9901 (2000).
- [330] L. J. Munro and D. J. Wales, *Phys. Rev. B* **59**, 3969 (1999).
- [331] G. Henkelman and H. Jónsson, *J. Chem. Phys.* **111**, 7010 (1999).
- [332] Y. Kumeda, L. J. Munro and D. J. Wales, *Chem. Phys. Lett.* **341**, 185 (2001).
- [333] Y. Zeng, P. Xiao and G. Henkelman, *J. Chem. Phys.* **140**, 044115 (2014).
- [334] F. H. Stillinger and T. A. Weber, *Science* **225**, 983 (1984).
- [335] D. J. Wales, *Mol. Phys.* **78**, 151 (1993).
- [336] F. H. Stillinger, *Science* **267**, 1935 (1995).

- [337] B. Strodel and D. J. Wales, Chem. Phys. Lett. **466**, 105 (2008).
- [338] V. A. Sharapov, D. Meluzzi and V. A. Mandelshtam, Phys. Rev. Lett. **98**, 105701 (2007).
- [339] J. P. K. Doye and D. J. Wales, J. Chem. Phys. **102**, 9659 (1995).
- [340] J. P. K. Doye and D. J. Wales, J. Chem. Phys. **102**, 9673 (1995).
- [341] F. Calvo, J. P. K. Doye and D. J. Wales, J. Chem. Phys. **115**, 9627 (2001).
- [342] I. Georgescu and V. A. Mandelshtam, J. Chem. Phys. **137**, 144106 (2012).
- [343] J. M. Carr and D. J. Wales, J. Phys. Chem. B **112**, 8760 (2008).
- [344] D. J. Wales and P. Salamon, Proc. Natl. Acad. Sci. USA **111**, 617 (2014).
- [345] K. Sutherland-Cash, D. Wales and D. Chakrabarti, Chem. Phys. Lett. **625**, 1 (2015).
- [346] K. H. Sutherland-Cash, R. G. Mantell and D. J. Wales, Chemical Physics Letters **685**, 288 (2017).
- [347] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry*, North-Holland, Amsterdam (1981).
- [348] R. E. Kunz, *Dynamics of First-Order Phase Transitions*, Deutsch, Thun (1995).
- [349] R. S. Berry and R. Breitengraser-Kunz, Phys. Rev. Lett. **74**, 3951 (1995).
- [350] R. E. Kunz and R. S. Berry, J. Chem. Phys. **103**, 1904 (1995).
- [351] K. D. Ball, R. S. Berry, R. E. Kunz, F.-Y. Li, A. Proykova and D. J. Wales, Science **271**, 963 (1996).
- [352] G. Di Gesù, T. Lelièvre, D. Le Peutrec and B. Nectoux, Faraday Discuss. **195**, 469 (2016).
- [353] D. A. Evans and D. J. Wales, J. Chem. Phys. **118**, 3891 (2003).
- [354] D. A. Evans and D. J. Wales, J. Chem. Phys. **119**, 9947 (2003).
- [355] D. A. Evans and D. J. Wales, J. Chem. Phys. **121**, 1080 (2004).
- [356] A. B. Bortz, M. H. Kalos and J. L. Lebowitz, J. Comput. Phys. **17**, 10 (1975).
- [357] A. F. Voter, Phys. Rev. B **34**, 6819 (1986).
- [358] K. A. Fichthorn and W. H. Weinberg, J. Chem. Phys. **95**, 1090 (1991).
- [359] D. J. Wales, J. Chem. Phys. **130**, 204111 (2009).
- [360] R. S. MacKay and J. D. Robinson, Philos. T. Roy. Soc. A **376** (2018).
- [361] D. J. Wales, Mol. Phys. **100**, 3285 (2002).
- [362] D. J. Wales, Mol. Phys. **102**, 891 (2004).
- [363] D. J. Wales, Curr. Opin. Struc. Biol. **20**, 3 (2010).
- [364] C. Dellago, P. G. Bolhuis and D. Chandler, J. Chem. Phys. **110**, 6617 (1999).

