Consensus tetratricopeptide repeat proteins are complex superhelical nanosprings

Marie Synakewicz,[1,*] Rohan S. Eapen,[1] Albert Perez-Riba,[1] Daniela Bauer,[2] Andreas Weißl,[2] Gerhard Fischer,[3] Marko Hyvönen,[3] Matthias Rief,[2] Laura S. Itzhaki,[1,*] and Johannes Stigler[4,*]

[1]*Department of Pharmacology, University of Cambridge, Tennis Court Road, Cambridge, CB2 1PD, United Kingdom*

[2]*Physik-Department, Technische Universität München, James-Franck-Straße 1, 85748 Garching, Germany*

[3]*Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge, CB2 1GA, United Kingdom*

[4]*Gene Center Munich, Ludwig-Maximilians-Universität München, Feodor-Lynen-Straße 25, 81377 München, Germany*

(Dated: March 27, 2021)

Abstract

Tandem-repeat proteins comprise small secondary structure motifs that stack to form one-dimensional arrays with distinctive dynamic properties that are proposed to direct their cellular functions. Here, we report the force response of consensus-designed tetratricopeptide repeats (CTPRs) - superhelical arrays of short helix-turn-helix motifs. Not only are we able to directly observe the repeat-protein superhelix in the force data, but we also find that individual repeats undergo rapid fluctuations between folded and unfolded conformations resulting in a spring-like mechanical response. Using protein engineering, we show how the superhelical geometry can be altered by carefully placed amino-acid substitution and subsequently employ Ising model analysis to examine how these changes affect repeat stability and inter-repeat coupling in ensemble and single-molecule experiments. The Ising models furthermore allow us to map the unfolding pathway of CTPRs under mechanical load revealing how the repeat arrays are unzipped from both ends at the same time. Our findings provide the means to dissect and modulate repeat-protein dynamics and stability, thereby supporting ongoing research efforts into their functional relevance and exploiting them for nanotechnology and biomedical applications.

Approximately one third of proteins in the human proteome contain tandem repeat units of varying sequence length and composition, and secondary and tertiary structures [1]. Tetratricopeptide repeats (TPRs), the subject of this study, belong to the so-called "tandem-repeat protein" class, members of which are characterised by the repetition of small secondary structure motifs of 20-40 amino acids in length that are stacked to form solenoidal arrays. TPRs are 34 amino acids in length, form a helix-turn-helix motif, and are found in arrays of 3 to 16 repeats in nature [2]. Long arrays of natural and designed TPRs can form superhelical structures with about 8 repeats per turn of the superhelix [3]. Many natural proteins and protein complexes contain TPR arrays, and their functions are diverse, ranging from scaffolding, regulating cell division, molecular chaperones to mediating bacterial quorum sensing [2, 4, 5]. Solenoid repeat proteins were previously shown to be flexible and have spring-like properties, both of which are thought to be crucial to their biological function [6–16].

To date, the force-induced unfolding of repeat proteins has only been studied by atomic force microscopy (AFM) and molecular dynamics (MD) simulations. However, AFM is of somewhat limited use when studying α-helical repeat proteins, as they appear to be mechanically weak and require a more sensitive detection method to fully resolve their force unfolding signature. The first experimental and MD studies suggested that a repeat array has the ability to stretch by rearranging its inter-repeat interfaces before breakage of the array or unfolding of the secondary structure elements occurs [8, 9, 17]. However, others report repeat proteins to unfold one or multiple repeats at a time upon application of an external force [10, 12, 18–21]. Studies of proteins comprising identical repeats (based on the consensus sequence of a repeat family) have been confined to ankyrin repeats. At the single molecule level, consensus ankyrin repeats were shown to unfold at 20 to 50 pN, one repeat at a time [18, 19], whereas their temperature and chemical-induced unfolding in ensemble experiments was found to be highly cooperative [22–25]. Both natural and consensus TPR structures have been the subject of intensive study, because their simple structures make them strikingly amenable to both the dissection and the redesign of their biophysical properties, with tremendous potential for exploitation in biotechnology and medicine [26, 27]. Their chemical stability has been described in depth [3, 28–36], whereas their mechanical properties remain unexplored even though they are one of the repeat protein types whose structures most closely resemble physical springs.

More recently, the unfolding behaviour of various natural and consensus repeat proteins has been explored using Ising models [25, 28]. In an Ising model, the free energy of the system can be linearly decomposed into intrinsic contributions from the repetitive units themselves and the interaction, or coupling, between neighbouring units. Furthermore, the repeating unit, a single helix or a whole repeat, can assume two states, folded and unfolded, and two neighbouring repeats can only be coupled when both are folded. It is important to note that this model implicitly includes a variety of possible folding intermediates, the abundance of which may vary between proteins and repeat types. This formalism has been applied successfully to equilibrium denaturation data of natural and consensus designed repeat proteins [3, 23, 24, 28, 31, 33–35, 37–41].

The stability of some repeat proteins, such as ankyrin repeats, are characterised by relatively weak intrinsic repeat stability and very strong inter-repeat coupling (Fig. S7), leading to very stable repeat arrays. We and others have shown that many α-helical repeats such as TPRs exhibit less of a difference between intrinsic and interfacial stabilities: only a single repeat can already be marginally stable but interfacial coupling is weaker than observed for ankyrin repeats [23, 24, 28, 32, 33, 35, 38, 40, 41]. By mutating single residues within the consensus sequence, the intrinsic and coupling contributions can be manipulated. Altogether, Ising analyses of repeat protein folding provided the scientific community with the theoretical framework to examine folding and cooperativity of both natural and designed repeat arrays: a protein is stabilised and folds with higher cooperativity as the coupling is increased, while the reverse is true when the intrinsic stability is increased [42].

Building on these findings on the mechanics and the energetic modularity of repeat proteins, we set out to examine how these two characteristics might be linked in the context of CTPRs. Unexpectedly, we found that CTPR proteins are strikingly different in their force response from other repeat proteins and indeed any protein reported to date. Using conservative mutations at the inter-repeat interface we show how such substitutions modulate the packing geometry and the energetics of the repeat array without influencing the force response of the structure itself. Lastly, we present evidence that mutagenesis of CTPRs is likely to always affect both the intrinsic and the interfacial stabilities of the repeats, which is contrary to our initial expectations and to previous findings, and will therefore have significant implications for the design and analysis of repeat proteins in the future.

RESULTS

Design and structure determination of a TPR array with altered superhelical geometry

We based our study of the structure and energetics of TPR proteins on previous research on a consensus TPR protein (referred to as CTPRa) that adopts a superhelical structure within the crystal lattice [3]. This superhelical geometry of TPRs is established through stacking of the A-helix (A_{i+1} in Fig. 1A) of a given repeat against both A and B helices of the preceding repeat (A_i and B_i in Fig. 1A) [27]. The resulting angles between repeat planes then give rise to the pitch and diameter of the superhelix (Fig. 1B) [11]. Interfaces between repeats are largely formed by bulky hydrophobic residues. Therefore, to create a CTPR variant with altered superhelical geometry, we identified four such interface residues based on sequence conservation in different TPR families [27, 43] and substituted them with polar or small hydrophobic side chains (W4L, Y5N, A10V, Y12R).

