Biochimica et Biophysica Acta 1850 (2015) 903–910

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbagen

Effect of external pulling forces on the length distribution of peptides ☆

Matthew Batchelor, James Gowdy, Emanuele Paci *

School of Molecular and Cellular Biology and the Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds LS2 9JT, UK

ARTICLE INFO

Article history:
Received 3 July 2014
Received in revised form 15 September 2014
Accepted 16 September 2014
Available online 28 September 2014

Keywords:
Peptides
Length distribution
Force
Freely jointed chain
Alpha helix
Random coil

ABSTRACT

Background: The distribution of the length of a polypeptide, or that of the distance between any two of its atoms, is an important property as it can be analytically or numerically estimated for a number of polymer models. Importantly, it is directly measurable through a number of different experimental techniques. Length distributions can be straightforwardly assessed from molecular dynamics simulation; however, true convergence through full accurate coverage of the length range is difficult to achieve.
Methods: The application of external constant force combined with the weighted-histogram analysis method (WHAM) is used to enhance sampling of unlikely 'long' or 'short' conformations and obtain the potential of mean force, while also collecting dynamic properties of the chain under variable tension.
Results: We demonstrate the utility of constant force to enhance the sampling efficiency and obtain experimentally measurable quantities on a series of short peptides, including charge-rich sequences that are known to be highly helical but whose properties are distinct from those of helical peptides undergoing helix–coil transitions.
Conclusions: Force-enhanced sampling enhances the range and accuracy of the length-based potential of mean force of the peptide, in particular those sequences that contain increased numbers of charged residues.
General significance: This approach allows users to simultaneously probe the force-dependent behaviour of peptides directly, enhance the range and accuracy of the length-based PMF of the peptide and also test the convergence of simulations by comparing the overlap of PMF profiles from different constant forces. This article is part of a special issue entitled Recent developments of molecular dynamics.

1. Introduction

The average and probability distribution of the end-to-end distance are important quantities for describing the physical properties of polymeric chains [1]. The simplest model to describe polymers is the freely-jointed chain (FJC) [2]. It only assumes a polymer as a random walk and neglects any kind of interactions among monomers. If each monomer is assumed to be a rigid rod of length d and N monomers form the polymer, the maximal polymer length is $L_{\max} = Nd$. The distribution of the length (or end-to-end distance, L) is

$$P(L) = 4\pi L^2 \left(\frac{3}{2\pi N d^2}\right)^{3/2} e^{-\frac{3L^2}{2Nd^2}}.$$

In the case of polypeptides and proteins, 'monomers' consist of a variety of amino acid residues, the sequence is variable in length, and the interactions between monomer units vary both in strength and specificity. Depending on factors such as length, temperature and solvent condition, they deviate more or less strongly from an ideal polymer. For example, at temperatures below the folding temperature, $P(L)$ for a single alpha-helical peptide will be peaked at about $N \times 1.5$ Å, while that of a perfect beta hairpin will be a few Å, i.e., the distance between two residues sharing a main-chain hydrogen bond. In both cases, above the folding temperature, $P(L)$ will be better approximated by an ideal chain result and increasingly so with increasing temperature, i.e., when intra-chain interactions become negligible. Under some conditions, such as for unfolded proteins under the effect of mechanical force, ideal models, and specifically the worm-like chain model adequately reproduce experimentally observed properties [3].

The potential of mean force (PMF) along L is simply related to the length distribution by $W(L) = -k_B T \ln P(L)$ [4]. The estimation of $W(L)$ from molecular dynamics simulation—or for that matter estimating the PMF associated with any parameter that is a function of the coordinates—is straightforward when all the values assumed by the function are accurately sampled during the simulation. This is not the case in many instances in which the process being studied is an activated one, i.e., when a sizeable free energy barrier appears in the potential of mean force and the transition between different states, identified by different values of the function, or reaction coordinate, are rare events that may not occur spontaneously during the simulation. The most broadly used approach to enhance the sampling of regions of conformation space that would not be accurately sampled otherwise, consists of adding a biasing potential that harmonically restrains the excursion of

☆ This article is part of a special issue entitled Recent developments of molecular dynamics.

* Corresponding author.
E-mail address: e.paci@leeds.ac.uk (E. Paci).

http://dx.doi.org/10.1016/j.bbagen.2014.09.019
0304-4165/

the reaction coordinate around chosen values along the whole range of values assumed by the reaction coordinate, thus enforcing an approximately uniform sampling [5]. The bias on the potential of mean force can be readily removed for the results, but general thermodynamic and kinetic properties of the system in the absence of the bias cannot be obtained. A variation of the umbrella sampling method has been recently proposed where the reaction coordinate is confined by reflecting boundaries [6]; a continuous PMF can be obtained by imposing that the forward and backward flux be equal at each boundary, with the advantage relative to umbrella sampling being that rates can also be obtained. A PMF can also be obtained by constraining the reaction coordinates at specific values [7–9].

In the specific case in which the reaction coordinate is the distance between two atoms, for example, the distance between the two ends of a polymer, a novel class of methods, based on the discoveries of Jarzynski [10] and Crooks [11], have shown that free energy differences can be obtained from non-equilibrium measurements. Such measurements have been made possible by single molecule manipulation techniques [12–15] and their simulation counterparts [16–18].

Experimental observation of the end-to-end distance under the application of a constant force is possible, thanks to techniques such as force-clamp atomic force spectroscopy and magnetic tweezers [19]. The potential of mean force along the extension may not be directly measurable, but it can be probed by single molecule force spectroscopy experiments that measure the force at which proteins 'snap'—when they cross the free energy barrier that separates the native compact state from the denatured extended state when the two ends are pulled apart at constant velocity (or equivalently, the average time it takes the native protein to snap when a constant force is applied to its ends) [20]. If the process of mechanical fracture is thought of as diffusion over a free energy barrier on the potential of mean force defined by the extension of the protein (i.e., distance between the points where force is applied), then the unfolding rate depends exponentially on the applied force [21]. This relation, which goes under the name 'Bell model', seems to be generally obeyed although a number of exceptions have been reported [22,23].

The application of a constant force, parallel to the vector joining any two atoms of a polypeptide, modifies the probability of different lengths of the vector, a positive force favouring longer conformations and a negative force favouring shorter conformations. When modified by a force F, the length distribution is given by the relation $P_F(L) = P(L)e^{FL/k_BT}$. This relation is simply obtained by observing that the PMF ($W(L) = -k_BT \ln P(L)$) is modified by the application of an external force parallel to the end-to-end vector by $W_F(L) = W_0(L)–FL$ (in units of k_BT). For the FJC this is $W_F(L) = \frac{3L^2}{2Nd^2} - 2 \ln L - FL$.

