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: August 20, 2021)

Abstract. Tandem-repeat proteins comprise small secondary structure motifs that stack to form one-dimensional arrays with distinctive mechanical properties that are proposed to direct their cellular functions. Here, we use single-molecule optical tweezers to study the folding of consensus-designed tetratricopeptide repeats (CTPRs) — superhelical arrays of short helix-turn-helix motifs. We find that CTPRs display a spring-like mechanical response in which individual repeats undergo rapid equilibrium fluctuations between folded and unfolded conformations. We rationalise the force response using Ising models and dissect the folding pathway of CTPRs under mechanical load, revealing how the repeat arrays form from the centre towards both termini simultaneously. Strikingly, we also directly observe the protein's superhelical tertiary structure in the force signal. Using protein engineering, crystallography and single-molecule experiments, we show how the superhelical geometry can be altered by carefully placed amino-acid substitutions and examine how these sequence changes affect intrinsic repeat stability and inter-repeat coupling. Our findings provide the means to dissect and modulate repeat-protein stability and dynamics, which will be essential for researchers to understand the function of natural repeat proteins and to exploit artificial repeats proteins in nanotechnology and biomedical applications.

Significance statement. Repetition of biological building blocks is crucial to modulating and diversifying structure and function of biomolecules across all organisms. In tandem-repeat proteins, the linear arrangement of small structural motifs leads to the formation of striking supramolecular shapes. Using a combination of single-molecule biophysical techniques and modelling approaches, we dissect the spring-like nature of a designed repeat protein and demonstrate how its shape and mechanics can be manipulated by design. These novel insights into the biomechanical and biochemical characteristics of this protein class give us a methodological basis from which to understand the biological functions of repeat proteins and to exploit them in nanotechnology and biomedicine.

Classification. Physical sciences – Biophysics and Computational Biology

Keywords. Repeat proteins | Protein folding | Ising models | Protein mechanics | Optical tweezers

INTRODUCTION

Approximately one third of proteins in the human proteome contain repetitive motifs of varying size and structural composition [1, 2]. Within this group, members of the tandem-repeat protein class stand out due to their striking three-dimensional shapes that arise from the stacking of small secondary structure motifs of 20 to 40 amino acids into either quasi-one-dimensional arrays (solenoids) or doughnut-like shapes (toroids). Examples include tetratricopeptide, ankyrin, HEAT and armadillo repeats. Due to their structural simplicity, repeat proteins have been recognised very early to have tremendous potential for applications in nanotechnology, e.g. as synthetic biomaterials, and in biomedicine as antibody alternatives [3, 4]. Previous experiments employed ensemble biochemical techniques to study the folding of repeat proteins and to address the question how alterations in protein sequence translate to changes in protein stability, dynamics and structure [5–7]. However, after two decades of research, it is not yet clear how exactly sequence, shape, stability and dynamics of individual repeats translate into the particular thermodynamic and mechanical properties of the whole array.

Many all-helical repeat proteins look like springs and have indeed been shown to be flexible molecules with spring-like properties, both of which are thought to be crucial to their biological functions [8–18]. Despite this defining feature, our understanding of the mechanics has been limited to date, because the methodology of choice — atomic force microscopy (AFM) — lacks sensitivity in the low-pN regime relevant for these these α-helical proteins. To address the limitations of current mechanical measurements of repeat proteins, we employ optical tweezers, which have been used to directly observe conformational transitions close to equilibrium in the low-pN range [19]. Long-term stability and high time-resolution enable us to simultaneously manipulate single repeat proteins, study their spring-like mechanics, and provide detailed information on their dynamics and equilibrium energetics.

Our research focuses on the tetratricopeptide repeat (TPR), which comprises a helix-turn-helix motif and is found in arrays of 3 to 16 repeats in nature [20, 21]. The packing of the TPR motif results in superhelical structures [22], which means that, of all the different repeat-protein types, TPR arrays most closely resemble a physical spring. The functions of TPR proteins are diverse, ranging from scaffolds of multi-protein assemblies and regulators of cell division to molecular chaperones and mediators of bacterial quorum sensing [21, 23–25]. Consensus-designed TPRs (CTPRs) are good candidates for building 'made to measure' proteins, because they form stable arrays and are very amenable to manipulation including loop insertions [26–28]. Their chemical stability has been characterised previously [22, 27, 29–36], whereas, in contrast to other repeat proteins [12, 14, 37–41], their mechanical properties remain unexplored. For the above reasons, we chose TPRs to examine how repeat energetics are connected to both the shape and the mechanics of the superhelix.

To achieve this goal, we rationally re-design the geometry of the TPR superhelix by substituting residues at the repeat interface, and then characterise their biophysical properties using Ising models and single-molecule force spectroscopy. Intriguingly, we find that the force response of CTPRs is very different from that of any other protein reported to date. In particular, their rapid dynamics allow them to unfold and refold at equilibrium over a large range of loading rates. We show that by using force we can access the full energy landscape of the CTPR array, which was previously impossible and which allows us to now accurately determine the effect of the chosen mutations. Collectively, our methodology and findings represent an important first step in formulating and redefining future research directions to link repeat protein structure to function.

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 of a consensus TPR protein (referred to as CTPRa) that adopts a superhelical structure within the crystal lattice [22]. 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) [26]. The resulting angles between repeat planes then give rise to the pitch and diameter of the superhelix (Fig. 1B) [13]. 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 [26, 42] 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 containing a C-terminal solvating helix, CTPRrv4, by X-ray crystallography to 3.0 Å resolution (see Tab. S3 and Fig. S6). 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 [22]. 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, when compared to other CTPRs, all of which exhibit backbone RMSDs within 0.72 Å, the CTPRrv backbone differs by 1.4 Å relative to the CTPRa backbone.