- [365] L. R. Pratt, J. Chem. Phys. **85**, 5045 (1986).
- [366] C. Dellago and P. G. Bolhuis, Adv. Polymer Sci. **221**, 167 (2009).
- [367] T. S. van Erp, D. Moroni and P. G. Bolhuis, J. Chem. Phys. **118**, 7762 (2003).
- [368] D. Moroni, T. S. van Erp and P. G. Bolhuis, Phys. A Stat. Mech. its Appl. **340**, 395 (2004).
- [369] T. S. van Erp and P. G. Bolhuis, J. Comput. Phys. **205**, 157 (2005).
- [370] G. A. Huber and S. Kim, Biophys. J. **70**, 97 (1996).
- [371] B. W. Zhang, D. Jasnow and D. M. Zuckerman, J. Chem. Phys. **132**, 054107 (2010).
- [372] L. T. Chong, A. S. Saglam and D. M. Zuckerman, Curr. Opin. Struct. Biol. **43**, 88 (2017).
- [373] D. Bhatt, B. W. Zhang and D. M. Zuckerman, J. Chem. Phys. **133**, 014110 (2010).
- [374] M. C. Zwier, A. J. Pratt, J. L. Adelman, J. W. Kaus, D. M. Zuckerman and L. T. Chong, J. Phys. Chem. Lett. **7**, 3440 (2016).
- [375] A. Nunes-Alves, D. M. Zuckerman and G. M. Arantes, Biophys. J. **114**, 1058 (2018).
- [376] R. J. Allen, D. Frenkel and P. R. t. Wolde, J. Chem. Phys. **124**, 024102 (2006).
- [377] R. J. Allen, D. Frenkel and P. R. t. Wolde, J. Chem. Phys. **124**, 194111 (2006).
- [378] K. Kratzer, A. Arnold and R. J. Allen, J. Chem. Phys. **138**, 164112 (2013).
- [379] M. J. Morelli, S. Tanase-Nicola, R. J. Allen and P. R. t. Wolde, Biophysical J. **94**, 3413 (2008).
- [380] A. K. Faradjian and R. Elber, J. Chem. Phys. **120**, 10880 (2004).
- [381] A. M. A. West, R. Elber and D. Shalloway, J. Chem. Phys. **126**, 145104 (2007).
- [382] E. Vanden-Eijnden, M. Venturoli, G. Ciccotti and R. Elber, J. Chem. Phys. **129**, 174102 (2008).
- [383] J. M. Bello-Rivas and R. Elber, J. Chem. Phys. **142**, 094102 (2015).
- [384] P. Ma, A. E. Cardenas, M. I. Chaudhari, R. Elber and S. B. Rempe, J. Am. Chem. Soc. **139**, 14837 (2017).
- [385] W. E, W. Ren and E. Vanden-Eijnden, Phys. Rev. B **66**, 052301 (2002).
- [386] B. Peters, A. Heyden, A. T. Bell and A. Chakraborty, J. Chem. Phys. **120**, 7877 (2004).
- [387] V. Ovchinnikov, M. Karplus and E. Vanden-Eijnden, J. Chem. Phys. **134**, 085103 (2011).
- [388] V. Ovchinnikov, M. Cecchini, E. Vanden-Eijnden and M. Karplus, Biophys. J. **101**, 2436 (2011).
- [389] V. Ovchinnikov and M. Karplus, J. Chem. Phys. **140**, 175103 (2014).
- [390] F. Noe, C. Schutte, E. Vanden-Eijnden, L. Reich and T. R. Weigl, Proc. Natl. Acad. Sci. USA **106**, 19011 (2009).

- [391] E. Vanden-Eijnden and M. Venturoli, J. Chem. Phys. **130**, 194101 (2009).
- [392] R. Elber, J. Bello-Rivas, P. Ma, A. Cardenas and A. Fathizadeh, Entropy **19**, 219 (2017).
- [393] R. Zwanzig, J. Stat. Phys. **30**, 255 (1983).
- [394] C. Schütte, A. Fischer, W. Huisinga and P. Deuffhard, J. Comput. Phys. **151**, 146 (1999).
