

<sup>1</sup> **Consensus tetratricopeptide repeat proteins are complex superhelical  
2 nanosprings**

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

<sup>5</sup> *1Department of Pharmacology, University of Cambridge,  
6 Tennis Court Road, Cambridge, CB2 1PD, United Kingdom*

<sup>7</sup> *2Physik-Department, Technische Universität München,  
8 James-Franck-Straße 1, 85748 Garching, Germany*

<sup>9</sup> *3Department of Biochemistry, University of Cambridge,  
10 Tennis Court Road, Cambridge, CB2 1GA, United Kingdom*

<sup>11</sup> *4Gene Center Munich, Ludwig-Maximilians-Universität München,  
12 Feodor-Lynen-Straße 25, 81377 München, Germany*

<sup>13</sup> (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

14 **INTRODUCTION**

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

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

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

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

59 **RESULTS**

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

61 We based our study of the structure and energetics of TPR proteins on previous research of  
62 a consensus TPR protein (referred to as CTPRa) that adopts a superhelical structure within the  
63 crystal lattice [22]. This superhelical geometry of TPRs is established through stacking of the A-  
64 helix ( $A_{i+1}$  in Fig. 1A) of a given repeat against both A and B helices of the preceding repeat ( $A_i$   
65 and  $B_i$  in Fig. 1A) [26]. The resulting angles between repeat planes then give rise to the pitch and  
66 diameter of the superhelix (Fig. 1B) [13]. Interfaces between repeats are largely formed by bulky  
67 hydrophobic residues. Therefore, to create a CTPR variant with altered superhelical geometry,  
68 we identified four such interface residues based on sequence conservation in different TPR families  
69 [26, 42] and substituted them with polar or small hydrophobic side chains (W4L, Y5N, A10V,  
70 Y12R).

71 We determined the structure of a 4-repeat construct of the new variant containing a C-terminal  
72 solvating helix, CTPRrv4, by X-ray crystallography to 3.0 Å resolution (see Tab. S3 and Fig. S6).  
73 CTPRrv4 crystallised in the  $P\ 3_1\ 2\ 1$  space group with two molecules per asymmetric unit. As was  
74 observed for CTPRa, the C-terminal solvating helix was not resolved due to a higher preference for  
75 end-to-end stacking between molecules in neighbouring asymmetric units [22]. The resolution was  
76 sufficient to determine the change in repeat plane angles using the  $C_\alpha$  coordinates: while the twist  
77 remained almost unchanged, an increase in curving angle is compensated by a similar decrease in  
78 the bending angle (Fig. 1A, Tab. S4). Although these changes may at first appear insignificant  
79 in the context of a 4-repeat array, they translate into clear differences in the longer superhelical  
80 arrays, leading to a decrease in helix length and increase in helix diameter (Fig. 1B). Furthermore,  
81 when compared to other CTPRs, all of which exhibit backbone RMSDs within 0.72 Å, the CTPRrv  
82 backbone differs by 1.4 Å relative to the CTPRa backbone.

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

84 To examine the mechanical folding and unfolding of the CTPR superhelix, we prepared CTPRrv  
85 arrays of  $N = 3, 5, 10, 20$  and 26 and CTPRa arrays of  $N = 5$  and 9 for force spectroscopy  
86 measurements using optical tweezers (Fig. 2A, [43]). Using the Sfp-enzyme, coenzyme-A modified  
87 ssDNA oligos were conjugated to N- and C-terminal ybbR tags on the protein [44]. These protein-  
88 DNA chimeras were then attached to micron-sized silica beads using dsDNA handles.

89 For an initial characterization of the CTPR force response, we recorded force-distance curves

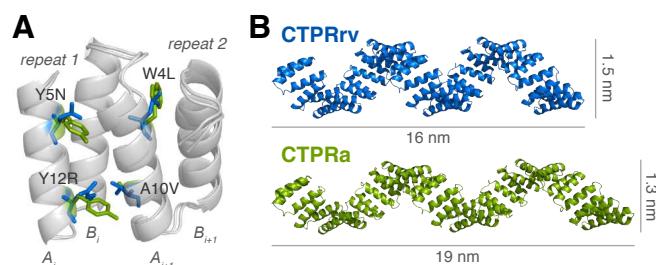


FIG. 1. Design of the CTPRrv variant (blue) based on the original CTPRa (green, PDB accession code: 2hyz [22]). (A) Structural representation of two neighbouring CTPRs highlighting the interfacial mutations introduced in CTPRa to form CTPRrv. (B) The slight alteration in repeat packing leads to changes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

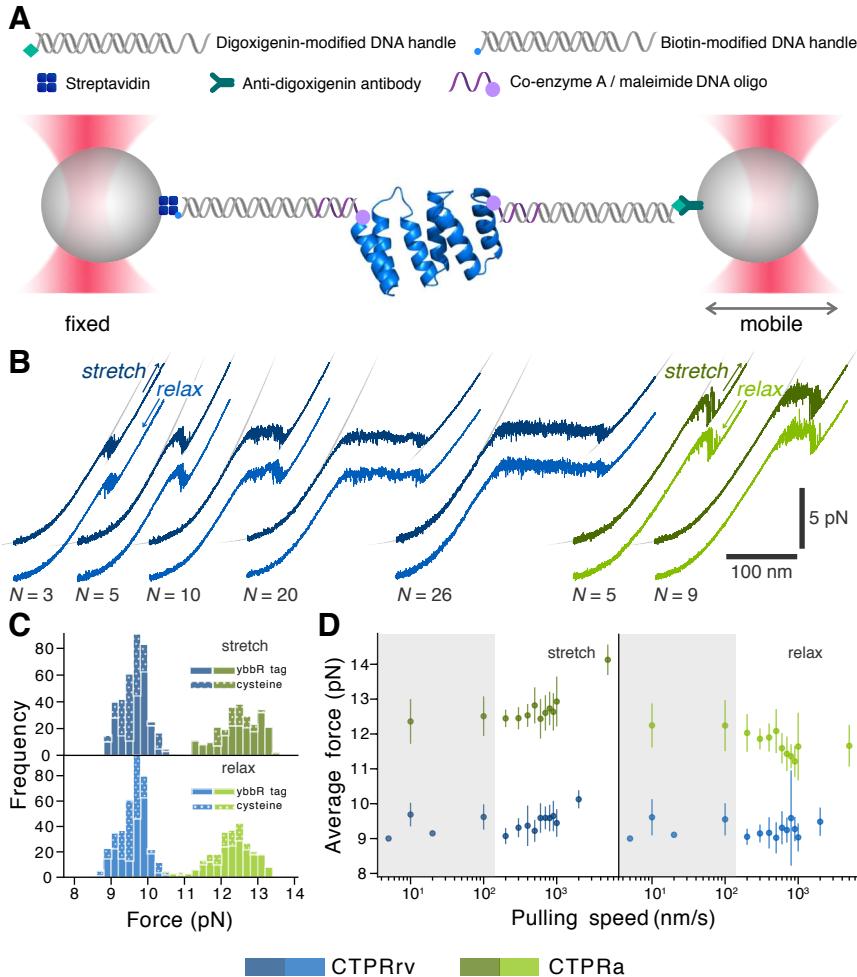


FIG. 2. Probing the mechanics of CTPRs using optical tweezers. (A) The protein of interest bears N- and C-terminal modifications that allow the site-specific attachment of ssDNA oligonucleotides that can hybridize with single-stranded overhangs of dsDNA handles tethered to micron-sized silica beads. 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 overlap 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  $\pm$  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.