Single-molecule force-distance data indicate equilibrium folding of CTPRs

To examine the mechanical folding and unfolding of the CTPR superhelix, we prepared CTPRrv arrays of $N = 3$, 5, 10, 20 and 26 and CTPRa arrays of $N = 5$ and 9 for force spectroscopy measurements using optical tweezers (Fig. 2A, [43]). Using the Sfp-enzyme, coenzyme-A modifed 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

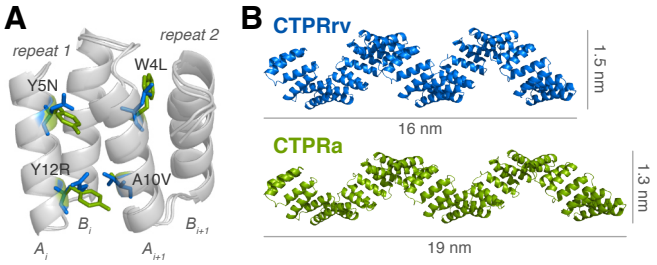

FIG. 1. Design of the CTPRrv variant (blue) based on the original CTPRa (green, PDB accession code: 2hyz [22]). (A) Structural representation of two neighbouring CTPRs highlighting the interfacial mutations introduced in CTPRa to form CTPRrv. (B) The slight alteration in repeat packing leads to changes in the diameter and the length of the superhelix, here shown with an array length of 20 repeats.

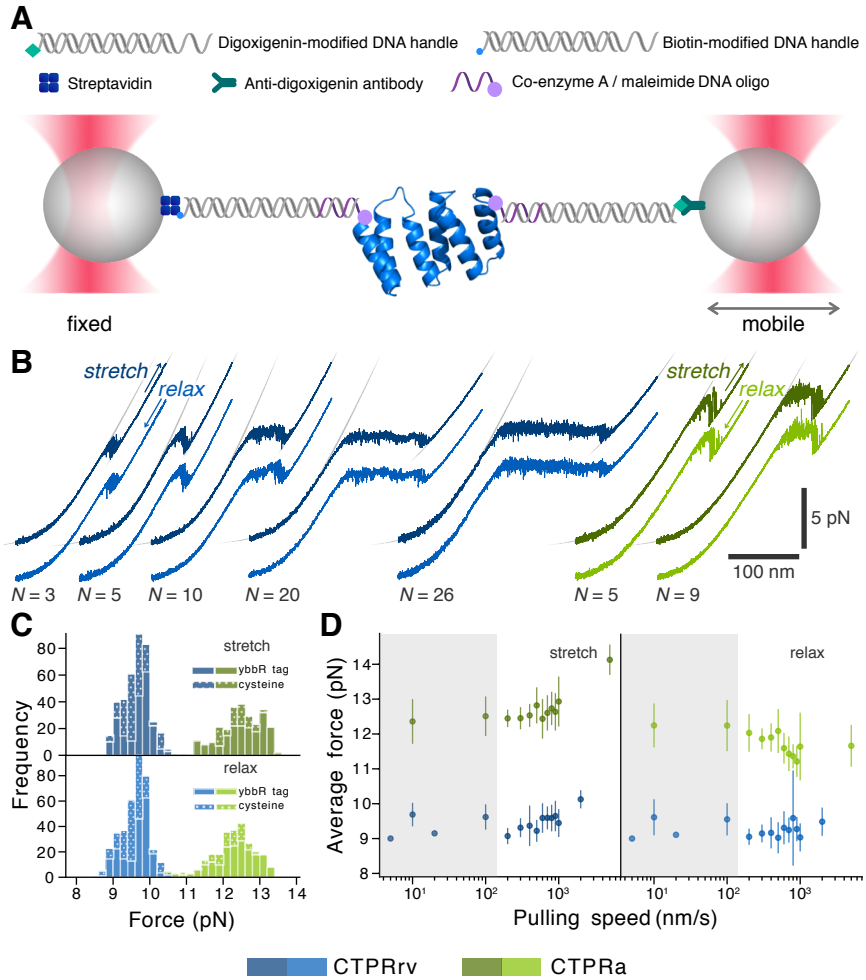

FIG. 2. Probing the mechanics of CTPRs using optical tweezers. (A) The protein of interest bears N- and C-terminal modifications that allow the site-specific attachment of ssDNA oligonucletides that can hybridize with single-stranded overhangs of dsDNA handles tethered to micron-sized silica beads. By 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. Reversely, if the distance is decreased, the force on the DNA-protein construct relaxes and the protein is allowed to refold. (B) Representative force-distance curves (FDCs) for CTPRrv (blue) and CTPRa proteins (green) of different array length pulled at 10 nm/s. 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. Fits for the DNA WLC and the polypeptide-DNA construct are indicated in grey, and the resulting contour lengths of the unfolded polypeptide are listed in Tab. S2. (C) Histograms of the plateau forces for stretch and relax FDCs of all molecules at 10 nm/s and 100 nm/s indicate that the transition remains unaffected by the attachment method. (D) The average plateau forces from stretch and relax FDCs at different pulling speeds (shown as mean ± standard deviation to highlight the variation between traces) show only a modest loading rate dependence. This is almost negligible in the rv-type, but a small but clear increase and decrease in unfolding and refolding plateau forces, respectively, can be observed for the a-type. For representative FDCs at different pulling speeds see Fig. S8. The grey shaded area highlights the low pulling speeds at which the unfolding and refolding plateau forces of both repeat types are indistinguishable, and hence the regime in which the respective system is at equilibrium.

(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 (WLC) behaviour of stretching the linker and the linker combined with the unfolded polypeptide. The plateau is preceded by a small and gradual transition that can no longer be described by a WLC. Since the plateau's length correlates 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 FDCs from stretch and relax cycles of a single 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".

The two CTPR variants exhibit the same unfolding and refolding behaviour, albeit at different forces. We estimated the plateau force of each FDC by fitting gaussians to histograms of the respective force data (see SI Methods and Fig. S3). Whereas CTPRa variants unfolded and refolded at a plateau force of ≈12.5 pN, the plateau of CRPRrv was significantly lower at ≈9.5 pN (Figs. 2B,C), indicating that the introduced mutations have a destabilising 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 to the more commonly used cysteine modifications for maleimide-based attachment strategies (Fig. 2C and Fig. S7). Although a slight stabilisation was observed in chemical denaturation experiments, the ybbR-tag does not discernibly influence the unfolding or refolding plateau 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 plateau forces are only marginally affected by the loading rate, leading to a slight increase in hysteresis at pulling speeds of >500 nm/s (Fig. 2D and Fig. S8). 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 FDCs from stretch and relax cycles at pulling speeds ≤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.