- [395] N. Singhal, C. D. Snow and V. S. Pande, J. Chem. Phys. **121**, 415 (2004).
- [396] W. C. Swope, J. W. Pitera and F. Suits, J. Phys. Chem. B **108**, 6571 (2004).
- [397] J.-H. Prinz, H. Wu, M. Sarich, B. Keller, M. Senne, M. Held, J. D. Chodera, C. Schütte and F. Noé, J. Chem. Phys. **134** (2011).
- [398] R. Zwanzig, A. Szabo and B. Bagchi, Proc. Natl. Acad. Sci. USA **89**, 20 (1992).
- [399] R. Zwanzig, Proc. Natl. Acad. Sci. USA **92**, 9801 (1995).
- [400] V. S. Pande, K. Beauchamp and G. R. Bowman, Methods **52**, 99 (2010).
- [401] T. J. Lane, D. Shukla, K. A. Beauchamp and V. S. Pande, Curr. Opin. Struct. Biol. **23**, 58 (2013).
- [402] J. D. Chodera and F. Noé, Curr. Opin. Struct. Biol. **25**, 135 (2014).
- [403] B. E. Husic and V. S. Pande, J. Am. Chem. Soc. **140**, 2386 (2018).
- [404] G. R. Bowman and V. S. Pande, Proc. Natl. Acad. Sci. USA **107**, 10890 (2010).
- [405] J. P. K. Doye, Phys. Rev. Lett. **88**, 238701 (2002).
- [406] J. W. R. Morgan, D. Mehta and D. J. Wales, Phys. Chem. Chem. Phys. **19**, 25498 (2017).
- [407] D. J. Watts and S. H. Strogatz, Nature **393**, 440 (1998).
- [408] A. Cavalli, U. Haberthür, E. Paci and A. Caflisch, Prot. Sci. **12**, 1801 (2003).
- [409] D. Gfeller, D. M. de Lachapelle, P. De Los Rios, G. Caldarelli and F. Rao, Phys. Rev. E **76**, 026113 (2007).
- [410] W. E and E. Vanden-Eijnden, J. Stat. Phys. **123**, 503 (2006).
- [411] P. Metzner, C. Schütte and E. Vanden-Eijnden, SIAM Multiscale Model. Simul. **7**, 1192 (2009).
- [412] A. Berezhkovskii, G. Hummer and A. Szabo, J. Chem. Phys. **130**, 205102 (2009).
- [413] W. E and E. Vanden-Eijnden, Annu. Rev. Phys. Chem. **61**, 391 (2010).
- [414] G. R. Bowman, V. S. Pande and F. Noé, *Yersinia pestis: Retrospective and perspective* (2014).
- [415] H. Wu, J.-H. Prinz and F. Noé, The Journal of Chemical Physics **143** (2015).
- [416] W. Wang, S. Cao, L. Zhu and X. Huang, WIREs Comput. Mol. Sci. **8**, e1343 (2018).
- [417] M. Cameron and E. Vanden-Eijnden, J. Stat. Phys. **156**, 427 (2014).

- [418] F. H. Stillinger and T. A. Weber, Phys. Rev. A **25**, 978 (1982).
- [419] D. J. Wales and J. P. K. Doye, J. Chem. Phys. **119**, 12409 (2003).
- [420] J. P. K. Doye and D. J. Wales, J. Chem. Phys. **116**, 3777 (2002).
- [421] P. Sibani, J. C. Schön, P. Salamon and J. Andersson, Europhys. Lett. **22**, 479 (1993).
- [422] J. C. Schön, Ber. Bunsen-Ges. Phys. Chem. Chem. Phys. **100**, 1388 (1996).
- [423] J. C. Schön, H. Putz and M. Jansen, J. Phys.: Condens. Matter **8**, 143 (1996).
- [424] J. C. Schön, J. Phys. Chem. A **106**, 10886 (2002).
- [425] T. Komatsuzaki, K. Hoshino, Y. Matsunaga, G. J. Rylance, R. L. Johnston and D. J. Wales, J. Chem. Phys. **122**, 084714 (2005).