We determined the structure of a 4-repeat construct of the new variant, CTPRrv4, containing a C-terminal solvating helix by X-ray crystallography to 3.0 Å resolution (see Tab. S3 and Fig. S8). CTPRrv4 crystallised in the P 3_1 2 1 space group with two molecules per asymmetric unit. As was observed for CTPRa, the C-terminal solvating helix was not resolved due to a higher preference for end-to-end stacking between molecules in neighbouring asymmetric units [3]. The resolution was sufficient to determine the change in repeat plane angles using the C_α coordinates: while the twist remained almost unchanged, an increase in curving angle is compensated by a similar decrease in the bending angle (Fig. 1A, Tab. S4). Although these changes may at first appear insignificant in the context of a 4-repeat array, they translate into clear differences in the longer superhelical arrays, leading to a decrease in helix length and increase in helix diameter (Fig. 1B). Furthermore,

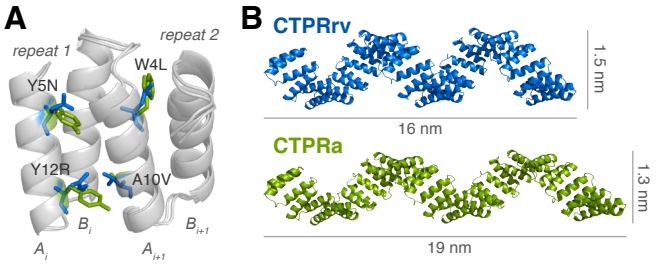

FIG. 1. Design of the CTPRrv variant (blue) based on the original CTPRa (green, PDB accession codes: 2hyz [3]). (A) Structural representation of two neighbouring CTPRs highlighting the interfacial mutations introduced in CTPRa to form CTPRrv. (B) The slight alteration in repeat packing leads to changes in the diameter and the length of the superhelix, here shown with an array of 20 repeats.

when compared to other CTPRs, all of which exhibit backbone RMSDs within 0.72 Å, the CTPRrv backbone differs by 1.4 Å when relative to the CTPRa backbone.

Single molecule force-distance data indicate equilibrium folding of CTPRs

Previously reported data on mechanical unfolding of repeat proteins was based on AFM experiments. Preliminary results in our laboratory and results reported by others [10] indicated that α-helical repeats proteins are much weaker than beta-sheet proteins such as immunoglobulin domains, and hence unfolding events occurred close to the noise limit of AFM. Therefore, we prepared CTPRrv arrays of $N = 3$, 5, 10, 20 and 26 and CTPRa arrays of $N = 5$ and 9 to be subjected to force spectroscopy using optical tweezers (Fig. 2A). Using the Sfp-enzyme, coenzyme-A modified ssDNA oligos were conjugated to N- and C-terminal ybbR tags on the protein [44]. These protein-DNA chimeras were then attached to micron-sized silica beads using dsDNA handles.

For an initial characterization of the CTPR force response, we recorded force-distance curves (FDCs) from stretch-relax cycles at a pulling velocity of 10 nm/s and 100 nm/s (Fig. 2B). The FDCs of all variants display a characteristic plateau region and a subsequent force dip, flanked by the characteristic worm-like-chain behavior of stretching the linker and the linker combined with the unfolded polypeptide. Since the length of the plateau correlated with CTPR array size, we attribute the plateau and dip regions to force-induced unfolding of CTPR repeats. The shape of the unfolding profile is indicative of sequential unfolding of repeats (plateau) until a minimally stable unit is reached, which appears to unravel in a more cooperative manner (dip). Furthermore, we noticed that unfolding and refolding FDCs of a molecule at these loading rates are almost indistinguishable when superimposed. The increased noise levels of plateau and dip indicate very fast unfolding and refolding transitions (Fig. 2B), which, together with the absence of hysteresis, suggests that CTPR unfolding and refolding occurs at equilibrium, i.e. the folding kinetics of the system under investigation are much faster than the pulling speed allowing it to"re-equilibrate instantly".

Both CTPR variants exhibit the same behaviour, albeit at different forces: whereas CTPRa variants unfolded at $\approx$12.5 pN, the force plateau of CRPRrv was significantly lower at $\approx$9.5 pN (Figs. 2B,C), indicating that the introduced mutations have a destabilizing effect. Furthermore, to ensure that the intrinsic α-helicity of the ybbR-tag [45] did not alter the stability of the repeat arrays, we compared the mechanical and chemical stability of proteins with ybbR-tags and the more commonly used cysteine modifications for maleimide-based attachment strategies (Fig. 2C and Fig. S9). Although a slight stabilisation was observed in chemical denaturation experiments, the ybbR-tag does not discernibly influence the unfolding or refolding forces, the unfolding and refolding energies (beyond the contribution of contour length), or the character of the mechanical response (Fig. 2C).

A systematic screen of pulling velocities ranging from 10 nm/s to 5 µm/s revealed that for both CTPRrv and CTPRa, the average folding and unfolding forces are marginally affected by the loading rate, leading to a slight increase in hysteresis at pulling speeds of >500 nm/s (Fig. 2D and Fig. S10). This effect is more pronounced in CTPRa arrays than in CTPRrv arrays. It is interesting to note that even at higher pulling speeds not all stretch-relax cycles exhibit hysteresis leading to a large variation within even a single molecule (e.g. see FDCs collected at 1 µm/s in Fig S9). However, most importantly, we found that there was no significant hysteresis between unfolding and refolding traces at pulling speeds $\leq$100 nm/s in both repeat types. The absence of a pulling speed-dependent folding/unfolding force is again evidence for rapid equilibrium fluctuations of CTPR subunits in this loading rate regime.