In this paper, we exploit the relationship $W_F(L) = W_0(L) - FL$ to accurately determine the equilibrium potential of mean force associated with the distance between two atoms of a polypeptide chain. Unlike the methods mentioned above where an artificial biasing term is added to the Hamiltonian, the application of a constant force is also possible experimentally, and the equilibrium and kinetic properties of the real system are not perturbed.

We focus on relatively short peptides that have high intrinsic helical propensity. The α helix is a ubiquitous motif found throughout the proteome. Its structure is stabilised by hydrogen bonds between the backbone carbonyl oxygen of residue i and the backbone N–H group of residue $i + 4$. This pattern causes the backbone of the polypeptide chain to form a right-handed helix, with side chains pointing out from the core and slightly toward the N-terminus. Most α helices are found within globular proteins where interactions between neighbouring secondary structure elements stabilise the structure. When investigated in isolation, the short peptide sequences that exist as helices within globular proteins are often not helical. However, certain short sequences, notably the C-peptide from RNase [24–26] and several synthetic peptides [27,28], have since been shown to form stable helices in solution.

Alanine-rich peptides are often used as the archetype of 'normal' alpha helices, and there exists a great deal of literature describing their properties [29]. The substitution of alanine for charged residues, glutamic acid (E), lysine (K) and arginine (R), originally to improve the solubility of these peptides, has been shown to have a significant effect on peptide helicity [27,28,30–32]. A dramatic example is the recent experimental comparison of short E-R/K-rich peptides using circular dichroism (CD) spectroscopy [33]. We use this study as an experimental touchstone for comparison with our simulations.

The peptides studied here are in general polyalanine based, 10 residues in length but have different numbers and patterns of charged residues (glutamic acid, arginine and lysine). We also use a glycine-rich sequence as a random coil model. Each peptide has an N-terminal YS 'tag', used in experiments as a means of accurately calculating peptide concentration, which is retained in simulations to allow for direct comparison, and for consistency, all other peptide sequences start in this manner. As expected, the glycine-rich sequence is completely disordered, alanine-rich sequences show a helical propensity, and that propensity increases as a pattern of charged residues sequentially replace alanine. The PMF properties of the charge-rich sequences are distinct from those of alanine-rich sequences. We also highlight some advantages of using mechanical force to enhance sampling relative to other computational methods, in part related to the fact that constant force can also be applied experimentally to individual proteins.

2. Methods

The equilibrium dynamic behaviour of the peptides was simulated using a united-atom force field (CHARMM19) and implicit solvent model (FACTS) [34]. The force field and the solvation model were chosen after testing several alternatives. Simulations performed using the CHARMM22/CMAP force field [35] with FACTS showed that helical conformations for charged sequences were excessively stable and reversible helix formation was never observed. This same behaviour was observed using CHARMM22/CMAP and explicit solvent (with TIP3P water). A similarly initiated 1.25 μs simulation using a fully solvated recently revised all-atom model of YSE_4R_4 (CHARMM36 [36], run using NAMD [37]) did not show any significant helicity (4%, most of which being 3_{10} helix). Simulations performed using CHARMM19 with the SASA implicit solvent model [38] showed, in clear disagreement with experiment, poor differentiation between charged and non-charged sequences, likely due to the neutralization of charged side chains imposed with the SASA implicit solvation model. The combination of CHARMM19 and FACTS was recently shown to give excellent agreement with experiment for a highly charged system [39]. The 'standard' CHARMM19 FACTS parameters for structured peptides were used: dielectric constant = 2.0, nonpolar surface tension coefficient = 0.015 kcal mol^{-1} $Å^{-2}$.

Unless otherwise stated, all simulations were performed at 300 K, with Langevin dynamics using the leapfrog integrator, a time step of 2 fs and a friction coefficient of 3 ps^{-1}, and run using CHARMM [38]. Trajectory frames (and associated analysis parameters) were recorded every 500 steps. A constant force of between −50 and 50 pN was applied between the N atom of the first residue and the carbonyl carbon atom of the last (tenth) residue. The N and C termini were capped with acetyl (ACE) and methylamine (CBX) groups, respectively [40]. Simulations lasted between 1 and 4 μs. Starting structures for simulations were prepared by performing a steepest descent minimisation (1000 steps) from an all-*trans* backbone conformer followed by a short (20 ps) dynamics run. The first 100 ns of each of the simulations was removed prior to analysis in an effort to remove starting structure bias.

Wordom (version 21) [41] was used to analyse the simulation trajectories. The secondary structure of the peptide was assigned for each timeframe using the DSSPcont [42] criteria. This was then used to calculate the helicity (or helical fraction) of the peptide overall, with helicity defined as the fraction of 3_{10}, α, and π residues (i.e., G + H + I) [43]. It

should be noted that the capping groups are included in the secondary structure assignment and are therefore a factor in the helicity fractions reported. If the capping group assignments are removed from the analysis, the helicity values all increase by a factor of 6/5, but the ratios between different peptides do not change. Sample analyses using the DSSP [44] criteria gave similar results (typically being 1–2% less helical at $F = 0$).

Wordom was also used to output the distance between the backbone N atom in the first residue and the carbonyl carbon in the last residue, i.e., the distance used to define the length (L). $P_F(L)$ and thereby $W_F(L)$ were calculated at all forces. WHAM [45] was then used to iteratively reweight histograms obtained at different forces, using $W_0(L) = W_F(L) + FL$, where F is the constant force. This yields a zero-force PMF with improved accuracy over the entire reaction coordinate range compared to a PMF direct from the $F = 0$ simulation alone. The code required to perform the constant-force bias WHAM analysis was written from scratch in Python based on Hub et al. [46] with the perturbation being a constant-force potential rather than an umbrella potential.

3. Results

3.1. Constant-force bias as a method for accurately estimating the PMF

To demonstrate the constant-force bias procedure used, $W_F(L)$ plots for the peptide YSAAARAARA at forces between −50 and 50 pN are shown in Fig. 1. At zero force, the PMF has its minimum at a length of ~9 Å. The thicker lines show how as the force increases from $F = 0$, the minimum energy length for the peptide moves to higher values. More than one minimum exists at intermediate forces; for instance, at 10–20 pN, a clear minimum can be seen at 14–15 Å. At high forces, the peptide is completely stretched with a length of ~30 Å. Negative forces cause the peptide to shorten. Two more minima can be seen with the length ultimately tending to 4 Å, the closest approach of the termini.

