1

2 Context-dependent energetics of loop extensions 3 in a family of tandem-repeat proteins

4

5 Albert Perez-Riba[1], Alan R. Lowe[2,3,4], Ewan R.G. Main*[5] and Laura S.
6 Itzhaki*[1]

7 *Corresponding authors: lsi10@cam.ac.uk, e.main@qmul.ac.uk.

8 [1]Department of Pharmacology, University of Cambridge
9 Tennis Court Road, Cambridge
10 CB2 1PD, UK
11
12 [2] London Centre for Nanotechnology
13 17-19 Gordon Street, London
14 WC1H 0AH, UK
15
16 [3] Structural & Molecular Biology, University College London,
17 Gower Street, London,
18 WC1E 6BT, UK
19
20 [4] Department of Biological Sciences, Birkbeck College, University of London
21 Malet Street, London
22 WC1E 7HX, UK
23
24 [5] School of Biological and Chemical Sciences, Queen Mary University of London
25 Mile End Road, London
26 E1 4NS, UK

27

28 ## Abstract

29 **Consensus-designed tetratricopeptide repeat proteins (CTPRs) are highly stable,**
30 **modular proteins that are strikingly amenable to rational engineering. They therefore**
31 **have tremendous potential as building blocks for biomaterials and biomedicine. Here we**
32 **explore the possibility of extending the loops between repeats to enable further**
33 **diversification, and we investigate how this modification affects stability and folding**
34 **cooperativity. We find that extending a single loop by up to 25 residues does not disrupt**
35 **the overall protein structure, but, strikingly, the effect on stability is highly context-**
36 **dependent: In a two-repeat array, destabilisation is relatively small and can be accounted**
37 **for purely in entropic terms, whereas extending a loop in the middle of a large array is**
38 **much more costly, due to weakening of the interaction between the repeats. Our findings**
39 **provide new insights into structure and folding that will be important both for**
40 **understanding the function of natural repeat proteins and for the design of artifical**
41 **repeat proteins in biotechnology.**
42

43 ## Introduction

44 Tandem-repeat arrays are one of the most common protein architectures. Their high frequency
45 is considered to be a result of DNA replication slippage and recombination events (1, 2). The
46 α-solenoids are one large family composed of such tandem-repeat arrays. Their repeats
47 comprise between 12 and 45 amino acids that form pairs of antiparallel α-helices. Examples
48 include ankyrin repeats, armadillo repeats and HEAT repeats (**H**untingtin, **e**longation factor 3,
49 protein phosphatase 2**A** subunit, and the yeast kinase **T**OR1) and TPRs (tetratricopeptide
50 repeats) (3–6). They function in mediating protein-protein interactions by providing extended
51 surfaces for molecular recognition. Moreover, the modularity of their architectures have
52 allowed the design of ultra-stable consensus repeat proteins by selecting the most conserved
53 residues in each family (7–11).

54 In contrast with globular proteins, repeat proteins have quasi-1D structures that are
55 stabilised exclusively by interactions between residues close in primary sequence. Despite the

56 lack of sequence-distant contacts, repeat proteins are able to fold in a cooperative manner. The
57 co-operativity arises due to the mismatch between the intrinsically unstable repeats and the
58 highly stabilising inter-repeat interfaces (12). Repeat-protein folding can be modelled using 1-
59 D Ising formalism (13), which assumes that each repeat is either folded or unfolded, and that
60 this state is determined by both the intrinsic repeat stability (ΔG_i) and energetic coupling
61 between the nearest neighbours, also referred to as the interface stability (ΔG_{ij}). The simplest
62 expression of the 1-D Ising model, the homopolymer model, assumes single values of intrinsic
63 and interfacial stabilities, and it has been shown to be valid for proteins comprising tandem
64 arrays of identical repeats. One of the most important implications of this description of repeat
65 protein folding is that the stability of the protein should scale linearly with the number of
66 repeating units, referred to as "additive rule" of the 1-D Ising model:

67 $$\Delta G_{\text{D-N}} = n\Delta G_{i} + (n-1)\Delta G_{ij}$$

68 where n is the number of repeating units (12, 13).

69 The folding of natural repeat proteins has been characterised both experimentally and
70 in silico (14–22). The best-studied consensus-designed repeat proteins are the consensus
71 ankyrin repeats (referred to as DARPins (8) or CARPs (7)) and consensus tetratricopeptide
72 repeats (CTPRs) (13, 23). Although both have repeat units composed of pairs of antiparallel α-
73 helices, they are structurally and energetically quite different. CARPs/DARPins are stabilised
74 by a much larger interfacial term than the CTPRs. This can be attributed in part to the long
75 semi-structured loops of the former that have extensive hydrogen-bonding networks (24, 25).
76 CTPRs, in contrast, have very short (four-residue) loops that are involved in a more limited,
77 though still significant, number of stabilising interactions (24, 25).