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

To obtain a more detailed picture of the underlying patters in the equilibrium force response of CTPRs, we binned FDCs of repeated stretch and relax cycles at pulling speeds ≤100 nm/s for each individual molecule. Figs. 3A,B show three representative FDCs of each repeat type 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 unconstrained DNA overstretching [46] or the force response of the myosin coiled-coil [47], but rather it showed highly reproducible force oscillations. This pattern was the clearest in the longer arrays of CTPRrv20 and CTPRrv26 which

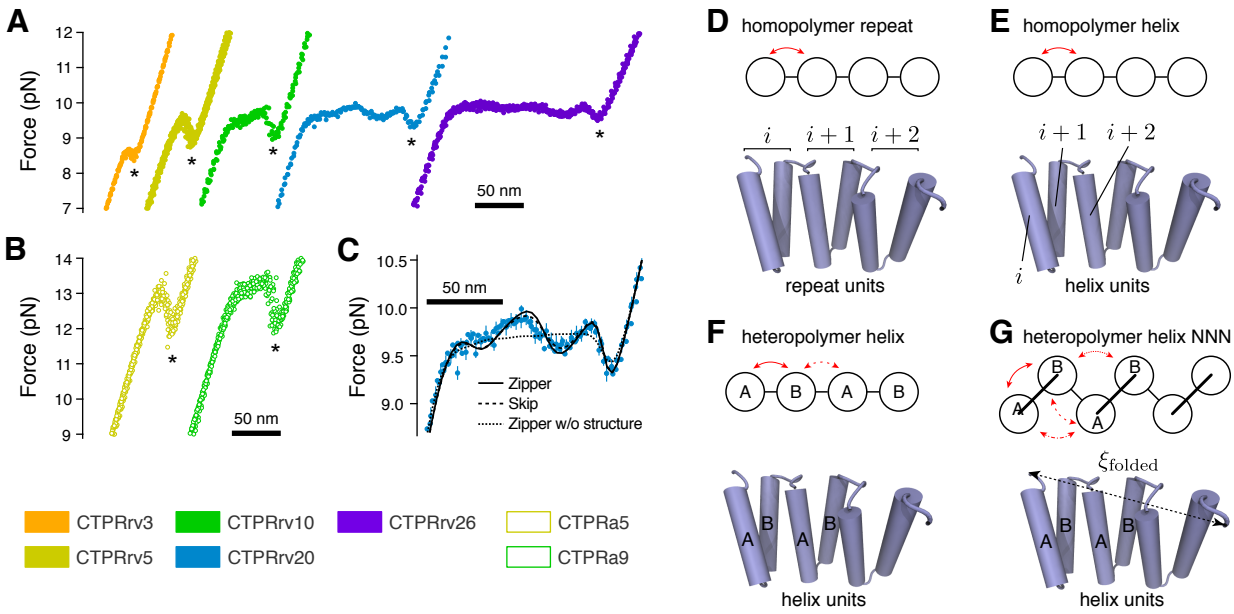

FIG. 3. Describing the average force response of CTPRs using Ising models. (A,B) Overlays of aligned equilibrium force-distance curves (FDCs) of CTPRrv and CTPRa variants in which asterisks mark the characteristic force dip after the plateau. (C) Overlay of CTPRrv20 FDCs (cyan) together with fits to 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 (A-C), average curves of three representative molecules are shown. (D-G) Different Ising models were tested to describe the folding of TPR proteins. In all models, red arrows indicate the interactions between respective subunits and ξ_{folded} represents the end-to-end distance of the folded portion. (D) In the homopolymer repeat model subunits consist of a whole repeats (i.e. two helices). (E) In the homopolymer helix model subunits consist of individual helices that are treated exactly the same. (F) In the heterpolymer helix model the structural repeat is divided into its A and B helices with respective energies. (G) The heteropolymer helix model can be extended to include nearest & next-nearest neighbour interactions (NNN) that may occur e.g. due to structural contacts.

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.

The force response of CTPRs can be modelled using mechanical Ising analysis

We next set out to model the phenomena observed in our data based on thermodynamic first principles [48–52]. For a given trap distance d (see Fig. S4), the total free energy stored in the system, $\mathcal{H}_d$ is the sum of the folding free energy of the protein, $\mathcal{H}^{\text{int}}$, and the mechanical energy for stretching the polymer linkers as well as the traps, $\mathcal{H}^{\text{mech}}$:

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

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, and the harmonic potential of deflecting the beads from the traps. 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).

Due to the linear stacking of repetitive motifs, we and others have used Ising models to describe the folding energy of repeat proteins [22, 29, 32–36, 53–60]. Hence, we explored various Ising models for $\mathcal{H}^{\text{int}}$. In general, the total energy of an Ising-like system is a linear addition of the energies of the individual units and the interaction energies between nearest neighbours [61, 62]. In repeat proteins, the intrinsic energy of the repeating motif, which can be a single helix or a whole repeat, is considered to be zero when it is unfolded and non-zero when it is folded. The interaction between two neighbouring units (coupling or interfacial stability) then depends on the states they are in: it is only non-zero when two neighbouring repeats are folded but zero when at least one of the neighbours is unfolded.

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. 3D). 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 folding free 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. S5A). 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. 3E) [29, 35]. Whereas this model described most of our data much better, we still observed significant deviations for CTPRrv5 molecules (Fig. S5B). 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. 3F), or (ii) can also form additional, next-nearest neighbour interactions to A- and B-helices in adjacent repeats (*heterpopolymer helix NNN model*, Figs. 1A, 3G). 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 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. This selection was confirmed by a direct comparison of the models based on the Akaike information criterion (Fig. S5C,D).

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 [56], could not be used to describe the mechanical unfolding because of the

non-linear contributions of the linker molecules (DNA and unfolded polypeptide) to the mechanical energy. 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 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 $\xi_{\text{folded}}(c) \approx n(c)/N \cdot \xi_{\max}$, where $n(c)$ is the number of folded subunits in conformation c and $\xi_{\max}$ 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*.