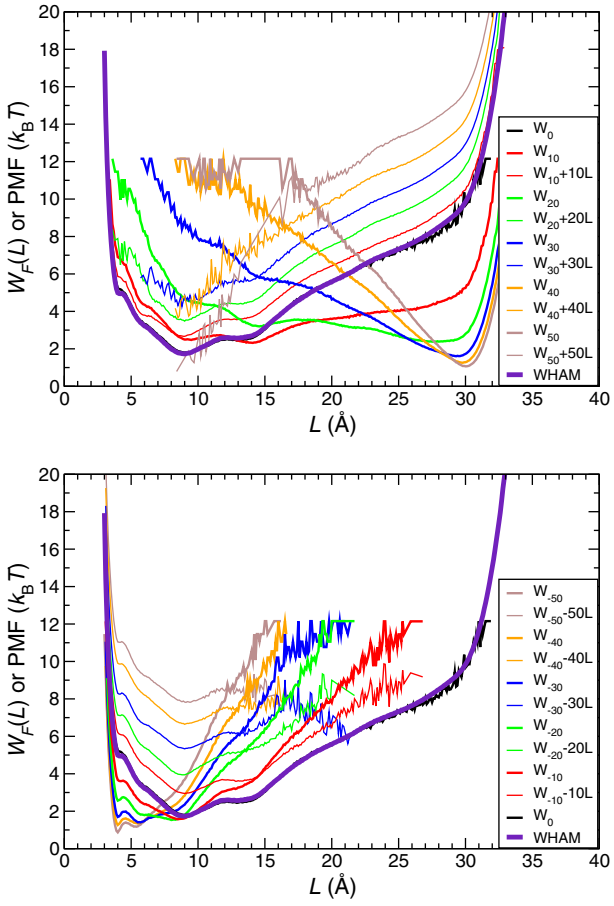

Fig. 1. Depiction of the constant-force bias procedure using $W_F(L)$ profiles for YSAAARAARA from a range of force-assisted 2 μs simulations at 300 K. For clarity, plots are split into positive (top) and negative (bottom) forces. $W_F(L)$ profiles at constant force (thick lines) are converted back to a zero-force case (thin lines) using $W_0(L) = W_F(L) + FL$, allowing for an offset in each case.

The thin lines in Fig. 1 show zero-force PMFs as converted from non-zero-force $W_F(L)$ plots by addition of the term $-FL$ to the potential energy function. The profiles have been manually offset to aid comparison. In all cases, the profiles match well with each other, and with the zero-force case, over a certain range of L, the negative forces matching well at low L and the positive forces matching well at larger L. Outside the matching regions, the converted profiles are noisy, representing poorly sampled conformers.

The weighted histogram analysis method (WHAM) procedure does the job of combining length distribution histograms from simulations at different constant forces to provide a PMF at zero force that is accurate at all values of L (shown as a thick purple line in Fig. 1). The purple line matches well with the $F = 0$ trajectory, indicating that, in this case, the zero-force trajectory does a good job of sampling the accessible conformers as gauged by trajectories at all other forces, suggesting good convergence, including at relatively large values of L. Fig. S1 shows in greater detail the corrected PMF profiles generated from trajectories at $F = -5$, 0 and 5 pN and compares with the WHAM PMF. All three trajectories provide PMFs that compare reasonably well with the WHAM PMF, which we assert represents the best approximation to the perfectly converged case. At lengths greater than 6 Å, the 5 pN PMF is in fact the best match to the WHAM PMF showing that in certain cases, it can be beneficial to generate trajectories under a small applied load and convert rather than collect directly at zero force. In order to probe the energy landscape in the region of a barrier, it makes sense to add a biasing force to allow transitions across the barrier to occur more frequently and thereby enhance conformation sampling; this is the basis of the extensively used experimental method of force spectroscopy [47,48].

$W_F(L)$ profiles are shown in Fig. 2 for a polyglycine (YSG_8) and polyalanine (YSA_8) peptide, which act as a random coil and 'typical' helix mimic, respectively. As judged by the overlap of the $F = 0$ PMF with the WHAM PMF, barring the 'noise' at high extensions, the conformations for both peptides are well sampled. A 1 μs simulation is enough to allow convergence; thus, this single simulation is all that is required to accurately predict the length distribution at any force. There is no significant benefit in using the constant-force bias procedure in this case other than for improved accuracy at high extension and a small increase in the range over which the PMF is described (beyond 30 Å).

The resulting PMFs for the two peptides show clear differences with YSA_8 having a more well-defined minimum at $L \sim 9$ Å. Further information on the differentiation of different 'states' that have the same length is not accessible from the PMF. The length convergence confirmed using WHAM gives us confidence that the properties resulting from the single zero-force simulation give an accurate representation of the equilibrium behaviour of our model. Analysis of the simulation carried out in the absence of force showed, as expected, that YSG_8 was completely non-helical (0.4%; α, π or 3_{10} helix). By contrast, YSA_8 shows significant helicity (21% helix). Fig. S2 shows example structures with lengths corresponding to minima in $W_F(L)$ plots for YSA_8. At zero force, the region around the minimum at $L = 9.0 \pm 0.5$ Å is composed of both helical and non-helical conformers, with completely non-helical conformers making up 29% of the total.

3.2. Convergence of the PMF

Moving now to more complicated peptides containing residues with potential for specific interactions, we consider the case of adding E and R

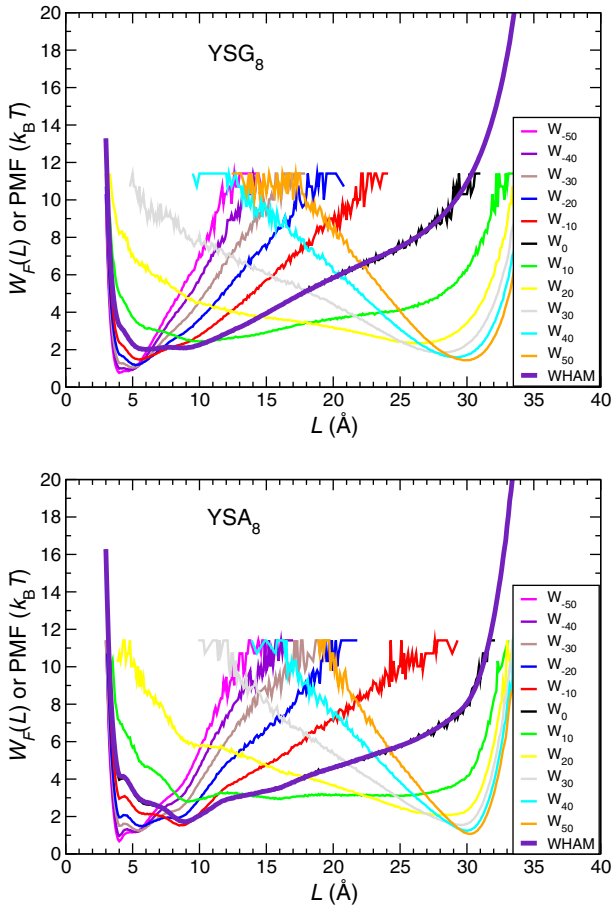

Fig. 2. $W_F(L)$ profiles from simulations at various forces and the PMF as matched using WHAM for YSG_8 (top) and YSA_8 (bottom) at 300 K.

residues at i and $i + 3$, positions, respectively, to our polyalanine peptide. This pattern has been implicated in the stabilisation of α helices since simultaneous formation of an $(i, i + 4)$ backbone hydrogen bond and $(i, i + 3)$ salt bridge are possible [33]. We consider the sequence of peptides YSAAAEAARA, YSAAEEARRA and YSAEEERRRA, which contain 1, 2 and 3 potential salt bridges between i and $i + 3$ positions. We therefore treat YSA_8 as our baseline.