78 The structural simplicity of consensus-designed repeat proteins makes them popular
79 systems to engineer for biotechnology purposes (10, 26–29). Two significant outputs from
80 these studies are the use of repeat proteins as building blocks for self-assembly systems and as
81 alternatives to antibodies. In such systems, an avenue for further functionalisation would be

82 the extension of the loops between repeats to enable additional materials diversification. To
83 this end, we created a series of 15 CTPR proteins that contained different numbers of repeats
84 of different sequences. Into two of these proteins (CTPR2 and CTPR6) we engineered a loop
85 between two adjacent repeats with a poly-GS linker of variable length between 10 and 25
86 residues. The loop-extension proteins together with the other proteins within the series were
87 assayed using equilibrium denaturation experiments and globally analysed using a
88 heteropolymer Ising model. This global analysis allowed us both to determine the energetic
89 contributions of non-identical repeat units and to dissect the contributions from the intrinsic
90 stability of each repeat and each interface between repeats. The results show that extending a
91 single inter-repeat loop by up to 25 amino acids can be tolerated within the overall native
92 structure. Moreover, although increasing the length of the inter-repeat loop weakens the
93 nearest-neighbour cooperativity, it does not completely abolish it. Importantly, therefore, our
94 results demonstrate that CTPR arrays are amenable to further functionalisation through both
95 large and small loop insertions. Strikingly, we find that the loss of stability associated with loop
96 insertion is highly context-dependent: When a loop is inserted into a two-repeat array the
97 destabilisation incurred is much smaller than the same loop inserted between the two central
98 repeats of a six-repeat array. These results indicate that loop insertion destabilises through both
99 the entropic cost of loop closure and also the decoupling of the adjacent repeat modules.

100 In summary, our study provides new insights into the TPR proteins, a family with over
101 500,000 sequences in which long inter-repeat loops are often observed (30). Our results show
102 that the insertion of a long loop between repeat motifs weakens the inter-repeat interface, which
103 could cause the repeats to decouple, thereby stabilising partly folded states. Such decoupling
104 would enable loop-containing proteins to display enhanced conformational dynamics and/or
105 mechanical flexibility. These properties may regulate the biological functions of natural repeat

106 proteins and should be considered when used as an avenue for functionalisation of artificial
107 repeat proteins for biotechnological applications.

108 **Materials and methods**

109

110 **Construction of tandem-repeat genes from individual repeat sequences**

111 **CTPRn, CTPR-YD and CTPR2-loop constructs:** All constructs were commercially
112 synthesised by GeneArt Invitrogen. Each construct was generated with a BamHI and a HindIII
113 site for subcloning into pRSet for His-tag purification.

114 **CTPRa2 construct:** The tandem repeat arrays of two repeats was constructed by
115 concatemerization of two individual CTPRa motifs using BamHI and BglII sites (31). Briefly,
116 a single consensus tetratricopeptide repeat (CTPRa1) was purchased as a "gBlock" oligo
117 (Figure S1) and inserted into the multi cloning site of the vector pRSET B between the BamHI
118 and HindIII restriction sites (ThermoFisher Scientific). An oligo consisting of the CTPRa1
119 "gBlock" was then PCR-amplified using primers complementary to the T7 promoter sites on
120 each side of the multi cloning site of pRSET B. This PCR product and the CTPRa1 gene in
121 the pRSET B vector were then digested with BamHI/HindIII and BglII/HindIII restriction
122 enzymes, respectively. The two digested products could then ligated to form a CTPRa2 gene
123 (as the BamH1 and BglII sites leave compatible ligation ends). The ligation of BamHI and
124 BglII leaves an Arg and a Ser after the Pro at position 31 of the CTPR sequence. This results
125 in a DPRS loop in the CTPRa2 (i.e. two-repeat array) (32). This process can be repeated as
126 many times as required to generate CTPRa arrays of different lengths.

127 **CTPR6-YD-loop constructs:** Loop extensions of different length were added to the C-
128 terminus of CTPR3n templates at the DNA-level by whole plasmid Round-the-Horn
129 polynucleotide chain reaction (PCR) (33). This method enables large insertions to be made in
130 a plasmid. Primers are designed so that they anneal back to back on the plasmid, with the
131 desired insertion on the 5'-end of one primer (or separated onto both primers for large inserts).

