Chemistry Europe
European Chemical Societies Publishing

WILEY VCH

Excellence in Chemistry Research

Announcing our new flagship journal

- Gold Open Access
- Publishing charges waived
- Preprints welcome
- Edited by active scientists

Meet the Editors of *ChemistryEurope*

Luisa De Cola
Università degli Studi di Milano Statale, Italy

Ive Hermans
University of Wisconsin-Madison, USA

Ken Tanaka
Tokyo Institute of Technology, Japan

ChemPhysChem

Research Article
doi.org/10.1002/cphc.202300351

Chemistry Europe
European Chemical Societies Publishing

www.chemphyschem.org

The Effect of Pulling and Twisting Forces on Chameleon Sequence Peptides**

James Meadows[a, c] and Konstantin Röder*[b, c]

Chameleon sequences are amino acid sequences found in several distinct configurations in experiment. They challenge our understanding of the link between sequence and structure, and provide insight into structural competition in proteins. Here, we study the energy landscapes for three such sequences, and interrogate how pulling and twisting forces impact the available structural ensembles. Chameleon sequences do not necessarily exhibit multiple structural ensembles on a multi-funnel energy landscape when we consider them in isolation. The application of even small forces leads to drastic changes in the energy landscapes. For pulling forces, we observe transitions from helical to extended structures in a very small span of forces. For twisting forces, the picture is much more complex, and highly dependent on the magnitude and handedness of the applied force as well as the reference angle for the twist. Depending on these parameters, more complex and more simplistic energy landscapes are observed alongside more and less diverse structural ensembles. The impact of even small forces is significant, confirming their likely role in folding events. In addition, small forces exerted by the remaining scaffold of a protein may be sufficient to lead to the adoption of a specific structural ensemble by a chameleon sequence.

Introduction

Proteins fulfil a vast range of functions in living organisms, and understanding this functionality and the regimes in which it fails has been central to chemical and structural biology for the last century. Fundamental to the function of proteins is the three dimensional structure they adopt, which in many cases is determined through the underlying amino acid sequence. Anfinsen postulated that the adaptation of a stable structure is based on thermodynamics, although how a protein is folded remained unknown.[1,2] A physical explanation of protein folding was provided through the principle of minimal frustration[3] and the concept of a folding funnel on the potential energy landscape.[4–6] Recently, advances in machine-learning techniques have allowed the development of methods to predict with reasonable confidence the structure of proteins from sequence.[7,8]

Importantly, while these breakthroughs have started a new era of structural biology, there are many factors that impact structure beyond the sequence. Common examples include the interactions with other molecules, from small co-factors to large protein complexes, post-translational modifications, for example glycosylation, as well as environmental conditions.

Another complication arises from a variety of proteins for which sequence and structure are not uniquely linked. The first group of these proteins includes metamorphic and transformer proteins.[9,10] Here, multiple stable structures are observed, and these structures can readily be adopted to fulfil different functions. This group of proteins still follows the principle of minimal frustration.[11] A second large group, which challenges even the most recent structure prediction algorithms, are intrinsically disordered peptides and proteins.[12,13] These molecules exhibit a range of structures without a preferred fold. This property can arise from a number of competing stable structures of similar energy or due to the absence of stable structures.[14] External conditions may lead to a folded structure.

A final group of interest are chameleon sequences.[15–18] These sequences are observed in multiple configurations in proteins, and exhibit structural ambivalence, i.e. no clear preference for either of their exhibited structure is observed. Importantly, these are only small segments of larger proteins, which distinguishes them from metamorphic and transformer proteins and their close relatives switch peptides.

A detailed review of these groups of proteins may be found elsewhere.[19] Here, we focus on chameleon sequences and their structural properties.

Due to their structural properties, these sequences have been used to evaluate the performance of sequence-based structure predictors.[20] Furthermore, chameleon sequences have also been implicated in misfolding diseases.[20,21]

[a] J. Meadows
Department of Chemistry, Durham University
Stockton Road, Durham DH1 3LE, UK

[b] Dr. K. Röder
Randall Centre for Cell & Molecular Biophysics
King's College London
Guy's Campus, Great Maze Pond, London SE1 1UL, UK
E-mail: konstantin.roeder@kcl.ac.uk

[c] J. Meadows, Dr. K. Röder
Previous affiliation: Yusuf Hamied Department of Chemistry,
University of Cambridge
Lensfield Road, Cambridge CB2 1EW, UK

[**] A previous version of this manuscript has been deposited on a preprint server (DOI: https://doi.org/10.26434/chemrxiv-2022-pr7wg).

Supporting Information for this article is available on the WWW under https://doi.org/10.1002/cphc.202300351

Previous work has revealed some reasons for this structural ambiguity. Early computational work highlights the inherent stabilisation of two distinct structural propensities on the energy landscape.[22] More recently, it was shown that the energy landscapes for such proteins may exhibit a large number of funnels, stabilising distinct structural motifs of comparable energies.[23] Within larger proteins, these sequences only form weak interactions with the surrounding scaffold, resulting in no clear preference for a specific structure.[24] This property is reflected in a limited number of amino acids found in such regions.[24,25] In addition, the sequences show a higher residue mobility than their non-ambiguous counterparts, leading to more likely fluctuations.[26] These observations highlight that ultimately the structure depends on the environment.[25]

A key environmental factor is the surrounding protein, often described in terms of flanking residues.[24,25] Not only are the chameleon sequences themselves affected by the surrounding, but their increased mobility also affects the surrounding protein, as these mobility islands may seed thermal denaturation. The flanking residues can also impact structural selection of the chameleon sequence itself.[27] Structural propensities of these flanking residues change the structure of the chameleon sequence and its aggregation properties. In part, this may be related to changes in flexibility that have been shown by NMR experiment to impact structural propensities of the chameleon sequence.[28]

Given the link between the flanking residues and the adopted local conformation of the chameleon sequence, it raises the question how this structural modulation occurs. As the number of contacts with the remaining scaffold is low,[24] it is unlikely that specific interactions are the root cause underlying structural propensities. The observation that different structural propensities for helices or sheets in the flanking residues impact structure[27] opens another possibility. Different structures adopt different backbone orientations, meaning that the structural propensities of the flanking residues may exert a force on the chameleon sequence. Such a force, which may be a pulling or twisting force, could then impact the structure of the chameleon sequence. The nature of this force is likely to vary based on the structure adopted by the protein overall.

Here, we explore this effect by considering how twisting and pulling forces affect the potential energy landscape of three chameleon sequences. The exploration of the energy landscape allows for a description of all structural ensembles,[29,30] and has been previously used to investigate the effect of forces on the folding landscape.[31] Not only has it been shown that forces have significant impact on the energy landscape,[31] but furthermore a switching in structural propensities due to applied force for a simplified landscape has been reported.[32] Previous studies of applying forces to unfold proteins focus on identifying unfolding pathways and characterising the folding landscape,[33,34] but less attention has been paid to smaller regions and their rearrangements rather than complete unfolding. Twisting forces have not been studied in this context either. Recent experiments on larger proteins highlight the potential of understanding such changes not only in the context of protein folding, but also as part of designing new biosensors.[35]

Results

This work leads to three sets of results, which we will consider in turn. First of all, we explored the energy landscapes of the chameleon sequences under zero force, which gives us a better idea what structures are more readily adopted by chameleon sequences. Second, we consider the impact of pulling forces on the energy landscapes. Finally, the results for applying twisting forces are presented.