As shown in Fig. 3, all the WHAM PMFs exhibit a clear minimum at 9 Å, but the gradient away from the minimum becomes increasingly steep as more charge-paired residues are added; it becomes increasingly

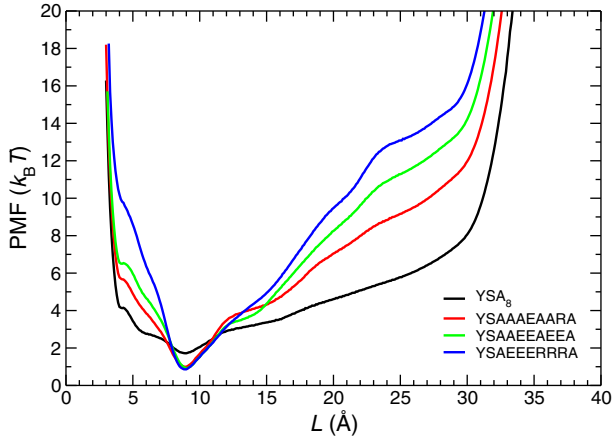

Fig. 3. WHAM PMFs for YSA_8 and E–R $(i, i + 3)$-containing peptides.

more difficult to either extend or compress these peptides. On comparison of the WHAM PMF with the PMF from the single trajectories (1 μs) carried out under no force (Fig. S3), the difference is initially confined to large extensions, but as more charged residues are added, the differences become significant at shorter and shorter extensions. The range of L being sampled also becomes increasingly confined with YSAEEERRRA only being sampled from 4 to 19 Å compared to a range of 3 to 33 Å sampled with WHAM. This clearly indicates that for the more charge-rich sequences convergence of the PMF is poor and therefore either much longer trajectories or an enhanced sampling method is required.

Possibly the most striking examples of non-convergent simulations are shown for peptides where all the alanine residues have been exchanged for glutamic acid and arginine/lysine pairs (whilst maintaining overall neutrality). Zero-force simulation and WHAM PMFs generated from longer (3 or 4 μs) simulations for YSE_4R_4 and YSE_4K_4 (and some other same-residue permutations) are given in Fig. S4; their properties have been previously investigated using CD spectroscopy. Fig. 4 shows the PMF profile generated from each microsecond of simulation individually for YSE_4K_4. All four profiles have the same minimum at 9 Å but differ significantly in the proportion of time spent as conformers with lengths longer than 10 Å. It is clear that 1 μs alone (black line, Fig. 4A) is not enough to allow convergence, with poor sampling beyond 10 Å and a very poor fit to the overall WHAM plot. As the simulation becomes longer (depicted in Fig. S5), the zero-force PMF begins to converge on a particular profile, with good matching to the WHAM plot for $L = 6.5$–16 Å. However, the profiles are a poor match outside these

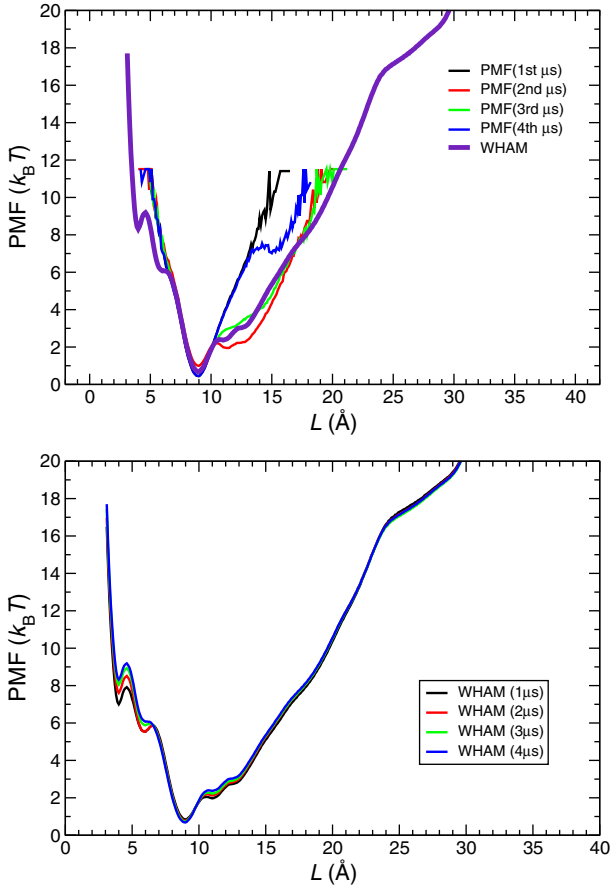

Fig. 4. (Top) PMF profiles from the first (after removing the first 100 ns), second, third and fourth microsecond of simulation for YSE_4K_4 with no applied force. Significant differences are observed between each part of the simulation for lengths beyond 10 Å. (Bottom) WHAM was used to combine histograms taken at forces between −50 and +50 pN for 1 μs, 2 μs, 3 μs and the entire 4 μs. More subtle differences are shown.

values, and due to the restrictions imposed by the temperature, the zero-force simulations do not probe any lengths beyond 21 Å (note: 3–33 Å are examined using the constant-force bias approach). It is unlikely that any reasonable length simulation will be able to access these other conformers to quantify their populations accurately.

When simulations at all forces are collected together and combined using the constant-force bias procedure, an improved PMF is provided. The second graph in Fig. 4 shows how little the resulting WHAM PMF is affected on changing from use of 1 μs simulations (at all forces) to 4 μs simulations. This indicates that even short, non-fully convergent simulations when taken at a range of constant-force biases can yield a good approximation of the PMF for a peptide. The constant-force-bias approach also reveals additional minima for YSE_4R_4 and YSE_4K_4 at extensions of ~4 and 6 Å. Representative WHAM PMFs that include an estimate of error are included in Fig. S6.