132

133 **Protein purification**

134 The pRSET B (His-tagged) constructs were transformed into chemically competent *E. coli* C41
135 cells by heat shock and plated on LB-Amp plates. Colonies were grown in 2TY media
136 containing ampicillin (50 μg/mL) at 37 ^{o}C, 220 rpm until the optical density (O.D.) at 600 nm
137 reached 0.6. Cultures were then induced with IPTG (0.5mM) for 16-20 h at 20°C. Cells were
138 pelleted by centrifugation at 3000 g (4 ^{o}C, 10 min) and resuspended in lysis buffer (10 mM
139 sodium phosphate pH 7.4, 150 mM NaCl, 1 tablet of SIGMAFAST protease inhibitor cocktail
140 (EDTA-free per 100 mL of solution), and lysed on an Emulsiflex C5 homogenizer at 15000
141 psi. Cell debris was pelleted by centrifugation at 15,000 g at 4 ^{o}C for 45 min. Ni-NTA beads
142 50% bed volume (GE Healthcare) (5 mL) were washed once with phosphate buffer (10 mM
143 sodium phosphate pH 7.4, 150 mM NaCl) before binding the supernatant from the cell lysate
144 for 1 hr at 4 ^{o}C in batch. The beads were washed three times with phosphate buffer (40 mL)
145 containing 30 mM of imidazole to prevent nonspecific interaction of lysate proteins with the
146 beads. Protein was eluted using phosphate buffer with 300 mM Imidazole and purified by size-
147 exclusion gel-filtration using a HiLoad 16/60 SuperdexG75 column (GE Life-Science) pre-
148 equilibrated in phosphate buffer (10 mM sodium phosphate, pH 7.4, 150 mM NaCl) and
149 proteins separated in isocratic conditions. Purity was checked by NuPage protein gel
150 (Invitrogen) and pure protein fractions were pooled. Purified protein was flash-frozen and
151 stored at -80 ^{o}C until further use. Concentrations were determined by absorbance at 280 nm
152 using a calculated extinction coefficient (ExPASy ProtParam) (34) for each variant. Protein
153 molecular weight and purity was confirmed using mass spectrometry (MALDI) (Mass
154 Spectrometry Facility, Department of Chemistry, or PNAC, Department of Biochemistry).

155

156

157 **Circular dichroism (CD) spectroscopy**

158 All CD measurements were made under the same configuration on a Chirascan CD
159 spectrometer (Applied Photophysics) in 1 mm pathlength Precision Cells (110-QS, Hellma
160 Analytics) at 25 °C. All protein samples (at 5- 20 μM concentration) were prepared in 50 mM
161 sodium phosphate buffer pH 6.8, 150 mM NaCl, and the CD spectrum was measured between
162 200 nm to 280 nm wavelengths using a 1 nm of bandwidth unless specified otherwise.
163 Measurements were taken at 1 nm intervals and were collected every 0.5 s; each reading was
164 repeated between three and five times and the data averaged.

165

166 **Equilibirum denaturation monitored by fluorescence spectroscopy**

167 High-throughput equilibrium denaturation experiments were carried as previously described
168 (35). Briefly, solutions were dispensed into Corning® 96-well, half area, black polystyrene
169 plates (CLS3993) with a Microlab ML510B dispenser. All plate measurements were carried on
170 a CLARIOstar Plate Reader (BMG labtech) with a tryptophan detection set consisting of three
171 filters, an excitation of 280-10 nm (275 nm to 285 nm), a dichroic PL325 nm and an emission
172 at 360-20 nm (350 nm to 370 nm) at 25 °C. Protein concentrations were 0.3-1 μM. For each
173 protein, three sets of serial dilutions were plated consecutively. Plates were covered with a
174 Corning® 96 Well Microplate Aluminium Sealing Tape to prevent evaporation, shaken for 30
175 s with the CLARIOstar double orbital shaking option, and incubated at 25 °C for 1 h. The
176 temperature was set at 25 °C for the duration of the experiment.

177

178 **Equilibirum denaturation monitored by CD**

179 Aliquots of GdmHCl (300 μL) were prepared by dispensing the appropriate volume of stock
180 solution of GdmHCl (7 M) in buffer (50 mM sodium phosphate buffer pH 6.8, 150 mM NaCl)
181 and sodium phosphate buffer (or otherwise indicated) using a Hamilton Microlab ML510B.

182 Samples were equilibrated at 25 °C for 2 hours. The α-helicity was monitored by ellipticity at
183 222 nm. Results were plotted using GraphPad Prism, and a two-state model was used to
184 describe the system and calculate the mid-points and the slope of the transition (*m* value).

185

186 **Equilibrium denaturation data analysis**

187 Data were analysed in two different ways as follows: They were either analysed with a two-
188 state model (36) or with a heteropolymer Ising model (12). Analysis of the data with the
189 heteropolymer Ising model is described below. In the case of two-state model analysis, the
190 protein chemical denaturations were fitted directly using equation 1.

