

UC Berkeley

UC Berkeley Previously Published Works

Title

Evolution of the Potential Energy Landscape with Static Pulling Force for Two Model Proteins

Permalink

<https://escholarship.org/uc/item/67p277z3>

Journal

The Journal of Physical Chemistry B, 116(29)

ISSN

1520-6106

Authors

Wales, David J
Head-Gordon, Teresa

Publication Date

2012-07-26

DOI

10.1021/jp211806z

Peer reviewed

Evolution of the Potential Energy Landscape with Static Pulling Force for Two Model Proteins

David J. Wales^{1*} and Teresa Head-Gordon^{2†}

¹ University Chemical Laboratories, Lensfield Road, Cambridge CB2 1EW, UK

² Department of Bioengineering, 274C Stanley Hall,

University of California, Berkeley Berkeley, CA 94720-1762, USA

March 3, 2012

*dw34@cam.ac.uk

†tlhead-gordon@lbl.gov

Abstract

The energy landscape is analysed for off-lattice bead models of protein L and protein G as a function of a static pulling force. Two different pairs of attachment points (pulling directions) are compared in each case, namely residues 1/56 and 10/32. For the terminal residue pulling direction 1/56 the distinct global minimum structures are all extended, aside from the compact geometry that correlates with zero force. The helical turns finally disappear at the highest pulling forces considered. For the 10/32 pulling direction the changes are more complicated, with a variety of competing arrangements for beads outside the region where the force is directly applied. These alternatives produce frustrated energy landscapes, with low-lying minima separated by high barriers. The calculated folding pathways in the absence of force are in good agreement with previous work. The N-terminal hairpin folds first for protein L and the C-terminal hairpin for protein G, which exhibits an intermediate. However, for a relatively low static force, where the global minimum retains its structure, the folding mechanisms change, sometimes dramatically, depending on the protein and the attachment points. The scaling relations predicted by catastrophe theory are found to hold in the limit of short path lengths.

1 Introduction

Minimalist models of proteins based on off-lattice bead models provide important physical insight into protein self-organisation.^{1,2} Formulations based on three different sorts of bead, namely hydrophobic (B), hydrophilic (L), and neutral (N), have been used in a wide variety of studies³⁻¹⁷ since they were first introduced.¹⁸⁻²¹ Some of these studies have considered the effect of a pulling force²²⁻²⁶ to model experiments that employ atomic force microscopy²⁷⁻³¹ (AFM) or optical^{32,33} and magnetic^{34,35} tweezers. These single molecule experiments can probe mechanical force as a denaturant to compare with studies that characterise folding/unfolding under chemical or thermal denaturing conditions.

In the present contribution we consider the potential energy landscapes of coarse-grained models of two small proteins, protein L and protein G, as a function of pulling force. These two IgG-binding proteins adopt the same topological fold with a central α -helix packed against a mixed four-strand β -sheet containing two hairpins,³⁶ although they have low sequence identity (16%). They suggest an interesting application for energy landscape analysis because experiments reveal differences in the folding pathways with no force applied.³⁷⁻⁴⁴ In particular, while protein L appears to fold via two-state kinetics, protein G may exhibit an early intermediate. For protein L, hairpin 1 (nearest the N-terminus) probably forms first,⁴⁵ while for protein G the early intermediate is followed by a rate-limiting step involving formation of hairpin 2.⁴⁶ In both cases the mechanism has been described as a nucleation-condensation or nucleation-collapse process,^{20,47,48} involving helix-assisted hairpin formation.⁴⁹ Previous simulations of protein L and G employing coarse-grained models have investigated the differences in folding mechanism, and suggest that protein G may fold through multiple pathways.⁴⁹⁻⁵¹ Furthermore, although neither protein performs a mechanical function in vivo, they both exhibit significant mechanical stability^{27,52-57} with associated unfolding forces of order 100 pN in AFM experiments involving pulling speeds of order 100 nm/s.

In the present work we investigate protein L and G using the BLN potential of Brown and Head-Gordon^{13,49} and consider the effect of a static pulling force on the energy landscape. We employ techniques based on the construction of a kinetic transition network from geometry optimisation,

avoiding the need for low-dimensional projections⁵⁸ or a reaction coordinate.^{49,50,59} Our results therefore provide insight into the potential energy landscape that underlies experiments conducted on proteins under tension. For simplicity we focus on the favoured structures and pathways for the quasistatic case at a series of fixed forces in this initial analysis. Experiments conducted at constant force under nonequilibrium conditions sample the same landscape, but the results will depend on how the initial state is prepared. Understanding the organisation of the energy landscape for fixed force can provide insight into both equilibrium and non-equilibrium experiments given appropriate initial conditions in the latter case.

The calculated folding pathways in the absence of force are in good agreement with previous theoretical studies. However, for a relatively low static force where the global minimum retains its structure, the folding mechanisms change so that the C-terminal hairpin folds first for protein L while the folding intermediate for protein G disappears altogether. For the 10/32 pulling direction the early part of the folding pathway is more dependent on the particular extended state and therefore distance to the transition state, especially for protein G.

2 Methods

2.1 Potentials

The potentials employed in the present work employ the force field and parameterisation described in previous work,^{13,49,60} which involves bond angle, dihedral angle, and non-bonded terms. Here these contributions were augmented by stiff springs to represent the covalent bonds.⁴ The additional bond length term is

$$\frac{1}{2}K_r \sum_{i=1}^{N-1} (R_{i,i+1} - R_e)^2, \quad (1)$$

where R_{ij} is the distance between beads i and j , $K_r = 231.2 \epsilon R_e^{-2}$ to maintain consistency with previous work.^{4,7,12,15} Since this force constant simply constrains the separation of consecutive beads it has no significant effect on the results within a wide range of values. ϵ and R_e are parameters of the potential,^{13,49} which provide natural reduced units of energy and distance, respectively. All the

results that follow will be reported in reduced units. The sequence dependence of the parameters was also taken from previous work,^{13,49} where it was found that appropriate changes in folding pathway could be modelled by changing only three beads. These models have also been used to investigate folding kinetics via distributed computing⁶¹ and protein aggregation.⁶² The coarse-graining provides a representation that is computationally tractable, but includes potentially favourable non-native interactions, which would be absent for a G \bar{o} model.

The sequence-dependent BLN potential was coded in our **GMIN**⁶³ and **OPTIM**⁶⁴ programs along with a general static pulling potential, V_{pull} , defined as

$$V_{\text{pull}} = -f(z_{\alpha} - z_{\beta}), \quad (2)$$

where z_{α} and z_{β} are the z coordinates of beads α and β , and f is the force. The additional gradient terms are then

$$\frac{\partial V}{\partial z_{\alpha}} = -f \quad \text{and} \quad \frac{\partial V}{\partial z_{\beta}} = f. \quad (3)$$

The second derivative matrix (Hessian) is unchanged by addition of the pulling potential, but for non-zero force the stationary points of the potential are different.

2.2 Exploring the Landscape

For both protein L and protein G we have examined the potential energy landscape as a function of the static force, f , for attachment points (pulling directions⁶⁵) at beads 1 and 56 and 10 and 32. We will identify these systems concisely as L/1/56/ f , L/10/32/ f , G/1/56/ f and G/10/32/ f for any given force, f . In each case we first identified the global minimum as a function of f using basin-hopping global optimisation,⁶⁶⁻⁶⁸ which has previously been employed for related BLN models.^{15,69} The range of f values considered was chosen to extend up to the point where no further change in the structure of the global minimum occurred. The number of different f values considered was 25 and 45 for the 1/56 and 10/32 attachment points, respectively. More values were needed for 10/32 because more distinct global minima were identified as a function of f for these systems, and larger values of f were investigated. In each case we started with two initial basin-hopping runs of 10^6 steps

each, seeded from different random geometries. Additional runs of up to 5×10^6 steps were performed with alternative parameter choices until most runs produced the same lowest-energy candidate. We did not attempt to optimise the parameters for these runs systematically, since this is only possible once the global minima are known. A fixed temperature of $k_B T / \epsilon = 0.75$ was employed throughout in the accept/reject tests, and the steps were proposed by perturbing all the Cartesian coordinates by random changes in the interval $[-\Delta, \Delta]$. The maximum step size, Δ , was increased or decreased by 10% every 50 basin-hopping steps, to maintain a local acceptance ratio of 0.5. In most runs Δ fluctuated in the range 0.4 to 0.5 in units of R_e .

Depending on the complexity of the landscape, some global minima were identified rapidly, while others took longer to locate. To improve the efficiency in the more difficult cases we employed two of the GMIN options that accelerate searches on multi-funnel^{70,71} landscapes. The first of these keywords, *NEWRESTART* x , reseeds the run if the energy of the lowest minimum has not improved in the last x steps. The configuration of the lowest minimum is added to a cyclic taboo list,⁷² which can be used to avoid revisiting areas of configuration space that have already been searched. Here the *AVOID* keyword was used with a distance threshold of 1.0 reduced units. Runs were then reseeded if the minimum Euclidean distance of the current configuration came within a distance of 1.0 to any members of the saved list. These reseeding procedures seemed to work effectively with a reseeding criterion of $x = 1000$ steps and a taboo list of ten configurations.

We scanned f in increments of 0.1 over the range where the global minimum structure was observed to change most rapidly. To establish the range for each structure we also relaxed the lowest minimum obtained for each value of f progressively at intervals of 0.1 down to zero force and up to the highest value considered. This relaxation is very fast compared to the basin-hopping runs themselves, and serves as a check on the global optimisation results as well as identifying common structures. In total, we identified 6, 10, 6, and 8 distinct global minima for L/1/56, L/10/32, G/1/56, and G/10/32, respectively. These values should be considered lower bounds, since it is possible that a change of structure occurring within an interval of 0.1 in f could have been missed. However, within the f range where the structures change rapidly the global minima are usually similar, so

missing one particular global minimum here would not affect our general conclusions.

To explore the potential energy landscape and the connectivity as a function of the static force and the two attachment patterns we constructed kinetic transition networks^{73–77} using geometry optimisation techniques. These networks consist of local minima and transition states on the potential energy surface, where we employ the geometrical definition of a transition state as a stationary point with precisely one negative eigenvalue.⁷⁸ Each transition state links two local minima via steepest-descent paths defined by the Hessian eigenvector corresponding to the unique negative Hessian eigenvalue, thus establishing the connectivity of the network. In the present work we employed energy minimisation to define approximations to the steepest-descent paths using a slightly modified version of the limited-memory Broyden–Fletcher–Goldfarb–Shanno (LBFGS) algorithm.^{79,80} Transition states were calculated using initial double-ended searches between pairs of local minima with the (doubly-nudged⁸¹ elastic band^{82–87}) approach. The band was optimised until distinct local maxima were obtained, which were then taken as candidates for refinement by hybrid eigenvector-following (EF).^{84,88,89} Complete paths between distant minima may involve many transition states, and pairs of minima were chosen to identify such paths using our missing connection algorithm.⁹⁰ The convergence condition for all stationary points was a root-mean-square gradient of less than $10^{-6} \epsilon/R_e$, and all geometry optimisations were performed with the OPTIM code.⁶⁴ To avoid chain crossings in the pathway calculations we used a new initial interpolation procedure, which will be described elsewhere.

Systems with non-vanishing values of f have only four zero Hessian eigenvalues at a stationary point, instead of the usual six for a molecule in field-free space. Since the force is applied along the z direction rotations around the x and y axes no longer conserve the total energy. Hybrid EF calculations involving the full Hessian therefore employed eigenvalue shifting only for the modes corresponding to translation and to rotation about the z axis. In fact, the gradient-only formulation of hybrid EF was found to be more efficient for these BLN systems, and this was the method used in all the production runs. In this approach the eigenvectors corresponding to zero Hessian eigenvalues are simply projected out of the search direction using the known analytical components.

Expanding a kinetic transition network through successive pathway calculations obtained via geometry optimisation corresponds to the discrete path sampling (DPS) approach.^{74,91,92} Similar networks can also be obtained using molecular dynamics if the states of interest interconvert sufficiently fast.^{93–95} Various strategies to refine DPS databases have been employed in previous work, depending on whether the objective is to improve the current fastest path (*SHORTCUT*^{96,97}) or remove artificial frustration (*FREPAIRS*⁹⁸). In the present work we seeded each database from paths between each pair of distinct global minimum structures for each system, as calculated by *OPTIM*. These databases were then refined using the *PATHSAMPLE* program⁹⁹ to choose new pairs of minima for connection attempts and run the corresponding *OPTIM* jobs. Refinement was continued until the low-lying portion of the landscape converged, as judged from the corresponding disconnectivity graphs.^{100,101} The last stages of refinement employed the *UNTRAP* keyword. Here we modified our previous scheme,⁹⁷ which considered only potential energy barriers, and selected pairs of minima according to the ratio of the barrier between them to the potential energy difference. This scheme helps to remove artificial frustration from a potential energy disconnectivity graph in much the same way as the *FREPAIRS* strategy⁹⁸ for free energy disconnectivity graphs.^{12,102}

Details of the DPS databases constructed in the present work are given in Tables 1-4. These tables include the number of minima and transition states in each database, and provide the energies of the distinct global minima at each selected f value for reference. The corresponding disconnectivity graphs are discussed in §3.2.

3 Results

3.1 Evolution of the Global Minimum with Static Force

Figure 1 summarises the global optimisation results. The values of f where systematic global optimisation was performed are indicated by squares, with the interpolated energies of the distinct global minimum structures represented as lines. The latter plots are almost linear functions of f , aside from panel (d) for L/10/32 in which two of the lines exhibit steps, indicating a stability limit for the cor-

responding structures, as discussed in §4. Changes in structure for the global minimum occur where the lines corresponding to different structures cross. A similar situation occurs when we consider the global minima of clusters or bulk systems bound by the Morse potential as a function of the range of the pair interaction.^{103–105} It is noteworthy that most of the global minima for these BLN models of protein L and protein G are also local minima over the full range of f values considered.

For each system we will label the distinct global minima as GM1, GM2, etc., as they appear for increasing force. Since GM1 corresponds to zero force it is the same for L/1/56 and L/10/32 and for G/1/56 and G/10/32. The other structures are all different, and each one is illustrated in Figure 2. The minima obtained by progressive relaxation of a given GMn geometry at different values of f are all associated with the original GMn structure. For both proteins the evolution of the global minimum with f is similar for the 1/56 attachment. As shown in Figures 2a and 2b, there is only one compact global minimum morphology, while the other structures are all extended. In each case four distinct global minima were identified that retain helical turns, before all secondary structure is lost in the limit of high force. The threshold at which the global minimum switches to an extended conformation is $f = 1.2$ for protein L and $f = 1.3$ for protein G. This difference is consistent with somewhat greater mechanical stability for protein G, which exhibits an average unfolding force of around 180 pN experimentally for a pulling velocity of 400 nm/s,⁵³ while the value for protein L is around 150 pN for a faster pulling velocity of 700 nm/s.⁵² The gradient of the potential energy of GM1 with respect to f , $\partial V_{\text{GM1}}/\partial f$, is -7.23 , -7.29 , -3.69 and -4.56 for L/1/56, G/1/56, L/10/32, and G/10/32, respectively. The stored potential energy is therefore greater for the 1/56 attachment pattern, and the gradients are higher in magnitude for protein G. These results are consistent with the mechanical stabilities observed experimentally, and with the identification of 1/56 and 10/32 as ‘strong’ and ‘weak’ coordinates.¹⁰⁶ Since the terminal β strands are parallel, the 1/56 attachment pattern corresponds to a longitudinal shearing force.

To provide a more quantitative connection with previous work we consider the G \bar{o} model employed by Graham and Best.¹⁰⁶ The mean mass per residue in this study is 109.4 amu and the mean well depth for pair contacts is -0.88 kcal/mol, corresponding to a distance of 3.8Å. For these parameters

the reduced units of force employed in the present work are $\epsilon/R_e \sim 1.6 \times 10^{-11}$ N or 16 pN. Graham and Best report a switch between unfolding mechanisms around 100 pN and 50 pN for the 1/56 and 10/48 attachment points in protein G, which correspond to $f \sim 6$ and $f \sim 3$ in reduced units for the present model. Precise agreement is not expected, but the correct order of magnitude is reassuring.