3.3. Uses and limitations of the constant-force bias approach

The extent of overlap of probability histograms generated using constant-force biased simulations is much greater than would be the case for a windowed sampling method, *e.g.*, umbrella sampling or boxed molecular dynamics. When the histograms collected at a wide range of forces all overlap well, we have greater confidence that adequate sampling is achieved and the resulting PMF is accurate. In umbrella sampling, the overlap of histograms is normally only with neighbouring simulations, and it is the manner in which histograms are joined that determines the PMF. Convergence is gauged only within each simulation, when the histograms no longer change on increasing the simulation time. If a transition between different conformational states at a particular L is more easily facilitated by first extending or contracting to an L that is not well represented due to the effect of the umbrella potential, then the rate of transition will be reduced and there is more likelihood of sampling being poor due to trapping in these artificially created kinetic traps. A careful balancing act between umbrella size, separation and duration of simulation must be performed. With the constant-force bias method, transitions between conformations of different L are actively encouraged by effectively flattening out the local energy landscape. One disadvantage of the constant-force approach is that due to the fact that a large conformational space is still sampled in each simulation, the 'enhancement' of sampling will be less efficient than with a windowed technique.

A further benefit of the constant-force-bias approach is that as simulations are at equilibrium, trajectory averages correspond to experimental observables, and this is true for *all* simulations since the application of fixed loads to peptides is becoming increasingly possible. Of particular interest is the spectrum of the fluctuations of the end-to-end distance and its dependence on the constant-force applied. Fluctuations of the end-to-end distance is a quantity that can be measured through atomic force spectroscopy [49], and that provides a dynamical fingerprint of the underlying free energy landscape of the polypeptide chain [50].

3.4. Further analysis of zero-force simulations

Although the PMF profiles indicate that YSE_4R_4, YSE_4K_4, the E–R $(i, i+3)$-containing peptides, YSAAARAARA and YSA_8 most commonly exist with $L \sim 9$ Å, they do not allow us to differentiate between molecular states and groups of states, of which many could exist at a particular L. Indeed the most common lengths of these 'helical' peptides overlap to a significant extent with those of the random-coil (exemplified by YSG_8) as shown in Fig. 2. An interesting point to note is that a length of 9 Å does not correspond to a fully helical 10-residue structure, which would be expected to have $L \sim 15$ Å. The distance is shorter because even for the most helical peptides, the YS tag is never helical and prefers to bend around and interact with the side of the helix, thus reducing the average L (see an example conformation for YSE_4K_4 in Fig. 5, and also for YSA_8 in Fig. S2). For comparison with the calculation shown above for YSA_8, the region around the minimum ($L = 9.0 \pm 0.5$ Å) for both YSE_4R_4 and YSE_4K_4 is almost completely devoid of non-helical conformers (<0.4% of the total).

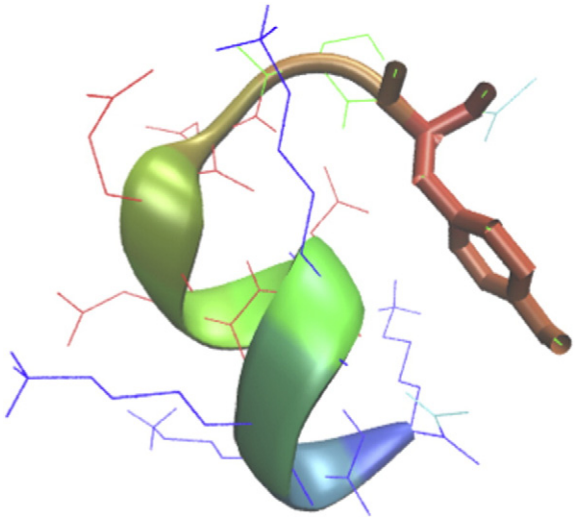

Fig. 5. A typical conformation observed during simulation of YSE_4K_4 under no force. The Y and S residues do not, in general, sit as a continuation of the helix pattern (i.e., there is no backbone hydrogen bonding between S ($i = 2$) and the fourth E ($i = 6$)) but rather tend to be bent around with the aromatic ring interacting with the flank of the helix.

Despite having shown above that the PMFs from simulations for the highly charge-rich sequences at $F = 0$ are not truly convergent, since the discrepancies are due to incorrect sampling of unlikely conformers, their omission from the analysis of *average* structural parameters will have a minor effect. We include here a description of the average helical properties of the peptides from our simulations.

The helicity of some of the short peptides (of course, only in the absence of applied force) are available from the literature [33]. In addition, longer sequences including $YSE_4K_4E_4K_4$ and $YSE_2K_2E_2K_2E_2K_2E_2K_2$ [31] and motifs $(EAAKA)_n$, $(EAAAK)_n$ and $(EAAAR)_n$ have been studied [27,30]. The average helicity over simulations in the absence of force for each of these sequences are qualitatively similar to the experimental values reported by Sommese et al. [33] in that YSE_4R_4 and YSE_4K_4 are more helical than YSAAARAARA (Table 1). However, the helicity values show some discrepancies. YSE_4K_4 is (rather serendipitously) a good match for simulation and experiment, whereas experimental helicities

Table 1
Comparison of helicity values at 300 K (or 350 K) and $F = 0$ for selected peptides. Simulation error estimates were achieved by dividing trajectories into 100 ns blocks and finding the standard deviation in helicity values from analysis of each block.

Peptide	Fraction of helical residues	
	Simulation	Experiment[a]
YSG_8	0.004	
YSA_8	0.21 ± 0.05	
YSAAARAARA	0.35 ± 0.06	0.15
YSAAARAARA (350 K)	0.10 ± 0.02	0.13
YSAAAEAARA	0.41 ± 0.05	
YSAAEEARRA	0.45 ± 0.02	
YSAEEERRRA	0.46 ± 0.004	
YSE_4R_4	0.44 ± 0.04	0.57
YSE_4R_4 (350 K)	0.37 ± 0.07	0.37
YSE_4K_4	0.45 ± 0.04	0.45
YSE_4K_4 (350 K)	0.41 ± 0.03	0.35
$YSE_2K_2E_2K_2$	0.27 ± 0.09	
YSK_4E_4	0.25 ± 0.09	
YSR_4E_4	0.11 ± 0.10	

[a] Values estimated from figure 1 in Sommese et al. [33].

are higher in the case of YSE_4R_4 and lower for YSAAARAARA. Simulation also yields a significant helicity value for $YSE_2K_2E_2K_2$ (0.27), which is unexpected given that the double length version ($YSE_2K_2E_2K_2E_2K_2E_2K_2$) has been shown to be non-helical [31]. The YSAAARAARA and $YSE_2K_2E_2K_2$ results point to a possible general overemphasis of the force field for helical conformations [51]. Interestingly, the helicity values from simulations and experiment are much closer when carried out at 350 K (Table 1). Also helicities for the 'oppositely oriented' YSK_4E_4 and YSR_4E_4 peptides are lower than the values for YSE_4R_4 and YSE_4K_4, in line with the finding that positively charged Lys and Arg residues are destabilizing at early positions in a helix whereas negatively charged Glu promotes helix formation [52–54].