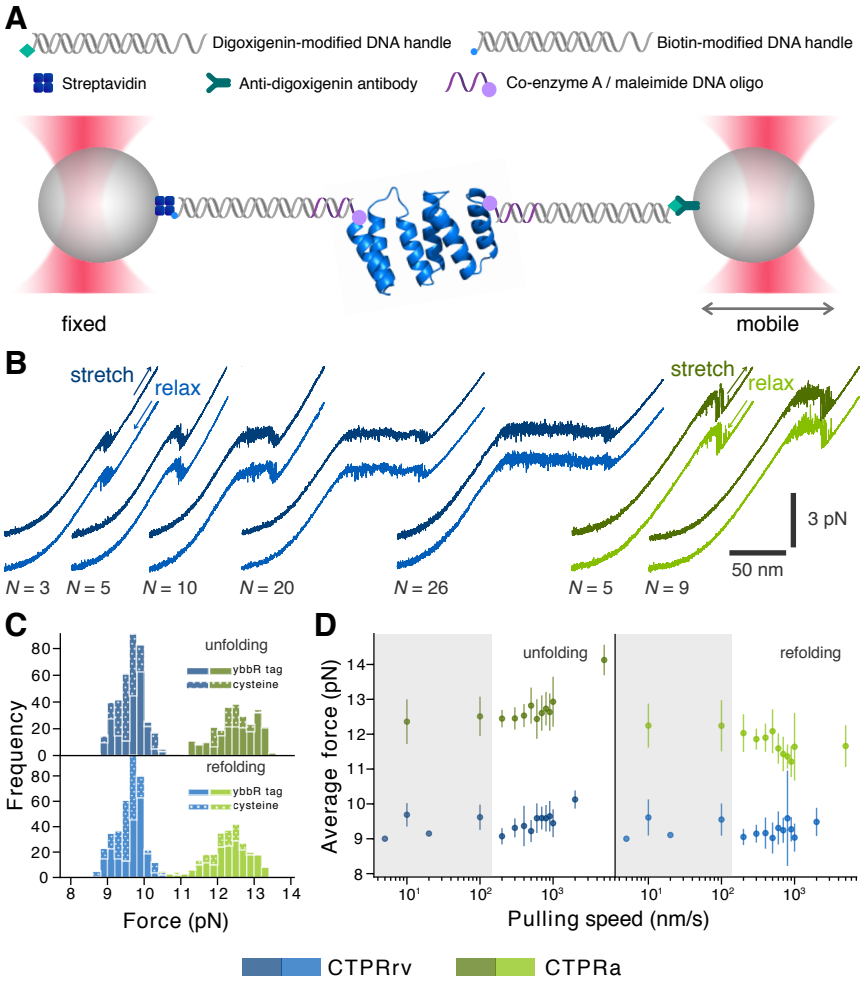

FIG. 2. Probing the mechanics of CTPRs using optical tweezers. (A) The protein of interest bears N- and C-terminal modifications that allow the site-specific attachment of ssDNA oligonucletides (cysteine-maleimide or ybbR-CoA) that can hybridize with the single-stranded overhang of dsDNA handles. The other end of these DNA handles are tethered to micron-sized silica beads using streptavidin-biotin and digoxigenin-antidigoxigenin interactions. By moving one trap away from the other force is exerted onto the protein-DNA construct resulting in the characteristic stretching of the dsDNA handles, followed by the unfolding of the protein of interest, and finally stretching of the DNA-polypeptide construct. (B) Representative force-distance curves (FDCs) for CTPRrv (blue) and CTPRa proteins (green) of different array length pulled at 10 nm/s, in which the force is plotted against the distance of the two traps. The FDCs display a characteristic force "plateau" and a final "dip" between the WLC-like stretching of the dsDNA at low forces and the DNA-polypeptide at high forces. The corresponding stretch and relax traces are offset for clarity as they would almost perfectly overlay otherwise. (C) Histograms of the unfolding and folding forces of all molecules at 10 nm/s and 100 nm/s indicate that the unfolding and refolding forces remain unaffected by the attachment method. (D) The average unfolding and refolding forces at different pulling speeds (shown as mean ± standard deviation highlight the variation between traces) show only a small loading rate the dependence of the forces. While the effect is almost negligible in the rv-type, a clear increase and decrease in unfolding and refolding forces, respectively, can be observed for the a-type. For representative FDCs at different pulling speeds see Fig. S10. The grey shaded area highlights the low pulling speeds at which the unfolding and refolding forces of both repeat types are indistinguishable, and hence the regime in which the respective system is at equilibrium.

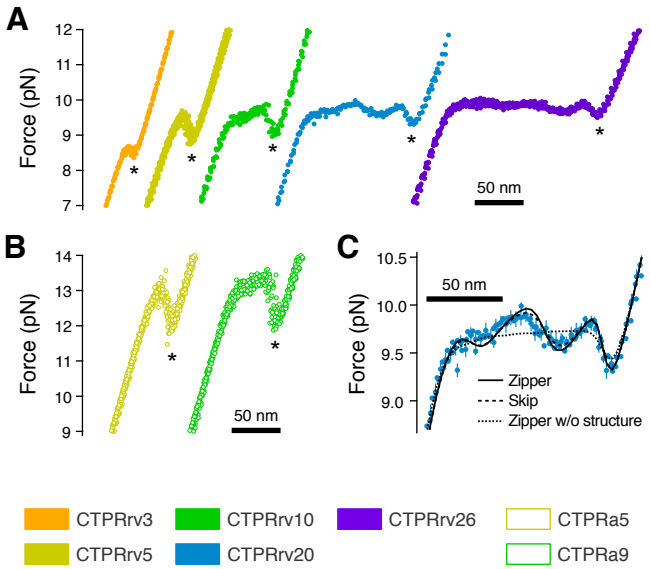

FIG. 3. The average unfolding traces of all CTPR variants show reproducible oscillations that are explained by their supramolecular structure. (A,B) Overlays of aligned equilibrium force-distance curves (FDCs) of CTPRrv and CTPRa variants in which asterisks mark the characteristic force dip after the plateau. (C) Overlay of CTPRrv20 FDCs (cyan) together with fits to the heteropolymer helix Ising models with zipper approximation (solid line), skip approximation (dashed line) and zipper approximation without taking into account structural parameters of the superhelix. In all overlays, average curves of three representative molecules are shown.

Averaged force-distance data hint at the folding mechanism of the CTPR arrays

To obtain a more detailed picture of the equilibrium folding/unfolding trajectories of CTPRs, we binned and averaged FDCs of repeated stretch and relax cycles at pulling speeds $\leq$100 nm/s for each individual molecule. Figs. 3A,B show averaged FDCs of multiple molecules for each repeat types at the chosen array lengths, overlaid and aligned along force and distance coordinates to avoid the introduction of common instrumental artefacts such as miscalibration of the trap stiffness or the zero distance point. We found the plateau region was not uniformly flat after averaging, as would be expected from other related phenomena such as DNA overstretching[46, 47], but rather it showed a highly reproducible force oscillations. This pattern was the clearest in the longer arrays of CTPRrv20 and CTPRrv26 which exhibit about 2 and 3 periods, respectively. Therefore, we reasoned that these oscillations arise directly from the structure of the superhelix. In contrast, the characteristic force dip at the end of each FDC was present in all arrays, and we hypothesise that it corresponds to the unfolding of a final stable unit.

Modelling the force response of CTPRs using mechanical Ising models

We next set out to model the phenomena observed in our data based on thermodynamic first principles [46–50]. For a given trap distance d (see Fig. S4), the Hamiltonian that describes the full energy stored in the system, $\mathcal{H}_d$ is the sum of the intrinsic folding energy of the protein and the mechanical energy for stretching the polymer linkers

$$\mathcal{H}_d(F, c) = \mathcal{H}^{\text{int}}(c) + \mathcal{H}_d^{\text{mech}}(F, c), \tag{1}$$

where F is the force at that distance and $c = \{c_1, \ldots, c_N\}$ represents a particular configuration of the 2^N-sized conformational space of an N-mer. $\mathcal{H}^{\text{mech}}$ is derived from polymer models of stretching dsDNA and unfolded polypeptide. Theoretical equilibrium FDCs, as are required for fitting data, can subsequently be obtained by numerically calculating the partition function and integrating over all degrees of freedom (see SI Methods for full detail).