- [426] S. V. Krivov and M. Karplus, J. Chem. Phys. **117**, 10894 (2002).
- [427] H. Eyring and M. Polanyi, Z. Phys. Chem. Abt. B **12**, 279 (1931).
- [428] M. G. Evans and M. Polanyi, Trans. Faraday Soc. **31**, 875 (1935).
- [429] H. Eyring, J. Chem. Phys. **3**, 107 (1935).
- [430] J. E. Jones and A. E. Ingham, Proc. R. Soc. A **107**, 636 (1925).
- [431] D. J. Wales, Phys. Rev. E **95**, 030105(R) (2017).
- [432] A. L. Mackay, Acta Cryst. **15**, 916 (1962).
- [433] J. N. Onuchic, P. G. Wolynes, Z. Luthey-Schulten and N. D. Socci, Proc. Natl. Acad. Sci. USA **92**, 3626 (1995).
- [434] J. N. Onuchic, H. Nymeyer, A. E. García, J. Chahine and N. D. Socci, Adv. Prot. Chem. **53**, 87 (2000).
- [435] D. J. Wales, Chem. Phys. Lett. **285**, 330 (1998).
- [436] D. J. Wales, Chem. Phys. Lett. **294**, 262 (1998).
- [437] D. J. Wales and J. P. K. Doye, J. Phys. Chem. A **101**, 5111 (1997).
- [438] J. P. K. Doye, D. J. Wales and M. A. Miller, J. Chem. Phys. **109**, 8143 (1998).
- [439] F. Calvo, J. P. Neirotti, D. L. Freeman and J. D. Doll, J. Chem. Phys. **112**, 10350 (2000).
- [440] J. P. Neirotti, F. Calvo, D. L. Freeman and J. D. Doll, J. Chem. Phys. **112**, 10340 (2000).
- [441] P. A. Frantsuzov and V. A. Mandelshtam, Phys. Rev. E **72**, 037102 (2005).
- [442] C. Predescu, P. A. Frantsuzov and V. A. Mandelshtam, J. Chem. Phys. **122**, 154305 (2005).
- [443] H. B. Liu and K. D. Jordan, J. Phys. Chem. B **109**, 5203 (2005).
- [444] V. A. Sharapov and V. A. Mandelshtam, J. Phys. Chem. A **111**, 10284 (2007).
- [445] D. J. Wales and H. A. Scheraga, Science **285**, 1368 (1999).

- [446] M. T. Oakley, R. L. Johnston and D. J. Wales, *Phys. Chem. Chem. Phys.* **15**, 3965 (2013).
- [447] M. Dittner and B. Hartke, *Computational and Theoretical Chemistry* **1107**, 7 (2017).
- [448] F. Calvo, *Phys. Rev. E* **82**, 046703 (2010).
- [449] M. Picciani, M. Athenes, J. Kurchan and J. Tailleur, *J. Chem. Phys.* **135**, 034108 (2011).
- [450] M. Cameron and E. Vanden-Eijnden, *Journal of Statistical Physics* **156**, 427 (2014).
- [451] M. J. A. Bailey, V. Koronakis, T. Schmoll and C. Hughes, *Mol. Microbiol.* **6**, 1003 (1992).
- [452] M. J. A. Bailey, C. Hughes and V. Koronakis, *Mol. Microbiol.* **26**, 845 (1997).
- [453] G. A. Belogurov, M. N. Vassilyeva, V. Svetlov, S. Klyuyev, N. V. Grishin, D. G. Vassilyev and I. Artsimovitch, *Mol. Cell* **26**, 117 (2007).
- [454] B. M. Burmann, S. H. Knauer, A. Sevostyanova, K. Schweimer, R. A. Mooney, R. Landick, I. Artsimovitch and P. Rösch, *Cell* **150**, 291 (2012).
- [455] M. J. A. Bailey, C. Hughes and V. Koronakis, *Mol. Gen. Genet.* **262**, 1052 (2000).
- [456] G. A. Belogurov, R. A. Mooney, V. Svetlov, R. Landick and I. Artsimovitch, *EMBO J.* **28**, 112 (2009).