Some difference in helicity values are to be expected given that the method by which they are calculated are completely different. The simulations use hydrogen-bonding patterns to assign a single secondary structure label to each residue in each frame of the trajectory, whereas CD relies on the mean residue ellipticity at 222 nm and an empirical relation. Both methods have potential problems and drawbacks. For instance, due to the secondary structure assignment criteria, the simulated peptides cannot ever be 100% helical because the first two residues and the last residue cannot be helical by definition; the upper limit for helicity for these peptides is therefore 75%.

The higher experimental helicity reported for YSE_4R_4 compared to YSE_4K_4 tallies with studies on longer alanine-based peptides with $(i, i + 4)$ spaced E and R/K residues, where the replacement of K for R has been reported to increase helicity [30]. The lack of discrimination between the helicity of YSE_4R_4 and YSE_4K_4 in our simulations is likely due to the simplicity of the model used. The simulations do not allow for any polarisation of bonds/groups in the peptide, nor are there explicit water molecules that could feasibly modify the salt bridge interactions (e.g., through stronger binding to K). It is possible that polarisation of the guanidinium group may enhance salt bridge interactions between R and E compared to K and E and potentially allow greater flexibility in bonding patterns for longer chains. This difference between the behaviour of the K- and R-containing peptides warrants further study with more advanced force fields.

Having simulated the peptides at different constant forces, in addition to improving the PMF accuracy and range at zero force, we are also able to characterise their force-dependent properties. Fig. 6 shows the helical fraction for a range of different sequences under the application of different constant forces. The glycine-rich sequence behaves as a random coil, exhibiting no helical character at any force. By contrast, the helicity for the alanine-rich sequences YSA_8 and YSAAARAARA, though significant at zero force, essentially drops to zero by increasing the force to 25 pN. YSE_4R_4 (and YSE_4K_4) are the most helical across all forces and the difference between these sequences and others is most apparent at moderate forces. The helical fraction only slightly decreases by 25 pN, whereas the alanine-rich sequences are essentially non-helical at this force. An equivalent plot is given for the E–R $(i, i + 3)$-containing peptides in Fig. S7; a step-wise increased resistance to force as extra $(i, i + 3)$ pairs are added is shown. For most peptides studied, the helicity is exclusively due to α helix; no significant 3_{10} or π helicity is present. Only for YSE_4R_4 and YSE_4K_4 do we see a contribution from π-helical residues as demonstrated in the case of YSE_4R_4 in Fig. S8. It is interesting to see that the π contribution only comes in at moderate forces, but this 'prolongs' the high overall helicity to higher force. The wider, flatter π-helix formed by $(i, i + 5)$ backbone hydrogen bonds appears to provide an additional conformational outlet for these charge-rich sequences that the peptide can adopt to delay transition to coil.

As alluded to above, force-dependent simulations were also carried out for a selection of peptides at 350 K. Fig. 7 shows the change in the force-dependent helicity of YSAAARAARA and YSE_4R_4 on increasing the temperature from 300 to 350 K. There is a 16% reduction in helicity at 0 pN for YSE_4R_4, whereas the same increase in T gives a far larger reduction for YSAAARAARA (73% at 0 pN). A comparatively small reduction in helicity was also observed for YSE_4K_4. The substitution to include more charged residues therefore increases the resistance not only to force but also to temperature-induced reduction in helical content.

4. Concluding discussion

We have introduced a simple way to efficiently sample conformations of polypeptides with unlikely end-to-end distances by introducing a constant force between the two ends (either positive, to sample 'longer' conformations, or negative, to sample 'shorter' ones). The effect of the force on the distribution can be then removed and simulations performed with different forces used simultaneously to obtain an accurate estimation of the potential of mean force for all end-to-end distances. Simulations with the application of a constant force could also provide the fluctuations of the end-to-end distance that are measurable and provide information on their viscoelastic properties and insight on the free-energy landscape of the system [50].

In this work, we have focussed on short peptides, using an implicit solvent model, and performed long simulations (several microsecond

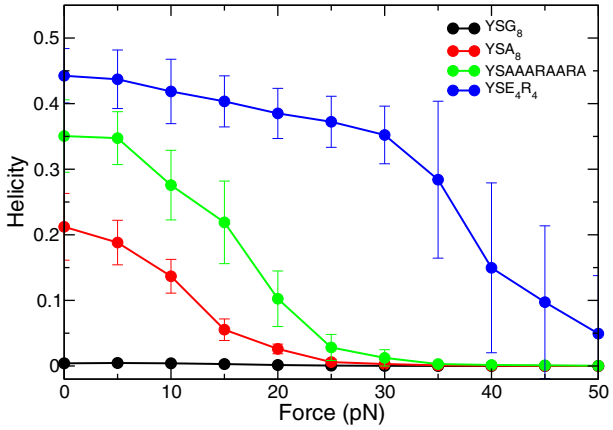

Fig. 6. Helicity of various peptides as a function of force. YSG_8 exhibits no helicity at any force. In the absence of force, YSE_4R_4 (also YSE_4K_4, not-shown) are the most helical (44%). The alanine-rich sequences are less helical and more interestingly their helicity drops rapidly with force (negligible at 25 pN), whereas the charge-rich peptides are still 35% helical at a force of 30 pN. Error bars represent standard deviation from analysis of the simulations in 100 ns blocks.

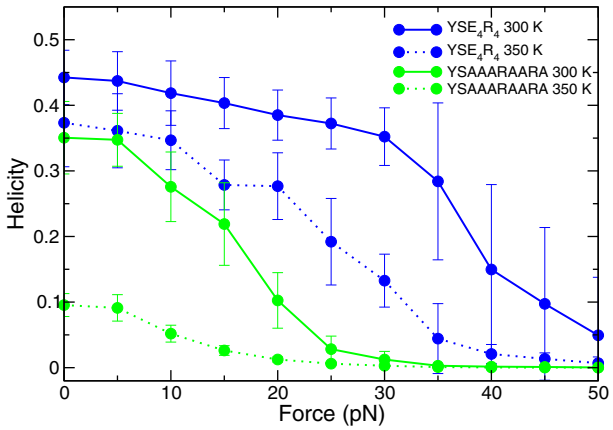

Fig. 7. The effect of temperature on the helicity of alanine-rich (YSAAARAARA) and charge-rich (YSE_4R_4) sequences. Increasing the simulation temperature to 350 K causes a small reduction in helicity for YSE_4R_4, whereas the same increase in T gives a large reduction for YSAAARAARA. Error bars represent standard deviation from analysis of the simulations in 100 ns blocks.

for each sequence). In most cases, the potential of mean force of the extension converges, except for the most unlikely extensions, with a single simulation. The force-enhanced sampling is still necessary to dissect the properties of sequences that contain mostly charged residues, and that, as experiments previously showed, are highly helical.