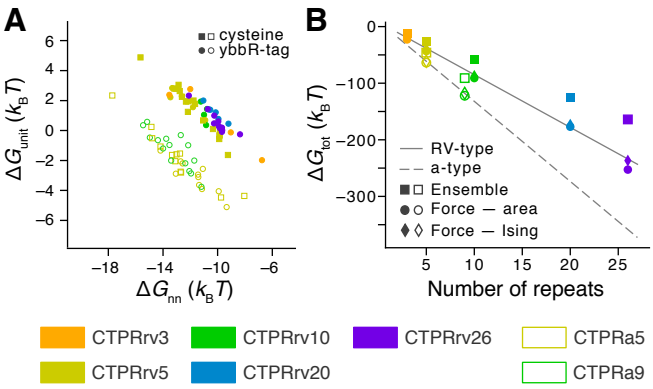

FIG. 4. Energetic contributions to the folding free energy derived from single-molecule and ensemble measurements. (A) Intrinsic repeat energy ΔG_{unit} and repeat next-neighbour energy ΔG_{nn} for individual CTPRrv (filled symbols) and CTPRa molecules (empty symbols) based on the zipper model. For a direct comparison between results derived from skip and zipper approximations see Fig. S9 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 deviate significantly. 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.

TABLE 1. Quantitative energetic description of CTPRs. All energies are reported in units of $k_{\mathrm{B}}T$ and N refers to the number of repeats. For parameters derived from Ising models, ΔG_{unit} is 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)). Alternatively, the folding energy can be approximated as the work done by the protein, W_F, from the area under the curve.

		Equilibrium denaturation[a]			Area	Heteropolymer Ising model (Zipper approximation)			Heteropolymer Ising model (Skip approximation)		
Type	N	ΔG_{tot}	ΔG_{unit}	ΔG_{nn}	W_F[b]	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[c]	N/A[c]	N/A[c]
	all[d]		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[d]		-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 [63], see Fig. S7. Values of the a-type are as reported previously [33]. The errors shown here were propagated from the global fit.
[b] See Eq. (7) in the SI. Values are reported as mean±s.e.m. for each repeat length.
[c] Values for the heteropolymer model in the skip approximation could not be computed because of limited computational capacity.
[d] Combined value for all repeat lengths.

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. S7 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 measurements severely underestimated the total free energy.

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. S10 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. S11), 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).

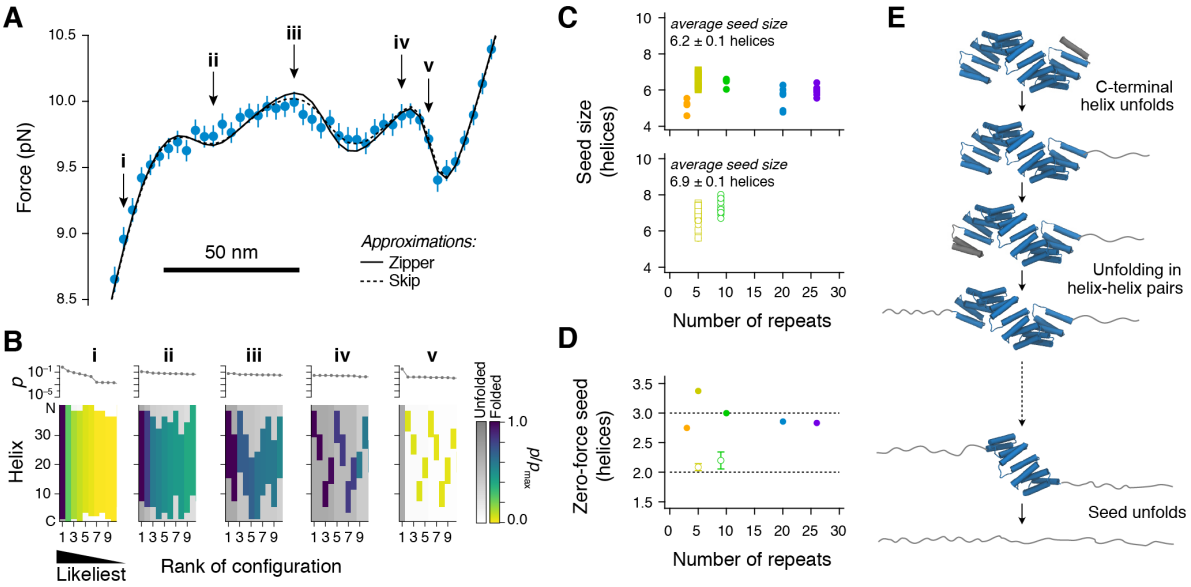

FIG. 5. Gaining insights into the folding pathway of CTPR proteins. (A) Average force-distance data for CTPRrv20 fitted to both zipper and skip approximations of the heteropolymer helix model. Roman numerals point to distances for which snapshots of the conformational ensemble are shown in B. (B) From 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.

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 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 conformation 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 estimates of the contour length increase extracted from the raw data (Fig. S12). However, it is important to stress 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 instead 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

CTPRs respond to force in a manner that has never been observed before: they fold and unfold in a single plateau-like transition with regular undulations 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 also the 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 with 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 "un-zipped" 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. S11). This directionality is a natural consequence of the array geometry itself, since repeats at the centre of the array are less likely to unfold than those at the termini, but it is also consistent with hydrogen-deuterium exchange experiments of CTPRs [27, 64] and further studies of other designed and natural repeat proteins. For example, consensus ankyrin repeats were shown to unfold from both ends using chemical denaturation [58] and from one end to the other under force [39]. In contrast, some natural repeat proteins evolved to have repeats (or repeat domains) of significantly different stability, the weakest of which unfold first even if they are located at the centre of the array [65–67].

To seed or not to seed?