90 (FDCs) from stretch-relax cycles at a pulling velocity of 10 nm/s and 100 nm/s (Fig. 2B). The FDCs  
91 of all variants display a characteristic plateau region and a subsequent force dip, flanked by the  
92 characteristic worm-like-chain (WLC) behaviour of stretching the linker and the linker combined  
93 with the unfolded polypeptide. The plateau is preceded by a small and gradual transition that  
94 can no longer be described by a WLC. Since the plateau's length correlates with CTPR array size,  
95 we attribute the plateau and dip regions to force-induced unfolding of CTPR repeats. The shape  
96 of the unfolding profile is indicative of sequential unfolding of repeats (plateau) until a minimally  
97 stable unit is reached, which appears to unravel in a more cooperative manner (dip). Furthermore,  
98 we noticed that FDCs from stretch and relax cycles of a single molecule at these loading rates  
99 are almost indistinguishable when superimposed. The increased noise levels of plateau and dip  
100 indicate very fast unfolding and refolding transitions (Fig. 2B), which, together with the absence  
101 of hysteresis, suggests that CTPR unfolding and refolding occurs at equilibrium, i.e. the folding  
102 kinetics of the system under investigation are much faster than the pulling speed allowing it to “re-  
103 equilibrate instantly”.

104 The two CTPR variants exhibit the same unfolding and refolding behaviour, albeit at different  
105 forces. We estimated the plateau force of each FDC by fitting gaussians to histograms of the  
106 respective force data (see SI Methods and Fig. S3). Whereas CTPRa variants unfolded and  
107 refolded at a plateau force of  $\approx$ 12.5 pN, the plateau of CTPRrv was significantly lower at  $\approx$ 9.5 pN  
108 (Figs. 2B,C), indicating that the introduced mutations have a destabilising effect. Furthermore,  
109 to ensure that the intrinsic  $\alpha$ -helicity of the ybbR-tag [45] did not alter the stability of the repeat  
110 arrays, we compared the mechanical and chemical stability of proteins with ybbR-tags to the more  
111 commonly used cysteine modifications for maleimide-based attachment strategies (Fig. 2C and  
112 Fig. S7). Although a slight stabilisation was observed in chemical denaturation experiments, the  
113 ybbR-tag does not discernibly influence the unfolding or refolding plateau forces, the unfolding and  
114 refolding energies (beyond the contribution of contour length), or the character of the mechanical  
115 response (Fig. 2C).

116 A systematic screen of pulling velocities ranging from 10 nm/s to 5  $\mu$ m/s revealed that for both  
117 CTPRrv and CTPRa, the average folding and unfolding plateau forces are only marginally affected  
118 by the loading rate, leading to a slight increase in hysteresis at pulling speeds of  $>500$  nm/s (Fig. 2D  
119 and Fig. S8). This effect is more pronounced in CTPRa arrays than in CTPRrv arrays. It is  
120 interesting to note that even at higher pulling speeds not all stretch-relax cycles exhibit hysteresis,  
121 leading to a large variation within even a single molecule (e.g. see FDCs collected at 1  $\mu$ m/s in Fig  
122 S9). However, most importantly, we found that there was no significant hysteresis between FDCs  
123 from stretch and relax cycles at pulling speeds  $\leq$ 100 nm/s in both repeat types. The absence of a  
124 pulling speed-dependent folding/unfolding force is again evidence for rapid equilibrium fluctuations  
125 of CTPR subunits in this loading rate regime.

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

127 To obtain a more detailed picture of the underlying patterns in the equilibrium force response  
128 of CTPRs, we binned FDCs of repeated stretch and relax cycles at pulling speeds  $\leq$ 100 nm/s  
129 for each individual molecule. Figs. 3A,B show three representative FDCs of each repeat type  
130 at the chosen array lengths, overlaid and aligned along force and distance coordinates to avoid  
131 the introduction of common instrumental artefacts such as miscalibration of the trap stiffness or  
132 the zero distance point. We found the plateau region was not uniformly flat after averaging, as  
133 would be expected from other related phenomena such as unconstrained DNA overstretching [46]  
134 or the force response of the myosin coiled-coil [47], but rather it showed highly reproducible force  
135 oscillations. This pattern was the clearest in the longer arrays of CTPRrv20 and CTPRrv26 which

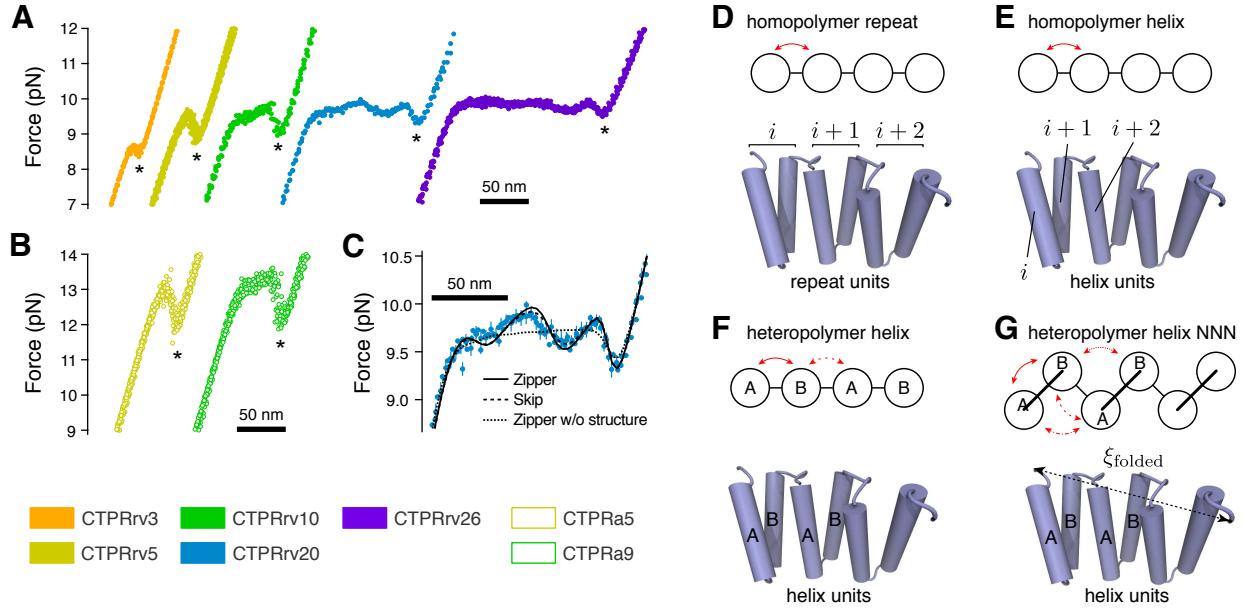


FIG. 3. Describing the average force response of CTPRs using Ising models. (A,B) Overlays of aligned equilibrium force-distance curves (FDCs) of CTPRrv and CTPRa variants in which asterisks mark the characteristic force dip after the plateau. (C) Overlay of CTPRrv20 FDCs (cyan) together with fits to 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  $\xi_{\text{folded}}$  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 heteropolymer 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.

136 exhibit about 2 and 3 periods, respectively. Therefore, we reasoned that these oscillations arise  
 137 directly from the structure of the superhelix. In contrast, the characteristic force dip at the end of  
 138 each FDC was present in all arrays, and we hypothesise that it corresponds to the unfolding of a  
 139 final stable unit.

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

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

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

145 where  $F$  is the force at that distance and  $c = \{c_1, \dots, c_N\}$  represents a particular configuration  
 146 of the  $2^N$ -sized conformational space of an  $N$ -mer.  $\mathcal{H}^{\text{mech}}$  is derived from polymer models of  
 147 stretching dsDNA and unfolded polypeptide, and the harmonic potential of deflecting the beads  
 148 from the traps. Theoretical equilibrium FDCs, as are required for fitting data, can subsequently

149 be obtained by numerically calculating the partition function and integrating over all degrees of  
 150 freedom (see SI Methods for full detail).