The underlying potential of mean force determined here reveals a difference between alanine-rich sequences such as YSA_8 and mostly charged sequences such as YSE_4R_4 and YSE_4K_4; for the latter, the PMF shows a deep minimum corresponding to helix-like extension and increases very steeply for both shorter and longer lengths.

A long, rod-like alpha helical spacer domain was found in smooth muscle caldesmon [55], and similar structures are found in some myosin proteins [56–58]. These domains contain a high proportion of charged residues, which generally appear as a pattern of alternating patches of like-charged residues, akin to repetitions of the E_4R_4/E_4K_4 pattern investigated here. We suggest here that the properties of the potential of mean force along the end-to-end distance for E_4R_4/E_4K_4 may be important for the biological function of naturally occurring helical EK/R-rich domains, if they are to have a mechanical role [59]. That is, the deep, steep-sided PMF helps to maintain/propagate the helix over long distances and also to recover rapidly, possibly without crossing a free-energy barrier from an elongated conformation to a helical one.

Acknowledgment

This work was partly supported by BBSRC (BB/I007423/1).

Appendix A. Supplementary data

Supplementary figures for this article can be found online at http://dx.doi.org/10.1016/j.bbagen.2014.09.019.

References

[1] P. Flory, Principles of Polymer Chemistry, Cornell University Press, 1953..
[2] S.B. Smith, L. Finzi, C. Bustamante, Direct mechanical measurements of the elasticity of single DNA molecules by using magnetic beads, Science 258 (1992) 1122–1126.
[3] M.S. Kellermayer, S.B. Smith, H.L. Granzier, C. Bustamante, Folding-Unfolding Transitions in Single Titin Molecules Characterized with Laser Tweezers, Science 276 (1997) 276, 1112–1116.
[4] B. Roux, The calculation of the potential of mean force using computer simulations, Comput. Phys. Commun. 91 (1995) 275–282.
[5] G. Torrie, J. Valleau, Nonphysical sampling distributions in Monte Carlo free-energy estimation: umbrella sampling, J. Comput. Phys. 23 (1977) 187.
[6] D.R. Glowacki, E. Paci, D.V. Shalashilin, Boxed molecular dynamics: a simple and general technique for accelerating rare event kinetics and mapping free energy in large molecular systems, J. Phys. Chem. B 113 (2009) 16603–16611.
[7] E. Carter, G. Ciccotti, J.T. Hynes, R. Kapral, Constrained reaction coordinate dynamics for the simulation of rare events, Chem. Phys. Lett. 156 (1989) 472–477.
[8] E. Paci, G. Ciccotti, M. Ferrario, R. Kapral, Activation energies by molecular dynamics with constraints, Chem. Phys. Lett. 176 (1990) 581–587.
[9] D. Trzesniak, A.-P.E. Kunz, W.F. van Gunsteren, A comparison of methods to compute the potential of mean force, ChemPhysChem 8 (2007) 162–169.
[10] C. Jarzynski, Equilibrium free-energy differences from nonequilibrium measurements: a master-equation approach, Phys. Rev. E. 56 (1997) 5018–5035.
[11] G. Crooks, Path-ensemble averages in systems driven far from equilibrium, Phys. Rev. E. 61 (2000) 2361–2366.
[12] G. Hummer, A. Szabo, Free energy reconstruction from nonequilibrium single-molecule pulling experiments, Proc. Natl. Acad. Sci. U. S. A. 98 (2001) 3658–3661.
[13] G. Hummer, A. Szabo, Free energy surfaces from single-molecule force spectroscopy, Acc. Chem. Res. 38 (2005) 504–513.
[14] G. Hummer, A. Szabo, Free energy profiles from single-molecule pulling experiments, Proc. Natl. Acad. Sci. U. S. A. 107 (2010) 21441–21446.
[15] J. Liphardt, S. Dumont, S.B. Smith, I. Tinoco, C. Bustamante, Equilibrium information from nonequilibrium measurements in an experimental test of Jarzynski's equality, Science 296 (2002) 1832–1835.
[16] J.R. Gullingsrud, R. Braun, K. Schulten, Reconstructing potentials of mean force through time series analysis of steered molecular dynamics simulations, J. Comput. Phys. 151 (1999) 190–211.
[17] S. Park, K. Schulten, Calculating potentials of mean force from steered molecular dynamics simulations, J. Chem. Phys. 120 (2004) 5946–5961.
[18] D.K. West, P.D. Olmsted, E. Paci, Free energy for protein folding from nonequilibrium simulations using the Jarzynski equality, J. Chem. Phys. 125 (2006) 204910.
[19] G. Zoldak, M. Rief, Force as a single molecule probe of multidimensional protein energy landscapes, Curr. Opin. Struct. Biol. 23 (2013) 48–57.
[20] B. Ramm, et al., Sequence-resolved free energy profiles of stress-bearing vimentin intermediate filaments, Proc. Natl. Acad. Sci. U. S. A. 111 (2014) 11359–11364.
[21] E. Evans, K. Ritchie, Dynamic strength of molecular adhesion bonds, Biophys. J. 72 (1997) 1541–1555.
[22] M. Schlierf, Z.T. Yew, M. Rief, E. Paci, Complex unfolding kinetics of single-domain proteins in the presence of force, Biophys. J. 99 (2010) 1620–1627.
[23] Z.T. Yew, M. Schlierf, M. Rief, E. Paci, Direct evidence of the multidimensionality of the free-energy landscapes of proteins revealed by mechanical probes, Phys. Rev. E. 81 (2010) 031923.
[24] J.E. Brown, W.A. Klee, Helix-coil transition of the isolated amino terminus of ribonuclease, Biochemistry 10 (1971) 470–476.
[25] A. Bierzynski, P.S. Kim, R.L. Baldwin, A salt bridge stabilizes the helix formed by isolated C-peptide of RNase A, Proc. Natl. Acad. Sci. U. S. A. 79 (1982) 2470–2474.
[26] K.R. Shoemaker, et al., Nature of the charged-group effect on the stability of the C-peptide helix, Proc. Natl. Acad. Sci. U. S. A. 82 (1985) 2349–2353.