We explored various Ising models to describe the intrinsic folding energy of the protein, $\mathcal{H}^{\text{int}}$. Building on recent work by us and others [24, 32, 38–41], we first generated models based on a whole repeat (i.e. one A- and B-helix) as the smallest independent protein unit (*homopolymer repeat model*, Fig. S6A, SI Methods). Here, the protein's internal energy in a given conformation is given by

$$\mathcal{H}^{\text{int}}(c) = \Delta G_{\text{unit}} \sum_{i=1}^{N} c_i + \Delta G_{\text{nn}} \sum_{i=1}^{N-1} c_i c_{i+1}, \tag{2}$$

where ΔG_{unit} is the intrinsic energy of a single subunit, ΔG_{nn} is the next-neighbour interaction, or coupling energy, between adjacent subunits, and c_i refers to a folded unit. However, this model failed to reproduce the curvature at the transition between the DNA stretch response and the protein unfolding plateau in the CTPRrv10 data (Fig. S7A). Second, we implemented a model that used individual helices as the smallest independent unit based on some of the earliest CTPR folding studies (*homopolymer helix model*, Fig. S6B, SI Methods) [28, 34]. Whereas this model described most of our data much better than did the homopolymer repeat model, we still observed significant deviations for CTPRrv5 molecules (Fig. S7B). Lastly, we expanded the homopolymer helix model to account for the differences between the A- and B-helices within a repeat, and furthermore considered the scenarios in which A and B helices can (i) only interact with their nearest neighbour (*heteropolymer helix model*, Fig. S6C, SI Methods), or (ii) can also form additional, next-nearest neighbour interactions to A- and B-helices in adjacent repeats (*heterpopolymer helix NNN model*, Figs. 1A, S6D, SI Methods). In the latter two models, the intrinsic energy is given by

$$\mathcal{H}^{\text{int}}(c) = n_{\text{AB}}\Delta G_{\text{AB}} + n_{\text{BA}}\Delta G_{\text{BA}} + n_{\text{AA}}\Delta G_{\text{AA}} + n_{\text{BB}}\Delta G_{\text{BB}} + n_{\text{A}}\Delta G_{\text{A}} + n_{\text{B}}\Delta G_{\text{B}}, \tag{3}$$

where n_{ij} count the number of adjacent folded i and j helices in a configuration c, ΔG_{ij} are the corresponding interaction energies and $\Delta G_{\text{A}}, \Delta G_B$ are the intrinsic energies of A- and B-helices (see SI Methods for details). The heteropolymer helix model without next-nearest neighbour interactions is recovered when $\Delta G_{\text{AA}} = \Delta G_{\text{BB}} = 0$. It is important to note that the both the heteropolymer helix model and the heteropolymer helix NNN model can be mapped to the repeat model using

$$\Delta G_{\text{unit}} = \Delta G_{\text{A}} + \Delta G_{\text{B}} + \Delta G_{\text{AB}} \quad \text{and} \quad \Delta G_{\text{nn}} = \Delta G_{\text{BA}} + \Delta G_{\text{AA}} + \Delta G_{\text{BB}}. \tag{4}$$

Both heteropolymer helix models were able to describe all data and yielded similar results (Tab. S5). Therefore, we proceeded with the heteropolymer helix model without next-nearest neighbour interaction to avoid over-parametrization.

All of the above models have exponential $\mathcal{O}(2^N)$ complexity and are computationally too expensive to be applied to CTPRrv20 and CTPRrv26 data. A matrix formalism, which was successfully employed to reduce the complexity of equivalent homopolymer and heteropolymer Ising models in chemical unfolding [25], could not be used to describe the mechanical unfolding because the contributions of the linker molecules (DNA and unfolded polypeptide) to the mechanical energy are non-linear. Instead, we considered alternative simplifications to reduce the conformational space by eliminating extremely unlikely high-energy configurations in two different ways. In the "skip" approximation, all configurations in which one or two folded helices are neighboured by unfolded

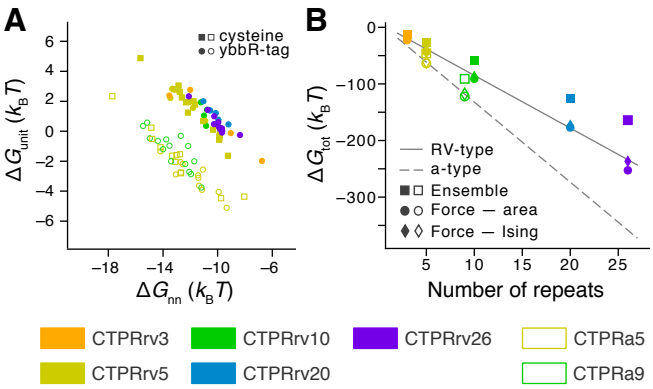

FIG. 4. Energetic contributions to the folding free energy derived from single molecule and ensemble measurements. (A) Intrinsic repeat energy ΔG_{unit} and repeat next-neighbour energy ΔG_{nn} for individual CTPRrv (filled symbols) and CTPRa molecules (empty symbols) based on the zipper model. For a direct comparison between results derived from skip and zipper approximations see Fig. S11 in the SI. (B) By calculating the total free energy using Eq. 5, a comparison between single molecule (circles, diamonds) and ensemble measurements (circle) is possible. While the values derived from our Ising models agree very well with those derived from calculating the work done by integration, it is clear that the energies from ensemble experiments increasingly deviate the longer the repeat array becomes. The solid and dashed lines indicate a simple linear regression fits to CTPRrv and CTPRa data obtained from Ising models, respectively, to guide the eye.

helices on either side are eliminated. This reduced the complexity from $\mathcal{O}(2^N)$ to $\mathcal{O}(1.65^N)$ but was still computationally too expensive to fit CTPRrv26 data. However, in the "zipper" approximation, in which unfolding occurs from the end(s), the conformational ensemble is reduced to $N(N+1)/2+1 = \mathcal{O}(N^2)$ states, and this simplification allowed us to fit data of CTPRrv26 arrays. Notably, the modelled FDCs based on these approximation differed only marginally from the data to which both approximations could be fitted (e.g. see dashed and continuous lines in Figs. 3C, S13A). Furthermore, the resulting energies of the "skip" and "zipper" approximations agreed within error (Tab. 1), and therefore, unless stated otherwise, we used the zipper approximation for all reported values.

Force oscillations are a consequence of the superhelical structure

During the initial rounds of model development we observed that none of our Ising models alone could account for the observed force oscillations in the plateau. At first, the molecular extension of the folded portion was approximated by $\xi_{\text{folded}}(c) \approx n(c)/N \cdot \xi_{\text{max}}$, where $n(c)$ is the number of folded subunits in conformation c and ξ_{max} is the end-to-end distance of the fully folded protein. However, this assumes an arrangement of subunits in the repeat array akin to beads-on-a-string. Although all models based on this assumption for the molecular extension correctly predicted the final dip, they falsely produced a flat plateau (dotted line in Fig. 3C). Only when structural parameters of the superhelix were included to account for the changes in the force vector across the folded remainder of the molecule as it unfolds, was it possible to reproduce these features. Given that our models describe the data well apart from some minor deviations, we can therefore conclude that the structure of CTPR arrays in solution is indeed very similar to that *in crystallo*.

TABLE 1. Fitted energy parameters in units of $k_{\mathrm{B}}T$, where N is the number of repeats, ΔG_{unit} the intrinsic repeat energy, ΔG_{nn} the next-neighbour interaction energy, and ΔG_{tot} the total energy for an N-mer (see Eqs. (4) and (5)).