The applied pulling forces are static, and as a result the situation described in this work corresponds to equilibrium shifts based on such a static force. Including the force as an additional term in the potential energy, it is possible to explore the energy landscape for a static force. This approach changes the topography of the potential energy landscape.[31] In contrast to other simulations where structures are dynamically evolved (e.g., molecular dynamics), the energy landscape explorations enables us to apply such static forces and maintain an equilibrium condition, albeit with an applied force. This approach is reasonable given that forces from scaffold orientation are effectively static themselves. The same equilibrium setup is achieved for the twisting forces by twisting towards a preferred dihedral angle. More details are given in the methods section and the Supporting Information.

Energy Landscapes for the Chameleon Sequences with No Applied Force

The disconnectivity graphs[37,38] for the potential energy landscapes of the three sequences without any applied force are shown in Figure 1. Somewhat surprisingly, the energy landscapes for sequence A and sequence C are single funnelled, and only show stabilisation for a single structural ensemble. For sequence A, the low energy structures are fully formed helical structures involving all residues (see Figure 1(a)). In sequence C, only a single helical turn is present, with the structural ensembles showing some extended character in the terminal residues of the sequence (see Figure 1(c)). The only system exhibiting a multifunnel energy landscape is sequence B (see Figure 1(b)). The sequence exhibits both, different helical configurations and more extended structures. A more diverse landscape is observed for both sequence B and C when we consider repeat sequences, i.e. two or three copies of the chameleon sequence in a row. Such diversity might be of interest to protein engineering, and we present some data for such landscapes in the Supporting Information (section S2, Figure S2). Due to computational cost, in the following sections, we only consider the chameleon sequences and not repeated sequences.

The potential of these energy landscapes to encode structural switches may be explored using the heat capacity curve for the individual systems. As sequence A and C do not

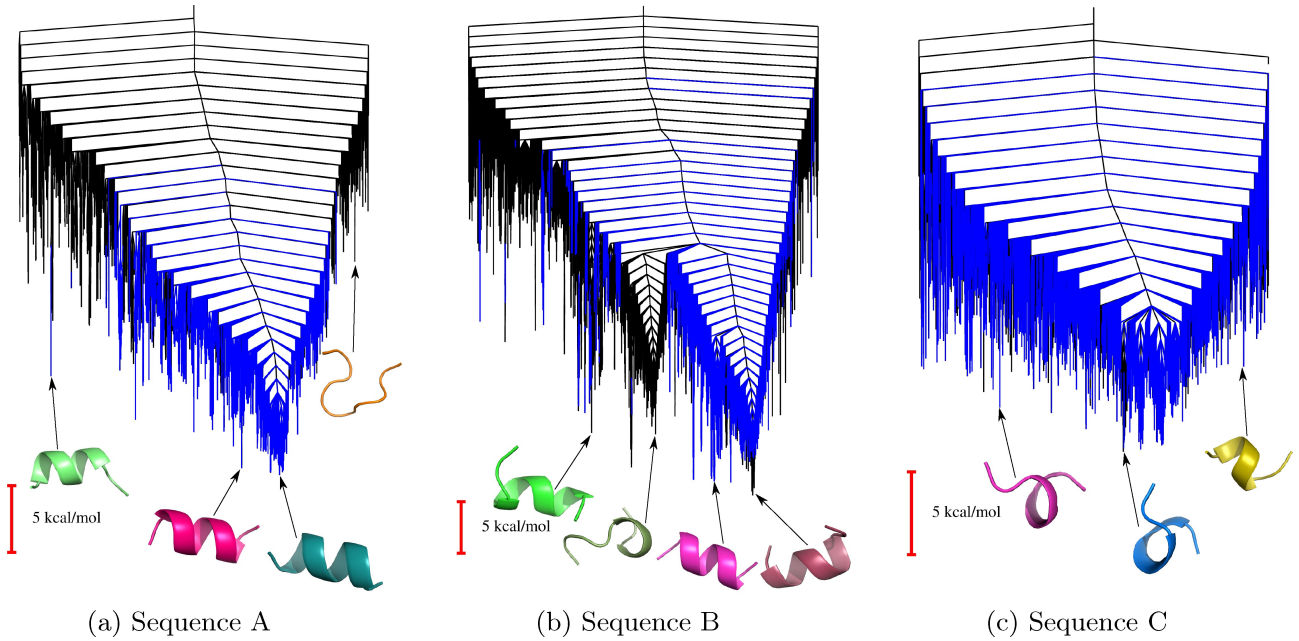


Figure 1. The unperturbed potential energy landscapes for three chameleon sequences. Lines coloured blue represent molecular structures with more than half of their residues determined to be helical by the DSSP algorithm.[36] Competing funnels are observed in the energy landscape of sequence B.

show competing funnels, it is unlikely they exhibit switching behaviour, and therefore we consider only sequence B for the moment. In Figure 2 the heat capacity curve for sequence B is shown alongside coloured disconnectivity graph representing the changes in occupation probability associated with specific features. The link between occupation probability of minima and heat capacity allows us to analyse which structural transitions are corresponding to phase transitions observed.[39] Generally, a switchable system will exhibit a peak in the heat capacity close to the relevant temperature, and shifts in the associated transition temperatures or the absence of peaks can be linked to experimental observations.[40] Here, as we analyse a biological system and our relevant energy scale (ϵ) is 1 kcal mol^{-1}, we would expect features at and below 0.6 to be relevant for the system's behaviour, with features at 1.0 or higher likely corresponding to melting peaks. For sequence B with no applied force, we only see a very weak feature at 0.5 corresponding to slight uncoiling of the helical structure (see

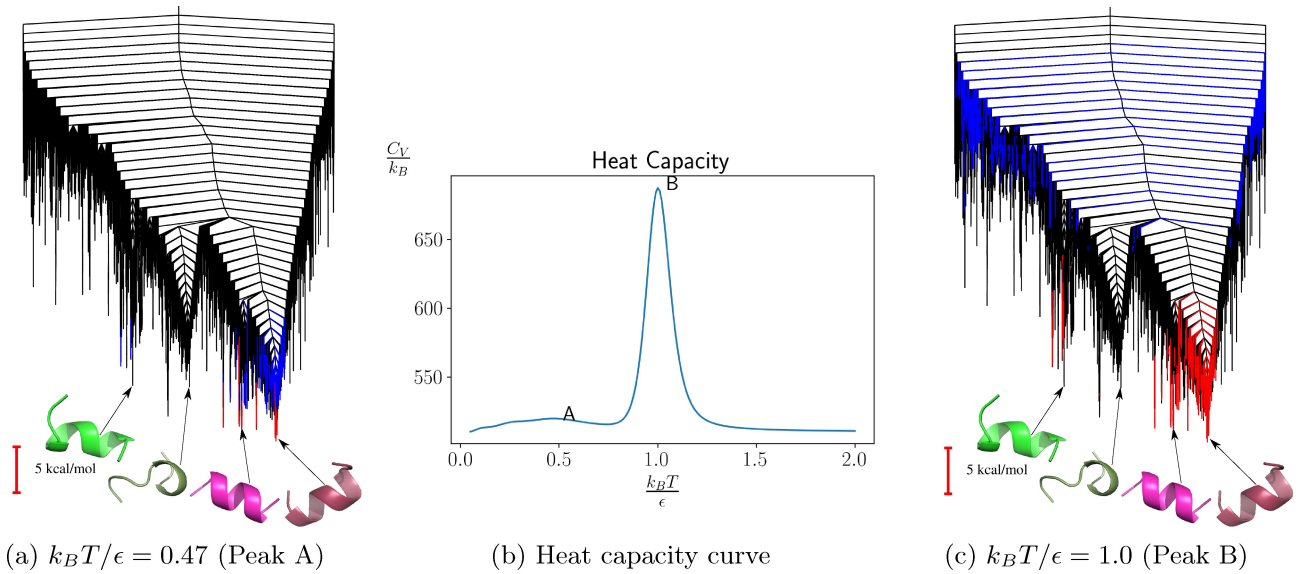


Figure 2. The heat capacity curve for sequence B with no applied force. Minima coloured in the disconnectivity graphs contribute to 99% of the total heat capacity of the system at their respective temperatures. Minima with negative and positive occupation probability gradients are coloured red and blue, respectively.