Many globular proteins and also natural repeat proteins have a region or a set of contacts that form a folding nucleus which then "seeds" folding of the remainder. In consensus repeat proteins, such a seed can theoretically form anywhere along the unfolded polypeptide and multiple seeds could form at the same time if the polypeptide is long enough [7]. Consequently, although it is very unlikely that multiple seeds can fold under force, the definition of a seed becomes blurred as we have shown above. However, the average "seed" size of 6 to 7 helices is, intriguingly, consistent with both a folding correlation length of roughly 3 repeats proposed by coarse-grained simulations [7] as well as a folding "nucleus" of 2.5 repeats as concluded from ensemble folding studies of a set of CTPR proteins [32]. In contrast, our estimate of the zero-force seed is much smaller as it only refers to the number of repeating units required for $\Delta G < 0$, which, we would like to point out, does not mean that such a structure is fully folded at all times (e.g. a single CTPRa repeat is still 13% unfolded).

The absence of saw-tooth-like unfolding is a consequence of tether elasticity

To our initial surprise, we did not observe the repeat-by-repeat, saw-tooth unfolding patterns observed for other solenoid repeat proteins, particularly ankyrins. However, using our mechanical Ising model, we increased the effective spring constant of the system connected to the protein (optical trap and linker molecules) to simulate the much stiffer compliance of surface and cantilever in AFM experiments. This modification 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. S13), highlighting that this behaviour is, at least in part, related to the stiffness of the experimental apparatus rather than an intrinsic characteristic of the protein. These findings raise important questions for future research of repeat proteins under load both *in vitro* and *in vivo* as the context of the set-up or the cellular environment may change how we perceive the force response of the protein of interest.

The intrinsic and interfacial stabilities are not modulated independently

We set out to create a CTPR array with a different superhelical geometry by making conservative substitutions at the interface between repeats. Our initial assumption 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 [35]. However, although the chosen mutations did alter the helix packing between repeats more significantly than that within a repeat, 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 propose 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 changes other side-chain interactions.

The nature of the unfolded state

Previously, the free energies of unfolding obtained by force spectroscopy were found to be consistent with those measured in ensemble studies or predicted by physical models [50, 51, 68]. Here, however, the differences between energies derived from ensemble and single-molecule force data are much larger than what could be expected from fitting errors and fit parameter correlation. Cortajarena and co-workers have shown that CTPRs can form polyproline-II (PPII) helices at high GdnHCl concentration and consequently are more compact than a random coil [69]. In fact, lattice simulations of simple polymer models were only able to achieve a compaction of the protein similar to that observed experimentally when attractive interactions between PPII helices were taken into account [69]. The presence of such secondary structures in the chemically denatured state can indeed explain the differences found here. In single-molecule force experiments, the formation of PPII helices is prevented and hence the completely unfolded state can be accessed (as judged by the contour length, Tab. S2).

Force response of the tertiary structure

Previously, several groups used steered MD simulations of natural repeat proteins to show that structural rearrangements at interfaces between repeats allowed the array to stretch as a whole before breaking of the array and unfolding of smaller structural elements occurred at higher forces [10, 11, 16]. However, at this time, we do not have evidence that 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, there is no indication 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. S14). 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, a conclusion supported by previous findings that describe packing between CTPRs as rather rigid [70]. 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 seen how careful modulation of intrinsic repeat stability and interfacial coupling can be used alter the overall stiffness of the tertiary structure to explain the flexibility of several natural repeat proteins observed in both simulations and experiments.

CONCLUSION

In summary, we have characterised the truly novel force response of a solenoid repeat protein at a quality that is high enough to resolve its tertiary structure. Using Ising models we furthermore have shown how two geometrically distinct CTPRs can differ in their thermodynamic and mechanical properties but still retain the same overall folding profile. Our approach circumvents current drawbacks of ensemble studies, as it only requires data of a single array length and has a clearly defined unfolded state. Hence, we envisage that our findings will be valuable for future endeavours to understand the mechanochemistry of repeat proteins.

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 denaturation experiments were performed using guanidine hydrochloride in sodium phosphate buffer pH 6.8, 150 mM NaCl in a 96-well plate format [71]. 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 data collection and processing can be found in the supplementary materials and methods section. Angles between repeat planes were calculated essentially as published previously [13]. Constructs were prepared for force spectroscopy using site-specific modification of either terminal ybbR-tags or cysteine residues [72, 73]. All single-molecule force spectroscopy data was collected on a custom-built instrument [74], processed using custom scripts developed in Igor Pro (WaveMetrics) and further analysed using Igor Pro or Python [75–81]. Theoretical FDCs were calculated using custom C++/CUDA software. Structural representations were generated using PyMol [82] or VMD [83].

Data availability

All scripts and data are available upon reasonable request. The structure of CTPRrv4 has been deposited in the RCSB PDB with the accession code 7obi.

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 single-molecule 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 of 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] Andrey V. Kajava, "Tandem repeats in proteins: From sequence to structure," Journal of Structural Biology **179**, 279 – 288 (2012).

[2] Matteo Pellegrini, Edward M. Marcotte, and Todd O. Yeates, "A fast algorithm for genome-wide analysis of proteins with repeated sequences," Proteins: Structure, Function, and Bioinformatics **35**, 440–446 (1999).

[3] D. Romera, P. Couleaud, S. H. Mejias, A. Aires, and A. L. Cortajarena, "Biomolecular templating of functional hybrid nanostructures using repeat protein scaffolds," Biochem Soc Trans **43**, 825–831 (2015).

[4] C. Jost and A. Plückthun, "Engineered proteins with desired specificity: DARPins, other alternative scaffolds and bispecific IgGs," Curr Opin Struct Biol **27**, 102–112 (2014).

[5] D. Barrick, D. U. Ferreiro, and E. A. Komives, "Folding landscapes of ankyrin repeat proteins: experiments meet theory," Curr Opin Struct Biol **18**, 27–34 (2008).

[6] Albert Perez-Riba, Marie Synakewicz, and Laura S. Itzhaki, "Folding cooperativity and allosteric function in the tandem-repeat protein class," Philosophical Transactions of the Royal Society of London B: Biological Sciences **373** (2018).

[7] D. U. Ferreiro, A. M. Walczak, E. A. Komives, and P. G. Wolynes, "The energy landscapes of repeat-containing proteins: topology, cooperativity, and the folding funnels of one-dimensional architectures," PloS Computational Biology **4**, e1000070 (2008).