For 10/32 attachment the evolution of the global minimum is more complicated, and there are some interesting differences between the two models. For low forces both systems exhibit a change in structure, but while protein L retains an α helix packed against two hairpins, protein G adopts a β sheet structure with a separate α turn (GM2 in Figure 2d). The 10/32 attachment means that the applied force is stretching the helix and the β_2 strand, which are contiguous. As the static force increases the global minima for both protein G and protein L correspond to structures with the helix and β_2 strand extended, the β_3 strand packed against the helix, and the β_1 and β_4 strands forming a β sheet with β_2 (Figure 2c and 2d). The threshold at which the global minimum changes to these partly extended structures is $f = 1.4$ for protein L and $f = 2.3$ for protein G, again consistent with somewhat greater mechanical stability for protein G. The formation of the α turn structure (GM2) for G/10/32 probably extends the stability range significantly in this case. For still larger forces the α helix turns are eventually lost, and the beads corresponding to residues 10 to 32 are extended in a zig-zag pattern, which is determined by the bond length and bond angle restraints in the potential. In contrast, the beads corresponding to the β_1 strand and the second hairpin are not subject to the pulling force directly, and can adopt a variety of different arrangements with comparable potential energies. This additional complexity for the 10/32 attachment is clearly visible in the energy landscape and in the folding/unfolding pathways discussed below. Such variations in mechanical resistance with the pulling direction are in line with previous studies.^{57,60,65,106–108} The predicted structural diversity might perhaps be verified in future experiments using additional structural probes, such as Förster resonance energy transfer (FRET). Modification of side chains through mutations might perhaps stabilise different subpopulations of structures that are revealed in the disconnectivity graphs discussed in the following section.

3.2 Visualising the Potential Energy Landscape

Disconnectivity graphs^{100,101} for selected DPS databases are illustrated in Figures 4-9. Since the main purpose of these graphs is to visualise the changes in frustration as a function of the protein, the attachment points, and the applied force, we have chosen a common energy range of 14ϵ for each graph. We also include only the lowest 1,000 local minima, along with any of the global minimum structures that appear in the chosen energy range. Selected structures are superimposed on the graphs using an alternative colouring scheme from Figure 2 to highlight the secondary structure rather than the bead sequence.

The disconnectivity graph for protein L in the absence of an applied force is coloured according to the classification scheme for ‘native’ and ‘unfolded’ structures described by Marianayagam et al.⁶¹ This scheme employs two order parameters for the classification, namely the radius of gyration, R_g , and a structural similarity parameter, χ , defined with respect to the global minimum. The separation into distinct sets is quite good, but a few local minima find themselves as minority components of the different regions. For example, minimum 33 in order of increasing energy is classified as ‘unfolded’, but its structure appears very similar to GM1. This problem may arise because the structural criteria in question were developed for instantaneous geometries, and different cutoffs might be required for the local minima considered here. Applying this classification scheme to calculate folding and unfolding rates leads to artificially high rate constants, because there are ‘native’ and ‘unfolded’ minima connected by low barriers. Hence we prefer to use a recursive regrouping scheme based on a threshold for the free energy barrier between states.^{96,98,109}

The remaining disconnectivity graphs are coloured according to the distance between the attachment points. The hue ranges from red to blue, scaled according to the smallest and largest separations present in the lowest 1000 minima. Table 5 collects some further structural information corresponding to the disconnectivity graphs illustrated. Here we report the mean, μ , and standard deviation, σ , of the distance between the attachment points, D , and of the minimised distance¹¹⁰ to the global minimum, G . These averages include a weighting corresponding to the estimated equilib-

rium occupation probability for each minimum at the folding temperatures in the absence of force. The approximate weights were calculated using harmonic vibrational densities of states obtained from normal mode analysis. Although much of the interesting structure is averaged out in these statistics there are still some clear trends. As expected, $\mu(D)$ increases systematically with f for both proteins and both attachment points. $\mu(G)$ ranges between about 2 and 4 reduced units for the 1/56 pulling direction, but exhibits much larger values for 10/32. This result reflects the greater structural diversity and frustration observed for the 10/32 attachment, which are discussed below.

The disconnectivity graphs for L/1/56 with non-zero force are all relatively unfrustrated. The extended minima that correlate with GM2, GM3, GM4 and GM5, where the helix turns are still present, can generally interconvert via relatively low barriers (Figure 5a). The fully extended minimum, where the helix has also been disrupted, is separated from these structures by a higher overall barrier for intermediate forces, and is not very different in energy from the extended minima with helical turns in the graph for $f = 5.0$ (Figure 5b). The reversal in energetics between fully extended and structures that still retain turns is clear from the colouring scheme. For $f = 5.0$ most of the branches in the graph are blue, corresponding to the longest end-to-end distances observed, while for $f = 2.5$ most branches are green, corresponding to distances below the maximum. In both cases the end-to-end distances in the lowest 1000 minima span a relatively narrow range. In the graphs for $f = 3.0$ and $f = 3.5$ GM3 and GM4 exchange places as the global minimum (Supplementary Information). The corresponding graphs for G/1/56 in Figures 6 and 7 are similar to those for L/1/56. The organisation of the low-lying minima is therefore not very different, and the reordering of the energetics for the fully and mostly extended structures from $f = 2.0$ to $f = 5.0$ is again clear from the colouring. Nevertheless, the folding and unfolding mechanisms differ in terms of the order of events, and protein G exhibits an intermediate in the absence of force, as discussed in §3.3.

Disconnectivity graphs for L/10/32 at low and high static forces are shown in Figures 8a and 8b. Both graphs exhibit significant frustration, with relatively low-lying minima separated by high barriers. At $f = 1.5$ structure GM3 lies lowest with residues 10 to 32 extended and distinct helix/ β_3 and $\beta_1/\beta_2/\beta_4$ sections. However, alternative conformations based on the native structure at zero force

are also relatively low in energy. The β sheet plus helical turn motif, which corresponds to GM2 for G/10/32, also appears as a favourable geometry for L/10/32/1.5 in Figure 8a, but does not appear to be the global minimum for L/10/32 in any force range. The diversity of the low-lying minima is evident from the geometrical parameters collected in Table 5, and from the colouring scheme based on the separation of the attachment points. This order parameter highlights the fact that the regions of configuration space corresponding to different morphologies are quite clearly separated by high barriers, but have overlapping energetics. The lowest-lying structures for L/10/32/7.5 all have fully extended chains for residues 10 to 32, with no remaining helical turns (Figure 8b). Hence most of the branches correspond to the largest distance between attachment points. However, the minimum that correlates with GM5 is not much higher in energy, but separated by a fairly large barrier, and the mean distance from the global minimum structure is larger than for $f = 1.5$. The structural diversity apparent in the L/10/32 disconnectivity graphs reflects the fact that the helix and β_2 part of the chain are constrained and extended by the static force, while the other parts of the protein are only affected indirectly. Hence the latter parts of the chain are relatively free to move, but are unable to achieve native-like conformations. Instead, a range of alternative possibilities exists with quite similar energies.

The disconnectivity graphs for G/10/32 are highly frustrated, with a variety of low-lying structures separated by high barriers. For example, with $f = 2.2$ structure GM2 is the global minimum, with a β sheet and a helical loop (Figure 9a). However, there are competitive structures lying almost as low in energy with alternative arrangements of the β strands, and the pathways between them involve quite lengthy rearrangements. There are also low-lying conformations corresponding to GM3 and GM4 with more extended structures consisting of helix/ β_3 plus a $\beta_1/\beta_2/\beta_4$ arrangement. Some of these subpopulations correspond to well defined distances between attachment points, as is evident from the colouring scheme. However, low-lying minima with similar 10/32 distances can also be separated by high barriers. For the landscape corresponding to $f = 6.5$ the lowest minima populate a relatively narrow range of 10/32 separations, but the average distance to the global minimum is still relatively large (Table 5).

When the static force increases to $f = 3.3$ GM3 is the global minimum structure, with GM4 very close in energy, along with GM5 and GM6, separated by high barriers. All the low-lying minima correspond to more extended structures in this regime, with all the helical turns intact (Figure 9b). When the force increases to the relatively high value of 6.5 structure GM8 becomes the global minimum, and the helix is disrupted in all the low-lying structures. There are significant barriers between some of these arrangements, even though they only appear to differ in the precise contacts made by strands β_3 and β_4 (Figure 9c).

From previous work^{13,62} we know that the unfolding temperatures for zero applied force occur for temperatures of 0.41 and 0.42 for protein G and L, respectively. Detailed analysis of the transition state ensembles and folding pathways in the absence of force can be found in these references, where a variety of structural order parameters were considered. For protein G various projections of the free energy surface indicate that the second β -hairpin forms first and then the second.¹³ The kinetics are best fitted by a double exponential, consistent with the formation of an intermediate, as discussed in §3.3. In contrast, the folding kinetics of protein L are well fitted by a single exponential.

The potential energy difference between compact and extended structures decreases roughly linearly with f (Figure 1 and Tables 1-4). Hence, for nonzero pulling forces the unfolding transition shifts to lower temperatures. Once a database of minima and transition states has been constructed the equilibrium occupation probabilities of all the minima can be calculated very rapidly as a function of temperature within the harmonic approximation. These probabilities can then be used to construct approximate probability distributions for any given structural order parameter. Examples are illustrated in Figure 3 for the G/10/32/2.2 database and an additional L/1/56/0.5 database containing 843246 minima. The equilibrium shifts from compact to extended over a temperature range of roughly 0.04 for L/1/56/0.5. In contrast, the distribution for G/10/32/2.2 is trimodal, resolving the more extended structures with and without the helical turns. In this case the population change occurs over a much wider temperature range, despite the higher applied force.

The frustration that appears as multiple funnels in the disconnectivity graphs for the 10/32 attachment pattern would probably be difficult to represent in projected free energy surfaces.^{75,77,95,111–114}

However, the corresponding barriers are evident in the pathways described below, where we focus on the order of events in the folding and unfolding pathways.

3.3 Folding and Unfolding Mechanisms

To examine the folding and unfolding mechanism as a function of static force we extracted the pathways corresponding to the largest contribution to the rate constant when intervening minima are placed in steady state^{74,91} using Dijkstra’s algorithm¹¹⁵ as implemented in the PATHSAMPLE program.⁹⁹ This approach accounts for all possible paths through the kinetic transition network, and does not require a reaction coordinate. For each kinetic transition network (DPS database) we compared the pathways to GM1 from the minima correlating with each of the other structures that are the global minimum for some force range. For brevity we will focus on the results in the absence of force, and for the lowest force value considered for each attachment pattern, where GM1 is still reasonably low in energy, but the applied force has a significant effect.

In the absence of force detailed comparisons are possible with previous simulations that employed the same potential with fixed bond lengths. The folding temperatures, T_f , were determined as 0.42 and 0.41 for protein L and G respectively,⁴⁹ with associated folding times of 1.57×10^4 for protein L and 1.37×10^4 and 4.64×10^4 from a double exponential fit for protein G (all in reduced units). The pathways discussed below correspond to rate constants calculated using transition state theory with harmonic vibrational densities of states at the corresponding folding temperatures.

For protein L in the absence of a static force the four pathways extracted for connections between GM1 and GM2 to GM5 correspond to very similar energy profiles and mechanisms (Figure 10; there is no connection to the even higher minimum that correlates with GM6 in this database). In the unfolding direction the β_4 strand separates first, followed by β_3 , and then the helix detaches from the β_1/β_2 hairpin. This hairpin then opens up, at first retaining some contacts in the loop region, but then all interstrand contacts are lost. The remaining snapshots in Figure 10 correspond to adjustments in the relative orientations of the β strands. The folding mechanism can be described as helix-assisted formation of hairpin one, in agreement with previous work.⁴⁹ The calculated folding rate constant

varies between 0.1×10^{-4} and 0.6×10^{-4} when states are regrouped recursively^{96,98,109} for a free energy barrier threshold of between 1.25 and 1.5. These calculations employed the lowest minimum GM1 as the product before regrouping, and a low-lying minimum from the region identified as non-native by the R_g and χ order parameters as the reactant. Smaller thresholds give rate constants about an order of magnitude lower, because some native-like states are not counted as product. Larger thresholds give values that are much too high, because folded and unfolded states become mixed.

In contrast to protein L, the pathways for protein G with $f = 0$ all involve a clear intermediate (Figure 10), which appears at a path length between 1500 and 2000 reduced units. The intermediate involves the helix and strands β_2 and β_3 , with β_3 associating first in the unfolding direction, in agreement with previous simulations for this model with fixed bond lengths.⁴⁹ This result also appears to be consistent with experimental results that indicate an early intermediate.³⁸ Alternative folding pathways have also been observed for a $G\bar{o}$ model representation of protein G.⁵⁰ The folding pathway illustrated in Figure 10 involves helix-assisted formation of the second hairpin after restructuring of the intermediate. Brown and Head-Gordon previously identified this as the fastest pathway, corresponding to some 80% of their folding population. The appearance of the intermediate in folding pathways will depend upon the precise geometry of the initial unfolded structure, especially the relative orientation of the β strands. The present results indicate that the intermediate provides a favourable three-state path for the most extended structures, which are expected to lie in the highest energy part of the population.

For L/1/56/1.0 the pathways are essentially the same from the relaxed GM1 structure at $f = 1$ to the structures that correlate with all the other global minima. Snapshots at intervals along the longest pathway to the GM6 minimum are shown in Figure 11a. The unfolding mechanism is qualitatively different from the $f = 0$ limit, and begins with the separation of the helix and β_2 strand from β_1 , which remains bound to the second hairpin. The next significant event is the dissociation of the contacts between β_1 and this C-terminal hairpin, followed by the separation of the helix from β_2 . The remainder of the path consists of reorientations of the β strands and loop regions, followed by sequential loss of the helix turns at the end. It is the loss of the helix that makes the pathway to

the GM6 structure longer than the paths to the other extended minima, which all retain the helix.

The rather different order of events for $f = 0$ and $f = 1$ in L/1/56 shows that the application of a static force can shift the balance between competing pathways. This observation is consistent with previous reports that unfolding paths may be different for chemical denaturation and mechanical disruption.^{107,114,116–120} In particular, previous simulations using a different potential found that unfolding of protein L proceeded via separate changes involving both hairpins, rather than just the C-terminal hairpin.⁵² In the case of the paths examined for G/1/56/1.5 the order of unfolding events is similar to the force-free scenario (Figure 11b). However, the overall path length is significantly shorter, and the intermediate described for $f = 0$ no longer appears, again illustrating the sensitivity of the landscape to perturbation.¹²¹ In the unfolding direction the β_1 strand first detaches, then β_4 , then β_2 , and eventually the remaining helix/ β_3 contacts are lost. The remainder of the path consists of reorientations of the β strands and sequential loss of the helix turns for the longest path that leads to the minimum correlating with GM6.

For L/10/32/1.5 the eight paths connecting the minimum correlating with GM1 to the other GM structures are very similar, except for the path to the minimum that correlates with GM5. This structure has an alternative arrangement of the three β strands, and the unfolding events are therefore different. For the majority pathways to the other GM structures the first major event in unfolding is the separation of the helix bound to β_3 from the N-terminal hairpin bound to β_4 (Figure 12a). The sequence containing residues 10 to 32 then straightens out further, before the helix turns are lost, the β_4 strand separates from the N-terminal hairpin, and β_3 dissociates from the helix. In the remainder of the path the C-terminal hairpin partly reforms and the β_1 strand forms contacts with β_2 . The rearrangements that occur beyond a path length of about 250 reduced units in Figure 12a may be driven by stretching of the helix/ β_2 residues, which makes this region incommensurate with favourable contacts from the other β strands.