Figure 2(a)) and a high energy peak (see Figure 2(c)). From these observations, we do not expect to see non-helical structures for this isolated sequence.

Large Changes of the Energy Landscapes from Small Pulling Forces

The application of pulling forces is known to affect energy landscapes significantly.[31,32] However, the required forces may differ significantly, with non-structural proteins showing significantly altered energy landscapes when some tens of pN are applied as pulling force,[31] while similar changes for structural proteins can go up to thousands of pN.[41] As reference, 1 kcal mol^{-1} Å^{-1} is roughly 69.5 pN, and forces of such magnitude are sufficient to introduce significant changes for the chameleon sequences. In Figure 3, the progressive changes to the energy landscape are shown for sequence A, as the force is increased step-wise. Already at 2 kcal mol^{-1} Å^{-1} (139 pN), the dominant structural ensemble consists of extended configurations, with only some helical content observed in higher energy structures (Figure 3(b)). At 4 kcal mol^{-1} Å^{-1} (278 pN), we exclusively observe extended configurations (see Figure 3(c)). A similar trend is observed for the repeated sequences (see Supporting Information S2, Figure S3).

Such a progression is not unexpected, and an analysis of the properties of all structures allows us better insight into the relevant force regimes. Two properties are representing good proxies for the structural changes observed. First, the helical content, as determined by DSSP,[36] characterises the helical structures well, in particular for sequence A and B. Secondly, the end-to-end distance gives a good indication on whether the structures are fully extended. Measuring these properties across all structures on a given landscape and applying thermal averaging allows a characterisation of the transition introduced by the forces. The thermal averaging here uses Boltzmann weights for each minimum. We observe a sharp drop off of the helical content between 1 and 2 kcal mol^{-1} Å^{-1}, the same regime in which we see a strong increase in the end-to-end distance (see Supporting Information S3, Figure S4).

Focusing on this force regime, further analysis of the energy landscape is possible, similar to the analysis presented for the energy landscape in absence of the pulling force shown in Figure 2. In contrast to the system in absence of pulling forces, the heat capacity curve for sequence B at a static pulling force of 1 kcal mol^{-1} Å^{-1} shows two low temperature features (see Figure 4(b)). In addition, lowest energy minima for extended and helical configurations are comparable in energy. Clearly, at this pulling force competition between structural ensembles exists and switching between them is a distinct possibility. Further evidence for this switching is seen in the frustration of the energy landscape as calculated with the frustration index.[42] The frustration is low for small and high forces, i.e., where a single structure is dominant, but increases for the medium forces where we observe competition between structures (see SI Figure S5). Experimental observations of the end to end distance of this peptide sequence under a small force should observe relatively rapid switching between a shorter and a longer form of the peptide. Importantly, given the difference between this energy landscape and the one in absence of a pulling force, once the force is removed we would expect rapid relaxation towards the original helical ensembles.

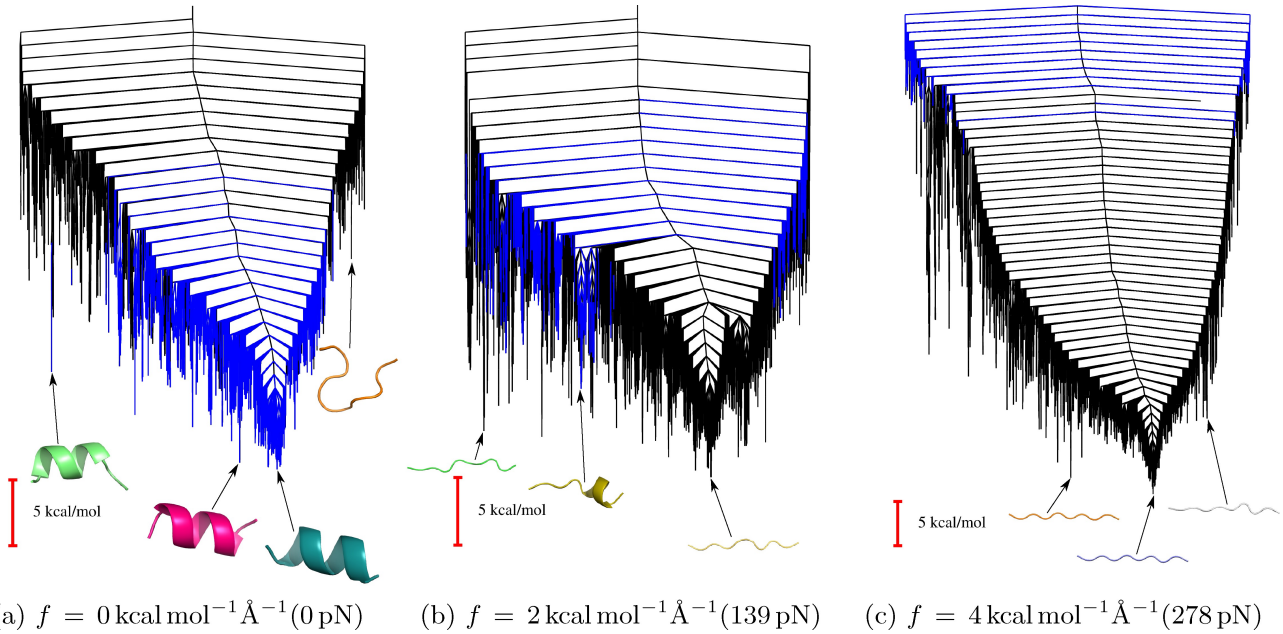
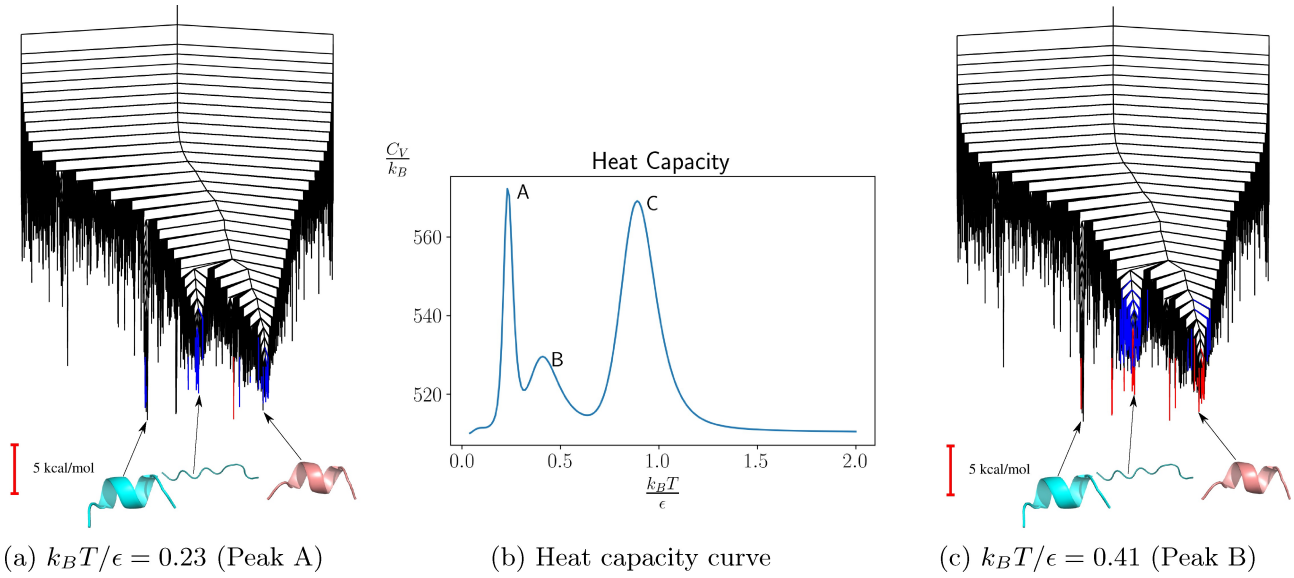