- [457] J. A. Joseph, D. Chakraborty and D. J. Wales, *J. Chem. Theor. Comput.* p. in press (2018).
- [458] J. Chen and R. W. Kriwacki, *J. Mol. Biol.* **430**, 2275 (2018).
- [459] G. T. Heller, M. Bonomi and M. Vendruscolo, *J. Mol. Biol.* **430**, 2288 (2018).
- [460] L. Dahal, T. O. Kwan, J. J. Hollins and J. Clarke, *J. Mol. Biol.* **430**, 2468 (2018).
- [461] S. Zhang, K. Iwata, M. Lachenmann, J. Peng, S. Li, E. Stimson, Y.-A. Lu, A. Felix, J. Maggio and J. Lee, *J. Struct. Biol.* **130**, 130 (2000).
- [462] J. Nasica-Labouze, P. H. Nguyen, F. Sterpone, O. Berthoumieu, N.-V. Buchete, S. Côté, A. De Simone, A. J. Doig, P. Faller, A. Garcia, A. Laio, M. S. Li, S. Melchionna, N. Mousseau, Y. Mu, A. Paravastu, S. Pasquali, D. J. Rosenman, B. Strodel, B. Tarus, J. H. Viles, T. Zhang, C. Wang and P. Derreumaux, *Chem. Rev.* **115**, 3518 (2015).
- [463] L. Nagel-Steger, M. C. Owen and B. Strodel, *ChemBioChem* **17**, 657 (2016).
- [464] T. Cragolini, D. Chakraborty, J. Sponer, P. Derreumaux, S. Pasquali and D. J. Wales, *J. Chem. Phys.* **147**, 152715 (2017).
- [465] J. Šponer, G. Bussi, P. Stadlbauer, P. Kührová, P. Banáš, B. Islam, S. Haider, S. Neidle and M. Otyepka, *Biochim. Biophys. Acta* **1861**, 1246 (2017).
- [466] V. K. de Souza and D. J. Wales, *J. Chem. Phys.* **129**, 164507 (2008).
- [467] V. K. de Souza and D. J. Wales, *J. Chem. Phys.* **130**, 194508 (2009).
- [468] S. P. Niblett, V. K. de Souza, J. D. Stevenson and D. J. Wales, *J. Chem. Phys.* **145**, 024505 (2016).

- [469] S. P. Niblett, M. Biedermann, D. J. Wales and V. K. de Souza, J. Chem. Phys. **147**, 152726 (2017).
- [470] V. Lubchenko and P. G. Wolynes, Ann. Rev. Phys. Chem. **58**, 235 (2007).
- [471] T. R. Kirkpatrick and D. Thirumalai, Rev. Mod. Phys. **87**, 183 (2015).
- [472] G. Tian, R. A. Broglia, H. E. Roman, E. Vigezzi and E. Shakhnovich, J. Chem. Phys. **108**, 757 (1998).
- [473] B. V. Reddy, S. Datta and S. Tiwari, Protein Eng. **11**, 1137 (1998).
- [474] B. Vekhter and R. S. Berry, J. Chem. Phys. **111**, 3753 (1999).
- [475] D. M. Taverna and R. A. Goldstein, J. Mol. Biol. **315**, 479 (2002).
- [476] D. M. Taverna and R. A. Goldstein, Proteins **46**, 105 (2002).
- [477] K. T. Sapra, G. P. Balasubramanian, D. Labudde, J. U. Bowie and D. J. Muller, J. Mol. Biol. **376**, 1076 (2008).
- [478] M. Held, P. Metzner, J. H. Prinz and F. Noé, Biophys. J. **100**, 701 (2011).
- [479] P. V. Banushkina and S. V. Krivov, J. Chem. Theory Comput. **9**, 5257 (2013).
- [480] W. Huang, K. M. Ravikumar and S. Yang, J. Chem. Theory Comput. **10**, 2897 (2014).
- [481] K. Röder, *Energy landscaping - on the relationship between functionality and sequence mutations for multifunctional biomolecules*, Ph.D. thesis, University of Cambridge (2018).