The seven pathways connecting GM1 to the minima that correlate with GM2 to GM8 for G/10/32/2.2 are more diverse in character, reflecting the diversity of structures. Figure 12b shows the shorter paths that link GM1 to the GM2, GM3 and GM4 structures, which coincide closely with

the first section of the unfolding path to GM8. The helix and strand β_3 dissociate from the other three β strands, and then β_4 detaches from the N-terminal hairpin. β_4 then binds to β_3 , which detaches from the helix, to give a structure composed of a β sheet and a helical loop, similar to GM2. This arrangement then reorganises to give a structure with residues 10 to 32 extended and two β hairpins. In the path that continues to the minimum correlating with GM8 the β_1 strand next detaches, followed by the C-terminal hairpin, which then separates into β_3 and β_4 strands. Strand β_3 next reassociates with the helix, and some contacts form with β_4 . Further details are omitted for brevity. The unfolding paths to the minima correlating with GM5 and GM6 (not illustrated) also begin with the helix and β_3 strand dissociating from the other strands. The extended structures consisting of helix plus β_3 and $\beta_1/\beta_2/\beta_4$ are reached without passing through conformations involving a helical loop. The path to the minimum that correlates with GM7 (not illustrated) begins with strand β_1 and the helix/ β_3 moiety separating from strands β_2 and β_4 , with β_1 then rebinding to β_2 and β_4 in a different arrangement. The β_1 strand then unbinds as residues 10 to 32 become extended, and the helix/ β_3 plus $\beta_1/\beta_2/\beta_4$ structure corresponding to GM7 is finally obtained by rearrangements of the various strands.

The greater complexity of the pathways obtained for the 10/32 attachment pattern for static force accords with the frustration that is clearly visible in the corresponding disconnectivity graphs. With residues 10 to 32 forced into a non-native extended state the remaining unconstrained residues have considerable freedom, but no native-like arrangements are available. A simple connection between the energy landscape and the pulling force¹²² is unlikely to hold in this situation. This conclusion agrees with previous simulations where cut-based free energy profiles¹¹³ were considered for a simplified equilibrium kinetic transition network. A pulling force was found to stabilise states that are not significantly populated for the unperturbed protein,¹¹⁴ as for the potential energy landscape in the present work. Experimental verification of the predicted complexity for 10/32 attachment would require structural probes that report on geometrical characteristics beyond the distance between attachment points. Mutations that would be expected to stabilise alternative populations could be designed based on the reported structures and pathways. FRET spectroscopy might be another

possibility.

4 Properties of Individual Pathways as a Function of Applied Force

The potential energy landscape is expected to exhibit fewer local minima and transition states at higher values of f . As f increases pairs of transition states and connected local minima become progressively closer together, until they merge at non-Morse points, which are stationary points that have an additional zero Hessian eigenvalue. The geometry in the neighbourhood of such points obeys certain universal geometrical properties corresponding to fold catastrophes.^{123,124}

The behaviour of potential energy landscapes as a function of a single control parameter has been investigated extensively for atomic and mesoscopic clusters bound by the Morse potential.^{103,104,125–129} A quantitative prediction of the theory is that the fold ratio, r_f , should tend to unity in the limit of short path lengths, where

$$r_f = \frac{6\Delta V}{\lambda(\Delta x)^2}, \quad (4)$$

where ΔV is the barrier height, λ is the modulus of the curvature corresponding to the unique vibrational normal mode with imaginary frequency for the transition state, and Δx is the Euclidean distance in $3N$ -dimensional space between the transition state and the minimum. This prediction has been verified for Morse clusters^{128,129} and for thermodynamic potentials corresponding to finite system analogues of first order phase transitions.^{71,130} It can be used to justify Hammond's postulate, that a transition state more closely resemble the higher energy minimum connected to it by steepest-descent pathways.¹²⁸

If the system approaches a fold catastrophe at applied force f_0 then the barrier height, path length, and curvature vanish as $(f_0 - f)^{3/2}$, $(f_0 - f)^{1/2}$ and $(f_0 - f)^{1/2}$, respectively,^{123,124} as f approaches f_0 from below. This behaviour has been investigated for model proteins in several previous studies,^{23,131,132} and the same scaling relations may apply for a projected free energy profile as a function of f .^{133–136} Since the predictions of catastrophe theory refer to the limit $f \rightarrow f_0$ it is particularly interesting to determine how well the relations hold away from f_0 .^{128,131,136}

To investigate the behaviour of the fold ratio for the model proteins considered in the present work we calculated r_f for all the transition states in the protein L database with $f = 0$ and the behaviour of one particular path as a function of f . Here Δx was obtained as the distance between connected minima and transition states, minimised with respect to overall translation and rotation.¹¹⁰ Figure 13 shows a plot of r_f values against Δx , which confirms that the fold ratio does indeed tend to unity as the path length decreases. This result should hold for any applied force.

Having examined the variation of r_f for a database of pathways at a fixed value of f we then considered one particular pathway as a function of f . The chosen pathway corresponds to the step in Figure 1(d) for the global minimum of protein L pulled between beads 10 and 32. This minimum merges with a transition state at $f \approx 9.621925$, and for smaller forces the corresponding pathway links GM1 with a minimum that exhibits a β -sheet formed from strands $\beta_1/\beta_2/\beta_4$ and β_3 packed against the α helix (Figure 14b). The convergence of r_f to unity for this pathway is illustrated in Figure 14a, while Figure 14b shows how the corresponding minimum and transition state merge as f increases.

The shortest pathways in any database for a given value of f are therefore expected to exhibit a direct correlation between the barrier height, which influences kinetics, and the smallest non-zero vibrational frequency of the minimum, which appears in the vibrational partition function. In previous work we have found that an approximate linear free energy relation can hold between the barrier height and the square of the displacement.¹³⁰ It remains to be seen whether this relation might be useful for analysing mechanical unfolding experiments, perhaps in comparing barriers for different mutants at constant force.

5 Conclusions

The potential energy landscape has been investigated for coarse-grained models of protein L and protein G using two distinct pulling directions, namely 1/56 and 10/32 attachments. The lowest-energy structure was first identified as a function of the pulling force f using basin-hopping global optimisation. When the pulling force is applied to the terminal residues in the 1/56 pattern the global

minimum changes abruptly from a compact structure, which correlates with the global minimum at zero force, to extended minima. This crossover occurs at a slightly larger force for protein G, in agreement with its greater mechanical stability in experiments. For the 10/32 attachment pattern the evolution of the landscape is more complicated, and significant differences arise between the two proteins. In particular, although both systems exhibit two compact global minima, the two structures for protein L are very similar, while the structure that appears at non-zero force for protein G has a helical turn. For intermediate forces the helix and β_2 strand regions are extended, with β_3 packed against the helix, and β_1 and β_4 forming a β sheet with β_2 . A number of different arrangements generally exist for the protein outside the region of applied force, leading to frustrated energy landscapes with high barriers between alternative low-lying minima. These results clearly show that pulling can simplify or cause significant frustration in the underlying landscape, depending on the pulling direction.

The folding pathways and rate constants in the absence of force are in good agreement with previous work. For protein L the N-terminal hairpin folds first, while for protein G it is the C-terminal hairpin, with a well-defined early intermediate involving the helix, β_3 and β_2 . However, for $f = 1$, where the global minimum still correlates with GM1, the C-terminal hairpin forms first on the folding path, followed by association of β_2 and the helix. In the last stages of the folding pathway β_1 associates with the C-terminal hairpin and the helix/ β_2 arrangement.

For protein G the intermediate disappears on the application of static force in the regime where the original global minimum is still relatively low in energy. The folding mechanism still involves the association of helix with β_3 and β_2 , with β_4 joining to make the C-terminal hairpin. Finally, the β_1 strand associates with the rest of the secondary structure elements, as in the limit of zero static force.

The pathways between the GM1 structure and higher-lying extended minima are qualitatively different for the 10/32 pulling direction, and are particularly diverse for protein G. For protein L the steps in folding for L/10/32/1.5 correspond to the dissociation of the β_4/β_3 arrangement in the extended structure and subsequent formation of contacts between β_4 and the N-terminal hairpin.

The extended structure corresponding to helix/ β_3 at one end and $\beta_1/\beta_2/\beta_4$ at the other then folds along the boundary between these elements to give GM1. For G/10/32/2.2 the N-terminal hairpin dissociates from the helix in GM3 and forms contacts with the C-terminal hairpin to give a β sheet structure and helical turn. The two hairpins then realign via structures with separate β_4 and β_3 strands to give GM1. Various other pathways arise depending on which extended minimum is chosen as the other endpoint. These results highlight the complexity that may arise for different pulling patterns, and show how the balance between different pathways may be shifted significantly by pulling.

Acknowledgements

We are grateful to Dr Robert Best for supplying the parameters used in the G \bar{o} model of Graham and Best.¹⁰⁶

References

- [1] Dill, K. A.; Chan, H. S. *Nature Struct. Biol.* **1997**, *4*, 10–19.
- [2] Head-Gordon, T.; Brown, S. *Curr. Op. Struct. Biol.* **2003**, *13*, 160–167.
- [3] Guo, Z.; Brooks III, C. L. *Biopolymers* **1997**, *42*, 745.
- [4] Berry, R. S.; Elmaci, N.; Rose, J. P.; Vekhter, B. *Proc. Natl Acad. Sci. USA* **1997**, *94*, 9520.
- [5] Nymeyer, H.; García, A. E.; Onuchic, J. N. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 5921.
- [6] Shea, J. E.; Nochomovitz, Y. D.; Guo, Z. Y.; Brooks, C. L. *J. Chem. Phys.* **1998**, *109*, 2895–2903.
- [7] Miller, M. A.; Wales, D. J. *J. Chem. Phys.* **1999**, *111*, 6610–6616.
- [8] Vekhter, B.; Berry, R. S. *J. Chem. Phys.* **1999**, *110*, 2195.
- [9] Vekhter, B.; Berry, R. S. *J. Chem. Phys.* **1999**, *111*, 3753.
- [10] Elmaci, N.; Berry, R. S. *J. Chem. Phys.* **1999**, *110*, 10606–10622.

- [11] Shea, J.-E.; Onuchic, J. N.; Brooks, C. L. *J. Chem. Phys.* **2000**, *113*, 7663.
- [12] Evans, D. A.; Wales, D. J. *J. Chem. Phys.* **2003**, *118*, 3891–3897.
- [13] Brown, S.; Fawzi, N. J.; Head-Gordon, T. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 10712–10717.
- [14] Stoycheva, A. D.; Onuchic, J. N.; Brooks, C. L. *J. Chem. Phys.* **2003**, *119*, 5722–5729.
- [15] Wales, D. J.; Dewsbury, P. E. J. *J. Chem. Phys.* **2004**, *121*, 10284–10290.
- [16] Kim, J.; Keyes, T. *J. Phys. Chem. B* **2007**, *111*, 2647–2657.
- [17] Kim, J.; Keyes, T.; Straub, J. E. *J. Chem. Phys.* **2009**, *130*, 124112.
- [18] Honeycutt, J. D.; Thirumalai, D. *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 3526.
- [19] Honeycutt, J. D.; Thirumalai, D. *Biopolymers* **1992**, *32*, 695.
- [20] Guo, Z. Y.; Thirumalai, D. *Biopolymers* **1995**, *36*, 83.
- [21] Guo, Z. Y.; Thirumalai, D. *J. Mol. Biol.* **1996**, *263*, 323.
- [22] Li, F.-Y.; Yuan, J.-M.; Mou, C.-Y. *Phys. Rev. E* **2001**, *63*, 021905.
- [23] Lacks, D. J. *Biophysical Journal* **2005**, *88*, 3494.
- [24] Imparato, A.; Luccioli, S.; Torcini, A. *Phys. Rev. Lett.* **2007**, *99*, 168101.
- [25] Luccioli, S.; Imparato, A.; Torcini, A. *Phys. Rev. E* **2008**, *78*, 031907.
- [26] Luccioli, S.; Imparato, A.; Mitternacht, S.; Irbäck, A.; Torcini, A. *Phys. Rev. E* **2010**, *81*, 13–16.
- [27] Carrion-Vazquez, M.; Oberhauser, A. F.; Fowler, S. B.; Marszalek, P. E.; Broedel, S. E.; Clarke, J.; Fernández, J. M. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 3694–3699.
- [28] Li, H.; Oberhauser, A. F.; Fowler, S. B.; Clarke, J.; Fernández, J. M. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 6527.

- [29] Oberhauser, A. F.; Hansma, P. K.; Carrion-Vazquez, M.; Fernández, J. M. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 6527–6531.
- [30] Best, R. B.; Fowler, S. B.; Toca-Herrera, J. L.; Clarke, J. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 12143–12148.
- [31] Brockwell, D. J. *Biochem. Soc. Trans.* **2007**, *35*, 1564–1568.
- [32] Liphardt, J.; Onoa, B.; Smith, S. B.; Tinoco, I. J.; Bustamante, C. *Science* **2001**, *292*, 733–737.
- [33] Ritort, F. *J. Phys. C* **2006**, *18*, R531–R583.
- [34] Greenleaf, W. J.; Woodside, M. T.; Block, S. M. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 171–190.
- [35] Liu, R.; Garcia-Manyes, S.; Sarkar, A.; Badilla, C. L.; Fernández, J. M. *Biophys. J.* **2009**, *96*, 3810–3821.
- [36] Wikstrom, M.; Drakenberg, T.; Forsen, S.; Sjobring, U.; Bjork, L. *Biochem.* **1994**, *33*, 14011–14017.
- [37] Gu, H. D.; Yi, Q. A.; Bray, S. T.; Riddle, D. S.; Shiau, A. K.; D., B. *Protein Sci.* **1995**, *4*, 1108–1117.
- [38] Park, S. H.; O’Neil, K. T.; Roder, H. *Biochem.* **1997**, *36*, 14277–14283.
- [39] Gu, H. D.; Kim, D.; D., B. *J. Mol. Biol.* **1997**, *274*, 588–596.
- [40] Plaxco, K. W.; Millett, I. S.; Segel, D. J.; Doniach, S.; D., B. *Nat. Struct. Biol.* **1999**, *6*, 554–556.
- [41] Park, S. H.; Shastry, M. C. R.; Roder, H. *Nature Struct. Biol.* **1999**, *6*, 943–947.
- [42] Scalley, M. L.; Yi, Q.; Gu, H. D.; McCormack, A.; Baker, D. *Biochem.* **1997**, *36*, 3373–3382.
- [43] Nauli, S.; Kuhlman, B.; Baker, D. *Nat. Struct. Biol.* **2001**, *8*, 602–605.

- [44] Sadler, D. P.; Petrik, E.; Taniguchi, Y.; Pullen, J. R.; Kawakami, M.; Radford, S. E.; Brockwell, D. J. *J. Mol. Biol.* **2009**, *393*, 237–248.
- [45] Kim, D. E.; Fisher, C.; Baker, D. *J. Mol. Biol.* **2000**, *298*, 971–984.
- [46] McCallister, E. L.; Alm, E.; Baker, D. *Nature Struct. Biol.* **2000**, *7*, 669–673.
- [47] Abkevich, V. I.; Gutin, A. M.; Shakhnovich, E. I. *Biochem.* **1994**, *33*, 10026.
- [48] Fersht, A. R. *Curr. Opin. Struct. Biol.* **1997**, *7*, 3–9.
- [49] Brown, S.; Head-Gordon, T. *Protein Sci.* **2004**, *13*, 958–970.
- [50] Shimada, J.; Shakhnovich, E. I. *Proc. Nat. Acad. Sci. USA* **2002**, *99*, 11175–11180.
- [51] Karanicolas, J.; Brooks, C. L. *Protein Sci.* **2002**, *11*, 2351–2361.
- [52] Brockwell, D. J.; Beddard, G. S.; Paci, E.; West, D. K.; Olmsted, P. D.; Smith, D. A.; Radford, S. E. *Biophys. J.* **2005**, *89*, 506–519.
- [53] Cao, Y.; Li, H. *Nature Materials* **2007**, *6*, 109–114.
- [54] Cao, Y.; Lam, C.; Wang, M.; Li, H. *Angew. Chem. Int. Ed. Engl.* **2006**, *45*, 642.
- [55] Li, P. C.; Makarov, D. E. *J. Phys. Chem. B* **2004**, *108*, 745–749.
- [56] Cao, Z.; Liu, H. *J. Chem. Phys.* **2008**, *129*, 01510–1.
- [57] Carrion-Vazquez, M.; Li, H.; Lu, H.; Marszalek, P. E.; Oberhauser, A. F.; Fernández, J. M. *Nat. Struct. Biol.* **2003**, *10*, 738.
- [58] Yew, Z. T.; Olmsted, P. D.; Paci, E. *Advances Chem. Phys.* **2011**, *146*, 395–418.
- [59] Sheinerman, F. B.; Brooks, C. L. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 1562–1567.
- [60] Yap, E.-H.; Fawzi, N. L.; Head-Gordon, T. *Proteins, Struct. Func.. Bioinf.* **2008**, *70*, 626–638.