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

160 We first generated models based on a whole repeat (i.e. one A- and B-helix) as the smallest  
 161 independent protein unit (*homopolymer repeat model*, Fig. 3D). Here, the protein’s internal energy  
 162 in a given conformation is given by

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

163 where  $\Delta G_{\text{unit}}$  is the folding free energy of a single subunit,  $\Delta G_{\text{nn}}$  is the next-neighbour interaction,  
 164 or coupling energy, between adjacent subunits, and  $c_i$  refers to a folded unit. However, this model  
 165 failed to reproduce the curvature at the transition between the DNA stretch response and the  
 166 protein unfolding plateau in the CTPRrv10 data (Fig. S5A). Second, we implemented a model  
 167 that used individual helices as the smallest independent unit based on some of the earliest CTPR  
 168 folding studies (*homopolymer helix model*, Fig. 3E) [29, 35]. Whereas this model described most of  
 169 our data much better, we still observed significant deviations for CTPRrv5 molecules (Fig. S5B).  
 170 Lastly, we expanded the homopolymer helix model to account for the differences between the A-  
 171 and B-helices within a repeat, and furthermore considered the scenarios in which A and B helices  
 172 can (i) only interact with their nearest neighbour (*heteropolymer helix model*, Fig. 3F), or (ii) can  
 173 also form additional, next-nearest neighbour interactions to A- and B-helices in adjacent repeats  
 174 (*heteropolymer helix NNN model*, Figs. 1A, 3G). In the latter two models, the intrinsic energy is  
 175 given by

$$\mathcal{H}^{\text{int}}(c) = n_{AB} \Delta G_{AB} + n_{BA} \Delta G_{BA} + n_{AA} \Delta G_{AA} + n_{BB} \Delta G_{BB} + n_A \Delta G_A + n_B \Delta G_B, \quad (3)$$

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

$$\Delta G_{\text{unit}} = \Delta G_A + \Delta G_B + \Delta G_{AB} \quad \text{and} \quad \Delta G_{\text{nn}} = \Delta G_{BA} + \Delta G_{AA} + \Delta G_{BB}. \quad (4)$$

181 Both heteropolymer helix models were able to describe all data and yielded similar results (Tab. S5).  
 182 Therefore, we proceeded with the heteropolymer helix model without next-nearest neighbour in-  
 183 teraction to avoid over-parametrization. This selection was confirmed by a direct comparison of  
 184 the models based on the Akaike information criterion (Fig. S5C,D).

185 All of the above models have exponential  $\mathcal{O}(2^N)$  complexity and are computationally too expen-  
 186 sive to be applied to CTPRrv20 and CTPRrv26 data. A matrix formalism, which was successfully  
 187 employed to reduce the complexity of equivalent homopolymer and heteropolymer Ising models  
 188 in chemical unfolding [56], could not be used to describe the mechanical unfolding because of the

189 non-linear contributions of the linker molecules (DNA and unfolded polypeptide) to the mechan-  
 190 ical energy. Instead, we considered alternative simplifications to reduce the conformational space  
 191 by eliminating extremely unlikely high-energy configurations in two different ways. In the “skip”  
 192 approximation, all configurations in which one or two folded helices are neighboured by unfolded  
 193 helices on either side are eliminated. This reduced the complexity from  $\mathcal{O}(2^N)$  to  $\mathcal{O}(1.65^N)$  but  
 194 was still computationally too expensive to fit CTPRrv26 data. However, in the “zipper” approx-  
 195 imation, in which unfolding occurs from the end(s), the conformational ensemble is reduced to  
 196  $N(N+1)/2+1 = \mathcal{O}(N^2)$  states, and this simplification allowed us to fit data of CTPRrv26 arrays.  
 197 Notably, the modelled FDCs based on these approximation differed only marginally from the data  
 198 to which both approximations could be fitted (e.g. see dashed and continuous lines in Figs. 3C,  
 199 S13A). Furthermore, the resulting energies of the “skip” and “zipper” approximations agreed within  
 200 error (Tab. 1), and therefore, unless stated otherwise, we used the zipper approximation for all  
 201 reported values.

202 **Force oscillations are a consequence of the superhelical structure**

203 During the initial rounds of model development we observed that none of our Ising models alone  
 204 could account for the observed force oscillations in the plateau. At first, the molecular extension  
 205 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  
 206 folded subunits in conformation  $c$  and  $\xi_{\text{max}}$  is the end-to-end distance of the fully folded protein.  
 207 However, this assumes an arrangement of subunits in the repeat array akin to beads-on-a-string.  
 208 Although all models based on this assumption for the molecular extension correctly predicted  
 209 the final dip, they falsely produced a flat plateau (dotted line in Fig. 3C). Only when structural  
 210 parameters of the superhelix were included to account for the changes in the force vector across  
 211 the folded remainder of the molecule as it unfolds, was it possible to reproduce these features.  
 212 Given that our models describe the data well apart from some minor deviations, we can therefore  
 213 conclude that the structure of CTPR arrays in solution is indeed very similar to that *in crystallo*.

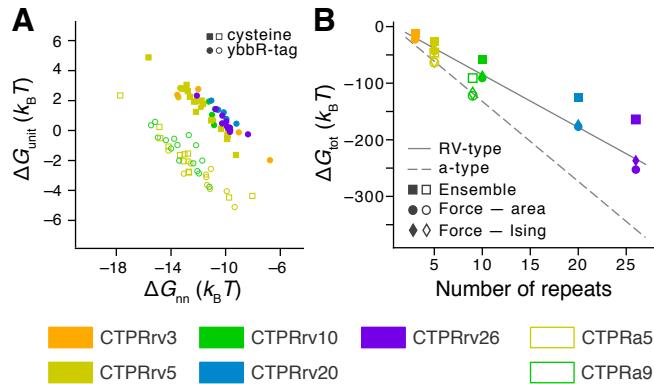


FIG. 4. Energetic contributions to the folding free energy derived from single-molecule and ensemble measurements. (A) Intrinsic repeat energy  $\Delta G_{\text{unit}}$  and repeat next-neighbour energy  $\Delta G_{\text{nn}}$  for individual CTPRrv (filled symbols) and CTPRa molecules (empty symbols) 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  $k_B T$  and  $N$  refers to the number of repeats. For parameters derived from Ising models,  $\Delta G_{\text{unit}}$  is the intrinsic repeat energy,  $\Delta G_{\text{nn}}$  the next-neighbour interaction energy, and  $\Delta G_{\text{tot}}$  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.

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

<sup>a</sup> 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.

<sup>b</sup> See Eq. (7) in the SI. Values are reported as mean $\pm$ s.e.m. for each repeat length.

<sup>c</sup> Values for the heteropolymer model in the skip approximation could not be computed because of limited computational capacity.

<sup>d</sup> Combined value for all repeat lengths.