		Equilibrium denaturation[a]			Area	Heteropolymer Ising model (Zipper approximation)			Heteropolymer Ising model (Skip approximation)		
Type	N	ΔG_{tot}	ΔG_{unit}	ΔG_{nn}	W_F[c]	ΔG_{tot}	ΔG_{unit}	ΔG_{nn}	ΔG_{tot}	ΔG_{unit}	ΔG_{nn}
rv	3	-13.0 ± 0.3			-22.7 ± 0.3	-18.4 ± 0.9	1 ± 1	-10 ± 2	-18 ± 1	1.1 ± 1.1	11 ± 2
	5	-26.1 ± 0.5			-43.4 ± 0.2	-39.7 ± 0.4	1.5 ± 0.3	-11.8 ± 0.3	-39.7 ± 0.4	1.5 ± 0.3	-11.8 ± 0.3
	10	-59 ± 1			-91 ± 1	-87 ± 3	1.2 ± 0.3	-11.0 ± 0.1	-87 ± 3	1.2 ± 0.3	-11.0 ± 0.1
	20	-125 ± 2			-177 ± 4	-173 ± 2	1.0 ± 0.3	-10.2 ± 0.2	-173 ± 2	1.1 ± 0.2	-10.3 ± 0.2
	26	-164 ± 3			-252.7 ± 0.6	-237 ± 2	0.5 ± 0.2	-10.0 ± 0.3	N/A[d]	N/A[d]	N/A[c]
	all[b]		0.20 ± 0.05	-6.8 ± 0.1			1.1 ± 0.2	-11.0 ± 0.2		1.3 ± 0.2	-11.3 ± 0.3
a	5	-46 ± 1			-63.7 ± 0.4	-61.3 ± 0.6	-2.4 ± 0.4	-12.4 ± 0.4	-61.4 ± 0.6	-2.4 ± 0.4	-12.4 ± 0.4
	9	-91 ± 2			-122.2 ± 0.9	-118 ± 2	-1.3 ± 0.3	-13.3 ± 0.4	-117 ± 2	-1.3 ± 0.3	-13.2 ± 0.4
	all[b]		-1.0 ± 0.2	-10.3 ± 0.1			-1.9 ± 0.3	-12.7 ± 0.3		-1.9 ± 0.3	-12.7 ± 0.3

[a] Results of the rv-type are based on a global fit to data of arrays with $N = 2$, 4, 5, 8 and 10 repeats [51], see Fig. S9. Values of the a-type are as reported previously [32]. The errors shown here were propagated from the global fit.

[c] Refolding energies as approximated from the work done by the protein, see Eq. (7) in the SI. Values are reported as mean±s.e.m. for each repeat length.

[d] Values for the heteropolymer model in the skip approximation could not be computed because of limited computational capacity.

[b] Combined value for all repeat lengths of a type.

Comparing intrinsic and interfacial energies between single molecule and ensemble data

Using our models, we could now determine the intrinsic and interfacial contributions to the free energy, and extrapolated the repeat stabilities and nearest-neighbour coupling from the individual terms of the A and B helices using Equation 4 (Fig. 4A and Tab. 1). Due to the interdependence of the model parameters the data cluster diagonally, i.e. destabilisation of one energetic term is compensated with stabilisation of the other. However, the data of the two variants (filled vs. empty symbols) clearly separate and are independent of attachment method (squares vs. circles). When averaging over the whole data set we found that the interfacial energy ($\Delta G_{\mathrm{nn}}^{\mathrm{CTPRrv}} = -11.0 \pm 0.2\, k_{\mathrm{B}}T$, $\Delta G_{\mathrm{nn}}^{\mathrm{CTPRa}} = -12.7 \pm 0.3\, k_{\mathrm{B}}T$) vastly outweighs the intrinsic energy ($\Delta G_{\mathrm{unit}}^{\mathrm{CTPRrv}} = 1.1 \pm 0.2\, k_{\mathrm{B}}T$, $\Delta G_{\mathrm{unit}}^{\mathrm{CTPRa}} = -1.9 \pm 0.3\, k_{\mathrm{B}}T$) for both repeat types. Next, we determined the total free energy derived from our Ising models using

$$\Delta G_{\mathrm{tot}} = N\Delta G_{\mathrm{unit}} + (N-1)\Delta G_{\mathrm{nn}} \tag{5}$$

for each variant and array length (diamonds in Fig. 4B and Tab. 1). These values could then be compared to (i) values derived from the area under the curve of FDCs (circles in Fig. 4B and Tab. 1), and (ii) those calculated from ensemble Ising models using Eq. 5 (squares in Fig. 4B, Fig. S9 and Tab. 1). While there is excellent agreement between methods based on single molecule data, validating our mechanical Ising model approach, energies derived from ensemble measurement severely underestimated the total free energy. Previous data published by Regan and co-workers indicated that CTPRs can form polyproline-II helices at high GdnHCl concentration [52]. The presence of such non-native structures indicates that the free energy derived from ensemble experiments may not be representative of the energetic difference between the folded and unfolded states but rather that between the folded state and a state with considerable secondary structure. In contrast, the completely unfolded state is reached in mechanical experiments, as can be judged from its contour length and entropic elasticity.

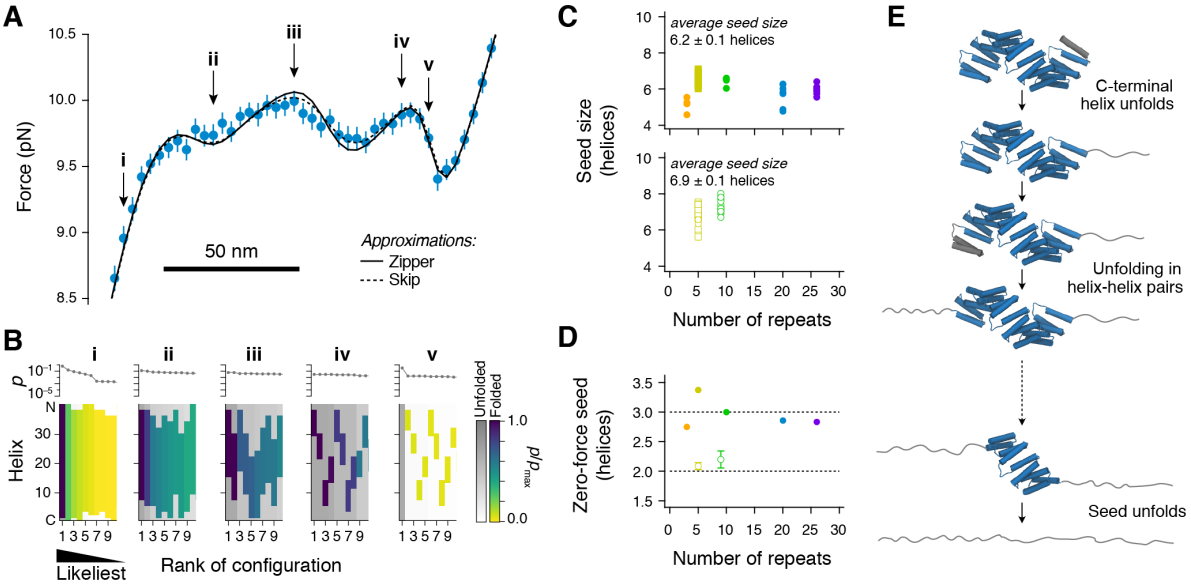

FIG. 5. Gaining insights into the folding pathway of CTPR proteins. (A) Average force-distance data for CTPRrv20 fitted to both zipper and skip approximations of the heteropolymer helix model. Roman numerals point to distances for which snapshots of the conformational ensemble are shown B. (B) From the Ising model, we can extract the ten likeliest configurations at each of the indicated distances in A, ranked according to their probability from highest (index 1) to lowest (shown above the colour maps). Shown are the results for the skip approximation, which does *not* explicitly enforce unfolding from the ends. Coloured bars in the map refer to segments of helices that are folded in a given configuration, with the exact shade giving the relative probability of being folded. Bars in grey-scale on the other hand represent helices that are unfolded, with the exact shade being the corresponding relative probability. Please note, that the N-C-terminal direction is numerically reversed, the C-terminal helix having the index 0 on the y-axis. (C) Average size of the minimally stable folding unit in force experiments for rv- (top) and a-type repeats (bottom). Symbols and colours are the same as those in Fig. 4A. (D) Inferred average minimal stable folding unit in the absence of force for rv-type (filled circles) and a-type variants (empty circles). Error bars show the standard error of the mean. (E) A model for the force-induced unfolding of CTPRs.