Figure 3. The potential energy landscapes for sequence A with a static pulling force of magnitude *f* applied to the ends of the carbon backbone. With a force of 4 kcal mol^{-1} Å^{-1} strand-like structures dominate the low-energy minima.



(a) $k_BT/\epsilon = 0.23$ (Peak A) (b) Heat capacity curve (c) $k_BT/\epsilon = 0.41$ (Peak B)

Figure 4. The heat capacity curve for sequence B with an applied force of 1 kcal mol^{-1} Å^{-1} (69.5 pN). Minima coloured in the disconnectivity graphs contribute to 99% of the total heat capacity of the system at their respective temperatures. Minima with negative and positive occupation probability gradients are coloured red and blue, respectively.

Diverse Structural Ensembles from Small Twisting Forces

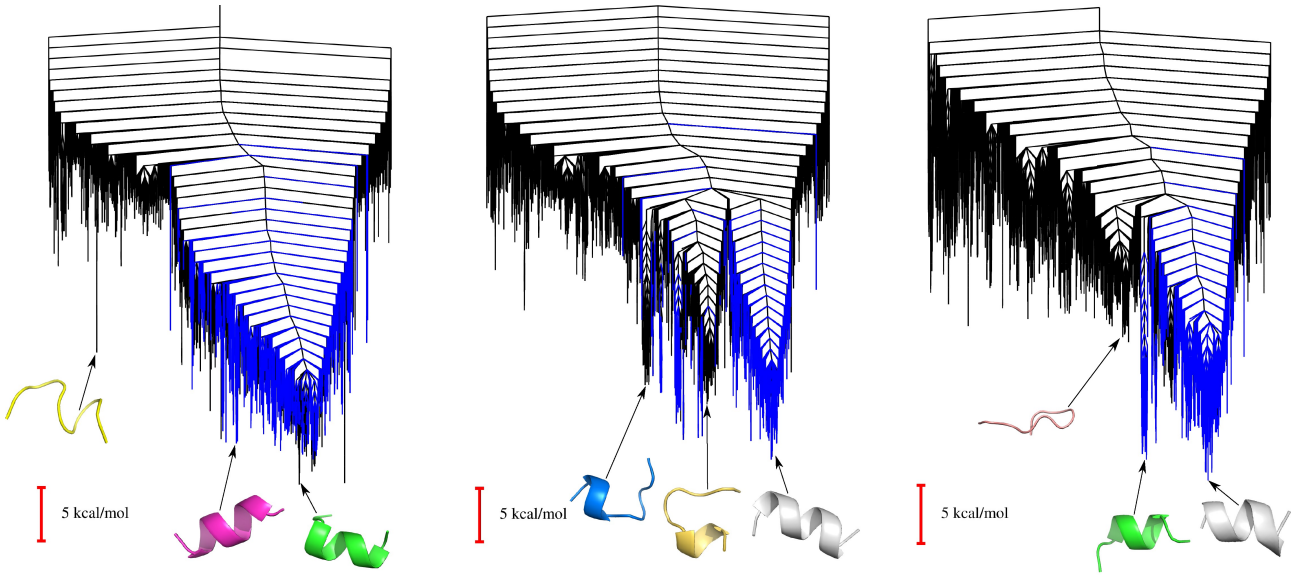
For the application of pulling forces, two parameters are required. Firstly, we need to determine the appropriate twisting force to be applied. We find that the twisting forces can be quite large in magnitude with respect to typical energy scales of biomolecules. Up to 20 kcal mol^{-1} rad^{-1} (139 pN·nm) of applied force does not result in loss of structure on the energy landscape. This observation likely stems from the fact that the force is counteracted by all dihedral angle preferences in the peptides, while the pulling forces are compensated by weaker non-covalent interactions. The more subtle point about the force is that it is signed. As a result, it is necessary to consider both, positive and negative twisting forces, which correspond to the different handedness of the applied twist.

The second parameter is the reference angle for the twist. The use of a reference in this way aims to reproduce the experimental setup for single molecules, and also is the situation found in the actual protein, i.e. the protein adopts a fixed configuration around the chameleon sequence, resulting in a force. This force, however, is not constant but depends on the relative configuration, i.e. how well the angle implied by the flanking structure matches the angle adopted by the chameleon sequence.

First, a progression of applied forces was consider with magnitudes of 1 and −1 kcal mol^{-1} rad^{-1} (±6.95 pN·nm), 5 and −5 kcal mol^{-1} rad^{-1} (±34.75 pN·nm), and 20 and −20 kcal mol^{-1} rad^{-1} (±139.0 pN·nm) with the reference chosen as the global minimum for the unperturbed energy landscape for sequence B. The key question is whether the multifunnel nature disappears as these forces are applied forcing the system more towards helical configurations. Figure 5 shows the three applied forces, and Figure 6 shows their negative counterparts. A number of observations are readily made. While larger forces do not remove all structure and still allow for alternative structures at higher energies (see Figure 5(c) and Figure 6(c)), which stands in contrast to the higher pulling forces (for example Figure 3(c)), a strong preference for helical structures emerges, although different helical architectures are supported. Interestingly, a similar energy landscape is observed for the smallest positive twisting force (see Figure 5(a)), but not for the corresponding negative force (see Figure 6(a)). Clearly, small twisting forces, depending on its handedness, can lead to increased structural diversity. The largest structural diversity is observed for intermediate forces. In all scenarios, helical structures are preserved as some of the lowest energy structures, but additional partly uncoiled structures are observed as well. Similar to the pulling forces, once these twisting forces are removed, we would expect relaxation to the unperturbed structural ensemble, and the non-helical features will not persist, unless stabilised in a different way.