## 214 Comparing intrinsic and interfacial energies between single-molecule and ensemble data

215 Using our models, we could now determine the intrinsic and interfacial contributions to the free  
216 energy, and extrapolated the repeat stabilities and nearest-neighbour coupling from the individual  
217 terms of the A and B helices using Equation 4 (Fig. 4A and Tab. 1). Due to the interdependence  
218 of the model parameters the data cluster diagonally, i.e. destabilisation of one energetic term  
219 is compensated with stabilisation of the other. However, the data of the two variants (filled  
220 vs. empty symbols) clearly separate and are independent of attachment method (squares vs.  
221 circles). When averaging over the whole data set we found that the interfacial energy ( $\Delta G_{\text{nn}}^{\text{CTPRrv}} =$   
222  $-11.0 \pm 0.2 k_B T$ ,  $\Delta G_{\text{nn}}^{\text{CTPRa}} = -12.7 \pm 0.3 k_B T$ ) vastly outweighs the intrinsic energy ( $\Delta G_{\text{unit}}^{\text{CTPRrv}} =$   
223  $1.1 \pm 0.2 k_B T$ ,  $\Delta G_{\text{unit}}^{\text{CTPRa}} = -1.9 \pm 0.3 k_B T$ ) for both repeat types. Next, we determined the total  
224 free energy derived from our Ising models using

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

225 for each variant and array length (diamonds in Fig. 4B and Tab. 1). These values could then be  
226 compared to (i) values derived from the area under the curve of FDCs (circles in Fig. 4B and Tab. 1),  
227 and (ii) those calculated from ensemble Ising models using Eq. 5 (squares in Fig. 4B, Fig. S7  
228 and Tab. 1). While there is excellent agreement between methods based on single-molecule data,  
229 validating our mechanical Ising model approach, energies derived from ensemble measurements  
230 severely underestimated the total free energy.

## 231 The (un)folding pathway of CTPR arrays

232 Based on our Ising analyses, we were able to develop a model of the likeliest unfolding (and  
233 folding) pathway for CTPR proteins under mechanical load. To this end, we chose to examine

234 results of the skip approximation in more detail to avoid possible bias introduced by the assumption  
 235 that arrays unfold from the end as it is done in the zipper approximation. For each trap distance  
 236 we recorded the likeliest conformations ranked by their probabilities. Examples at six different  
 237 distances along the average unfolding profile of CTPRrv20 are shown in Fig. 5 (and Fig. S10 for  
 238 CTPRrv26), along with the probabilities of a given configuration and the relative probabilities of  
 239 individual helices being folded or unfolded. Our data show that unfolding preferentially occurs  
 240 from the ends in a zipper-like fashion. However, at increasing distances, there are several, almost  
 241 equally likely conformations that have different segments of folded and unfolded helices. When  
 242 plotting the probability of being folded for all helices as a function of distance (Fig. S11), we  
 243 furthermore noticed that unfolding starts with the most C-terminal helix, and then propagates one  
 244 or two helices at a time from the C-terminus (i.e. A and B helices alternating, or almost in pairs  
 245 of  $B_{i-1}A_i$ ) and one repeat at a time from the N-terminus (i.e. in pairs of  $A_iB_i$ ).

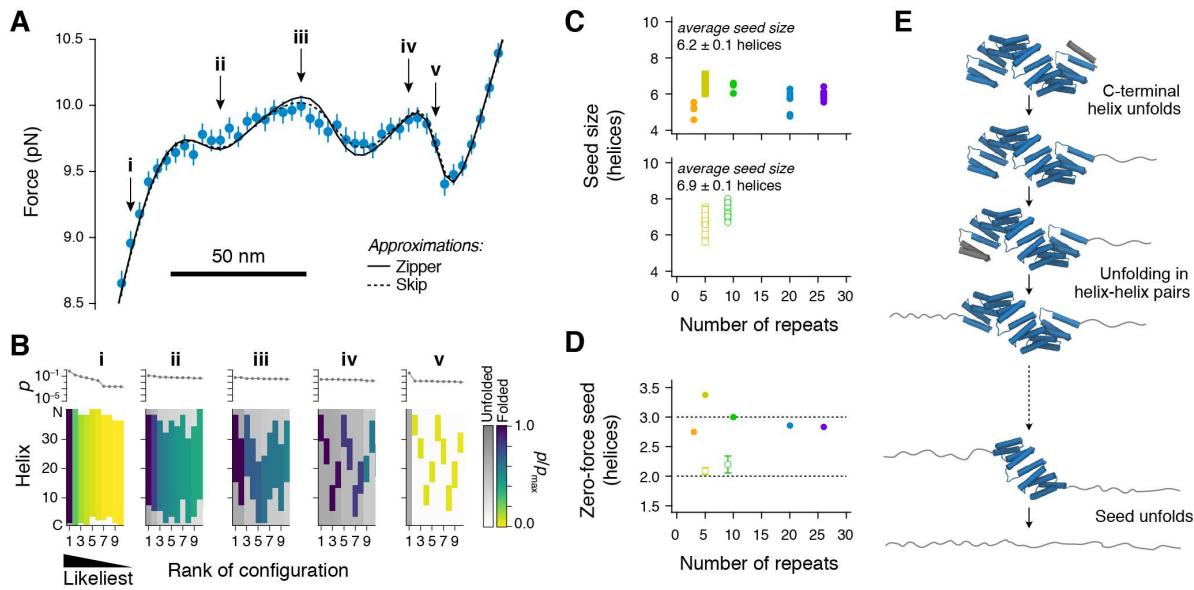


FIG. 5. Gaining insights into the folding pathway of CTPR proteins. (A) Average force-distance data for CTPRrv20 fitted to both zipper and skip approximations of the heteropolymer helix model. Roman numerals point to distances for which snapshots of the conformational ensemble are shown in B. (B) From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

246 Next, we extracted the microscopic information on the characteristic “dip” feature at the end of  
 247 the plateau region. Traditionally, one would have expected that the dip is caused by the cooperative  
 248 folding/unfolding of a well-defined minimally stable unit of a certain length that exchanges in  
 249 a two-state manner with the unfolded state. However, our model indicates that instead of a  
 250 clear two-state transition, the dip ((iv)–(v) in Fig. 5A,B) represents inter-conversions between a

251 large ensemble of marginally stable conformations of varying size between one another and the  
252 unfolded conformation. For example, at position (v) in Fig. 5A, the likeliest conformation is the  
253 unfolded state. However, this state exchanges rapidly with an ensemble of conformations with  
254  $\approx 6$  to 7 consecutively folded helices that individually are less populated than the unfolded state,  
255 but together amount to about 50 % of the conformational space (Fig. 5B(v)). We then calculated  
256 the average number of folded helices once the fully unfolded conformation reaches a likelihood  
257 of 50 % to extract the length of the (un)folding “seed”. As expected, the size of this seed was  
258 independent of array length, and comprised approximately  $6.2 \pm 0.1$  and  $6.9 \pm 0.1$  helices for the  
259 rv- and a-type arrays, respectively (Fig. 5C), which agrees with estimates of the contour length  
260 increase extracted from the raw data (Fig. S12). However, it is important to stress that these  
261 seeds do not cooperatively exchange with the unfolded conformation in a two-state manner, nor  
262 should they be mistaken for the minimal folded unit under zero-force such as one would obtain in  
263 ensemble measurements. To get an estimate of the minimal folding unit at zero force, we instead  
264 calculated the minimal number of consecutive folded helices needed in order to achieve a negative  
265  $\Delta G_{\text{tot}}$  using Eq. (5). The resulting inferred zero-force “seeds” were independent of variant length  
266 and were close to 3 helices for CTPRrv and 2 helices for CTPRa (Fig. 5D). Therefore, based on  
267 our results, a single CTPRa repeat is weakly stable ( $-1.9 \pm 0.3 k_B T$ ) and may fold transiently on  
268 its own. On the contrary, a single CTPRrv repeat is unstable ( $1.1 \pm 0.2 k_B T$ ) and requires the  
269 energy from the interface with at least one more helix to fold. These results show the same trend  
270 as our ensemble denaturation data, even if they differ numerically.