- [61] Marianayagam, N. J.; Fawzi, N. L.; Head-Gordon, T. *Proc. Nat. Acad. Sci. USA* **2005**, *102*, 16684–16689.
- [62] Fawzi, N. L.; Chubukov, V.; Brown, S.; ; Head-Gordon, T. *Protein Sci.* **2005**, *14*, 993–1003.
- [63] Gmin: A program for basin-hopping global optimisation. Wales, D. J.; <http://www-wales.ch.cam.ac.uk/software.html>.
- [64] Optim: A program for optimising geometries and calculating pathways. Wales, D. J.; <http://www-wales.ch.cam.ac.uk/software.html>.
- [65] West, D. K.; Brockwell, D. J.; Olmsted, P. D.; Radford, S. E.; Paci, E. *Biophys. J.* **2006**, *90*, 287–297.
- [66] Li, Z.; Scheraga, H. A. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 6611.
- [67] Li, Z.; Scheraga, H. A. *J. Mol. Struct.* **1988**, *179*, 333.
- [68] Wales, D. J.; Doye, J. P. K. *J. Phys. Chem. A* **1997**, *101*, 5111–5116.
- [69] Oakley, M. T.; Wales, D. J.; Johnston, R. L. *J. Phys. Chem. B* **2011**, *115*, 11525–11529.
- [70] Doye, J. P. K.; Miller, M. A.; Wales, D. J. *J. Chem. Phys.* **1999**, *110*, 6896–6906.
- [71] Wales, D. J. *Energy Landscapes*; Cambridge University Press: Cambridge, 2003.
- [72] Cvijovic, D.; Klinowski, J. *Science* **1995**, *267*, 664–666.
- [73] Noé, F.; Krachtus, D.; Smith, J. C.; Fischer, S. *J. Chem. Theory Comput.* **2006**, *2*, 840–857.
- [74] Wales, D. J. *Int. Rev. Phys. Chem.* **2006**, *25*, 237–282.
- [75] Noé, F.; Fischer, S. *Curr. Op. Struct. Biol.* **2008**, *18*, 154–162.
- [76] Prada-Gracia, D.; Gómez-Gardenes, J.; Echenique, P.; Fernando, F. *PLoS Comput. Biol.* **2009**, *5*, 1–9.

- [77] Wales, D. J. *Curr. Op. Struct. Biol.* **2010**, *20*, 3–10.
- [78] Murrell, J. N.; Laidler, K. J. *Trans. Faraday. Soc.* **1968**, *64*, 371.
- [79] Nocedal, J. *Mathematics of Computation* **1980**, *35*, 773–782.
- [80] Liu, D.; Nocedal, J. *Math. Program.* **1989**, *45*, 503–528.
- [81] Trygubenko, S. A.; Wales, D. J. *J. Chem. Phys.* **2004**, *120*, 2082.
- [82] Elber, R.; Karplus, M. *Chem. Phys. Lett.* **1987**, *139*, 375.
- [83] Czerminski, R.; Elber, R. *J. Chem. Phys.* **1990**, *92*, 5580.
- [84] Henkelman, G.; Jónsson, H. *J. Chem. Phys.* **1999**, *111*, 7010–7022.
- [85] Henkelman, G.; Uberuaga, B. P.; Jónsson, H. *J. Chem. Phys.* **2000**, *113*, 9901–9904.
- [86] Henkelman, G.; Jónsson, H. *J. Chem. Phys.* **2000**, *113*, 9978–9985.
- [87] Henkelman, G.; Jónsson, H. *J. Chem. Phys.* **2001**, *115*, 9657–9666.
- [88] Munro, L. J.; Wales, D. J. *Phys. Rev. B* **1999**, *59*, 3969–3980.
- [89] Kumeda, Y.; Munro, L. J.; Wales, D. J. *Chem. Phys. Lett.* **2001**, *341*, 185–194.
- [90] Carr, J. M.; Trygubenko, S. A.; Wales, D. J. *J. Chem. Phys.* **2005**, *122*, 234903.
- [91] Wales, D. J. *Mol. Phys.* **2002**, *100*, 3285–3306.
- [92] Wales, D. J. *Mol. Phys.* **2004**, *102*, 891–908.
- [93] Pande, V. S.; Rokhsar, D. S. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 1273–1278.
- [94] Swope, W. C.; Pitner, J. W.; Suits, F. *J. Phys. Chem. B* **2004**, *108*, 6571–6581.
- [95] Krivov, S. V.; Karplus, M. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 14766–14770.
- [96] Carr, J. M.; Wales, D. J. *J. Chem. Phys.* **2005**, *123*, 234901.

- [97] Strodel, B.; Whittleston, C. S.; Wales, D. J. *J. Am. Chem. Soc.* **2007**, *129*, 16005–16014.
- [98] Carr, J. M.; Wales, D. J. *J. Phys. Chem. B* **2008**, *112*, 8760–8769.
- [99] Pathsample: A program for generating connected stationary point databases and extracting global kinetics. Wales, D. J.; <http://www-wales.ch.cam.ac.uk/software.html>.
- [100] Becker, O. M.; Karplus, M. *J. Chem. Phys.* **1997**, *106*, 1495.
- [101] Wales, D. J.; Miller, M. A.; Walsh, T. R. *Nature* **1998**, *394*, 758–760.
- [102] Krivov, S. V.; Karplus, M. *J. Chem. Phys.* **2002**, *117*, 10894–10903.
- [103] Doye, J. P. K.; Wales, D. J. *Science* **1996**, *271*, 484–487.
- [104] Doye, J. P. K.; Wales, D. J. *J. Phys. B* **1996**, *29*, 4859–4894.
- [105] Miller, M. A.; Doye, J. P. K.; Wales, D. J. *Phys. Rev. E* **1999**, *60*, 3701–3718.
- [106] Graham, T. G. W.; Best, R. B. *J. Phys. Chem. B* **2011**, *115*, 1546–1561.
- [107] Brockwell, D. J.; Paci, E.; Zinober, R. C.; Beddard, G. S.; Olmsted, P. D.; Smith, D. A.; Perham, R. N.; Radford, S. E. *Nat. Struct. Biol.* **2003**, *10*, 731–737.
- [108] Dietz, H.; Berkemeier, F.; Rief, M. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 12724–12728.
- [109] Carr, J. M.; Wales, D. J. *Phys. Chem. Chem. Phys.* **2009**, *11*, 3341–3354.
- [110] Kearsley, S. K. *Acta Cryst. A* **1989**, *45*, 208–210.
- [111] Krivov, S. V.; Karplus, M. *J. Phys. Chem. B* **2006**, *110*, 12689–12698.
- [112] Muff, S.; Caffisch, A. *Proteins: Struct., Func. and Bioinf.* **2008**, *70*, 1185–1195.
- [113] Krivov, S. V.; Karplus, M. *Proc. Nat. Acad. Sci. USA* **2008**, *105*, 13841–13846.
- [114] Yew, Z. T.; Krivov, S.; Paci, E. *J. Phys. Chem. B* **2008**, *112*, 16902–16907.

- [115] Dijkstra, E. W. *Numerische Math.* **1959**, *1*, 269–271.
- [116] Paci, E.; Karplus, M. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 6521.
- [117] Best, R. B.; Li, B.; Steward, A.; Daggett, V.; Clarke, J. *Biophys. J.* **2001**, *81*, 2344–2356.
- [118] Brockwell, D. J.; Beddard, G. S.; Clarkson, J.; Zinober, R. C.; Blake, A. W.; Trinick, J.; Olmsted, P. D.; Smith, D. A.; Radford, S. E. *Biophys. J.* **2002**, *83*, 458–472.
- [119] Fowler, S. B.; Best, R. B.; Herrera, J. L. T.; Rutherford, T. J.; Steward, A.; Paci, E.; Karplus, M.; Clarke, J. *J. Mol. Biol.* **2002**, *322*, 841–849.
- [120] West, D. K.; Olmsted, P. D.; Paci, E. *J. Chem. Phys.* **2006**, *124*, 154909.
- [121] Brockwell, D. J.; Radford, S. E. *Curr. Op. Struct. Biol.* **2007**, *17*, 30–37.
- [122] Bell, G. I. *Science* **1978**, *200*, 618.
- [123] Thom, R. *Stabilité Structurale et Morphogénèse*; Benjamin: New York, 1972.
- [124] Gilmore, R. *Catastrophe theory for scientists and engineers*; Wiley: New York, 1981.
- [125] Braier, P. A.; Berry, R. S.; Wales, D. J. *J. Chem. Phys.* **1990**, *93*, 8745–8756.
- [126] Doye, J. P. K.; Wales, D. J.; Berry, R. S. *J. Chem. Phys.* **1995**, *103*, 4234–4249.
- [127] Doye, J. P. K.; Wales, D. J. *J. Chem. Soc., Faraday Trans.* **1997**, *93*, 4233–4243.
- [128] Wales, D. J. *Science* **2001**, *293*, 2067–2069.
- [129] Calvo, F.; Doye, J. P. K.; Wales, D. J. *Nanoscale* **2011**, *0*, 0000.
- [130] Bogdan, T. V.; Wales, D. J. *J. Chem. Phys.* **2004**, *120*, 11090–11099.
- [131] Maloney, C. E.; Lacks, D. J. *Phys. Rev. E* **2006**, *73*, 061106.
- [132] Tshiprut, Z.; Urbakh, M. *J. Chem. Phys.* **2009**, *130*, 084703.

- [133] Garg, A. *Phys. Rev. B* **1995**, *51*, 15592–15595.
- [134] Dudko, O. K.; Filippov, A. E.; Klafter, J.; Urbakh, M. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 11378–11381.
- [135] Dias, C. L.; Dube, M.; Oliveira, F. A.; Grant, M. *Phys. Rev. E* **2005**, *72*, 011918.
- [136] Lacks, D. J.; Willis, J.; Robinson, M.-P. *J. Phys. Chem. B* **2010**, *114*, 10821–10825.
- [137] Humphrey, W.; Dalke, A.; Schulten, K. *J. Molec. Graphics* **1996**, *14*, 33–38.

f	0.0	1.0	1.5	2.5	3.0	3.5	5.0
min	190714	1007353	74013	181140	92085	81512	104568
ts	206163	772810	80985	185530	91226	71072	95662
GM1	-13.774117	-20.924895	-24.522881	-31.763396	-35.405778	-39.062872	-50.122897
GM2	23.590422	-14.822502	-34.908882	-75.746112	-96.448782	-117.324116	-180.917609
GM3	24.844548	-13.900260	-34.245806	-75.554097	-96.464264	-117.530951	-181.621425
GM4	27.311100	-12.47503	-33.172210	-75.092067	-96.282195	-117.616193	-182.446267
GM5	29.76259	-11.397321	-32.297439	-74.582686	-95.950267	-117.460543	-182.812571
GM6	39.629570	-4.174596	-26.243915	-70.692796	-93.070535	-115.549688	-183.591190

Table 1: Number of minima and transition states in the DPS databases constructed for L/1/56 as a function of the static force, f . The energies of the six distinct global minimum structures that arise for different values of f are reported for the relaxed structures. The f values at which the global minimum changes from GM1 to GM2, etc. are approximately 1.2, 3.0, 3.4, 4.0, and 4.6.

f	0.0	1.5	7.5
min	190714	892228	248988
ts	206163	673501	218983
GM1	-13.774116	-19.310864	
GM2	-13.732850	-20.300141	-50.263786
GM3	-3.498625	-21.663330	-101.889300
GM4	-3.365967	-21.630401	-102.280377
GM5	2.355742	-18.930808	-110.949664
GM6	6.909812	-15.822193	-112.745199
GM7	17.939084	-7.817254	-113.967281
GM8	21.267683	-5.005314	-113.357006
GM9	21.677542	-4.532854	-113.056340
GM10	25.685065	-0.817348	-110.698589

Table 2: Number of minima and transition states in the DPS databases constructed for L/10/32 as a function of the static force, f . The energies of the ten distinct global minimum structures that arise for different values of f are reported for the relaxed structures. The GM1 structure does not exist as a minimum on the PES at $f = 7.5$. The f values at which the global minimum changes from GM1 to GM2, etc. are approximately 0.1, 1.4, 2.0, 2.9, 5.2, 6.7, 9.1, 11.8, and 16.4.

f	0.0	1.5	2.0	2.5	5.0
min	182544	33872	30307	35681	28350
ts	175290	45244	44355	53417	32657
GM1	-16.164846	-26.898684	-30.516517	-34.153568	-52.626753
GM2	23.758413	-34.828123	-55.148704	-75.670588	-180.857162
GM3	24.968053	-34.160495	-54.730050	-75.479607	-181.563282
GM4	27.425023	-33.125844	-54.012679	-75.058767	-182.432800
GM5	29.878532	-32.231099	-53.305153	-74.532622	-182.789185
GM6	39.580602	-26.296445	-48.468047	-70.742259	-183.633277

Table 3: Number of minima and transition states in the DPS databases constructed for G/1/56 as a function of the static force, f . The energies of the six distinct global minimum structures that arise for different values of f are reported for the relaxed structures. The f values at which the global minimum changes from GM1 to GM2, etc. are approximately 1.3, 3.0, 3.3, 4.0, and 4.6.

f	0.0	2.2	3.3	3.4	4.0	5.5	6.5
min	182544	608892	669423	599638	587728	568181	533009
ts	175290	482011	540711	482977	511702	461731	476345
GM1	-16.164846	-25.899406	-30.912781	-31.373383	-34.154071	-41.235973	-46.065480
GM2	-7.489007	-27.119149	-37.192639	-38.116132	-43.683099	-57.788356	-67.334678
GM3	-1.283862	-26.985670	-40.286835	-41.508045	-48.874683	-67.579700	-80.684017
GM4	4.216722	-25.220851	-40.243663	-41.618038	-49.893374	-70.793443	-84.891806
GM5	7.755061	-23.870611	-40.099398	-41.586597	-50.549003	-73.233475	-88.566137
GM6	9.501782	-23.256687	-40.007664	-41.541234	-50.778825	-74.135310	-89.909222
GM7	12.238124	-21.667282	-39.157868	-40.760495	-50.417694	-74.855058	-91.367248
GM8	23.298156	-15.601070	-35.362937	-37.169469	-48.043306	-75.485236	-93.981685

Table 4: Number of minima and transition states in the DPS databases constructed for G/10/32 as a function of the static force, f . The energies of the eight distinct global minimum structures that arise for different values of f are reported for the relaxed structures. The f values at which the global minimum changes from GM1 to GM2, etc. are approximately 2.0, 2.3, 3.4, 3.5, 3.6, 4.6, and 5.2.

system	separation of attachment points		distance to global minimum	
	$\mu(D)$	$\sigma(D)$	$\mu(G)$	$\sigma(G)$
L	7.21	0.22	3.33	1.80
L/1/56/2.5	41.42	0.42	4.25	1.47
L/1/56/5.0	44.64	0.65	4.32	2.28
L/10/32/1.5	13.16	0.92	9.15	7.81
L/10/32/7.5	18.58	0.13	26.10	6.55
G	7.14	0.26	2.39	1.82
G/1/56/2.0	40.81	0.40	4.30	1.64
G/1/56/5.0	44.72	0.64	4.00	2.31
G/10/32/2.2	15.32	0.71	45.88	6.15
G/10/32/3.3	16.15	0.44	39.77	2.98
G/10/32/6.5	18.56	0.06	15.82	4.29

Table 5: The mean, μ , and standard deviation, σ , of the distance between the attachment points, D , and the minimised distance to the global minimum, G , for selected databases. G corresponds to the shortest distance in the 168-dimensional configuration space, minimised with respect to overall translation and rotation.¹¹⁰ The averages include a weight corresponding to the equilibrium occupation probability of each minimum calculated using harmonic vibrational densities of states.