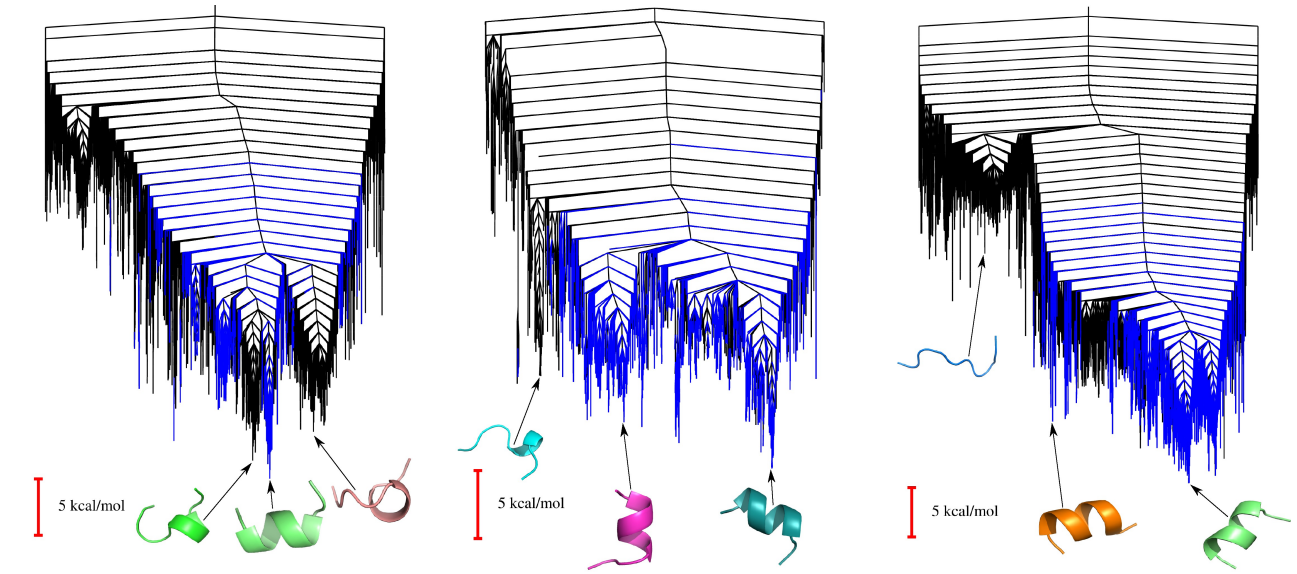
Strong Impact from Different Twist Reference Angles

Fixing the force at 20 kcal mol^{-1} rad^{-1} (−139.0 pN·nm), a survey of energy landscapes was undertaken, where we looked at 30° intervals of the reference dihedral angles for sequence B, i.e. values for ϕ_r of −150°, −120° … −30°, 0°, 30° … 150°, 180°. Within this series we observe no clear trends, and changes to the energy landscape seem unpredictable. The energy landscapes include single funnel energy landscapes, multifunnel energy landscapes, where competing funnels are high in energy compared to the global minimum, and true multifunnel energy landscapes with competition between ensembles. All of these effects can be seen in Figure 7, where we present three of twelve energy landscapes surveyed for this part of the study. These changes highlight the sensitivity of chameleon sequences



(a) $1\,\mathrm{kcal\,mol^{-1}rad^{-1}}$ (6.95 pN·nm) (b) $5\,\mathrm{kcal\,mol^{-1}rad^{-1}}$ (34.75 pN·nm) (c) $20\,\mathrm{kcal\,mol^{-1}rad^{-1}}$ (139.0 pN·nm)

Figure 5. The disconnectivity graphs for sequence B with a fixed reference dihedral angle $\phi_r = -1.083$ rad and increasing magnitude of positive twisting force.



(a) $-1\,\mathrm{kcal\,mol^{-1}rad^{-1}}$ (-6.95 pN·nm) (b) $-5\,\mathrm{kcal\,mol^{-1}rad^{-1}}$ (-34.75 pN·nm) (c) $-20\,\mathrm{kcal\,mol^{-1}rad^{-1}}$ (-139.0 pN·nm)

Figure 6. The disconnectivity graphs for sequence B with a fixed reference dihedral angle $\phi_r = -1.083$ rad and increasing magnitude of negative twisting force.

with respect to twisting forces. To quantify these preferences, we calculated the frustration indices[42] for this series of landscapes (see SI Figure S6). From the frustration index, we observe that different twist references lower the frustration significantly, with non-ideal twist angles raising the frustration further than we see for the pulling forces. Clearly, twisting forces applied in such a fashion through the scaffold can define structural propensities of chameleon sequences.

A final consideration is reserved to the actual availability of helical and extended structures within a simplified model of backbone dihedrals, i.e. when we assume that the reference angle for the twisting force matches the dihedral angle required for the helical and extended structures, respectively. The associated free energy landscapes at 298 K are presented in Figure 8. Perhaps not surprisingly, considering the results for the interval changes to the reference angle, the landscapes appear fairly different. Importantly, in both cases the helical structures sit at the bottom of the energy landscape. While in the helical case there is more of an energy difference between the helical structures and the extended ones, it is clear that the backbone configuration has an impact on the observed structural ensembles, but does not explain the structural

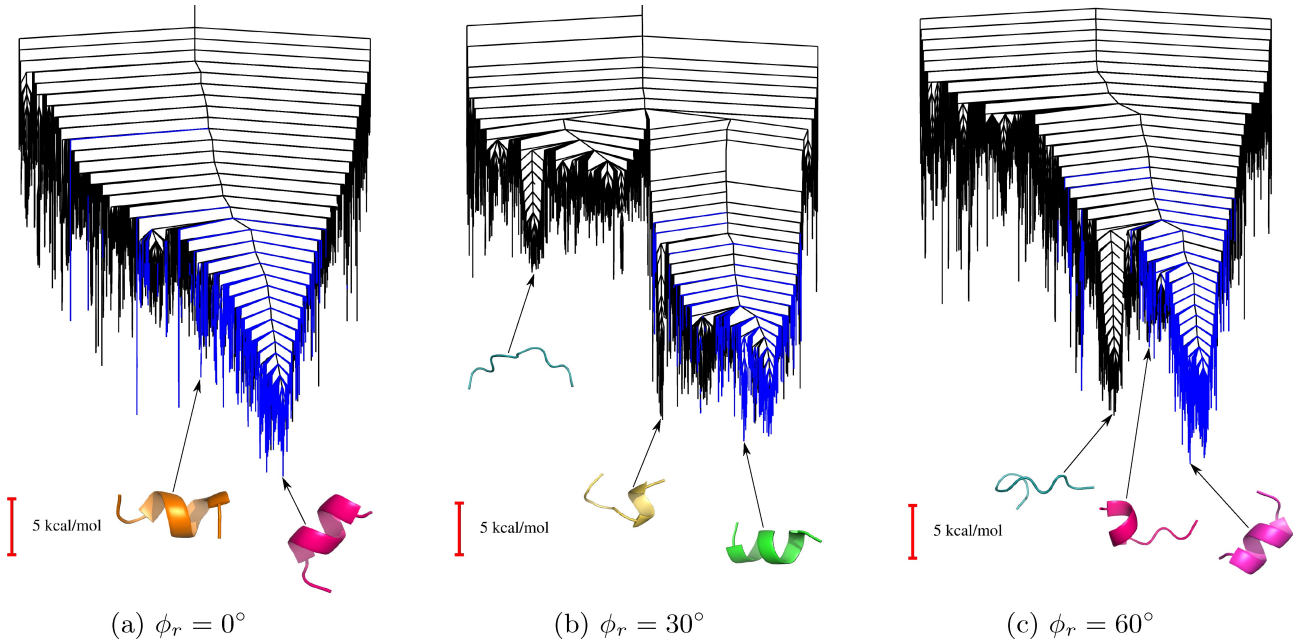


Figure 7. The disconnectivity graphs for sequence B with a dihedral angle varying from 0° to 60°.