- [482] K. Röder and D. J. Wales, J. Chem. Theory Comput. **13**, 1468 (2017).
- [483] K. Röder and D. J. Wales, Biochemistry p. in press (2018).
- [484] F. H. C. Crick, Acta Crystallogr. **6**, 689 (1953).
- [485] T. Alber, Curr. Biol. **3**, 182 (1993).
- [486] P. B. Harbury, T. Zhang, P. S. Kim and T. Alber, Science **262**, 1401 (1993).
- [487] L. Gonzalez, J. J. Plecs and T. Alber, Nat. Struct. Biol. **3**, 510 (1996).
- [488] L. Gonzalez, R. A. Brown, D. Richardson and T. Alber, Nat. Struct. Biol. **3**, 1002 (1996).
- [489] G. Grigoryan and A. Keating, Curr. Opin. Struc. Biol. **18**, 477 (2008).
- [490] Y. Gambin, A. Schug, E. A. Lemke, J. J. Lavinder, A. C. M. Ferreón, T. J. Magliery, J. N. Onuchic and A. A. Deniz, Proc. Natl. Acad. Sci. USA **106**, 10153 (2009).
- [491] Y. Levy, S. S. Cho, T. Shen, J. N. Onuchic and P. G. Wolynes, Proc. Natl Acad. Sci. USA **102**, 2373 (2004).
- [492] M. K. Yadav, L. J. Leman, D. J. Price, C. L. Brooks, C. D. Stout and M. R. Ghadiri, Biochemistry **45**, 4463 (2006).
- [493] T. Hunter, Mol. Cell **28**, 730 (2007).
- [494] D. Komander and M. Rape, Annu. Rev. Biochem. **81**, 203 (2012).

- [495] J. S. Harrison, T. M. Jacobs, K. Houlihan, K. Van Doorslaer and B. Kuhlman, *J.Mol. Biol.* **428**, 679 (2016).
- [496] O. F. Lange, N.-A. Lakomek, C. Fares, G. F. Schröder, K. F. A. Walter, S. Becker, J. Meiler, H. Grubmüller, C. Griesinger and B. L. de Groot, *Science* **320**, 1471 (2008).
- [497] M. K. Hospenthal, S. M. V. Freund and D. Komander, *Nat. Struct. Mol. Biol.* **20**, 555 (2013).
- [498] A. H. Phillips and J. E. Corn, *J. Biol. Chem.* **290**, 26437 (2015).
- [499] P. O. Craig, J. Lätzer, P. Weinkam, R. M. B. Hoffman, D. U. Ferreira, E. A. Komives and P. G. Wolynes, *J. Am. Chem. Soc.* **133**, 17463 (2011).
- [500] T. Wauer, K. N. Swatek, J. L. Wagstaff, C. Gladkova, J. N. Pruneda, M. A. Michel, M. Gersch, C. M. Johnson, S. M. Freund and D. Komander, *EMBO J.* **34**, 307 (2015).
- [501] X. Dong, Z. Gong, Y.-B. Lu, K. Liu, L.-Y. Qin, M.-L. Ran, C.-L. Zhang, Z. Liu, W.-P. Zhang and C. Tang, *Proc. Natl. Acad. Sci. USA* **114**, 6770 (2017).
- [502] C. Gladkova, A. F. Schubert, J. L. Wagstaff, J. N. Pruneda, S. M. V. Freund and D. Komander, *EMBO J.* p. e201797876 (2017).
- [503] A. Pickrell and R. Youle, *Neuron* **85**, 257 (2015).
- [504] T. N. Nguyen, B. S. Padman and M. Lazarou, *Trends Cell Biol.* **26**, 733 (2016).
- [505] G. S. Hammond, *J. Am. Chem. Soc.* **77**, 334 (1955).
- [506] R. P. Bell, *Proc. Roy. Soc. London A* **154**, 414 (1936).
- [507] D. J. Wales, *Science* **293**, 2067 (2001).
- [508] K. Röder and D. J. Wales, *J.Phys. Chem. Lett.* **9**, 6169 (2018).