[8] Sharona S. Cohen, Inbal Riven, Aitziber L. Cortajarena, Lucia De Rosa, Luca D. D'Andrea, Lynne Regan, and Gilad Haran, "Probing the molecular origin of native-state flexibility in repeat proteins," Journal of the American Chemical Society **137**, 10367–10373 (2015).

[9] J. A. Lamboy, H. Kim, K. S. Lee, T. Ha, and E. A. Komives, "Visualization of the nanospring dynamics of the IkappaBalpha ankyrin repeat domain in real time," Proceedings of the National Academy of Sciences **108**, 10178–10183 (2011).

[10] Alison Grinthal, Ivana Adamovic, Beth Weiner, Martin Karplus, and Nancy Kleckner, "PR65, the HEAT-repeat scaffold of phosphatase PP2A, is an elastic connector that links force and catalysis," Proceedings of the National Academy of Sciences **107**, 2467–2472 (2010).

[11] Christian Kappel, Ulrich Zachariae, Nicole Dölker, and Helmut Grubmüller, "An unusual hydrophobic core confers extreme flexibility to HEAT repeat proteins," Biophysical Journal **99**, 1596–1603 (2010).

[12] Minkyu Kim, Khadar Abdi, Gwangrog Lee, Mahir Rabbi, Whasil Lee, Ming Yang, Christopher J. Schofield, Vann Bennett, and Piotr E. Marszalek, "Fast and forceful refolding of stretched α-helical solenoid proteins," Biophysical Journal **98**, 3086–3092 (2010).

[13] Jade K. Forwood, Allison Lange, Ulrich Zachariae, Mary Marfori, Callie Preast, Helmut Grubmüller, Murray Stewart, Anita H. Corbett, and Bostjan Kobe, "Quantitative structural analysis of importin-β flexibility: Paradigm for solenoid protein structures," Structure **18**, 1171–1183 (2010).

[14] David Serquera, Whasil Lee, Giovanni Settanni, Piotr E. Marszalek, Emanuele Paci, and Laura S. Itzhaki, "Mechanical unfolding of an ankyrin repeat protein," Biophysical Journal **98**, 1294–1301 (2010).

[15] Marian Baclayon, Peter van Ulsen, Halima Mouhib, Maryam Hashemi Shabestari, Timo Verzijden, Sanne Abeln, Wouter H. Roos, and Gijs J. L. Wuite, "Mechanical unfolding of an autotransporter passenger protein reveals the secretion starting point and processive transport intermediates," ACS Nano **10**, 5710–5719 (2016).

[16] M. Sotomayor, D. P. Corey, and K. Schulten, "In search of the hair-cell gating spring elastic properties of ankyrin and cadherin repeats," Structure **13**, 669–682 (2005).

[17] Noemi Fukuhara, Elena Fernandez, Judith Ebert, Elena Conti, and Dmitri Svergun, "Conformational variability of nucleo-cytoplasmic transport factors," Journal of Biological Chemistry **279**, 2176–2181 (2004).

[18] Ulrich Zachariae and Helmut Grubmüller, "Importin-β: Structural and dynamic determinants of a molecular spring," Structure **16**, 906–915 (2008).

[19] Gabriel Žoldák and Matthias Rief, "Force as a single molecule probe of multidimensional protein energy landscapes," Current Opinion in Structural Biology **23**, 48–57 (2013).