Figures

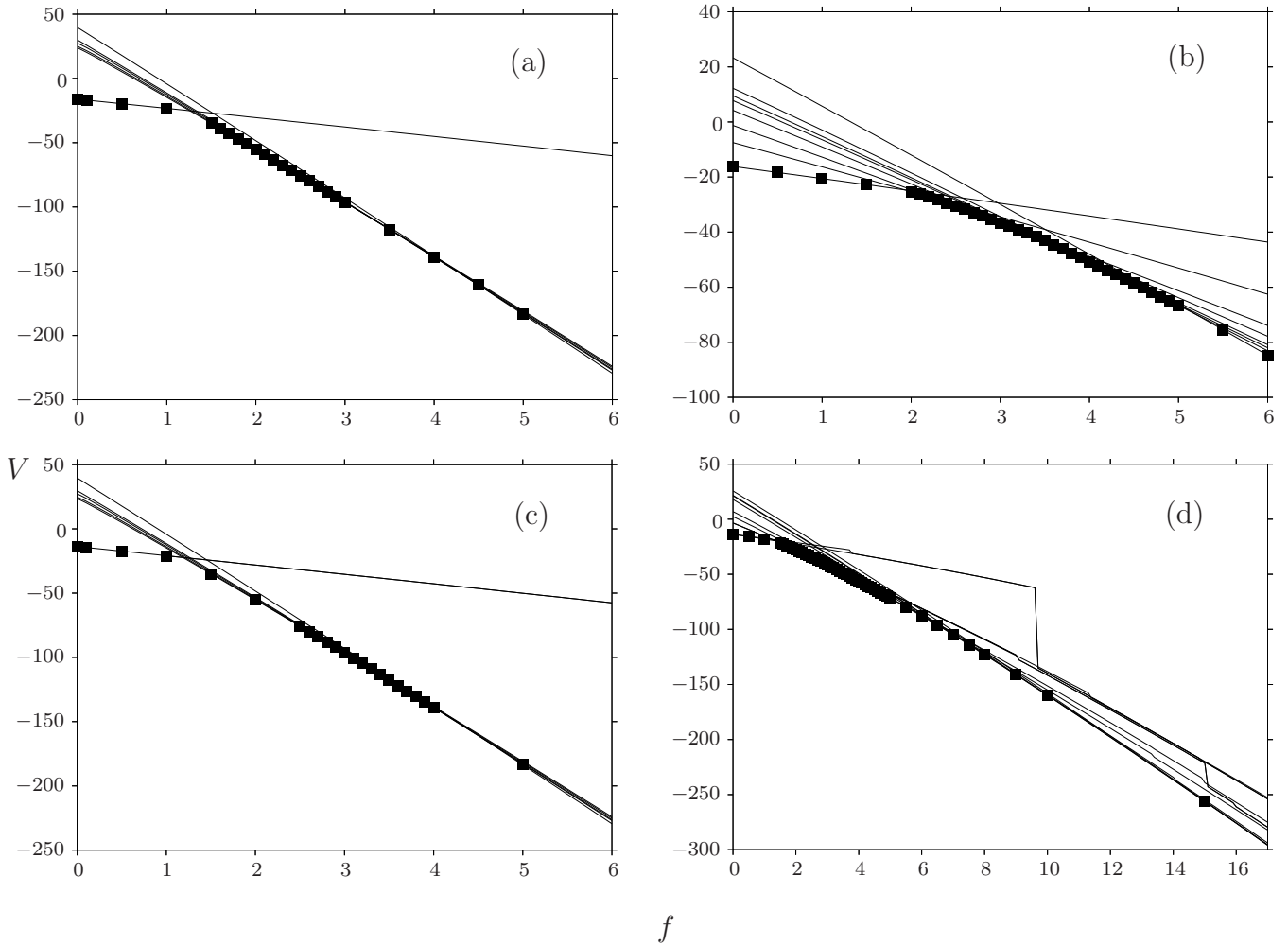


Figure 1: Global minimum potential energy, V/ϵ , as a function of pulling force, f , for (a) G/1/56, (b) G/10/32, (c) L/1/56, and (d) L/10/32. The black squares indicate the f values for which basin-hopping runs were conducted. The lines connect the relaxed potential energy values for each distinct global minimum structure, relaxed progressively over the full range of f considered.

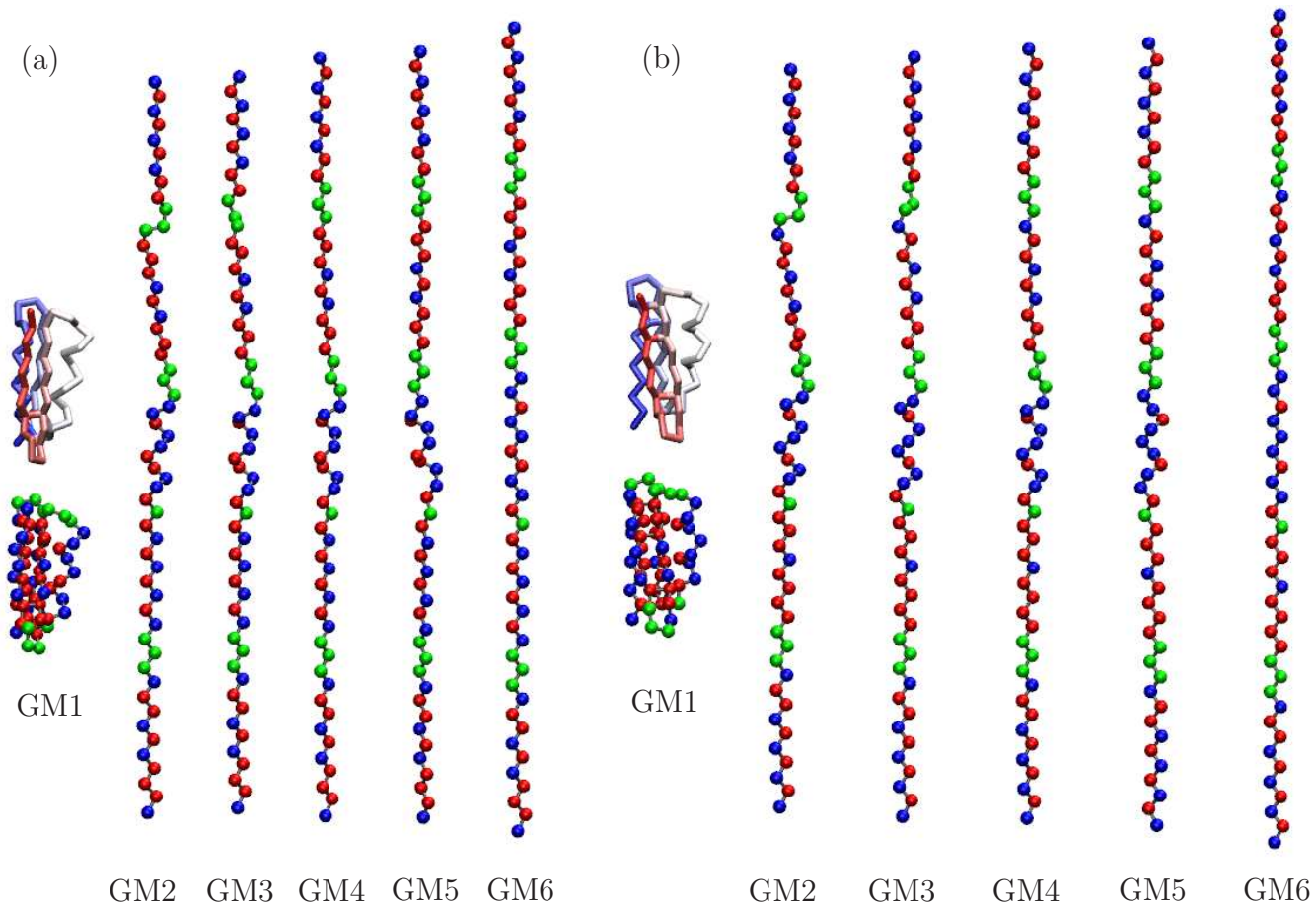


Figure 2

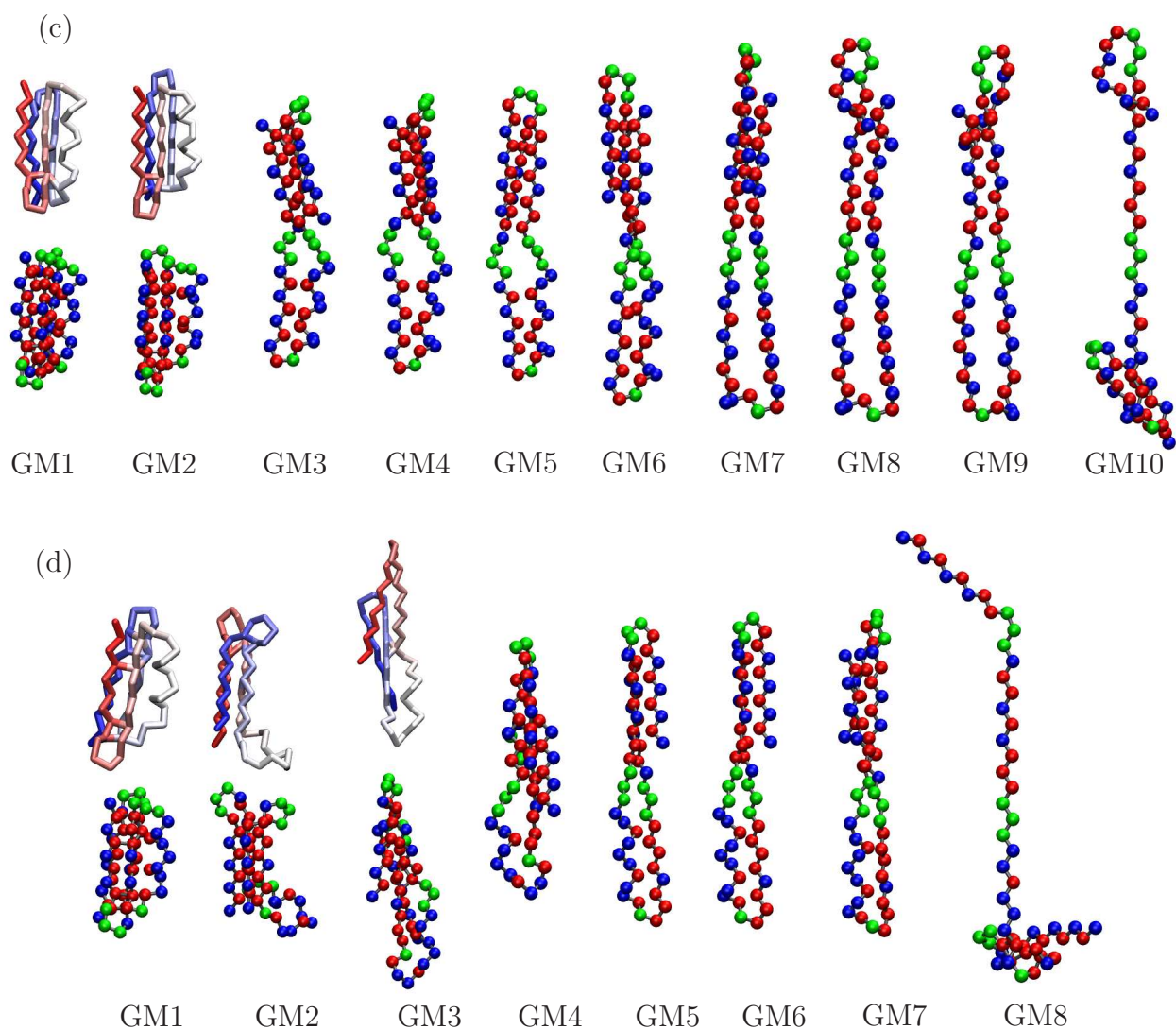


Figure 2: Structures of (a) the six distinct global minima identified for L/1/56, (b) the six distinct global minima identified for G/1/56, (c) the ten distinct global minima identified for L/10/32, (d) the eight distinct global minima identified for G/10/32, visualised using the VMD program.¹³⁷ Hydrophobic beads are coloured red, hydrophilic beads blue, and neutral beads green. The alternative views of GM1 [and GM2 in (c) and (d)] have the beads coloured from red to blue (N-terminus, β_1 to C-terminus, β_4) according to the position in the chain.

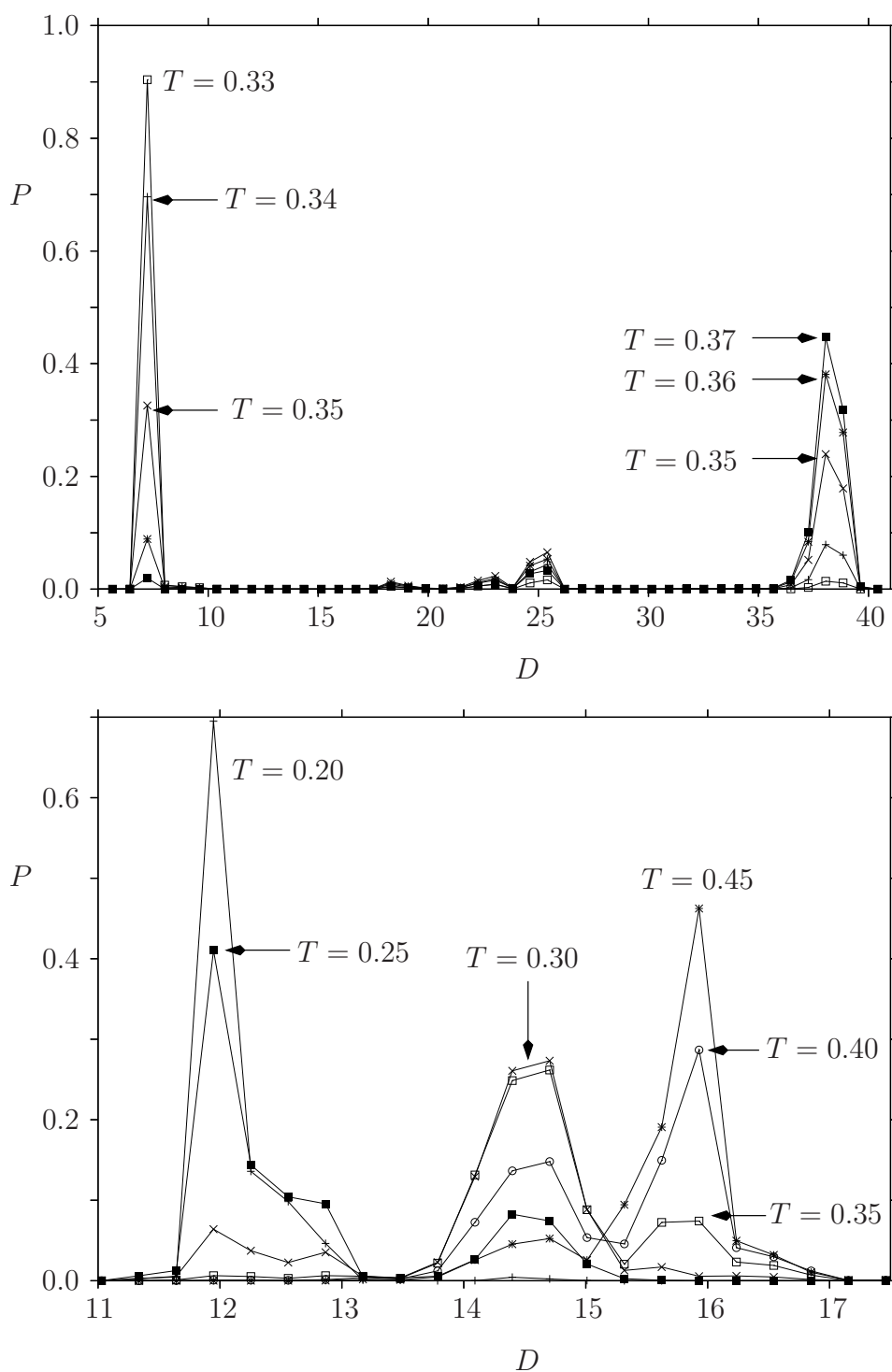


Figure 3: Probability distributions for the distance between attachment points. Top: for L/1/56/0.5 the distribution shifts from compact to extended over the temperature range 0.33 to 0.37. Bottom: for G/10/32/2.2 a trimodal distribution appears. The probability shifts from compact at the lowest temperatures to extended at the highest temperatures, with significant population of intermediate values, where the helical turns are retained, in between. Some of the corresponding structures are illustrated for the disconnectivity graphs shown in Figure 9.