## 271 DISCUSSION

### 272 CTPR solenoids unzip from both ends under force

273 CTPRs respond to force in a manner that has never been observed before: they fold and unfold  
274 in a single plateau-like transition with regular undulations that ends in a dip. The plateau arises  
275 from the asymmetric unzipping of helices or repeats from either end of the repeat array, starting  
276 with the C-terminal helix. The very fast equilibrium fluctuations of the C-terminal helix, and  
277 possibly also the N-terminal repeat, are beyond the time resolution of our instrument and hence  
278 the transition from DNA stretching into the plateau is rather smooth (i.e. “averaged”), particularly  
279 in FDCs of the rv-type arrays. The asymmetry of the unfolding pathway arises directly from the  
280 structure: (i) the C-terminal B-helix is exposed without any interactions beyond those with the  
281 corresponding A-helix of the last repeat of the folded protein/remainder; (ii) the force vector aligns  
282 the molecule such that the C-terminal helix can get “un-zipped” while at the N-terminus a whole  
283 repeat experiences shear forces; (iii) a  $B_{i-1}A_i$  repeat is structurally different from an  $A_iB_i$  repeat,  
284 leading to different unfolding patterns from either end as the protein is unzipped by force (Fig.  
285 S11). This directionality is a natural consequence of the array geometry itself, since repeats at the  
286 centre of the array are less likely to unfold than those at the termini, but it is also consistent with  
287 hydrogen-deuterium exchange experiments of CTPRs [27, 64] and further studies of other designed  
288 and natural repeat proteins. For example, consensus ankyrin repeats were shown to unfold from  
289 both ends using chemical denaturation [58] and from one end to the other under force [39]. In  
290 contrast, some natural repeat proteins evolved to have repeats (or repeat domains) of significantly  
291 different stability, the weakest of which unfold first even if they are located at the centre of the  
292 array [65–67].

293 **To seed or not to seed?**

294 Many globular proteins and also natural repeat proteins have a region or a set of contacts that  
295 form a folding nucleus which then “seeds” folding of the remainder. In consensus repeat proteins,  
296 such a seed can theoretically form anywhere along the unfolded polypeptide and multiple seeds  
297 could form at the same time if the polypeptide is long enough [7]. Consequently, although it is  
298 very unlikely that multiple seeds can fold under force, the definition of a seed becomes blurred as  
299 we have shown above. However, the average “seed” size of 6 to 7 helices is, intriguingly, consistent  
300 with both a folding correlation length of roughly 3 repeats proposed by coarse-grained simulations  
301 [7] as well as a folding “nucleus” of 2.5 repeats as concluded from ensemble folding studies of a  
302 set of CTPR proteins [32]. In contrast, our estimate of the zero-force seed is much smaller as it  
303 only refers to the number of repeating units required for  $\Delta G < 0$ , which, we would like to point  
304 out, does not mean that such a structure is fully folded at all times (e.g. a single CTPRa repeat  
305 is still 13% unfolded).

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

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

318 **The intrinsic and interfacial stabilities are not modulated independently**

319 We set out to create a CTPR array with a different superhelical geometry by making conservative  
320 substitutions at the interface between repeats. Our initial assumption was that by changing only  
321 interface interactions, we would affect solely the interfacial coupling energy — akin to how the  
322 intrinsic repeat stability was modified in a previous study [35]. However, although the chosen  
323 mutations did alter the helix packing between repeats more significantly than that within a repeat,  
324 ensemble and single-molecule data show that these mutations affected both energetic parameters,  
325 i.e. coupling and intrinsic stability. In fact, the mechanical Ising models indicate that the overall  
326 lower stability of the rv-type arrays relative to the original CTPR arrays is due to an almost equal  
327 destabilization of both intrinsic and interfacial energies ( $\Delta\Delta G_{\text{unit}} \approx 3 k_B T$  and  $\Delta\Delta G_{\text{nn}} \approx 2 k_B T$ ,  
328 respectively). We therefore propose that due to the tight interdependence of these two energetic  
329 parameters in combination with the structural flexibility of CTPR arrays, most mutations will be  
330 compensated by rearrangements in packing geometry and will therefore affect both intrinsic and  
331 interfacial energies. Modification of residues on the outside of helices that are not involved in any  
332 inter-repeat interactions may simply have a much smaller or no effect on the interfacial coupling,  
333 as long as such a modification does not cause any structural rearrangements or changes other  
334 side-chain interactions.

335     **The nature of the unfolded state**

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

348     **Force response of the tertiary structure**

349     Previously, several groups used steered MD simulations of natural repeat proteins to show that  
350 structural rearrangements at interfaces between repeats allowed the array to stretch as a whole  
351 before breaking of the array and unfolding of smaller structural elements occurred at higher forces  
352 [10, 11, 16]. However, at this time, we do not have evidence that the CTPR superhelix stretches  
353 before unfolding starts at the ends. On the one hand, we did not incorporate an additional term  
354 into our models to account for this, primarily to avoid over-parameterization. On the other hand,  
355 there is no indication that such a response is of similar compliance to DNA and could therefore  
356 be hidden in the DNA response. Indeed, it appears to be the reverse: since DNA parameters  
357 compensate for the dimension of the folded construct, we can estimate the end-to-end distances  
358 prior to unfolding using a linear regression and obtain values that agree well with our structural  
359 data (Fig. S14). Given that we can clearly “see” the superhelix in the force plateau, we believe  
360 that the interfaces in CTPR arrays are coupled too strongly to rearrange, a conclusion supported  
361 by previous findings that describe packing between CTPRs as rather rigid [70]. This coupling is  
362 likely much stronger than the intrinsic stability of the helices and repeats at either end of the  
363 array, and therefore unzipping occurs before any stretching can be observed. It remains to be seen  
364 how careful modulation of intrinsic repeat stability and interfacial coupling can be used alter the  
365 overall stiffness of the tertiary structure to explain the flexibility of several natural repeat proteins  
366 observed in both simulations and experiments.

367     **CONCLUSION**

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

375 **MATERIALS AND METHODS**

376 For a detailed description see the supplementary materials and methods section. In brief,  
377 repeat arrays were constructed in the background of a pRSET vector and expressed in *E. coli*.  
378 Equilibrium denaturation experiments were performed using guanidine hydrochloride in sodium  
379 phosphate buffer pH 6.8, 150 mM NaCl in a 96-well plate format [71]. CTPRrv4 with a C-terminal  
380 solvating helix was crystallized in a solution containing 0.2 M MgCl<sub>2</sub>, 0.1 M sodium cacodylate  
381 pH 6.5 and 50 % v/v PEG200. Further details on data collection and processing can be found in  
382 the supplementary materials and methods section. Angles between repeat planes were calculated  
383 essentially as published previously [13]. Constructs were prepared for force spectroscopy using site-  
384 specific modification of either terminal ybbR-tags or cysteine residues [72, 73]. All single-molecule  
385 force spectroscopy data was collected on a custom-built instrument [74], processed using custom  
386 scripts developed in Igor Pro (WaveMetrics) and further analysed using Igor Pro or Python [75–81].  
387 Theoretical FDCs were calculated using custom C++/CUDA software. Structural representations  
388 were generated using PyMol [82] or VMD [83].

389 **Data availability**

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

392 **Declaration of conflict of interest**

393 The authors declare no conflicting interests.

394 **Author contributions**

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

399 **Acknowledgements**

400 MS and AP acknowledge PhD studentship funding from the BBSRC Doctoral Training Part-  
401 nerships, Cambridge. MS acknowledges additional support from a travel grant provided by the  
402 Non-globular Protein Network (COST Action BM1405-39176) and the Eric Reid Fund for Method-  
403 ology from the British Biochemical Society, and AP from an Oliver Gatty Studentship. RE ac-  
404 knowledges funding of a PhD studentship from AstraZeneca. MR acknowledges funding by the  
405 Deutsche Forschungsgemeinschaft (DFG) through SFB863-A2 1111 6624. LSI acknowledges the  
406 support of an Senior Fellowship from the Medical Research Foundation of the UK. JS acknowl-  
407 edges the support of the LMU Center for Nanoscience CeNS and a DFG Emmy Noether grant  
408 STI673/2-1.