The (un)folding pathway of CTPR arrays

Based on our Ising analyses, we were able to develop a model of the likeliest unfolding (and folding) pathway for CTPR proteins under mechanical load. To this end, we chose to examine results of the skip approximation in more detail to avoid possible bias introduced by the assumption that arrays unfold from the end as it is done in the zipper approximation. For each trap distance we recorded the likeliest conformations ranked by their probabilities. Examples at six different distances along the average unfolding profile of CTPRrv20 are shown in Fig. 5 (and Fig. S12 for CTPRrv26), along with the probabilities of a given configuration and the relative probabilities of individual helices being folded or unfolded. Our data show that unfolding preferentially occurs from the ends in a zipper-like fashion. However, at increasing distances, there are several, almost equally likely conformations that have different segments of folded and unfolded helices. When plotting the probability of being folded for all helices as a function of distance (Fig. S13), we furthermore noticed that unfolding starts with the most C-terminal helix, and then propagates one or two helices at a time from the C-terminus (i.e. A and B helices alternating, or almost in pairs of $B_{i-1}A_i$) and one repeat at a time from the N-terminus (i.e. in pairs of A_iB_i).

Next, we extracted the microscopic information on the characteristic "dip" feature at the end of the plateau region. Traditionally, one would have expected that the dip is caused by the cooperative

folding/unfolding of a well-defined minimally stable unit of a certain length that exchanges in a two-state manner with the unfolded state. However, our model intriguingly indicates that instead of a clear two-state transition, the dip ((iv)–(v) in Fig. 5A,B) represents inter-conversions between a large ensemble of marginally stable conformations of varying size between one another and the unfolded conformation. For example, at position (v) in Fig. 5A, the likeliest conformation is the unfolded state. However, this state exchanges rapidly with an ensemble of conformations with ≈6 to 7 consecutively folded helices that individually are less populated than the unfolded state, but together amount to about 50 % of the conformational space (Fig. 5B(v)). We then calculated the average number of folded helices once the fully unfolded conformations reaches a likelihood of 50 % to extract the length of the (un)folding "seed". As expected, the size of this seed was independent of array length, and comprised approximately 6.2 ± 0.1 and 6.9 ± 0.1 helices for the rv- and a-type arrays, respectively (Fig. 5C), which agrees with the estimates of the contour length increase based on the raw data (Fig. S14). However, it is important to stress again, that these seeds do not cooperatively exchange with the unfolded conformation in a two-state manner, nor should they be mistaken for the minimal folded unit under zero-force such as one would obtain in ensemble measurements. To get an estimate of the minimal folding unit at zero force, we calculated the minimal number of consecutive folded helices needed in order to achieve a negative ΔG_{tot} using Eq. (5). The resulting inferred zero-force "seeds" were independent of variant length and were close to 3 helices for CTPRrv and 2 helices for CTPRa (Fig. 5D). Therefore, based on our results, a single CTPRa repeat is weakly stable ($-1.9 \pm 0.3\, k_{\mathrm{B}}T$) and may fold transiently on its own. On the contrary, a single CTPRrv repeat is unstable ($1.1 \pm 0.2\, k_{\mathrm{B}}T$) and requires the energy from the interface with at least one more helix to fold. These results show the same trend as our ensemble denaturation data, even if they differ numerically.

DISCUSSION

CTPR solenoids unzip from both ends under force

We report the first high-resolution force spectroscopy data of a solenoid repeat protein obtained using optical tweezers, giving insights into their thermodynamic properties and possible folding pathways. Our protein model of choice was the CTPR, which has never been studied by force spectroscopy but whose thermodynamics and kinetics have been thoroughly characterised by chemical denaturation studies. In contrast to earlier AFM studies of other repeat proteins that showed that their unfolding proceeds one repeat at a time, we show that CTPRs unfold in a single plateau-like transition that ends in a dip. The plateau arises from the asymmetric unzipping of helices or repeats from either end of the repeat array, starting with the C-terminal helix. The very fast equilibrium fluctuations of the C-terminal helix, and possibly the first N-terminal repeat, are beyond the time resolution of our instrument and hence the transition from DNA stretching into the plateau is rather smooth (i.e. "averaged"), particularly in FDCs of the rv-type arrays. The asymmetry of the unfolding pathway arises directly from the structure: (i) the C-terminal B-helix is exposed without any interactions beyond those to the corresponding A-helix of the last repeat of the folded protein/remainder; (ii) the force vector aligns the molecule such that the C-terminal helix can get "zipped off" while at the N-terminus a whole repeat experiences shear forces; (iii) a $\mathrm{B}_{i-1}\mathrm{A}_i$ repeat is structurally different from an $\mathrm{A}_i\mathrm{B}_i$ repeat, leading to different unfolding patterns from either end as the protein is unzipped by force (Fig. S13). As such, our findings on the unfolding pathway are consistent with the literature showing that unfolding of various consensus repeat arrays occurs from either end, albeit with different probabilities. They are furthermore supported by the fact that the terminal repeats of CTPRs are subject to higher hydrogen-deuterium exchange

rates [36, 53], and hence they are more likely to explore the unfolded state than more central repeats. Our Ising models even confirm the folding correlation length of 3 repeats and the possibility of multiple folding "seeds" that were previously determined based on coarse-grained simulations [42].

To our own initial surprise, we were not able to capture the repeat-by-repeat, saw-tooth unfolding patterns observed for other solenoid repeat proteins. However, using our mechanical Ising model, we increased the effective spring constant of the system connected to the protein of interest (optical trap and linker molecules) to simulate the much stiffer compliance of surface and cantilever in AFM experiments. This allowed us to reproduce the characteristic saw-tooth pattern of repeats unfolding one at a time for a consensus ankyrin-repeat protein with five repeats (Fig. S15), highlighting that this behaviour is, at least in part, related to the stiffness of the experimental apparatus.

The tight interdependence of intrinsic and interfacial stabilities makes CTPRs highly malleable

Originally, we set out to create a CTPR array with a different superhelical geometry by making conservative substitutions at the interface between repeats. Our initial assumptions was that by changing only interface interactions, we would affect solely the interfacial coupling energy - akin to how the intrinsic repeat stability was modified in a previous study [34]. However, although the chosen mutations did alter the helix packing between repeats more significantly than that within a repeat, both ensemble and single molecule data show that these mutations affected both energetic parameters, i.e. coupling and intrinsic stability. In fact, the mechanical Ising models indicate that the overall lower stability of the rv-type arrays relative to the original CTPR arrays is due to an almost equal destabilization of both intrinsic and interfacial energies ($\Delta\Delta G_{\mathrm{unit}} \approx 3\,k_{\mathrm{B}}T$ and $\Delta\Delta G_{\mathrm{nn}} \approx 2\,k_{\mathrm{B}}T$, respectively). We therefore suspect that due to the tight interdependence of these two energetic parameters in combination with the structural flexibility of CTPR arrays, most mutations will be compensated by rearrangements in packing geometry and will therefore affect both intrinsic and interfacial energies. Modification of residues on the outside of helices that are not involved in any inter-repeat interactions may simply have a much smaller or no effect on the interfacial coupling (as long as such a modification does not cause any structural rearrangements or change other side-chain interactions).