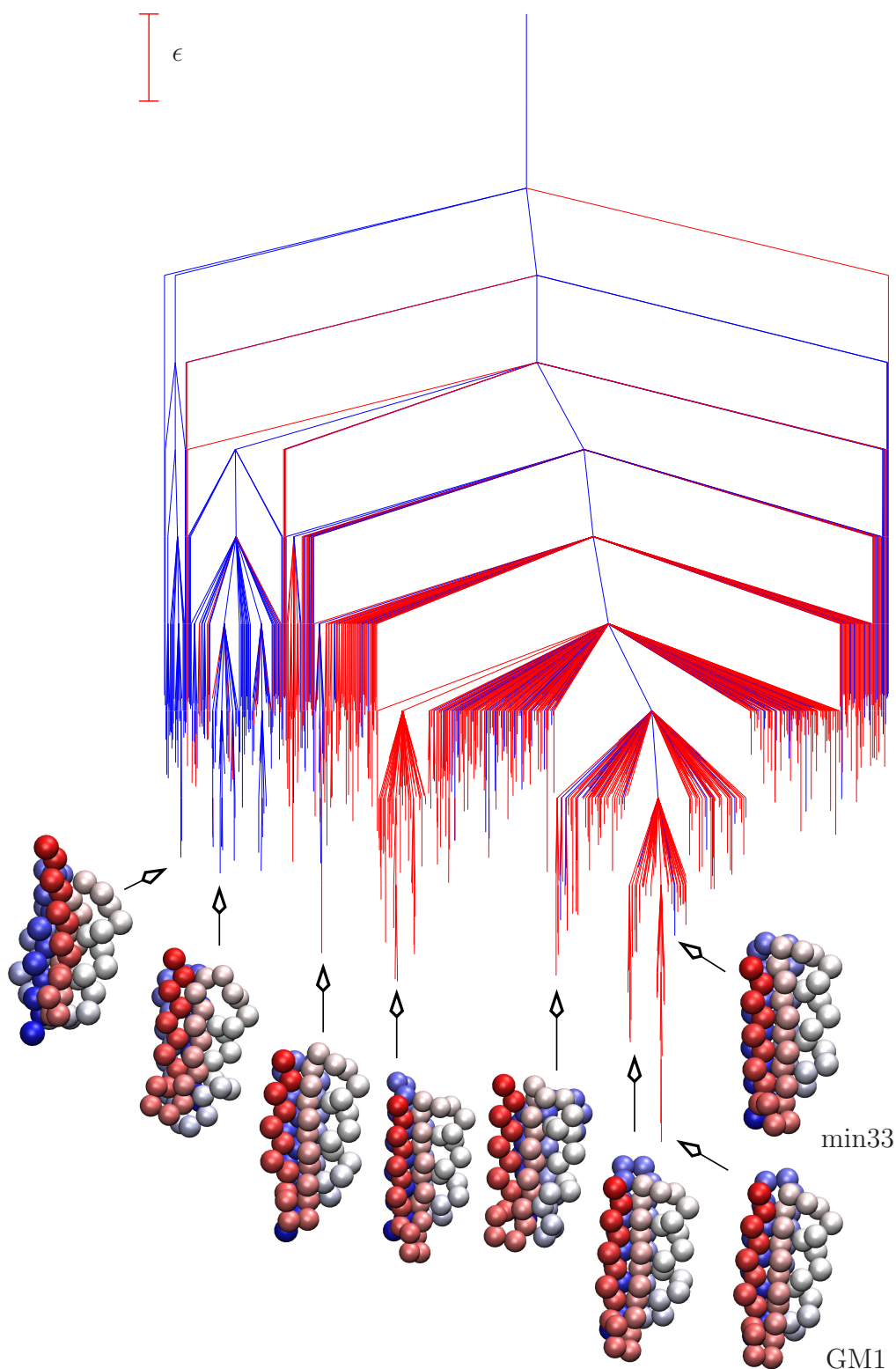


Figure 4: Disconnectivity graph for protein L in the absence of a static force including only the lowest 1000 minima with an energy range of 14ϵ . Selected minima are illustrated using the VMD program¹³⁷ to generate representations coloured from red to blue (N-terminus, β_1 to C-terminus, β_4) according to the position in the chain. Of the six distinct global minima characterised for L/1/56 only GM1 appears in this energy range. The red and blue branches correspond to ‘native’ and ‘unfolded’ structures when classified according to the order parameter description discussed in the text.⁶¹ min33 is the lowest minimum classified as unfolded by this scheme.

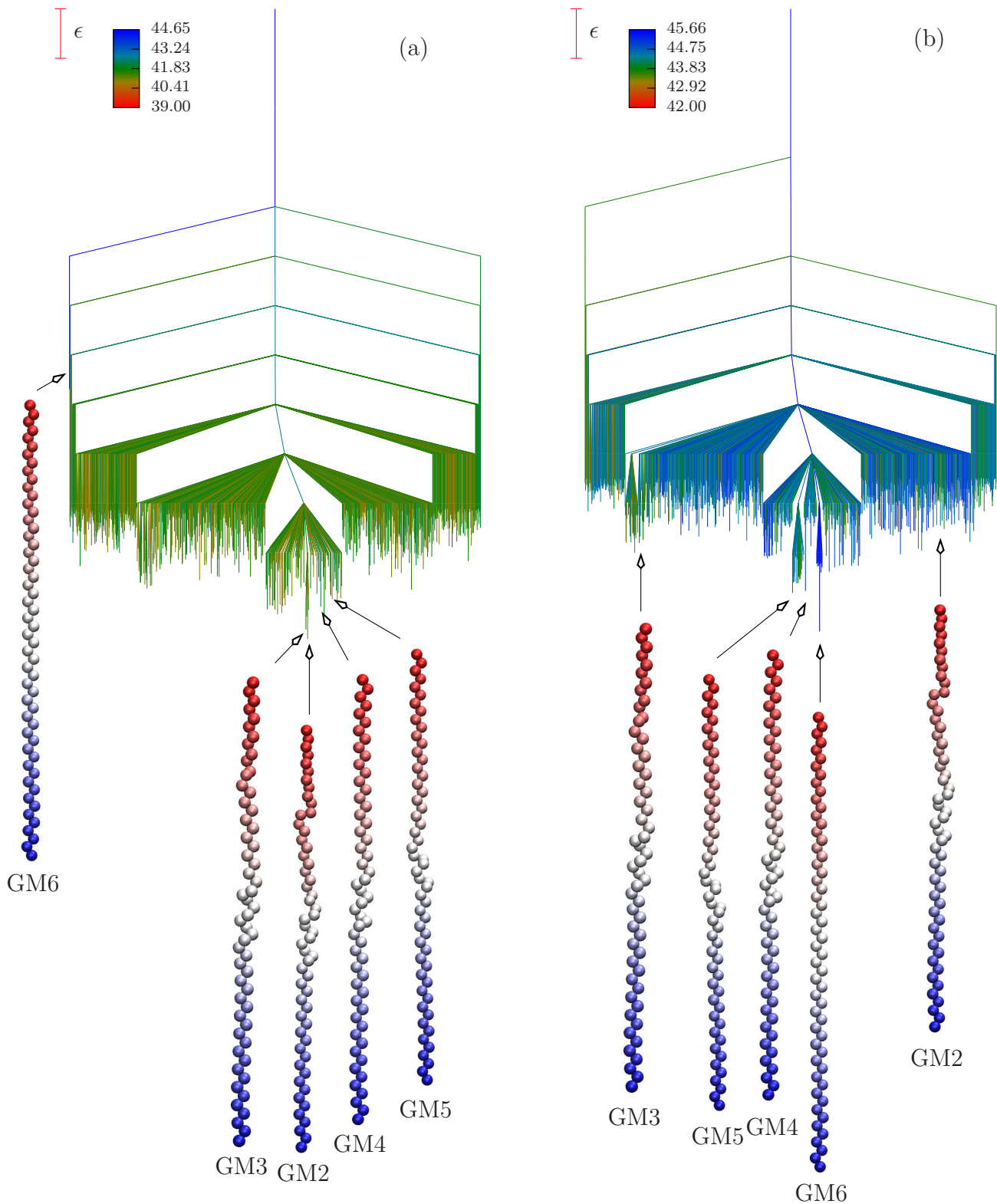


Figure 5: Disconnectivity graphs for (a) L/1/56/2.5 and (b) L/1/56/5.0 including only the lowest 1000 minima with an energy range of 14ϵ . The graphs are coloured according to the distance between residues 1 and 56 in the corresponding minimum, as indicated by the scale bar. Selected minima are illustrated using the VMD program¹³⁷ to generate representations coloured from red to blue (N-terminus, β_1 to C-terminus, β_4) according to the position in the chain. All of the six distinct global minima characterised for L/1/56 appear in this energy range aside from GM1.

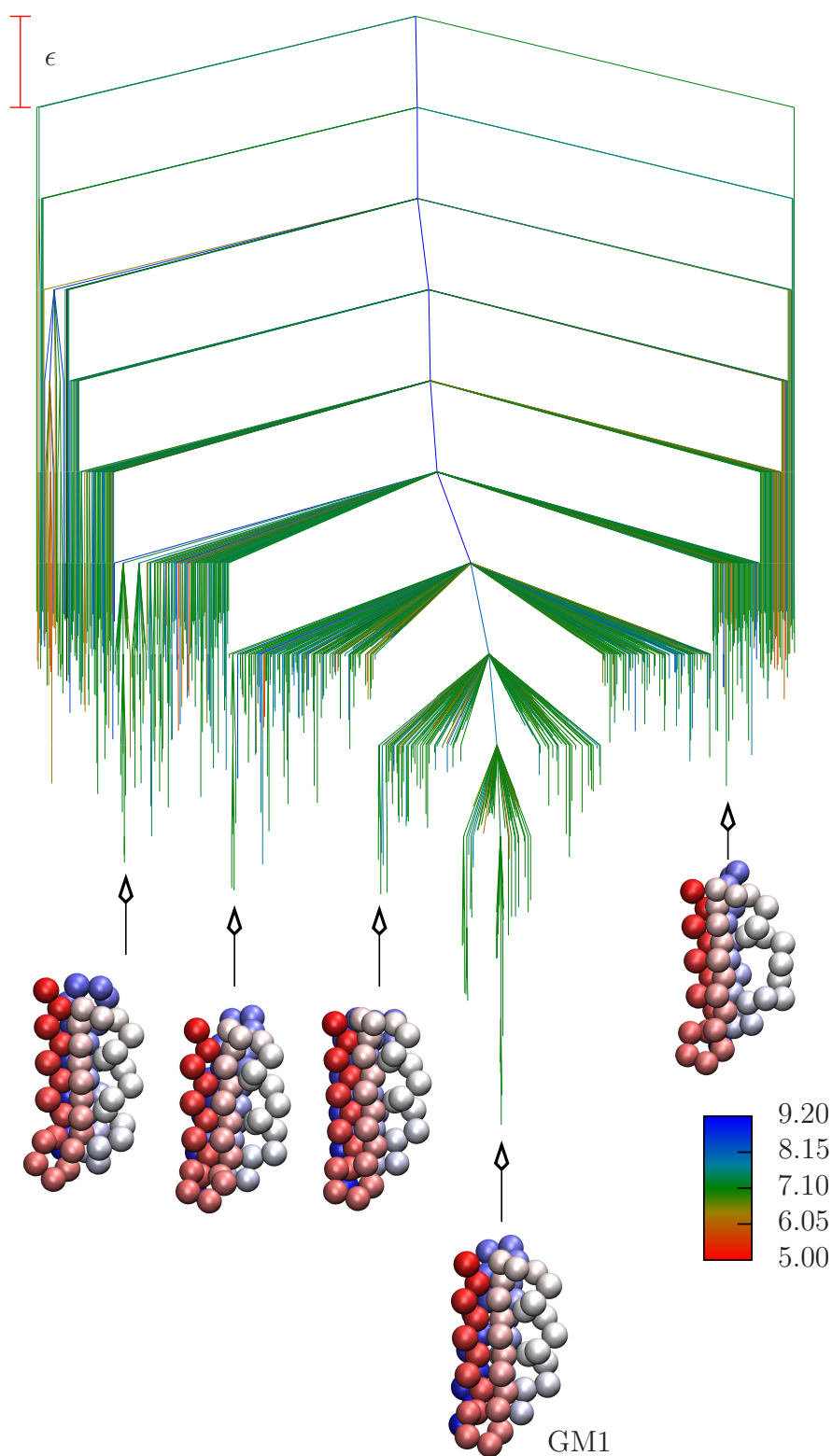


Figure 6: Disconnectivity graph for protein G in the absence of a static force including only the lowest 1000 minima with an energy range of 14ϵ . The graph is coloured according to the distance between residues 1 and 56 in the corresponding minimum, as indicated by the scale bar. Selected minima are illustrated using the VMD program¹³⁷ to generate representations coloured from red to blue (N-terminus, β_1 to C-terminus, β_4) according to the position in the chain. Only one of the six distinct global minima characterised for G/1/56 appears in this energy range.

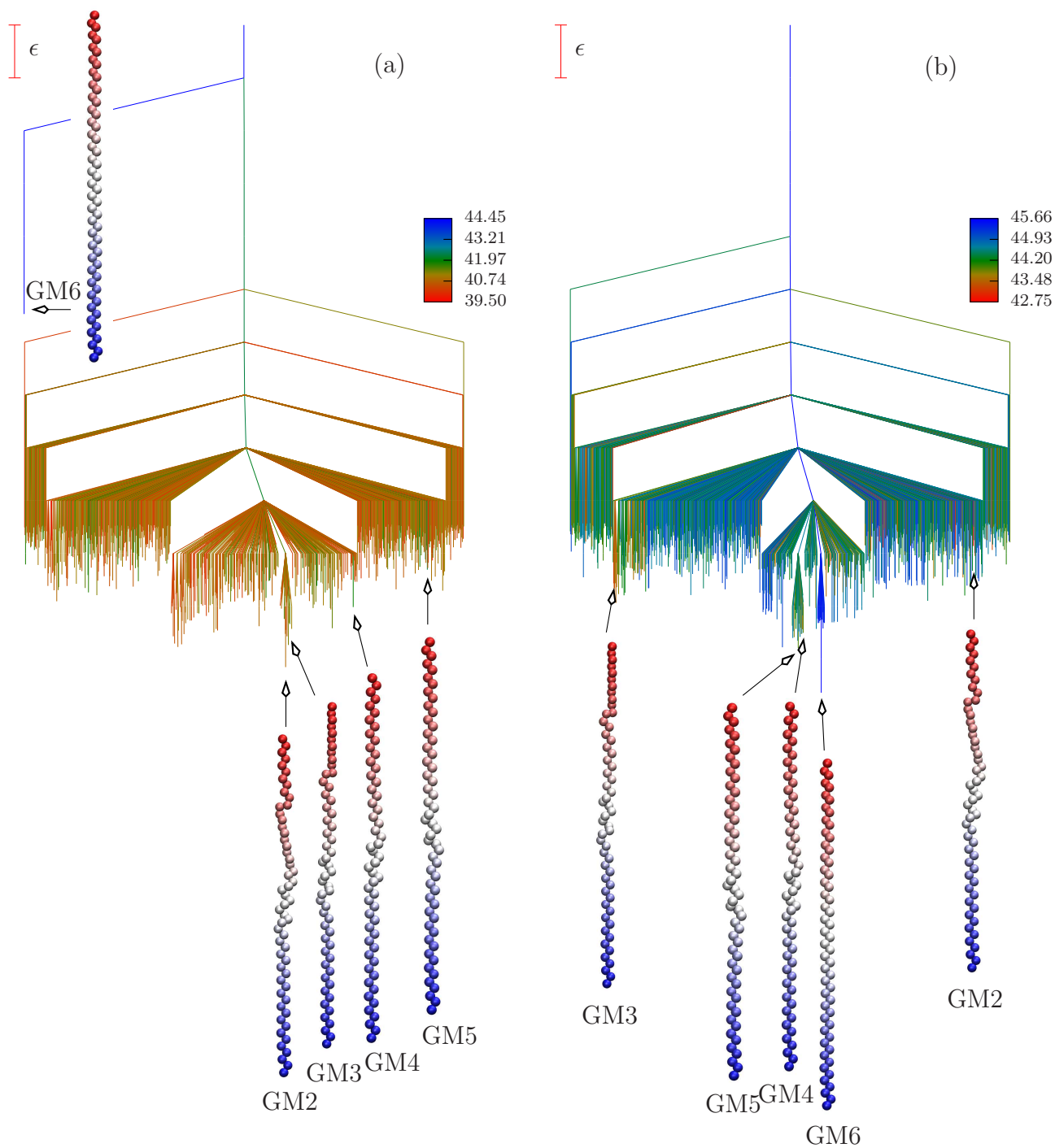


Figure 7: Disconnectivity graphs for a(a) G/1/56/2.0 and (b) G/1/56/5.0 including only the lowest 1000 minima with an energy range of 14ϵ . The graphs are coloured according to the distance between residues 1 and 56 in the corresponding minimum, as indicated by the scale bar. The red and blue colouring is reversed in the two panels because global minimum changes from an extended structure containing remnants of the α helix to a fully extended structure. Selected minima are illustrated using using the VMD program¹³⁷ to generate representations coloured from red to blue (N-terminus, β_1 to C-terminus, β_4) according to the position in the chain. All six of the distinct global minima characterised for G/1/56 appear in this energy range aside from GM1.