[27] S. Marqusee, R.L. Baldwin, Helix stabilization by Glu-…Lys+ salt bridges in short peptides of de novo design, Proc. Natl. Acad. Sci. U. S. A. 84 (1987) 8898–8902.
[28] S. Marqusee, V.H. Robbins, R.L. Baldwin, Unusually stable helix formation in short alanine-based peptides, Proc. Natl. Acad. Sci. U. S. A. 86 (1989) 5286–5290.
[29] J.M. Scholtz, R.L. Baldwin, The mechanism of alpha-helix formation by peptides, Annu. Rev. Biophys. Biomol. Struct. 21 (1992) 95–118.
[30] G. Merutka, W. Shalongo, E. Stellwagen, A model peptide with enhanced helicity, Biochemistry 30 (1991) 4245–4248.
[31] P.C. Lyu, P.J. Gans, N.R. Kallenbach, Energetic contribution of solvent-exposed ion pairs to alpha-helix structure, J. Mol. Biol. 223 (1992) 343–350.
[32] J.M. Scholtz, H. Qian, V.H. Robbins, R.L. Baldwin, The energetics of ion-pair and hydrogen-bonding interactions in a helical peptide, Biochemistry 32 (1993) 9668–9676.
[33] R.F. Sommese, S. Sivaramakrishnan, R.L. Baldwin, J.A. Spudich, Helicity of short E-R/K peptides, Protein Sci. 19 (2010) 2001–2005.
[34] U. Habertür, A. Caflisch, FACTS: fast analytical continuum treatment of solvation, J. Comput. Chem. 29 (2008) 701–715.
[35] A.D. Mackerell, M. Feig, C.L. Brooks III, Extending the Treatment of Backbone Energetics in Protein Force Fields: Limitations of Gas-Phase Quantum Mechanics in Reproducing Protein Conformational, J. Comput. Chem. 25 (2004) 1400.
[36] R.B. Best, et al., Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-Chain $\chi1$ and $\chi2$ dihedral angles, J. Chem. Theory Comput. 8 (2012) 3257.
[37] J.C. Phillips, et al., Scalable molecular dynamics with NAMD, J. Comput. Chem. 26 (2005) 1781–1802.
[38] B.R. Brooks, et al., CHARMM: the biomolecular simulation program, J. Comput. Chem. 30 (2009) 1545–1614.
[39] L. Adler-Abramovich, et al., Phenylalanine assembly into toxic fibrils suggests amyloid etiology in phenylketonuria, Nat. Chem. Biol. 8 (2012) 701–706.
[40] R. Fairman, K.R. Shoemaker, E.J. York, J.M. Stewart, R.L. Baldwin, Further studies of the helix dipole model: effects of a free alpha-NH_3+ or alpha-COO- group on helix stability, Proteins 5 (1989) 1–7.
[41] M. Seeber, M. Cecchini, F. Rao, G. Settanni, A. Caflisch, Wordom: a program for efficient analysis of molecular dynamics simulations, Bioinformatics 23 (2007) 2625–2627.
[42] P. Carter, C.A.F. Andersen, B. Rost, DSSPcont: continuous secondary structure assignments for proteins, Nucleic Acids Res. 31 (2003) 3293–3295.
[43] B. Zagrovic, G. Jayachandran, I.S. Millett, S. Doniach, V.S. Pande, How large is an alpha-helix? Studies of the radii of gyration of helical peptides by small-angle X-ray scattering and molecular dynamics, J. Mol. Biol. 353 (2005) 232–241.
[44] W. Kabsch, C. Sander, Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features, Biopolymers 22 (1983) 2577–2637.
[45] S. Kumar, D. Bouzida, R.H. Swendsen, P.A. Kollman, J.M. Rosenberg, The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method, J. Comput. Chem. 13 (1992) 1011–1021.
[46] J.S. Hub, B.L. de Groot, D. van der Spoel, g_wham—a free weighted histogram analysis implementation including robust error and autocorrelation estimates, J. Chem. Theory Comput. 6 (2010) 3713–3720.
[47] M. Rief, M. Gautel, F. Oesterhelt, J.M. Fernandez, H.E. Gaub, Reversible unfolding of individual Titin immunoglobulin domains by AFM, Science 276 (1997) 1109–1112.
[48] E. Paci, M. Karplus, Unfolding proteins by external forces and temperature: the importance of topology and energetics, Proc. Natl. Acad. Sci. U. S. A. 97 (2000) 6521–6526.
[49] Y. Taniguchi, B.S. Khatri, D.J. Brockwell, E. Paci, M. Kawakami, Dynamics of the coiled-coil unfolding transition of myosin rod probed by dissipation force spectrum, Biophys. J. 99 (2010) 257–262.
[50] B. Khatri, Z.T. Yew, S. Krivov, T. McLeish, E. Paci, Fluctuation power spectra reveal dynamical heterogeneity of peptides, J. Chem. Phys. 133 (2010) 015101.
[51] R.B. Best, N.-V. Buchete, G. Hummer, Are current molecular dynamics force fields too helical? Biophys. J. 95 (2008) L07–L09.
[52] D.A.E. Cochran, S. Penel, A.J. Doig, Effect of the N1 residue on the stability of the α-helix for all 20 amino acids, Protein Sci. 10 (2001) 463–470.
[53] D.A.E. Cochran, A.J. Doig, Effect of the N2 residue on the stability of the α-helix for all 20 amino acids, Protein Sci. 10 (2001) 1305–1311.
[54] T.M. Iqbalsyah, A.J. Doig, Effect of the N3 residue on the stability of the α-helix, Protein Sci. 3 (2004) 32–39.
[55] C.-L.A. Wang, et al., A long helix from the central region of smooth muscle Caldesmon, J. Biol. Chem. 266 (1991) 13958–13963.

[56] P.J. Knight, et al., The predicted coiled-coil domain of myosin 10 forms a novel elongated domain that lengthens the head, J. Biol. Chem. 280 (2005) 34702–34708.
[57] M. Peckham, P.J. Knight, When a predicted coiled coil is really a single α-helix, in myosins and other proteins, Soft Matter 5 (2009) 2493–2503.
[58] B.J. Spink, S. Sivaramakrishnan, J. Lipfert, S. Doniach, J.A. Spudich, Long single α-helical tail domains bridge the gap between structure and function of myosin VI, Nat. Struct. Mol. Biol. 15 (2008) 591–597.
[59] M. Wolny, et al., Stable single alpha helices are constant-force springs in proteins, J. Biol. Chem. (2014), http://dx.doi.org/10.1074/jbc.M114.585679.