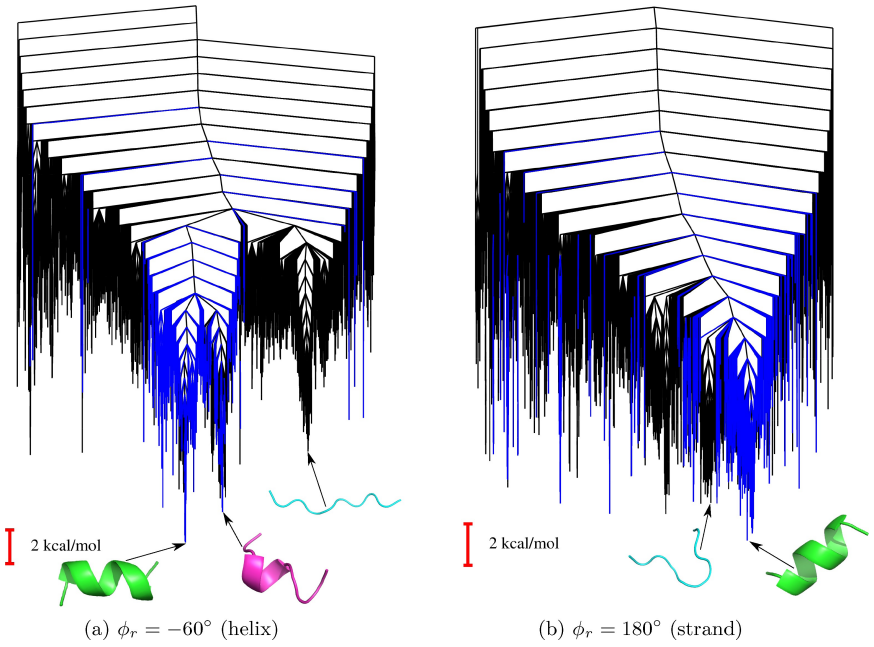


Figure 8. Free energy landscapes for sequence B, where the twisting force reference angle is set to match specific secondary structural elements found in the reported experimental structures in the PDB.

preferences. Nevertheless, chameleon sequences show different structural ensembles depending on the context they are located in, meaning structures adopted these sequences are context-specific, an observation certainly not unique to these peptides.

Discussion

Our energy landscape explorations have led to three key observations: i) Chameleon sequences do not necessary exhibit multifunnel energy landscapes, and hence their preference for multiple structures is not necessarily encoded in the native structural ensemble of these segments. ii) Pulling forces result in rapid and abrupt shifts in the structural ensembles, where competition between structures is effectively lost. iii) Twisting

forces lead to more diverse structural ensembles. However, the changes to the energy landscape are complex, and follow, at least for chameleon sequences, no clear pattern.

Considering the first point, we may consider the actual sequences more closely, and identify their helical propensities.[43] The average $\Delta\Delta G$ compared to Ala for sequence A is 0.486 kcal mol^{-1}, 0.434 kcal mol^{-1} for sequence B, and 0.278 kcal mol^{-1} for sequence C. For chameleon sequences, a high proportion of valine and serine is observed alongside residues that prefer helices (such as Leu, Ile and Ala), providing the structural competition.[20] While such considerations are somewhat simplified, as they ignore flanking residues and possible other interactions (for example the hydrogen bonding with distant-in-sequence residues), there is some indication that these sequences are likely to form helices. Importantly, force fields and implicit solvent can enhance such trends, though it should be noticed that in such cases extended structures are still found on the energy landscape. In any case, the unperturbed energy landscapes do not support the view that there is a structural competition for chameleon sequences in-built into the sequence. Instead, it is clear that such observation must depend strongly on the environment.

Abrupt Structural Changes from Pulling Forces

Pulling forces have a strong impact on the structures observed on the energy landscape, and lead to the loss of structural competition at higher forces. For smaller pulling forces, increased competition may be observed, with heat capacity features indicating transitions between the different structural ensembles. The strong impact of pulling forces in this case is not necessarily surprising, as these sequences are not part of structural proteins. There is no need for the sequences to resist large forces. Importantly, the change in structural ensembles has the clear signatures of a force-induced structural phase transition, which might be useful in biotechnology.

Structural Diversity from Twisting

The impact of twisting forces is more difficult to assess and analyse than the impact of pulling forces. The twisting forces can lead to significant simplification of the energy landscape into a single funnel energy landscape. At the same time, more complex structural ensembles can also result from the application of a twisting force. While changes from pulling are fairly clear and will be obvious to an observer, i.e. higher forces lead to extension of the structure, this relationship between applied force and observed structure is much more complicated. Firstly, the twisting can result in one structure or another based on the applied force and the chosen reference angle. Secondly, the structural ensemble may become more simplistic, likely leading to observations of the correct structural ensemble, but it also can become more complex, which will lead to difficulties resolving such changes in experiment. As a result, we are left with two competing observations: a) Twisting forces can have a significant impact on observed structural ensembles, and may play important roles. b) The resulting changes to structural ensembles are difficult to predict and assess, and hence it is challenging to understand the impact of twisting forces on protein structure. A likely source of these observation is the competition between the underlying biomolecular potential (i.e. the interaction due to the given sequence) and the twisting potential. This competition likely results in a preference for incompatible structural motifs. Similar effects give rise to complex landscapes for example in Stockmayer clusters, where dipole and isotropic pair potentials compete.[44,45]

Implications for Protein Structures and Potential Uses of Chameleon Sequences

The considerations so far leave us with an important question: What are the likely impacts of pulling and twisting forces in the context of adopting structures for chameleon sequences? This question is not only important for the specific case of chameleon sequences, but furthermore leads to questions about the impact of forces on the folding pathways of proteins. It is clear from the data presented in this work that the competition between structural ensembles is not necessarily encoded in the isolated sequence of chameleon sequences, and hence this competition must arise from external factors. This observation may also explain why chameleon sequences prove difficult in structure prediction efforts. The structure adopted is not necessarily explained solely by the sequence, nor do thermodynamic arguments hold. Importantly, it is questionable why there should be a structural competition – after all the sequences are not in regions that switch structure. It is therefore likely desirable that there is as little competition as possible, so folding occurs reliable. Both pulling and twisting forces exerted by the surrounding structure will have significant impact on the adopted structure of chameleon sequences, and in all likelihood of other short segments that have no strong preference for a specific configuration. Such arrangements cannot solely be explained by the backbone configuration as seen from the twisting forces applied to sequence B. Of course, such forces during folding might also be exerted from the environment, for example from chaperones. The significant changes of the energy landscapes observed even for small forces show clearly the importance of mechanics in protein structure beyond the well-known use in structural probing for example with atomic force microscopy. In this context, it is important to appreciate the non-permanent nature of force-induced structural changes in these sequences. We would not expect such structural changes to persist once a force is removed. There is, at least from our data, no energy barrier in the unperturbed energy landscape that would prevent fast relaxation towards helical structures. Any force-induced bias must therefore either persist or be replaced with other constraints on the structure. The first stabilising effect might come from the structural constraint of the folded protein, keeping the structure tensioned. The second one may come from possible new interactions. For example, an extended

structure might not be stable in absence of the force bias, but the presence of an extended structure due to the force may enable the formation of interactions leading to β-sheets, which are then stable.