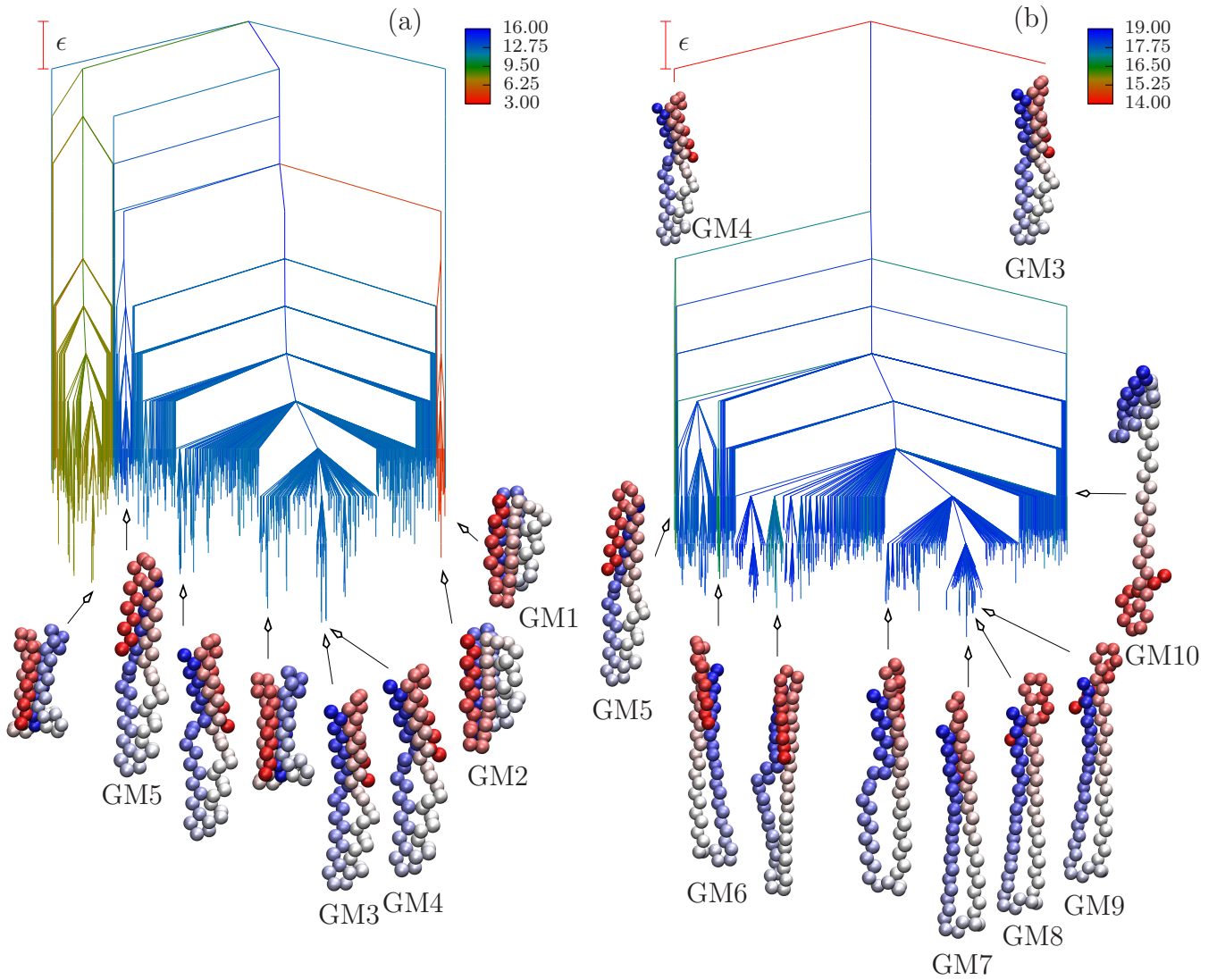


Figure 8: Disconnectivity graphs for (a) $L/10/32/1.5$ and (b) $L/10/32/7.5$ including only the lowest 1000 minima with an energy range of 14ϵ . The graphs are coloured according to the distance between residues 10 and 32 in the corresponding minimum, as indicated by the scale bar. Selected minima are illustrated using the VMD program¹³⁷ to generate representations coloured from red to blue (N-terminus, β_1 to C-terminus, β_4) according to the position in the chain. Five of the ten distinct global minima characterised for $L/10/32$ appear in this energy range for (a) and all ten for part (b) except for GM1 and GM2.

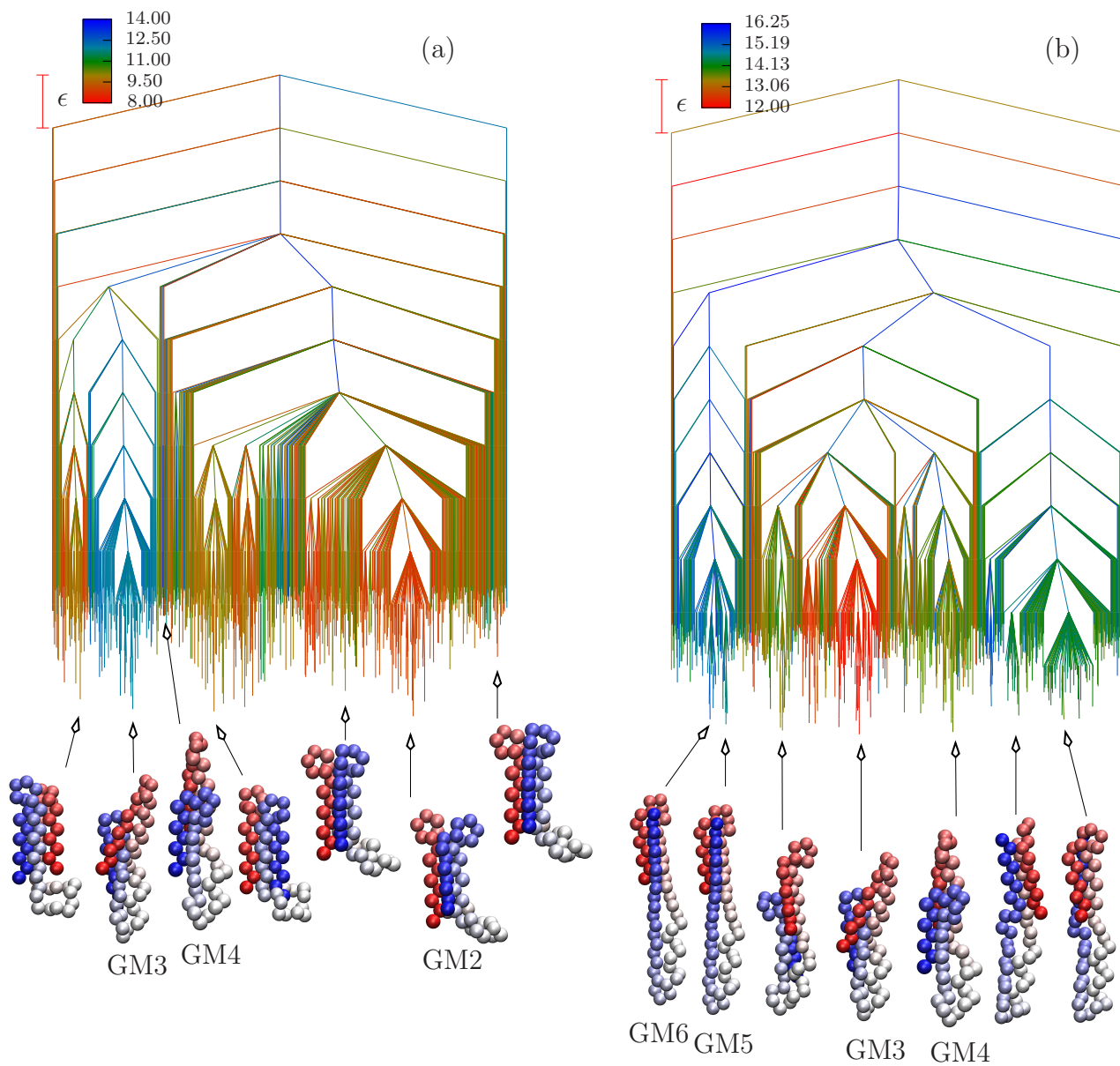


Figure 9

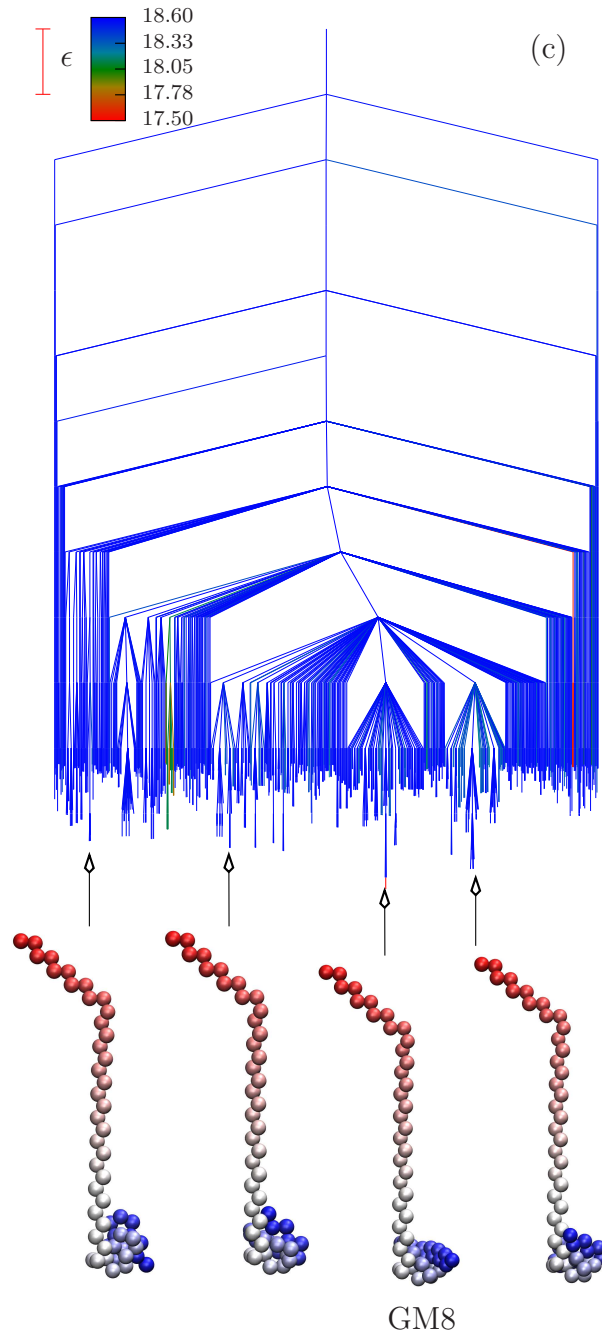


Figure 9: Disconnectivity graphs for (a) G/10/32/2.2, (b) G/10/32/3.3, (c) G/10/32/6.5 including only the lowest 1000 minima with an energy range of 14ϵ . The graphs are coloured according to the distance between residues 10 and 32 in the corresponding minimum, as indicated by the scale bar. Selected minima are illustrated using the VMD program¹³⁷ to generate representations coloured from red to blue (N-terminus, β_1 to C-terminus, β_4) according to the position in the chain. Three of the eight distinct global minima characterised for G/10/32 appear in this energy range for (a), four for (b) and only GM8 for (c).

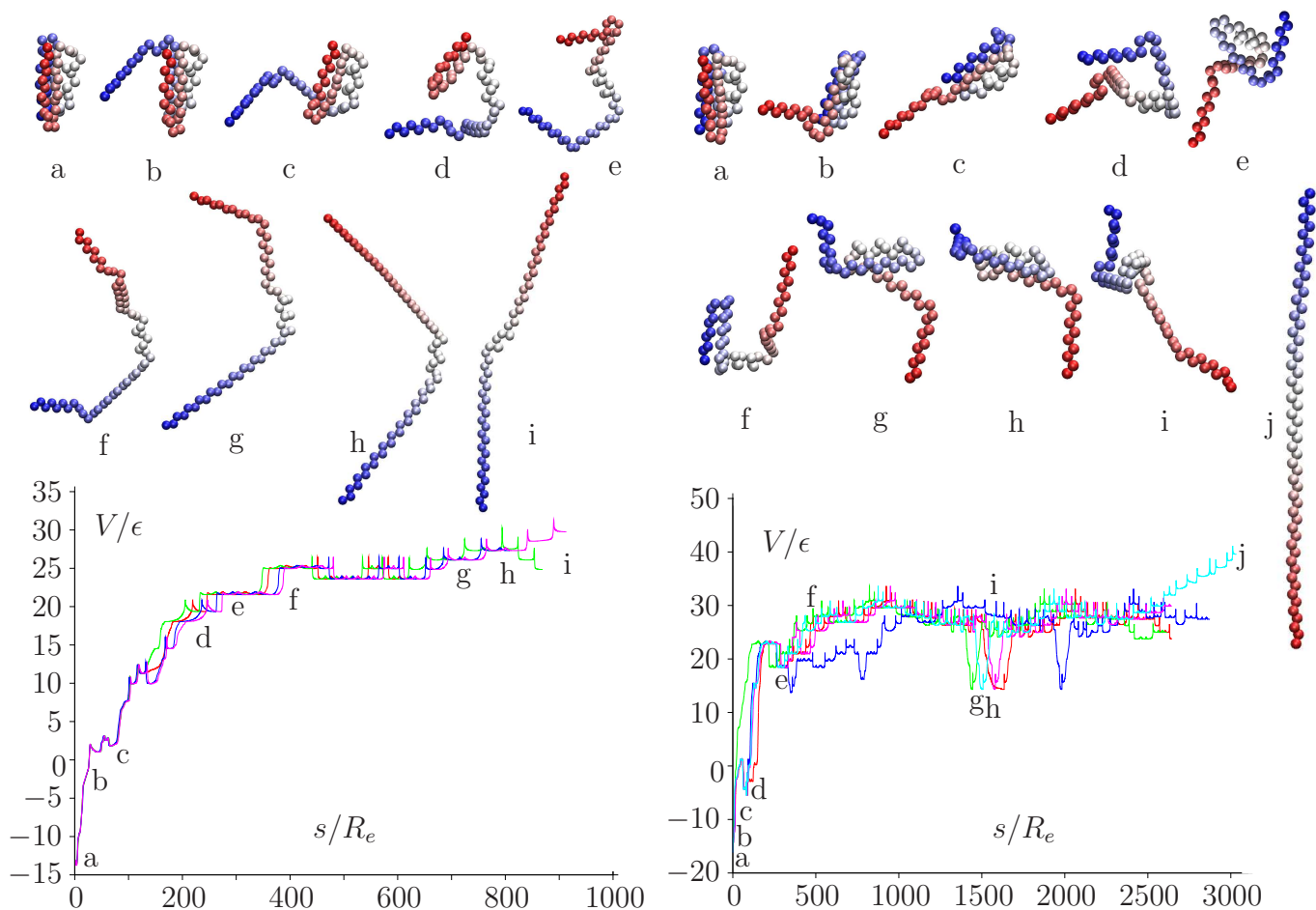


Figure 10: Potential energy, V , as a function of integrated path length, s , for paths connecting GM1 to the other structures that are global minima for $f \neq 0$ for (left) protein L and (right) protein G in the absence of a static pulling force. The overall mechanism is very similar within each set of paths, and the structures illustrated are taken from the longest path (purple for L and light blue for G), which connects GM1 to the minimum obtained by successively relaxing GM5 to $f = 0$. These paths involve 32 local minima linked by 31 transition states for L and 92 local minima linked by 91 transition states for G. Structures a–i and a–j correspond to the indicated positions along this energy profiles; for protein G the intermediate appears at a path length around 1500.