- 
- 409 \* To whom correspondence should be addressed: lsi10@cam.ac.uk, m.synakewicz@bioc.uzh.ch and  
410 stigler@genzentrum.lmu.de
- 411 [1] Andrey V. Kajava, “Tandem repeats in proteins: From sequence to structure,” *Journal of Structural*  
412 *Biology* **179**, 279 – 288 (2012).
- 413 [2] Matteo Pellegrini, Edward M. Marcotte, and Todd O. Yeates, “A fast algorithm for genome-wide  
414 analysis of proteins with repeated sequences,” *Proteins: Structure, Function, and Bioinformatics* **35**,  
415 440–446 (1999).
- 416 [3] D. Romera, P. Couleaud, S. H. Mejias, A. Aires, and A. L. Cortajarena, “Biomolecular templating  
417 of functional hybrid nanostructures using repeat protein scaffolds,” *Biochem Soc Trans* **43**, 825–831  
418 (2015).
- 419 [4] C. Jost and A. Plückthun, “Engineered proteins with desired specificity: DARPins, other alternative  
420 scaffolds and bispecific IgGs,” *Curr Opin Struct Biol* **27**, 102–112 (2014).
- 421 [5] D. Barrick, D. U. Ferreiro, and E. A. Komives, “Folding landscapes of ankyrin repeat proteins: exper-  
422 iments meet theory,” *Curr Opin Struct Biol* **18**, 27–34 (2008).
- 423 [6] Albert Perez-Riba, Marie Synakewicz, and Laura S. Itzhaki, “Folding cooperativity and allosteric  
424 function in the tandem-repeat protein class,” *Philosophical Transactions of the Royal Society of London*  
425 *B: Biological Sciences* **373** (2018).
- 426 [7] D. U. Ferreiro, A. M. Walczak, E. A. Komives, and P. G. Wolynes, “The energy landscapes of repeat-  
427 containing proteins: topology, cooperativity, and the folding funnels of one-dimensional architectures,”  
428 *PloS Computational Biology* **4**, e1000070 (2008).
- 429 [8] Sharona S. Cohen, Inbal Riven, Aitziber L. Cortajarena, Lucia De Rosa, Luca D. D’Andrea, Lynne  
430 Regan, and Gilad Haran, “Probing the molecular origin of native-state flexibility in repeat proteins,”  
431 *Journal of the American Chemical Society* **137**, 10367–10373 (2015).
- 432 [9] J. A. Lamboy, H. Kim, K. S. Lee, T. Ha, and E. A. Komives, “Visualization of the nanospring dynamics  
433 of the IkappaBalphalpha ankyrin repeat domain in real time,” *Proceedings of the National Academy of*  
434 *Sciences* **108**, 10178–10183 (2011).
- 435 [10] Alison Grinthal, Ivana Adamovic, Beth Weiner, Martin Karplus, and Nancy Kleckner, “PR65, the  
436 HEAT-repeat scaffold of phosphatase PP2A, is an elastic connector that links force and catalysis,”  
437 *Proceedings of the National Academy of Sciences* **107**, 2467–2472 (2010).
- 438 [11] Christian Kappel, Ulrich Zachariae, Nicole Dölker, and Helmut Grubmüller, “An unusual hydrophobic  
439 core confers extreme flexibility to HEAT repeat proteins,” *Biophysical Journal* **99**, 1596–1603 (2010).
- 440 [12] Minkyu Kim, Khadar Abdi, Gwangrog Lee, Mahir Rabbi, Whasil Lee, Ming Yang, Christopher J.  
441 Schofield, Vann Bennett, and Piotr E. Marszalek, “Fast and forceful refolding of stretched  $\alpha$ -helical  
442 solenoid proteins,” *Biophysical Journal* **98**, 3086–3092 (2010).
- 443 [13] Jade K. Forwood, Allison Lange, Ulrich Zachariae, Mary Marfori, Callie Preast, Helmut Grubmüller,  
444 Murray Stewart, Anita H. Corbett, and Bostjan Kobe, “Quantitative structural analysis of importin- $\beta$   
445 flexibility: Paradigm for solenoid protein structures,” *Structure* **18**, 1171–1183 (2010).
- 446 [14] David Serquera, Whasil Lee, Giovanni Settanni, Piotr E. Marszalek, Emanuele Paci, and Laura S.  
447 Itzhaki, “Mechanical unfolding of an ankyrin repeat protein,” *Biophysical Journal* **98**, 1294–1301 (2010).
- 448 [15] Marian Baclayon, Peter van Ulsen, Halima Mouhib, Maryam Hashemi Shabestari, Timo Verzijden,  
449 Sanne Abeln, Wouter H. Roos, and Gijs J. L. Wuite, “Mechanical unfolding of an autotransporter  
450 passenger protein reveals the secretion starting point and processive transport intermediates,” *ACS*  
451 *Nano* **10**, 5710–5719 (2016).
- 452 [16] M. Sotomayor, D. P. Corey, and K. Schulten, “In search of the hair-cell gating spring elastic properties  
453 of ankyrin and cadherin repeats,” *Structure* **13**, 669–682 (2005).
- 454 [17] Noemi Fukuhara, Elena Fernandez, Judith Ebert, Elena Conti, and Dmitri Svergun, “Conformational  
455 variability of nucleo-cytoplasmic transport factors,” *Journal of Biological Chemistry* **279**, 2176–2181  
456 (2004).
- 457 [18] Ulrich Zachariae and Helmut Grubmüller, “Importin- $\beta$ : Structural and dynamic determinants of a  
458 molecular spring,” *Structure* **16**, 906–915 (2008).
- 459 [19] Gabriel Žoldák and Matthias Rief, “Force as a single molecule probe of multidimensional protein energy  
460 landscapes,” *Current Opinion in Structural Biology* **23**, 48–57 (2013).