A final note is reserved for the potential of chameleon sequences as molecular switches. This work highlights how structural ensembles can be tweaked with small forces both to be more and less diverse. Such effects might be utilised in protein design and biotechnology. For example, the sharp transition between a helical and extended structure over a small force span would lend itself to applications as molecular sensors for force detection.

Conclusions

Chameleon sequence, despite the relative small system size, are complex molecules in isolation, which do not necessarily display the diverse structural ensemble ascribed to them in the absence of certain environmental factors. The application of pulling and twisting forces has significant impact on the structural ensembles. When pulling chameleon sequences, we observe sharp transitions from mostly helical structural ensembles to almost exclusively extended structures within a small range of applied forces. In contrast, twisting forces are much more difficult to predict in their impact. We observe more complex and more simplistic structures of the energy landscape and observe both more and less diverse structural ensembles, depending on the reference for the twisting and the magnitude and handedness of the applied force.

In both scenarios, we see a significant impact of small forces on the structure of chameleon sequences, and in all likelihood such effects are applicable to any similar sized sequence. It is likely that such forces play an important role, both along folding paths to allow for reliable folding and in folded structures to provide lower structural competition. These observations may be utilised in the design of proteins, and could also lead to interesting applications in molecular sensors.

Methods

The energy landscapes were explored using the the computational potential energy landscape framework.[29,30] The three sequences were taken from Li et al.[18] and are listed in Table 1 together with distinct reported experimental structures. An illustration of the distinct experimental structures for sequence B is provided in the Supporting Material (section S1, Figure S1). The surrounding protein was removed in structures taken from the protein databank and the isolated sequences minimised. A second minimisation was then run for each structure with the applied forces to yield two minima for each sequence for the different applied forces. These two minima for each condition then seeded the exploration of the energy landscapes. Additional low energy minima were obtained using basin-hopping global minimisation (BH).[46] Discrete pathsampling[47,48] was employed to construct a kinetic transition network[49,50] of local minima and transition states. The doubly-nudged elastic band (DNEB)[51–54] algorithm with a quasi-continuous interpolation (QCI)[55,56] was used to locate transition state candidates, which were then refined using hybrid eigenvector-following.[57] In brief, instead of considering a dynamic evolution of the peptides as it is common through MD simulations, we aim to characterise the energy landscape itself, which determines structural, thermodynamic and kinetic features for a given molecule. The landscape can be represented as a set of local minima (stable structures) and the transition states that connect them. After finding low energy structure using BH, we connect these minima with transition states, by interpolating between pairs of minima and selecting candidate structures from the interpolation for actual transition state characterisation. More details on the methodology and a description of how the heat capacity is calculated is provided in the Supporting Information (see section S4).

Table 1. The three sequences considered in this study and the protein databank identifiers for the two distinct structures.

	Sequence	PDB identifier 1	PDB identifier 2
A	VSVLTSKVLD	4jhw	3rki
B	MDSKLRCVFE	3ikk	2mdk
C	IKASQELV	3n4p	2kn8

The AMBER ff19SB force field[58] was used with implicit solvent. The pulling potential is:

$$V_{\text{pull}} = -f(z_1 - z_2), \tag{1}$$

where f is the applied force and z_1 and z_2 are the z coordinates of the atoms the pulling force is applied to. The twisting potential is:

$$V_{\text{twist}} = f\sin(\phi - \phi_r), \tag{2}$$

where ϕ is the dihedral angle the force is applied to and ϕ_r is the reference dihedral angle (offset by 90 degrees).

An initial survey of pulling forces used forces up to 50 kcal/(mol Å). Competition between structural ensembles was lost around 5 kcal/(mol Å), and hence the analysis in this paper focuses on the lower range of forces. For twisting forces, a larger force range was considered (up to 20 kcal/(mol rad)), and the following reference angles were chosen: (a) the global minimum, (b) the least occupied configuration in the absence of forces and (c) a scan in steps of 30 degrees. Forces were applied with a positive and negative sign for f, leading to twisting in both directions.

For the pulling simulations, forces are applied to the terminal heavy backbone atom on either end of the structure. For the twisting, the dihedral angle was defined by the two terminal heavy backbone atoms on either end of the structure.

The databases, extracted structures and scripts are available here: 10.5281/zenodo.7108327.

Supporting Information

The Supporting Information contains the energy landscapes for repeated sequences, graphs showing the force dependent properties for sequence 2, more details on the methodology, and frustration index results for sequence 2. The authors have cited additional references within the Supporting Information.[59–71]

Acknowledgements

The authors thank Prof. David Wales for his implementation of the twisting potential in GMIN and OPTIM and for access to computational resources.

Conflict of Interests

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are openly available in Zenodo at https://doi.org/10.5281/zenodo.7108327, reference number 7108327.

Keywords: chameleon sequences · energy landscapes · mechanical forces · structural ensembles