[20] A. K. Das, P. W. Cohen, and D. Barford, "The structure of the tetratricopeptide repeats of protein phosphatase 5: implications for TPR-mediated protein-protein interactions," EMBO J. **17**, 1192–1199 (1998).
[21] Luca D. D'Andrea and Lynne Regan, "TPR proteins: the versatile helix," Trends in Biochemical Sciences **28**, 655 – 662 (2003).
[22] Tommi Kajander, Aitziber L. Cortajarena, Simon Mochrie, and Lynne Regan, "Structure and stability of designed TPR protein superhelices: unusual crystal packing and implications for natural TPR proteins," Acta Crystallographica Section D **63**, 800–811 (2007).
[23] Hackwon Do and Muthiah Kumaraswami, "Structural mechanisms of peptide recognition and allosteric modulation of gene regulation by the rrnpp family of quorum-sensing regulators," Journal of Molecular Biology **428**, 2793 – 2804 (2016).
[24] Z. Zhang, L. Chang, J. Yang, N. Conin, K. Kulkarni, and D. Barford, "The Four Canonical TPR Subunits of Human APC/C Form Related Homo-Dimeric Structures and Stack in Parallel to Form a TPR Suprahelix," Journal of Molecular Biology **425**, 4236–4248 (2013).
[25] A. Perez-Riba and L. S. Itzhaki, "The tetratricopeptide-repeat motif is a versatile platform that enables diverse modes of molecular recognition," Curr Opin Struct Biol **54**, 43–49 (2019).
[26] E. R. Main, Y. Xiong, M. J. Cocco, L. D'Andrea, and L. Regan, "Design of stable α-helical arrays from an idealized TPR motif," Structure **11**, 497–508 (2003).
[27] A. Perez-Riba, E. Komives, E. R. G. Main, and L. S. Itzhaki, "Decoupling a tandem-repeat protein: Impact of multiple loop insertions on a modular scaffold," Sci Rep **9**, 15439 (2019).
[28] A. Diamante, P. K. Chaturbedy, P. J. E. Rowling, J. R. Kumita, R. S. Eapen, S. H. McLaughlin, M. de la Roche, A. Perez-Riba, and L. S. Itzhaki, "Engineering mono- and multi-valent inhibitors on a modular scaffold," Chem Sci **12**, 880–895 (2021).
[29] Tommi Kajander, Aitziber L. Cortajarena, Ewan R. G. Main, Simon G. J. Mochrie, and Lynne Regan, "A new folding paradigm for repeat proteins," Journal of the American Chemical Society **127**, 10188–10190 (2005).
[30] Aitziber L. Cortajarena, Simon G.J. Mochrie, and Lynne Regan, "Mapping the energy landscape of repeat proteins using NMR-detected hydrogen exchange," Journal of Molecular Biology **379**, 617 – 626 (2008).
[31] A. L. Cortajarena and L. Regan, "Calorimetric study of a series of designed repeat proteins: modular structure and modular folding," Protein Sci. **20**, 336–340 (2011).
[32] Yalda Javadi and Ewan R. G. Main, "Exploring the folding energy landscape of a series of designed consensus tetratricopeptide repeat proteins," Proceedings of the National Academy of Sciences **106**, 17383–17388 (2009).
[33] A. Perez-Riba, A. R. Lowe, E. R. G. Main, and L. S. Itzhaki, "Context-Dependent Energetics of Loop Extensions in a Family of Tandem-Repeat Proteins," Biophysical Journal **114**, 2552–2562 (2018).
[34] C. Millership, J. J. Phillips, and E. R. Main, "Ising Model Reprogramming of a Repeat Protein's Equilibrium Unfolding Pathway," Journal of Molecular Biology **428**, 1804–1817 (2016).
[35] Aitziber L. Cortajarena, Simon G. J. Mochrie, and Lynne Regan, "Modulating repeat protein stability: The effect of individual helix stability on the collective behavior of the ensemble," Protein Science **20**, 1042–1047 (2011).
[36] J. J. Phillips, Y. Javadi, C. Millership, and E. R. Main, "Modulation of the multistate folding of designed TPR proteins through intrinsic and extrinsic factors," Protein Sci. **21**, 327–338 (2012).
[37] Gwangrog Lee, Khadar Abdi, Yong Jiang, Peter Michaely, Vann Bennett, and Piotr E. Marszalek, "Nanospring behaviour of ankyrin repeats," Nature **440**, 246–249 (2006).
[38] Lewyn Li, Svava Wetzel, Andreas Plückthun, and Julio M. Fernandez, "Stepwise unfolding of ankyrin repeats in a single protein revealed by atomic force microscopy," Biophysical Journal **90**, L30–L32 (2006).
[39] Q. Li, Z. N. Scholl, and P. E. Marszalek, "Capturing the mechanical unfolding pathway of a large protein with coiled-coil probes," Angewandte Chemie International Edition in English **53**, 13429–13433 (2014).
[40] W. Lee, X. Zeng, H. X. Zhou, V. Bennett, W. Yang, and P. E. Marszalek, "Full reconstruction of a vectorial protein folding pathway by atomic force microscopy and molecular dynamics simulations," J Biol Chem **285**, 38167–38172 (2010).
[41] Giovanni Settanni, David Serquera, Piotr E. Marszalek, Emanuele Paci, and Laura S. Itzhaki, "Effects of ligand binding on the mechanical properties of ankyrin repeat protein gankyrin," PLoS Comput Biol

9, e1002864 (2013).
[42] T. J. Magliery and L. Regan, "Beyond consensus: statistical free energies reveal hidden interactions in the design of a TPR motif," Journal of Molecular Biology **343**, 731–745 (2004).
[43] Bharat Jagannathan and Susan Marqusee, "Protein folding and unfolding under force." Biopolymers **99**, 860–869 (2013).
[44] Jun Yin, Alison J. Lin, David E. Golan, and Christopher T. Walsh, "Site-specific protein labeling by Sfp phosphopantetheinyl transferase," Nature Protocols **1**, 280 (2006).
[45] Jun Yin, Paul D. Straight, Shaun M. McLoughlin, Zhe Zhou, Alison J. Lin, David E. Golan, Neil L. Kelleher, Roberto Kolter, and Christopher T. Walsh, "Genetically encoded short peptide tag for versatile protein labeling by Sfp phosphopantetheinyl transferase," Proceedings of the National Academy of Sciences **102**, 15815–15820 (2005).
[46] Joost van Mameren, Peter Gross, Geraldine Farge, Pleuni Hooijman, Mauro Modesti, Maria Falkenberg, Gijs J L Wuite, and Erwin J G Peterman, "Unraveling the structure of DNA during overstretching by using multicolor, single-molecule fluorescence imaging." Proceedings of the National Academy of Sciences **106**, 18231–18236 (2009).
[47] Ingo Schwaiger, Clara Sattler, Daniel R Hostetter, and Matthias Rief, "The myosin coiled-coil is a truly elastic protein structure." Nature materials **1**, 232–235 (2002).
[48] U Bockelmann, B Essevaz-Roulet, and F Heslot, "Molecular Stick-Slip Motion Revealed by Opening DNA with Piconewton Forces," Physical Review Letters **79**, 4489–4492 (1997).
[49] Thomas Bornschlögl and Matthias Rief, "Single molecule unzipping of coiled coils: sequence resolved stability profiles." Physical Review Letters **96**, 118102 (2006).
[50] Josep Maria Huguet, Cristiano V Bizarro, Núria Forns, Steven B Smith, Carlos Bustamante, and Felix Ritort, "Single-molecule derivation of salt dependent base-pair free energies in DNA." Proceedings of the National Academy of Sciences **107**, 15431–15436 (2010).
[51] Gabriel Žoldák, Johannes Stigler, Benjamin Pelz, Hongbin Li, and Matthias Rief, "Ultrafast folding kinetics and cooperativity of villin headpiece in single-molecule force spectroscopy," Proceedings of the National Academy of Sciences **110**, 18156 (2013).
[52] Allison Solanki, Krishna Neupane, and Michael T. Woodside, "Single-molecule force spectroscopy of rapidly fluctuating, marginally stable structures in the intrinsically disordered protein α-synuclein," Physical Review Letters **112**, 158103 (2014).
[53] Cecilia C. Mello and Doug Barrick, "An experimentally determined protein folding energy landscape," Proceedings of the National Academy of Sciences **101**, 14102–14107 (2004).
[54] Jacob D. Marold, Jennifer M. Kavran, Gregory D. Bowman, and Doug Barrick, "A naturally occurring repeat protein with high internal sequence identity defines a new class of tpr-like proteins," Structure **23**, 2055 – 2065 (2015).
[55] Svava K. Wetzel, Giovanni Settanni, Manca Kenig, H. Kaspar Binz, and Andreas Plückthun, "Folding and unfolding mechanism of highly stable full-consensus ankyrin repeat proteins," Journal of Molecular Biology **376**, 241 – 257 (2008).
[56] Tural Aksel and Doug Barrick, "Analysis of repeat-protein folding using nearest-neighbor statistical mechanical models," in *Biothermodynamics, Part A*, Methods in Enzymology, Vol. 455, edited by Michael L. Johnson, Jo M. Holt, and Gary K. Ackers (Academic Press, 2009) Chap. 4, pp. 95–125.
[57] T. Aksel, A. Majumdar, and D. Barrick, "The contribution of entropy, enthalpy, and hydrophobic desolvation to cooperativity in repeat-protein folding," Structure **19**, 349–360 (2011).
[58] T. Aksel and D. Barrick, "Direct observation of parallel folding pathways revealed using a symmetric repeat protein system," Biophysical Journal **107**, 220–232 (2014).
[59] Kathryn Geiger-Schuller and Doug Barrick, "Broken TALEs: Transcription Activator-like Effectors Populate Partly Folded States," Biophysical Journal **111**, 2395–2403 (2016).
[60] K. Geiger-Schuller, K. Sforza, M. Yuhas, F. Parmeggiani, D. Baker, and D. Barrick, "Extreme stability in de novo-designed repeat arrays is determined by unusually stable short-range interactions," Proceedings of the National Academy of Sciences **115**, 7539–7544 (2018).
[61] Ernst Ising, "Beitrag zur Theorie des Ferromagnetismus," Zeitschrift für Physik **31**, 253–258 (1925).
[62] Wilhelm Lenz, "Beitrag zum Verständnis der magnetischen Erscheinungen in festen Körpern," Physikalische Zeitschrift **21**, 613–615 (1920).
[63] Alan R. Lowe, Albert Perez-Riba, Laura S. Itzhaki, and Ewan R.G. Main, "Pyfolding: Open-source graphing, simulation, and analysis of the biophysical properties of proteins," Biophysical Journal **114**,