- 461 [20] A. K. Das, P. W. Cohen, and D. Barford, “The structure of the tetratricopeptide repeats of protein  
462 phosphatase 5: implications for TPR-mediated protein-protein interactions,” *EMBO J.* **17**, 1192–1199  
463 (1998).
- 464 [21] Luca D. D’Andrea and Lynne Regan, “TPR proteins: the versatile helix,” *Trends in Biochemical  
465 Sciences* **28**, 655 – 662 (2003).
- 466 [22] Tommi Kajander, Aitziber L. Cortajarena, Simon Mochrie, and Lynne Regan, “Structure and stabili-  
467 ty of designed TPR protein superhelices: unusual crystal packing and implications for natural TPR  
468 proteins,” *Acta Crystallographica Section D* **63**, 800–811 (2007).
- 469 [23] Hackwon Do and Muthiah Kumaraswami, “Structural mechanisms of peptide recognition and allosteric  
470 modulation of gene regulation by the rrnpp family of quorum-sensing regulators,” *Journal of Molecular  
471 Biology* **428**, 2793 – 2804 (2016).
- 472 [24] Z. Zhang, L. Chang, J. Yang, N. Conin, K. Kulkarni, and D. Barford, “The Four Canonical TPR  
473 Subunits of Human APC/C Form Related Homo-Dimeric Structures and Stack in Parallel to Form a  
474 TPR Suprahelix,” *Journal of Molecular Biology* **425**, 4236–4248 (2013).
- 475 [25] A. Perez-Riba and L. S. Itzhaki, “The tetratricopeptide-repeat motif is a versatile platform that enables  
476 diverse modes of molecular recognition,” *Curr Opin Struct Biol* **54**, 43–49 (2019).
- 477 [26] E. R. Main, Y. Xiong, M. J. Cocco, L. D’Andrea, and L. Regan, “Design of stable  $\alpha$ -helical arrays  
478 from an idealized TPR motif,” *Structure* **11**, 497–508 (2003).
- 479 [27] A. Perez-Riba, E. Komives, E. R. G. Main, and L. S. Itzhaki, “Decoupling a tandem-repeat protein:  
480 Impact of multiple loop insertions on a modular scaffold,” *Sci Rep* **9**, 15439 (2019).
- 481 [28] A. Diamante, P. K. Chaturbedy, P. J. E. Rowling, J. R. Kumita, R. S. Eapen, S. H. McLaughlin,  
482 M. de la Roche, A. Perez-Riba, and L. S. Itzhaki, “Engineering mono- and multi-valent inhibitors on  
483 a modular scaffold,” *Chem Sci* **12**, 880–895 (2021).
- 484 [29] Tommi Kajander, Aitziber L. Cortajarena, Ewan R. G. Main, Simon G. J. Mochrie, and Lynne  
485 Regan, “A new folding paradigm for repeat proteins,” *Journal of the American Chemical Society* **127**,  
486 10188–10190 (2005).
- 487 [30] Aitziber L. Cortajarena, Simon G.J. Mochrie, and Lynne Regan, “Mapping the energy landscape of  
488 repeat proteins using NMR-detected hydrogen exchange,” *Journal of Molecular Biology* **379**, 617 – 626  
489 (2008).
- 490 [31] A. L. Cortajarena and L. Regan, “Calorimetric study of a series of designed repeat proteins: modular  
491 structure and modular folding,” *Protein Sci.* **20**, 336–340 (2011).
- 492 [32] Yalda Javadi and Ewan R. G. Main, “Exploring the folding energy landscape of a series of designed  
493 consensus tetratricopeptide repeat proteins,” *Proceedings of the National Academy of Sciences* **106**,  
494 17383–17388 (2009).
- 495 [33] A. Perez-Riba, A. R. Lowe, E. R. G. Main, and L. S. Itzhaki, “Context-Dependent Energetics of Loop  
496 Extensions in a Family of Tandem-Repeat Proteins,” *Biophysical Journal* **114**, 2552–2562 (2018).
- 497 [34] C. Millership, J. J. Phillips, and E. R. Main, “Ising Model Reprogramming of a Repeat Protein’s  
498 Equilibrium Unfolding Pathway,” *Journal of Molecular Biology* **428**, 1804–1817 (2016).
- 499 [35] Aitziber L. Cortajarena, Simon G. J. Mochrie, and Lynne Regan, “Modulating repeat protein stability:  
500 The effect of individual helix stability on the collective behavior of the ensemble,” *Protein Science* **20**,  
501 1042–1047 (2011).
- 502 [36] J. J. Phillips, Y. Javadi, C. Millership, and E. R. Main, “Modulation of the multistate folding of  
503 designed TPR proteins through intrinsic and extrinsic factors,” *Protein Sci.* **21**, 327–338 (2012).
- 504 [37] Gwangrog Lee, Khadar Abdi, Yong Jiang, Peter Michael, Vann Bennett, and Piotr E. Marszalek,  
505 “Nanospring behaviour of ankyrin repeats,” *Nature* **440**, 246–249 (2006).
- 506 [38] Lewyn Li, Svava Wetzel, Andreas Plückthun, and Julio M. Fernandez, “Stepwise unfolding of ankyrin  
507 repeats in a single protein revealed by atomic force microscopy,” *Biophysical Journal* **90**, L30–L32  
508 (2006).
- 509 [39] Q. Li, Z. N. Scholl, and P. E. Marszalek, “Capturing the mechanical unfolding pathway of a large protein  
510 with coiled-coil probes,” *Angewandte Chemie International Edition in English* **53**, 13429–13433 (2014).
- 511 [40] W. Lee, X. Zeng, H. X. Zhou, V. Bennett, W. Yang, and P. E. Marszalek, “Full reconstruction of a  
512 vectorial protein folding pathway by atomic force microscopy and molecular dynamics simulations,” *J  
513 Biol Chem* **285**, 38167–38172 (2010).
- 514 [41] Giovanni Settanni, David Serquera, Piotr E. Marszalek, Emanuele Paci, and Laura S. Itzhaki, “Effects  
515 of ligand binding on the mechanical properties of ankyrin repeat protein gankyrin,” *PLoS Comput Biol*