Our data furthermore support previous findings by Cortajarena *et al.* [52] that the denatured states of CTPRs cannot be described as random coils. The mismatch observed between total free energies derived from ensemble and single molecule measurements is unexpectedly high, suggesting that Ising models of chemical denaturation data of CTPRs should be interpreted with care as they cannot access the fully unfolded state. The transient formation of non-native contacts, such as poly-proline-II helices, could indeed be the reason for the increase in hysteresis observed at higher pulling speeds. Future research efforts should therefore explore whether there is a correlation between Ising parameters obtained from ensemble and single-molecule experiments.

Force response of the tertiary structure

Previously, several groups used steered MD simulations of natural repeat proteins to show that the repeat array as a whole could stretch significantly by rearranging interfaces, i.e. without unfolding of any secondary structure [8, 9, 14]. However, at this time, we cannot say conclusively whether or not the CTPR superhelix stretches before unfolding starts at the ends. On the one hand, we did not incorporate an additional term into our models to account for this, primarily

to avoid over-parameterization. On the other hand, we did not find sufficient evidence that such a response is of similar compliance to DNA and could therefore be hidden in the DNA response. Indeed, it appears to be the reverse: since DNA parameters compensate for the dimension of the folded construct, we can estimate the end-to-end distances prior to unfolding using a linear regression and obtain values that agree well with our structural data (Fig. S16). Given that we can clearly "see" the superhelix in the force plateau, we believe that the interfaces in CTPR arrays are coupled too strongly to rearrange, which is supported by previous findings that describe packing between CTPRs as rather rigid [54]. This coupling is likely much stronger than the intrinsic stability of the helices and repeats at either end of the array, and therefore unzipping occurs before any stretching can be observed. It remains to be shown how careful modulation of intrinsic repeat stability and interfacial coupling will alter the overall stiffness of the tertiary structure to account for the flexibility of many natural repeat proteins observed in both simulations and experiments.

MATERIALS AND METHODS

For a detailed description see the supplementary materials and methods section. In brief, repeat arrays were constructed in the background of a pRSET vector and expressed in *E. coli*. Equilibrium denaturations were performed using guanidine hydrochloride in sodium phosphate buffer pH 6.8, 150 mM NaCl in a 96-well plate format [55]. CTPRrv4 with a C-terminal solvating helix was crystallized in a solution containing 0.2 M $MgCl_2$, 0.1 M sodium Cacodylate pH 6.5 and 50% v/v PEG200. Further details on collection and data processing can be found in the supplementary materials and methods section. Angles between repeat planes were calculated essentially as published previously [11]. Constructs were prepared for force spectroscopy using site-specific modification of either terminal ybbR-tags or cysteine residues [56, 57]. All single molecule force spectroscopy data was collected on a custom-built instrument [58], processed using custom scripts developed in Igor Pro (WaveMetrics) and further analysed using Igor Pro or Python [59–65]. Theoretical FDCs were calculated using custom C++/CUDA software. Structural representations were generated using PyMol [66] or VMD [67].

Data availability

All scripts and data are available upon reasonable request. The structure of CTPRrv4 is will be made available on the PDB prior to publication.

Declaration of conflict of interest

The authors declare no conflicting interests.

Author contributions

A.P.-R. designed the protein variants. A.P.-R., G.F and R.S.E. performed the crystallography and structural refinement. M.S. conducted all other experiments, and M.S. and J.S. analysed the data. D.B. and A.W. provided experimental guidance and maintained the equipment. M.S., L.S.I. and J.S. wrote the manuscript. All authors reviewed the manuscript.

Acknowledgements

MS and AP acknowledge PhD studentship funding from the BBSRC Doctoral Training Partnerships, Cambridge. MS acknowledges additional support from a travel grant provided by the Non-globular Protein Network (COST Action BM1405-39176) and the Eric Reid Fund for Methodology from the British Biochemical Society, and AP from an Oliver Gatty Studentship. RE acknowledges funding of a PhD studentship from AstraZeneca. MR acknowledges funding by the Deutsche Forschungsgemeinschaft (DFG) through SFB863-A2 1111 6624. LSI acknowledges the support of an Senior Fellowship from the Medical Research Foundation of the UK. JS acknowledges the support by the LMU Center for Nanoscience CeNS and a DFG Emmy Noether grant STI673/2-1.

---

* To whom correspondence should be addressed: lsi10@cam.ac.uk, m.synakewicz@bioc.uzh.ch and stigler@genzentrum.lmu.de

[1] M. Pellegrini, E. M. Marcotte, and T. O. Yeates, Proteins: Structure, Function, and Bioinformatics **35**, 440 (1999).
[2] L. D. D'Andrea and L. Regan, Trends in Biochemical Sciences **28**, 655 (2003).
[3] T. Kajander, A. L. Cortajarena, S. Mochrie, and L. Regan, Acta Crystallographica Section D **63**, 800 (2007).
[4] H. Do and M. Kumaraswami, Journal of Molecular Biology **428**, 2793 (2016).
[5] Z. Zhang, L. Chang, J. Yang, N. Conin, K. Kulkarni, and D. Barford, Journal of Molecular Biology **425**, 4236 (2013).
[6] S. S. Cohen, I. Riven, A. L. Cortajarena, L. De Rosa, L. D. D'Andrea, L. Regan, and G. Haran, Journal of the American Chemical Society **137**, 10367 (2015).
[7] J. A. Lamboy, H. Kim, K. S. Lee, T. Ha, and E. A. Komives, Proceedings of the National Academy of Sciences **108**, 10178 (2011).
[8] A. Grinthal, I. Adamovic, B. Weiner, M. Karplus, and N. Kleckner, Proceedings of the National Academy of Sciences **107**, 2467 (2010).
[9] C. Kappel, U. Zachariae, N. Dölker, and H. Grubmüller, Biophysical Journal **99**, 1596 (2010).
[10] M. Kim, K. Abdi, G. Lee, M. Rabbi, W. Lee, M. Yang, C. J. Schofield, V. Bennett, and P. E. Marszalek, Biophysical Journal **98**, 3086 (2010).
[11] J. K. Forwood, A. Lange, U. Zachariae, M. Marfori, C. Preast, H. Grubmüller, M. Stewart, A. H. Corbett, and B. Kobe, Structure **18**, 1171 (2010).
[12] D. Serquera, W. Lee, G. Settanni, P. E. Marszalek, E. Paci, and L. S. Itzhaki, Biophysical Journal **98**, 1294 (2010).
[13] M. Baclayon, P. v. Ulsen, H. Mouhib, M. H. Shabestari, T. Verzijden, S. Abeln, W. H. Roos, and G. J. L. Wuite, ACS Nano **10**, 5710 (2016).
[14] M. Sotomayor, D. P. Corey, and K. Schulten, Structure **13**, 669 (2005).
[15] N. Fukuhara, E. Fernandez, J. Ebert, E. Conti, and D. Svergun, Journal of Biological Chemistry **279**, 2176 (2004).
[16] U. Zachariae and H. Grubmüller, Structure **16**, 906 (2008).
[17] G. Lee, K. Abdi, Y. Jiang, P. Michaely, V. Bennett, and P. E. Marszalek, Nature **440**, 246 (2006).
[18] L. Li, S. Wetzel, A. Plückthun, and J. M. Fernandez, Biophysical Journal **90**, L30 (2006).
[19] Q. Li, Z. N. Scholl, and P. E. Marszalek, Angewandte Chemie International Edition in English **53**, 13429 (2014).
[20] W. Lee, X. Zeng, H. X. Zhou, V. Bennett, W. Yang, and P. E. Marszalek, J Biol Chem **285**, 38167 (2010).
[21] G. Settanni, D. Serquera, P. E. Marszalek, E. Paci, and L. S. Itzhaki, PLoS Comput Biol **9**, e1002864 (2013).
[22] H. Binz, M. T. Stumpp, P. Forrer, P. Amstutz, and A. Plückthun, Journal of Molecular Biology **332**, 489 (2003).