511–521 (2018).
[64] Ewan R. G. Main, Katherine Stott, Sophie E. Jackson, and Lynne Regan, "Local and long-range stability in tandemly arrayed tetratricopeptide repeats," Proceedings of the National Academy of Sciences **102**, 5721–5726 (2005).
[65] Maksym Tsytlonok, Patricio O. Craig, Elin Sivertsson, David Serquera, Sarah Perrett, Robert B. Best, Peter G. Wolynes, and Laura S. Itzhaki, "Complex energy landscape of a giant repeat protein," Structure **21**, 1954–1965 (2013).
[66] N. D. Werbeck, P. J. Rowling, V. R. Chellamuthu, and L. S. Itzhaki, "Shifting transition states in the unfolding of a large ankyrin repeat protein," Proceedings of the National Academy of Sciences **105**, 9982–9987 (2008).
[67] Katherine W. Tripp and Doug Barrick, "Rerouting the folding pathway of the notch ankyrin domain by reshaping the energy landscape," Journal of the American Chemical Society **130**, 5681–5688 (2008).
[68] Johannes Stigler and Matthias Rief, "Calcium-dependent folding of single calmodulin molecules," Proceedings of the National Academy of Sciences **109**, 17814–17819 (2012).
[69] A. L. Cortajarena, G. Lois, E. Sherman, C. S. O'Hern, L. Regan, and G. Haran, "Non-random-coil behavior as a consequence of extensive PPII structure in the denatured state," Journal of Molecular Biology **382**, 203–212 (2008).
[70] C. Y. Cheng, V. A. Jarymowycz, A. L. Cortajarena, L. Regan, and M. J. Stone, "Repeat motions and backbone flexibility in designed proteins with different numbers of identical consensus tetratricopeptide repeats," Biochemistry **45**, 12175–12183 (2006).
[71] A. Perez-Riba and L. S. Itzhaki, "A method for rapid high-throughput biophysical analysis of proteins," Scientific Reports **7**, 9071 (2017).
[72] M. Synakewicz, D. Bauer, M. Rief, and L. S. Itzhaki, "Bioorthogonal protein-DNA conjugation methods for force spectroscopy," Sci Rep **9**, 13820 (2019).
[73] Ann Mukhortava and Michael Schlierf, "Efficient formation of site-specific protein-dna hybrids using copper-free click chemistry," Bioconjugate Chemistry **27**, 1559–1563 (2016).
[74] Yann von Hansen, Alexander Mehlich, Benjamin Pelz, Matthias Rief, and Roland R. Netz, "Auto- and cross-power spectral analysis of dual trap optical tweezer experiments using bayesian inference," Review of Scientific Instruments **83**, 095116 (2012).
[75] K. Jarrod Millman and Michael Aivazis, "Python for scientists and engineers," Computing in Science & Engineering **13**, 9–12 (2011).
[76] Travis E. Oliphant, "Python for scientific computing," Computing in Science & Engineering **9**, 10–20 (2007).
[77] Stéfan van der Walt, S. Chris Colbert, and Gaël Varoquaux, "The numpy array: A structure for efficient numerical computation," Computing in Science & Engineering **13**, 22–30 (2011).
[78] J. D. Hunter, "Matplotlib: A 2D graphics environment," Computing In Science & Engineering **9**, 90–95 (2007).
[79] Michael Waskom and the seaborn development team, "mwaskom/seaborn," (2020).
[80] Eric Jones, Travis Oliphant, Pearu Peterson, *et al.*, "SciPy: Open source scientific tools for Python," (2001), http://www.scipy.org/.
[81] Wes McKinney, "Data Structures for Statistical Computing in Python," in *Proceedings of the 9th Python in Science Conference*, edited by Stéfan van der Walt and Jarrod Millman (2010) pp. 56 – 61.
[82] Schrödinger, LLC, "The PyMOL molecular graphics system, version 1.8," (2015).
[83] W. Humphrey, A. Dalke, and K. Schulten, "VMD: visual molecular dynamics," Journal of Molecular Graphics **14**, 33–38 (1996).