- 516 9, e1002864 (2013).
- 517 [42] T. J. Magliery and L. Regan, "Beyond consensus: statistical free energies reveal hidden interactions in  
518 the design of a TPR motif," *Journal of Molecular Biology* **343**, 731–745 (2004).
- 519 [43] Bharat Jagannathan and Susan Marqusee, "Protein folding and unfolding under force." *Biopolymers*  
520 **99**, 860–869 (2013).
- 521 [44] Jun Yin, Alison J. Lin, David E. Golan, and Christopher T. Walsh, "Site-specific protein labeling by  
522 Sfp phosphopantetheinyl transferase," *Nature Protocols* **1**, 280 (2006).
- 523 [45] Jun Yin, Paul D. Straight, Shaun M. McLoughlin, Zhe Zhou, Alison J. Lin, David E. Golan, Neil L.  
524 Kelleher, Roberto Kolter, and Christopher T. Walsh, "Genetically encoded short peptide tag for ver-  
525 satile protein labeling by Sfp phosphopantetheinyl transferase," *Proceedings of the National Academy  
526 of Sciences* **102**, 15815–15820 (2005).
- 527 [46] Joost van Mameren, Peter Gross, Geraldine Farge, Pleuni Hooijman, Mauro Modesti, Maria Falkenberg,  
528 Gijs J L Wuite, and Erwin J G Peterman, "Unraveling the structure of DNA during overstretching  
529 by using multicolor, single-molecule fluorescence imaging." *Proceedings of the National Academy of  
530 Sciences* **106**, 18231–18236 (2009).
- 531 [47] Ingo Schwaiger, Clara Sattler, Daniel R Hostetter, and Matthias Rief, "The myosin coiled-coil is a  
532 truly elastic protein structure." *Nature materials* **1**, 232–235 (2002).
- 533 [48] U Bockelmann, B Essevaz-Roulet, and F Heslot, "Molecular Stick-Slip Motion Revealed by Opening  
534 DNA with Piconewton Forces," *Physical Review Letters* **79**, 4489–4492 (1997).
- 535 [49] Thomas Bornschlögl and Matthias Rief, "Single molecule unzipping of coiled coils: sequence resolved  
536 stability profiles." *Physical Review Letters* **96**, 118102 (2006).
- 537 [50] Josep Maria Huguet, Cristiano V Bizarro, Núria Forns, Steven B Smith, Carlos Bustamante, and Felix  
538 Ritort, "Single-molecule derivation of salt dependent base-pair free energies in DNA." *Proceedings of  
539 the National Academy of Sciences* **107**, 15431–15436 (2010).
- 540 [51] Gabriel Žoldák, Johannes Stigler, Benjamin Pelz, Hongbin Li, and Matthias Rief, "Ultrafast folding  
541 kinetics and cooperativity of villin headpiece in single-molecule force spectroscopy," *Proceedings of the  
542 National Academy of Sciences* **110**, 18156 (2013).
- 543 [52] Allison Solanki, Krishna Neupane, and Michael T. Woodside, "Single-molecule force spectroscopy of  
544 rapidly fluctuating, marginally stable structures in the intrinsically disordered protein  $\alpha$ -synuclein,"  
545 *Physical Review Letters* **112**, 158103 (2014).
- 546 [53] Cecilia C. Mello and Doug Barrick, "An experimentally determined protein folding energy landscape,"  
547 *Proceedings of the National Academy of Sciences* **101**, 14102–14107 (2004).
- 548 [54] Jacob D. Marold, Jennifer M. Kavran, Gregory D. Bowman, and Doug Barrick, "A naturally occurring  
549 repeat protein with high internal sequence identity defines a new class of tpr-like proteins," *Structure*  
550 **23**, 2055 – 2065 (2015).
- 551 [55] Svava K. Wetzel, Giovanni Settanni, Manca Kenig, H. Kaspar Binz, and Andreas Plückthun, "Folding  
552 and unfolding mechanism of highly stable full-consensus ankyrin repeat proteins," *Journal of Molecular  
553 Biology* **376**, 241 – 257 (2008).
- 554 [56] Tural Aksel and Doug Barrick, "Analysis of repeat-protein folding using nearest-neighbor statistical  
555 mechanical models," in *Biothermodynamics, Part A, Methods in Enzymology*, Vol. 455, edited by  
556 Michael L. Johnson, Jo M. Holt, and Gary K. Ackers (Academic Press, 2009) Chap. 4, pp. 95–125.
- 557 [57] T. Aksel, A. Majumdar, and D. Barrick, "The contribution of entropy, enthalpy, and hydrophobic  
558 desolvation to cooperativity in repeat-protein folding," *Structure* **19**, 349–360 (2011).
- 559 [58] T. Aksel and D. Barrick, "Direct observation of parallel folding pathways revealed using a symmetric  
560 repeat protein system," *Biophysical Journal* **107**, 220–232 (2014).
- 561 [59] Kathryn Geiger-Schuller and Doug Barrick, "Broken TALEs: Transcription Activator-like Effectors  
562 Populate Partly Folded States," *Biophysical Journal* **111**, 2395–2403 (2016).
- 563 [60] K. Geiger-Schuller, K. Sforza, M. Yuhas, F. Parmeggiani, D. Baker, and D. Barrick, "Extreme sta-  
564 bility in de novo-designed repeat arrays is determined by unusually stable short-range interactions,"  
565 *Proceedings of the National Academy of Sciences* **115**, 7539–7544 (2018).
- 566 [61] Ernst Ising, "Beitrag zur Theorie des Ferromagnetismus," *Zeitschrift für Physik* **31**, 253–258 (1925).
- 567 [62] Wilhelm Lenz, "Beitrag zum Verständnis der magnetischen Erscheinungen in festen Körpern,"  
568 *Physikalische Zeitschrift* **21**, 613–615 (1920).
- 569 [63] Alan R. Lowe, Albert Perez-Riba, Laura S. Itzhaki, and Ewan R.G. Main, "Pyfolding: Open-source  
570 graphing, simulation, and analysis of the biophysical properties of proteins," *Biophysical Journal* **114**,

- 571 511–521 (2018).
- 572 [64] Ewan R. G. Main, Katherine Stott, Sophie E. Jackson, and Lynne Regan, “Local and long-range sta-  
573 bility in tandemly arrayed tetratricopeptide repeats,” *Proceedings of the National Academy of Sciences*  
574 **102**, 5721–5726 (2005).
- 575 [65] Maksym Tsytlonok, Patricio O. Craig, Elin Sivertsson, David Serquera, Sarah Perrett, Robert B.  
576 Best, Peter G. Wolynes, and Laura S. Itzhaki, “Complex energy landscape of a giant repeat protein,”  
577 *Structure* **21**, 1954–1965 (2013).
- 578 [66] N. D. Werbeck, P. J. Rowling, V. R. Chellamuthu, and L. S. Itzhaki, “Shifting transition states in the  
579 unfolding of a large ankyrin repeat protein,” *Proceedings of the National Academy of Sciences* **105**,  
580 9982–9987 (2008).
- 581 [67] Katherine W. Tripp and Doug Barrick, “Rerouting the folding pathway of the notch ankyrin domain  
582 by reshaping the energy landscape,” *Journal of the American Chemical Society* **130**, 5681–5688 (2008).
- 583 [68] Johannes Stigler and Matthias Rief, “Calcium-dependent folding of single calmodulin molecules,” *Pro-  
584 ceedings of the National Academy of Sciences* **109**, 17814–17819 (2012).
- 585 [69] A. L. Cortajarena, G. Lois, E. Sherman, C. S. O’Hern, L. Regan, and G. Haran, “Non-random-coil  
586 behavior as a consequence of extensive PPII structure in the denatured state,” *Journal of Molecular  
587 Biology* **382**, 203–212 (2008).
- 588 [70] C. Y. Cheng, V. A. Jarymowycz, A. L. Cortajarena, L. Regan, and M. J. Stone, “Repeat motions and  
589 backbone flexibility in designed proteins with different numbers of identical consensus tetratricopeptide  
590 repeats,” *Biochemistry* **45**, 12175–12183 (2006).
- 591 [71] A. Perez-Riba and L. S. Itzhaki, “A method for rapid high-throughput biophysical analysis of proteins,”  
592 *Scientific Reports* **7**, 9071 (2017).
- 593 [72] M. Synakiewicz, D. Bauer, M. Rief, and L. S. Itzhaki, “Bioorthogonal protein-DNA conjugation methods  
594 for force spectroscopy,” *Sci Rep* **9**, 13820 (2019).
- 595 [73] Ann Mukhortava and Michael Schlierf, “Efficient formation of site-specific protein-dna hybrids using  
596 copper-free click chemistry,” *Bioconjugate Chemistry* **27**, 1559–1563 (2016).
- 597 [74] Yann von Hansen, Alexander Mehlich, Benjamin Pelz, Matthias Rief, and Roland R. Netz, “Auto-  
598 and cross-power spectral analysis of dual trap optical tweezer experiments using bayesian inference,”  
599 *Review of Scientific Instruments* **83**, 095116 (2012).
- 600 [75] K. Jarrod Millman and Michael Aivazis, “Python for scientists and engineers,” *Computing in Science  
601 & Engineering* **13**, 9–12 (2011).
- 602 [76] Travis E. Oliphant, “Python for scientific computing,” *Computing in Science & Engineering* **9**, 10–20  
603 (2007).
- 604 [77] Stéfan van der Walt, S. Chris Colbert, and Gaël Varoquaux, “The numpy array: A structure for  
605 efficient numerical computation,” *Computing in Science & Engineering* **13**, 22–30 (2011).
- 606 [78] J. D. Hunter, “Matplotlib: A 2D graphics environment,” *Computing In Science & Engineering* **9**, 90–95  
607 (2007).
- 608 [79] Michael Waskom and the seaborn development team, “mwaskom/seaborn,” (2020).
- 609 [80] Eric Jones, Travis Oliphant, Pearu Peterson, *et al.*, “SciPy: Open source scientific tools for Python,”  
610 (2001), <http://www.scipy.org/>.
- 611 [81] Wes McKinney, “Data Structures for Statistical Computing in Python,” in *Proceedings of the 9th  
612 Python in Science Conference*, edited by Stéfan van der Walt and Jarrod Millman (2010) pp. 56 – 61.
- 613 [82] Schrödinger, LLC, “The PyMOL molecular graphics system, version 1.8,” (2015).
- 614 [83] W. Humphrey, A. Dalke, and K. Schulten, “VMD: visual molecular dynamics,” *Journal of Molecular  
615 Graphics* **14**, 33–38 (1996).