[1] C. B. Anfinsen, *Biochem. J.* **1972**, *128*, 737.
[2] C. B. Anfinsen, *Science* **1973**, *181*, 223.
[3] J. D. Bryngelson, P. G. Wolynes, *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 7524.
[4] P. E. Leopold, M. Montal, J. N. Onuchic, *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 8721.
[5] J. D. Bryngelson, J. N. Onuchic, N. D. Socci, P. G. Wolynes, *Proteins* **1995**, *21*, 167.
[6] J. N. Onuchic, P. G. Wolynes, *Curr. Opin. Struct. Biol.* **2004**, *14*, 70.
[7] K. Tunyasuvunakool, J. Adler, Z. Wu, T. Green, M. Zielinski, A. Žídek, A. Bridgland, A. Cowie, C. Meyer, A. Laydon, S. Velankar, G. J. Kleywegt, A. Bateman, R. Evans, A. Pritzel, M. Figurnov, O. Ronneberger, R. Bates, S. A. A. Kohl, A. Potapenko, A. J. Ballard, B. Romera-Paredes, S. Nikolov, R. Jain, E. Clancy, D. Reiman, S. Petersen, A. W. Senior, K. Kavukcuoglu, E. Birney, P. Kohli, J. Jumper, D. Hassabis, *Nature* **2021**, *596*, 590.
[8] M. Baek, F. DiMaio, I. Anishchenko, J. Dauparas, S. Ovchinnikov, G. R. Lee, J. Wang, Q. Cong, L. N. Kinch, R. D. Schaeffer, C. Millán, H. Park, C. Adams, C. R. Glassman, A. DeGiovanni, J. H. Pereira, A. V. Rodrigues, A. A. van Dijk, A. C. Ebrecht, D. J. Opperman, T. Sagmeister, C. Buhlheller, T. Pavkov-Keller, M. K. Rathinaswamy, U. Dalwadi, C. K. Yip, J. E. Burke, K. C. Garcia, N. V. Grishin, P. D. Adams, R. J. Read, D. Baker, *Science* **2021**, *373*, 871.
[9] A. G. Murzin, *Science* **2008**, *320*, 1725.
[10] M. Lella, R. Mahalakshmi, *Biochem. J.* **2017**, *56*.
[11] K. Röder, D. J. Wales, *J. Phys. Chem. B* **2018**, *122*, 10989.
[12] B. Strodel, *J. Mol. Biol.* **2021**, *433*, 167182.
[13] F. Pinheiro, J. S. S. Ventura, *J. Mol. Biol.* **2021**, *433*, 167059.
[14] K. Röder, D. J. Wales, *Front. Mol. Biosci.* **2022**, *9*, 820792.
[15] D. L. Minor, P. S. Kim, *Nature* **1996**, *380*, 730.
[16] M. Mezei, *Protein Eng. Des. Sel.* **1998**, *11*, 411.
[17] A. Ghozlane, A. P. Joseph, A. Bornot, A. G. de Brevern, *Bioinformatics* **2009**, *3*, 367.
[18] W. Li, L. N. Kinch, P. A. Karplus, N. V. Grishin, *Protein Sci.* **2015**, *24*, 1075.
[19] H. Zamora-Carreras, B. Maestro, J. M. Sanz, M. A. Jiménez, *ChemBioChem* **2020**, *21*, 432.
[20] J. T. Guo, J. W. Jaromczyk, Y. Xu, *Proteins* **2007**, *67*, 548.
[21] D. M. Gendoo, P. M. Harrison, *Protein Sci.* **2011**, *20*, 567.
[22] K. Ikeda, J. Higo, *Protein Sci.* **2003**, *12*, 2542.
[23] D. Chakraborty, Y. Chebaro, D. J. Wales, *Phys. Chem. Chem. Phys.* **2020**, *22*, 1359.
[24] A. P. Joseph, A. G. de Brevern, *J. R. Soc. Interface* **2014**, *11*, 20131147.
[25] I. B. Kuznetsov, S. Rackovsky, *Protein Sci.* **2003**, *12*, 2420.
[26] A. M. Ruvinsky, I. A. Vakser, *J. Chem. Phys.* **2010**, *133*, 155101.
[27] B. Kim, T. D. Do, E. Y. Hayden, M. T. Teplow, D. B. Bowers, J.-E. Shea, *J. Phys. Chem. B* **2016**, *120*, 8574.
[28] V. A. Jarymowycz, E. Krupinska, M. J. Stone, *Biochemistry* **2006**, *45*, 11179.
[29] J. A. Joseph, K. Röder, D. Chakraborty, R. G. Mantell, D. J. Wales, *Chem. Commun.* **2017**, *53*, 6974.
[30] K. Röder, J. A. Joseph, B. E. Husic, D. J. Wales, *Adv. Theory Simul.* **2019**, *2*, 1800175.
[31] D. J. Wales, T. Head-Gordon, *J. Phys. Chem. B* **2012**, *116*, 8394.
[32] D. J. Lacks, *Biophys. J.* **2005**, *88*, 3494.
[33] V. Barsegov, D. K. Klimov, D. Thirumalai, *Biophys. J.* **2006**, *90*, 3827.
[34] B. Mondal, D. Thirumalai, G. Reddy, *J. Phys. Chem. B* **2021**, *125*, 8682.
[35] A. Stannard, M. Mora, A. E. M. Beedle, M. Castro-López, S. Board, S. Garcia-Manyes, *Nano Lett.* **2021**, *21*, 2953.
[36] W. Kabsch, C. Sander, *Biopolymers* **1983**, *22*, 2577.
[37] O. M. Becker, M. Karplus, *J. Chem. Phys.* **1998**, *106*, 1495.
[38] D. J. Wales, M. A. Miller, T. R. Walsh, *Nature* **1998**, *394*, 758.
[39] D. J. Wales, *Phys. Rev. E* **2017**, *95*, 030105.
[40] K. Röder, D. J. Wales, *J. Chem. Theory Comput.* **2017**, *13*, 1468.
[41] J. Rowe, K. Röder, *Phys. Chem. Chem. Phys.* **2023**, *25*, 2331.
[42] V. K. de Souza, J. D. Stevenson, S. P. Niblett, J. D. Farrell, D. J. Wales, *J. Chem. Phys.* **2017**, *146*, 124103.
[43] C. N. Pace, J. M. Scholtz, *Biophys. J.* **1998**, *75*, 422.
[44] M. A. Miller, D. J. Wales, *J. Phys. Chem. B* **2005**, *109*, 23109.
[45] J. D. Farrell, C. Lines, J. J. Shepherd, D. Chakrabarti, M. A. Miller, D. J. Wales, *Soft Matter* **2013**, *9*, 5407.
[46] D. J. Wales, J. P. K. Doye, *J. Chem. Phys. A* **1997**, *101*, 5111.
[47] D. J. Wales, *Mol. Phys.* **2002**, *100*, 3285.
[48] D. J. Wales, *Mol. Phys.* **2004**, *102*, 891.
[49] F. Noé, S. Fischer, *Curr. Opin. Struct. Biol.* **2008**, *18*, 154.
[50] D. J. Wales, *Curr. Opin. Struct. Biol.* **2010**, *20*, 3.
[51] G. Henkelman, H. Jónsson, *J. Chem. Phys.* **1999**, *111*, 7010.
[52] G. Henkelman, H. Jónsson, *J. Chem. Phys.* **2000**, *113*, 9978.
[53] S. A. Trygubenko, D. J. Wales, *J. Chem. Phys.* **2004**, *120*, 2082.
[54] D. Sheppard, R. Terrell, G. Henkelman, *J. Chem. Phys.* **2008**, *128*, 134106.
[55] D. J. Wales, J. M. Carr, *J. Chem. Theory Comput.* **2012**, *8*, 5020.
[56] K. Röder, D. J. Wales, *J. Chem. Theory Comput.* **2018**, *14*, 4271.
[57] L. J. Munro, D. J. Wales, *Phys. Rev. B* **1999**, *59*, 3969.
[58] C. Tian, K. Kasavajhala, K. A. Belfon, L. Raguette, H. Huang, A. N. Migues, J. Bickel, Y. Wang, J. Pincay, Q. Wu, C. Simmerling, *J. Chem. Theory Comput.* **2020**, *16*, 528.
[59] Z. Li, H. A. Scheraga, *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 6611.
[60] Z. Li, H. A. Scheraga, *J. Mol. Struct.* **1988**, *48*, 333.
[61] K. Mochizuki, C. S. Whittleston, S. Somani, H. Kusumaatmaja, D. J. Wales, *Phys. Chem. Chem. Phys.* **2014**, *16*, 2842.
[62] M. T. Oakley, R. L. Johnston, *J. Chem. Theory Comput.* **2014**, *10*, 1810.
[63] J. Nocedal, *Math. Comput.* **1980**, *35*, 773.
[64] D. Liu, J. Nocedal, *Math. Prog.* **1989**, *45*, 503.
[65] G. Henkelman, B. P. Uberuaga, H. Jónsson, *J. Chem. Phys.* **2000**, *113*, 9901.
[66] D. J. Wales, J. M. Carr, M. Khalili, V. K. de Souza, B. Strodel, C. S. Whittleston, Pathways and rates for structural transformations of peptides and proteins, in D. Leitner, J. Straub (Editors), *Energy Flow in Proteins*, Taylor and Francis/CRC Press **2009** pages 315–340.
[67] J. M. Carr, S. A. Trygubenko, D. J. Wales, *J. Chem. Phys.* **2005**, *122*, 234903.
[68] B. Strodel, C. S. Whittleston, D. J. Wales, *J. Am. Chem. Soc.* **2007**, *129*, 16005.
[69] K. Röder, D. J. Wales, *J. Am. Chem. Soc.* **2018**, *140*, 4018.
[70] B. Strodel, D. J. Wales, *Chem. Phys. Lett.* **2008**, *466*, 105.
[71] D. J. Wales, *Energy Landscapes*, Cambridge University Press, Cambridge **2003**.

Manuscript received: May 16, 2023
Revised manuscript received: October 6, 2023
Accepted manuscript online: October 11, 2023
Version of record online: November 13, 2023