[23] S. K. Wetzel, G. Settanni, M. Kenig, H. K. Binz, and A. Plückthun, Journal of Molecular Biology **376**, 241 (2008).
[24] T. Aksel, A. Majumdar, and D. Barrick, Structure **19**, 349 (2011).
[25] T. Aksel and D. Barrick, in *Biothermodynamics, Part A*, Methods in Enzymology, Vol. 455, edited by M. L. Johnson, J. M. Holt, and G. K. Ackers (Academic Press, 2009) Chap. 4, pp. 95–125.
[26] A. K. Das, P. W. Cohen, and D. Barford, EMBO J. **17**, 1192 (1998).
[27] E. R. Main, Y. Xiong, M. J. Cocco, L. D'Andrea, and L. Regan, Structure **11**, 497 (2003).
[28] T. Kajander, A. L. Cortajarena, E. R. G. Main, S. G. J. Mochrie, and L. Regan, Journal of the American Chemical Society **127**, 10188 (2005).
[29] A. L. Cortajarena, S. G. Mochrie, and L. Regan, Journal of Molecular Biology **379**, 617 (2008).
[30] A. L. Cortajarena and L. Regan, Protein Sci. **20**, 336 (2011).
[31] Y. Javadi and E. R. G. Main, Proceedings of the National Academy of Sciences **106**, 17383 (2009).
[32] A. Perez-Riba, A. R. Lowe, E. R. G. Main, and L. S. Itzhaki, Biophysical Journal **114**, 2552 (2018).
[33] C. Millership, J. J. Phillips, and E. R. Main, Journal of Molecular Biology **428**, 1804 (2016).
[34] A. L. Cortajarena, S. G. J. Mochrie, and L. Regan, Protein Science **20**, 1042 (2011).
[35] J. J. Phillips, Y. Javadi, C. Millership, and E. R. Main, Protein Sci. **21**, 327 (2012).
[36] A. Perez-Riba, E. Komives, E. R. G. Main, and L. S. Itzhaki, Sci Rep **9**, 15439 (2019).
[37] C. C. Mello and D. Barrick, Proceedings of the National Academy of Sciences **101**, 14102 (2004).
[38] J. Marold, J. Kavran, G. Bowman, and D. Barrick, Structure **23**, 2055 (2015).
[39] T. Aksel and D. Barrick, Biophysical Journal **107**, 220 (2014).
[40] K. Geiger-Schuller and D. Barrick, Biophysical Journal **111**, 2395 (2016).
[41] K. Geiger-Schuller, K. Sforza, M. Yuhas, F. Parmeggiani, D. Baker, and D. Barrick, Proceedings of the National Academy of Sciences **115**, 7539 (2018).
[42] D. U. Ferreiro, A. M. Walczak, E. A. Komives, and P. G. Wolynes, PloS Computational Biology **4**, e1000070 (2008).
[43] T. J. Magliery and L. Regan, Journal of Molecular Biology **343**, 731 (2004).
[44] J. Yin, A. J. Lin, D. E. Golan, and C. T. Walsh, Nature Protocols **1**, 280 (2006).
[45] J. Yin, P. D. Straight, S. M. McLoughlin, Z. Zhou, A. J. Lin, D. E. Golan, N. L. Kelleher, R. Kolter, and C. T. Walsh, Proceedings of the National Academy of Sciences **102**, 15815 (2005).
[46] J. M. Huguet, C. V. Bizarro, N. Forns, S. B. Smith, C. Bustamante, and F. Ritort, Proceedings of the National Academy of Sciences **107**, 15431 (2010).
[47] U. Bockelmann, B. Essevaz-Roulet, and F. Heslot, Physical Review Letters **79**, 4489 (1997).
[48] T. Bornschlögl and M. Rief, Physical Review Letters **96**, 118102 (2006).
[49] G. Zoldák, J. Stigler, B. Pelz, H. Li, and M. Rief, Proceedings of the National Academy of Sciences **110**, 18156 (2013).
[50] A. Solanki, K. Neupane, and M. T. Woodside, Physical Review Letters **112**, 161 (2014).
[51] A. R. Lowe, A. Perez-Riba, L. S. Itzhaki, and E. R. Main, Biophysical Journal **114**, 511 (2018).
[52] A. L. Cortajarena, G. Lois, E. Sherman, C. S. O'Hern, L. Regan, and G. Haran, Journal of Molecular Biology **382**, 203 (2008).
[53] E. R. G. Main, K. Stott, S. E. Jackson, and L. Regan, Proceedings of the National Academy of Sciences **102**, 5721 (2005).
[54] C. Y. Cheng, V. A. Jarymowycz, A. L. Cortajarena, L. Regan, and M. J. Stone, Biochemistry **45**, 12175 (2006).
[55] A. Perez-Riba and L. S. Itzhaki, Scientific Reports **7**, 9071 (2017).
[56] M. Synakewicz, D. Bauer, M. Rief, and L. S. Itzhaki, Sci Rep **9**, 13820 (2019).
[57] A. Mukhortava and M. Schlierf, Bioconjugate Chemistry **27**, 1559 (2016).
[58] Y. von Hansen, A. Mehlich, B. Pelz, M. Rief, and R. R. Netz, Review of Scientific Instruments **83**, 095116 (2012).
[59] K. J. Millman and M. Aivazis, Computing in Science & Engineering **13**, 9 (2011).
[60] T. E. Oliphant, Computing in Science & Engineering **9**, 10 (2007).
[61] S. v. d. Walt, S. C. Colbert, and G. Varoquaux, Computing in Science & Engineering **13**, 22 (2011).
[62] J. D. Hunter, Computing In Science & Engineering **9**, 90 (2007).
[63] M. Waskom and the seaborn development team, "mwaskom/seaborn," (2020).
[64] E. Jones, T. Oliphant, P. Peterson, *et al.*, "SciPy: Open source scientific tools for Python," (2001), http://www.scipy.org/.

[65] Wes McKinney, in *Proceedings of the 9th Python in Science Conference*, edited by Stéfan van der Walt and Jarrod Millman (2010) pp. 56 – 61.
[66] Schrödinger, LLC, "The PyMOL molecular graphics system, version 1.8," (2015).
[67] W. Humphrey, A. Dalke, and K. Schulten, Journal of Molecular Graphics **14**, 33 (1996).