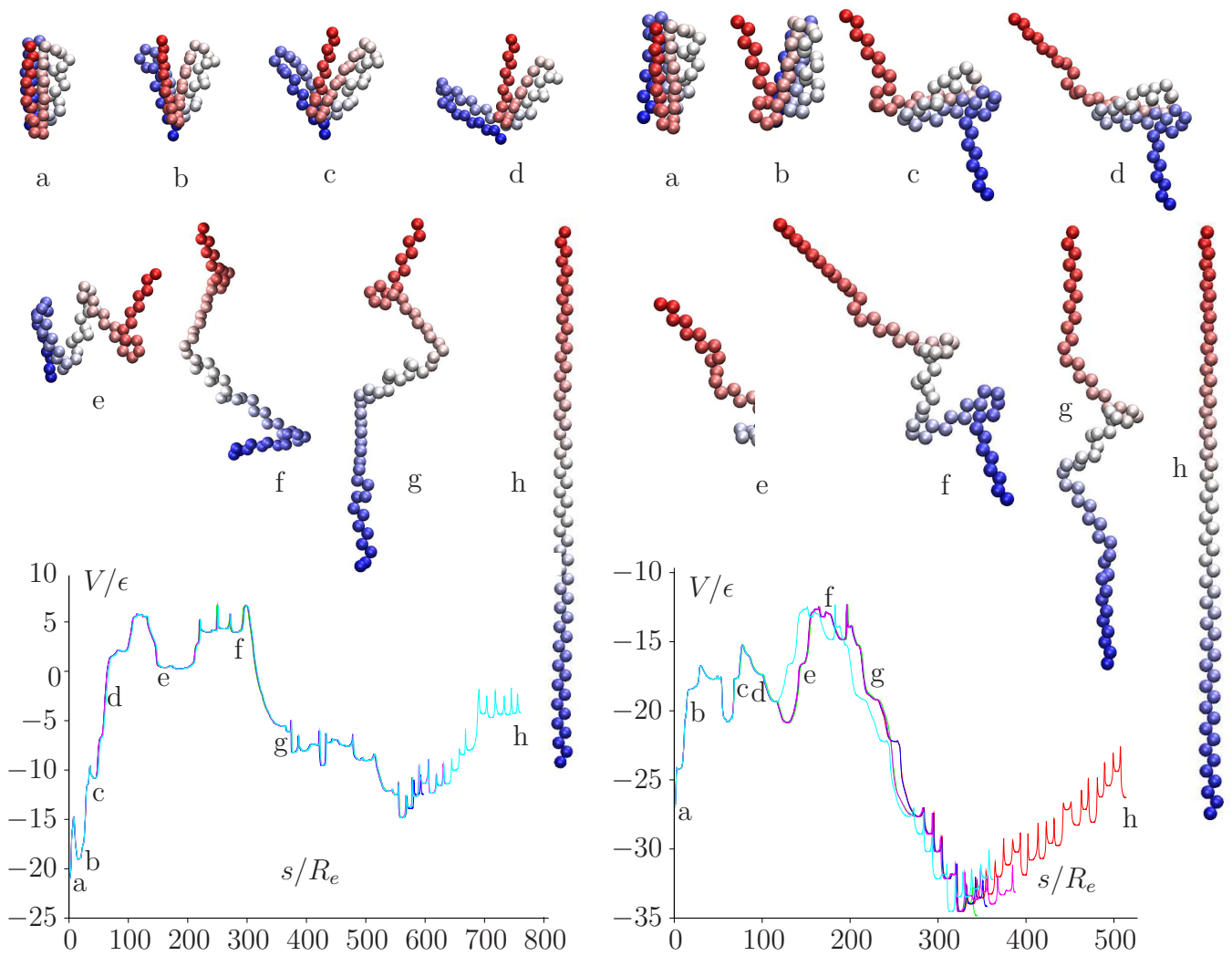


Figure 11: Potential energy, V , as a function of integrated path length, s , for paths connecting GM1 to the other structures that are global minima for $f \neq 0$ in (left) L/1/56/1.0 and (right) G/1/56/1.5. The overall mechanism is essentially the same within each set of paths, and the structures illustrated are taken from the longest path (light blue for L and red for G), which connects GM1 to the minimum obtained by successively relaxing GM6 to $f = 0$. These paths involves 43 local minima linked by 42 transition states for L and 39 local minima linked by 38 transition states for G. Structures a–h correspond to the indicated positions along these potential energy profiles.

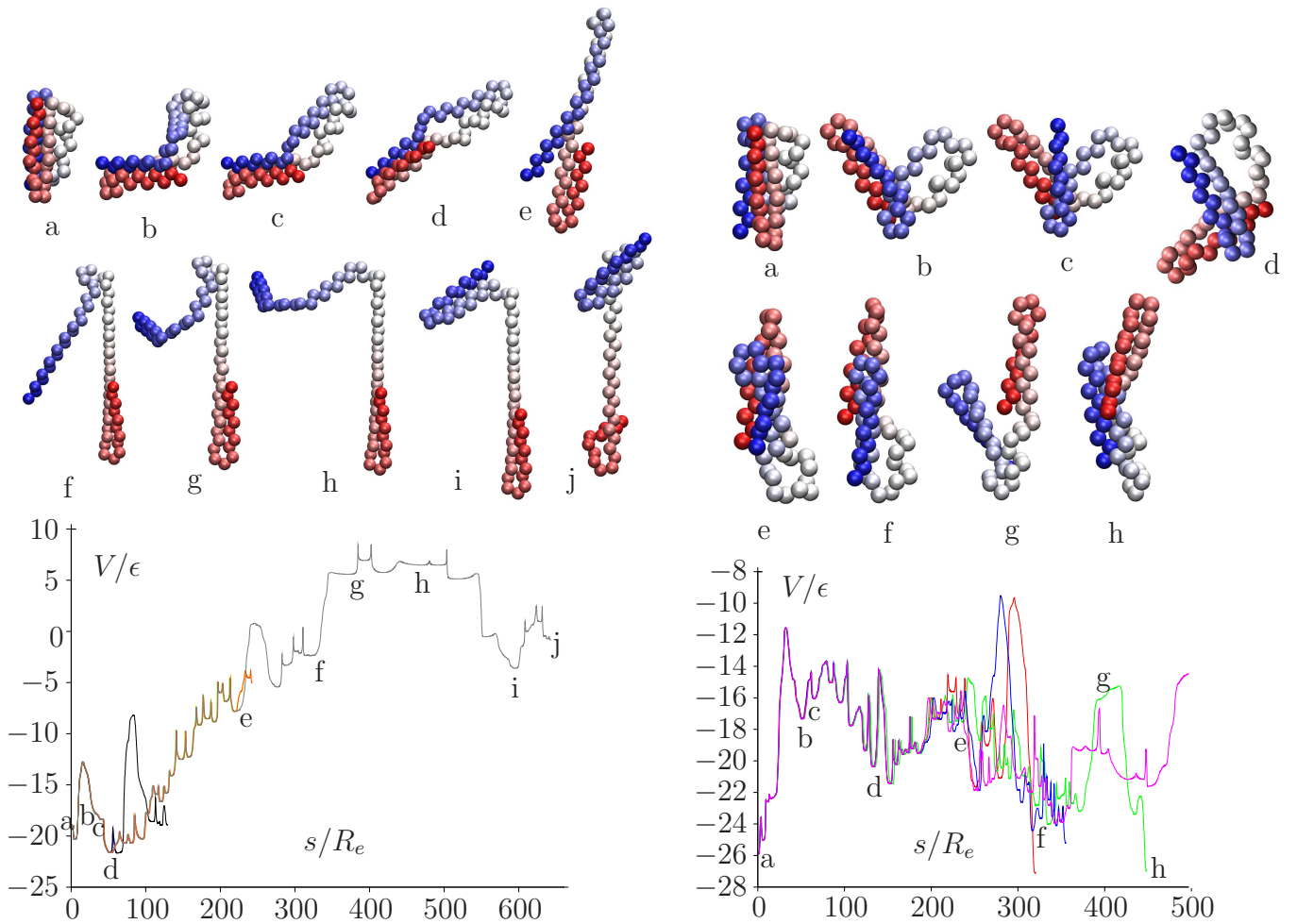


Figure 12: Potential energy, V , as a function of integrated path length, s , for (left) paths connecting GM1 to the other structures that are global minima for $f \neq 0$ in L/10/32/1.5 and (right) for paths connecting GM1 to the minima that correlate with GM2, GM3, GM4 and GM8 in G/10/32/2.2. For protein L the overall mechanism is similar in each case except for the short path that leads to the minimum correlating with GM5. The structures illustrated are taken from the longest path (blue), which connects GM1 to the minimum obtained by successively relaxing GM10 to $f = 1.5$ and involves 40 local minima linked by 39 transition states. For protein G the structures correspond to the path connecting GM1 to the minimum obtained by successively relaxing GM3 to $f = 2.2$ (green), which involves 43 local minima linked by 42 transition states. Structures a-i and a-h correspond to the indicated positions along the corresponding potential energy profiles.

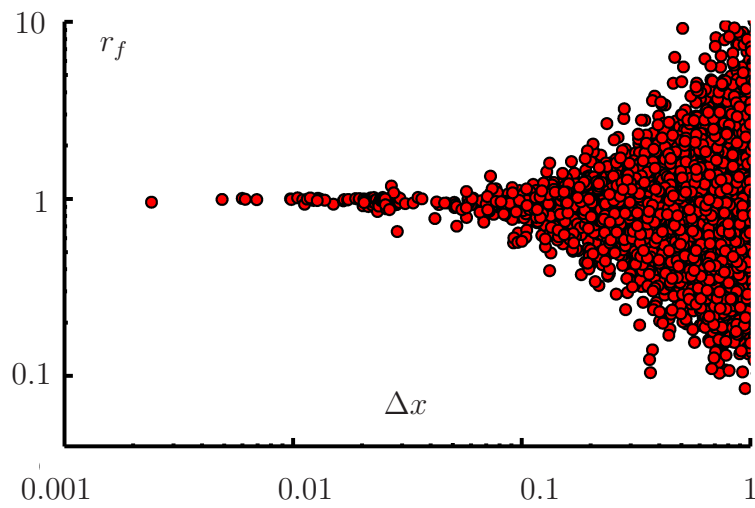


Figure 13: Fold ratio for protein L pathways obtained in the absence of an applied force. r_f is plotted as a function of Δx on log scales. The unit of length for the path length Δx is R_e ; r_f is dimensionless.

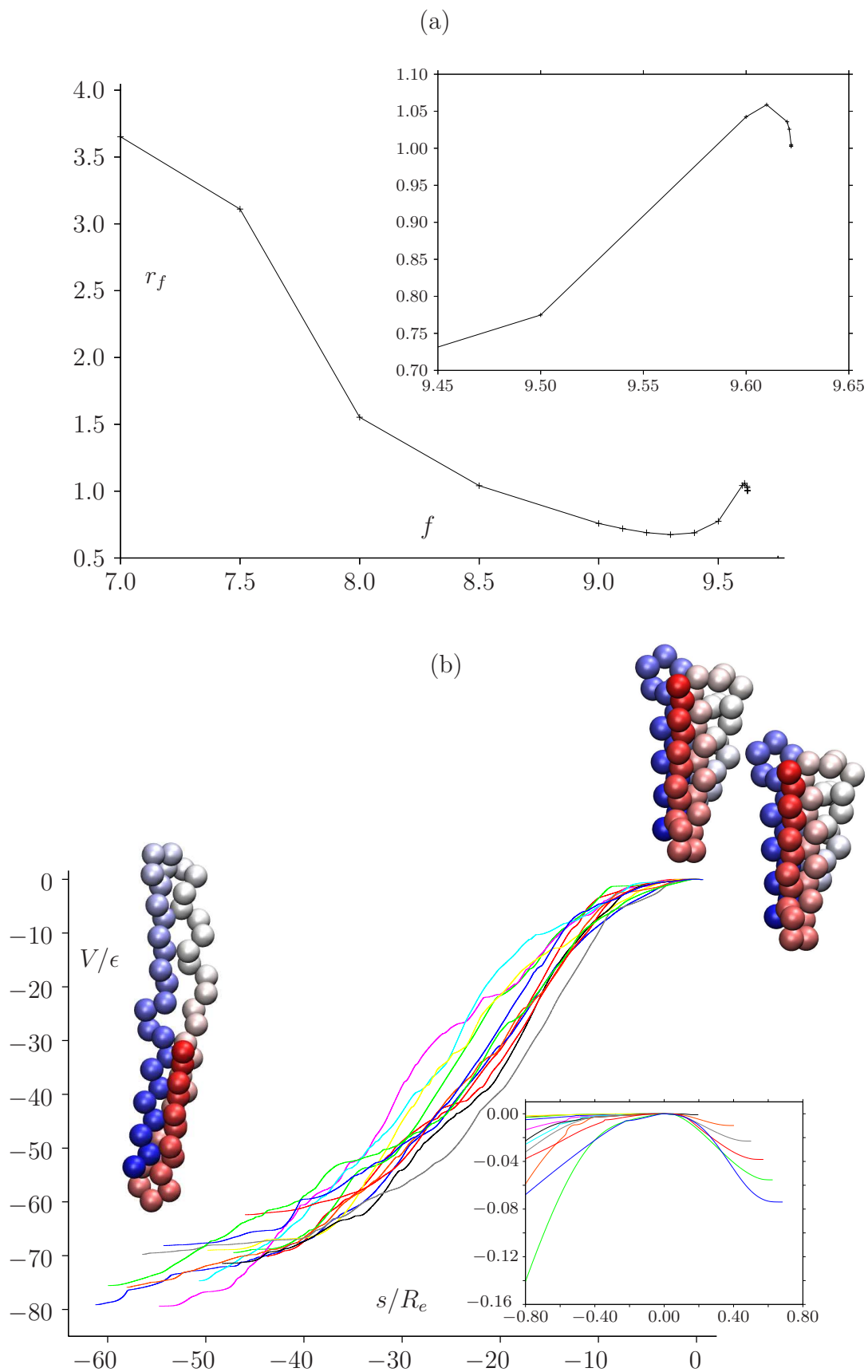


Figure 14: (a) Convergence of the fold ratio to unity for one particular transition state and pathway of L/10/32. (b) Potential energy, V , as a function of integrated path length, s , for one particular transition state at f values of 9.621925, 9.62192, 9.6219, 9.621, 9.62, 9.61, 9.6, 9.5, 9.4, 9.3, 9.2 and 9.1. The stationary point structures illustrated are for $f = 9.1$.

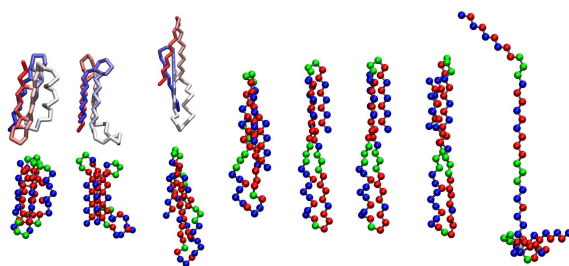


Figure 15: TOC graphic. Structures of the eight distinct global minima identified for G/10/32, visualised using the VMD program.¹³⁷ Hydrophobic beads are coloured red, hydrophilic beads blue, and neutral beads green. The alternative views of GM1 [and GM2 in (c) and (d)] have the beads coloured from red to blue (N-terminus, β_1 to C-terminus, β_4) according to the position in the chain.