191 **Equation 1:** $$\lambda_{obs} = \frac{\alpha_N+\beta_N[\mathrm{D}]+(\alpha_D+\beta_D[\mathrm{D}]).\exp[m_{D-N}([\mathrm{D}]-[\mathrm{D}]_{50\%})]/_{RT}}{1+\exp[m_{D-N}([\mathrm{D}]-[\mathrm{D}]_{50\%})]}$$

192 where λ_{obs} is the observed fluorescence, α_N and α_D are the intercepts, and β_N and β_D are the
193 slopes of the baselines at the lo (N) and high (D) denaturant concentrations, $[D]_{50\%}$ is the
194 midpoint of unfolding, [D] is the concentration of denaturant and $m_{D\text{-}N}$ is a constant that is
195 related to the increase in solvent exposure of the protein upon unfolding (37).

196 Equation 1 is based on a two-state model of denaturation where only the native and the
197 denatured states are populated, and assumes that the signal of the native state, λ_N, and the
198 denatured state, λ_D, are linearly dependent on the denaturant concentration ($\lambda_N = \alpha_N + \beta_N[D]$,
199 $\lambda_D = \alpha_D + \beta_D[D]$); for a detailed derivation see (36). Values for $[D]_{50\%}$ and $m_{D\text{-}N}$ are obtained
200 with their standard errors. The free energy of unfolding in water can then be calculated using
201 equation 2:

202 **Equation 2:** $$\Delta G_{D-N}^{H_2O} = m_{D-N}.[D]_{50\%}$$

203 where $\Delta G_{D-N}^{H_2O}$ is the free energy of unfolding in water, m_{D-N} is the m-value and $[D]_{50\%}$ is the
204 equilibrium midpoint.

205

Heteropolymer Ising model.

For the Ising analysis, each equilibrium denaturation curve was individually converted to fraction unfolded (λ_U) using Equation 3:

Equation 3: $$\lambda_U = \frac{\lambda_{obs} - (\alpha_N + \beta_N[D])}{(\alpha_D - \alpha_N) + (\beta_D - \beta_N)[D]}$$

where α_D / α_N are the y-intercept values of the denatured / native baselines and β_D / β_N are the slopes of the denatured /native baselines.

After normalization, all of the curves were globally fitted to a heteropolymer Ising model using the PyFolding package (38). We constructed the one-dimensional heteropolymer Ising model using a matrix formulation as previously described (12). Briefly, the model comprises a one-dimensional linear series of equilibrium constants. These account for the intrinsic folding stability (ΔG_i) and the interfacial energy ($\Delta G_{i-1,i}$) for each repeated unit in a nearest-neighbour TPR array. The intrinsic stability of the repeating unit has an associated coefficient (m) to represent its sensitivity to the external stimulus – in this case chemical denaturant.

In previous studies on CTPR proteins the repeating Ising unit used has been at the level of individual helices within each array (13, 32, 39). Here, the CTPR series were fitted to both (i) different repeating units of individual helices and (ii) different repeating units of TPR motifs. The fits showed that the model with different repeating unit of TPR motifs gives better agreement to the experimental data. This is most likely due to the nature of the input protein series used. i.e. the input proteins differ in number of TPR motifs as opposed to one with differing numbers of helices. Thus, asymmetry of CTPR proteins was modelled *via* unique sets of parameters to represent a "standard" CTPR motif (ΔG_i^{CTPR}, $\Delta G_{i-1,i}^{CTPR}$ and m^{CTPR}), a CTPR motif with the D to Y mutation ($\Delta G_i^{CTPR-Y91D}$, $\Delta G_{i-1,i}^{CTPR-Y91D}$ and $m^{CTPR-Y91D}$) and inserted single loops with the CTPR motif proceeding it ($\Delta G_i^{loop-CTPR}$, $\Delta G_{i-1,i}^{Loop-CTPR}$ and $m^{Loop-CTPR}$). The m parameters (m^{CTPR}, $m^{CTPR-Y91D}$ and $m^{Loop-CTPR}$) gave a denaturant dependence to the

231 intrinsic stabilities. The expressions defining the equilibrium constants (Equations 4 and 5) and

232 the protein partition function, *q(n)* are given below (Equation 6):

233 **Equation 4:** $\kappa_i = e^{[-(\Delta G_i - mx)/RT]}$

234 **Equation 5:** $\tau_{i-1,i} = e^{[-\Delta G_{i-1,i}/RT]}$

235 where ΔG_i is the free energy of folding for the domain at position i, with denaturant sensitivity

236 *m* and at denaturant concentration *x*. $\Delta G_{i-1,i}$ is the free energy for the interface between domains

237 at positions i-1 and i. *R* is the gas constant and *T* is experimental temperature in Kelvin.

238 The full partition function of the protein with *n* repeat motifs is given by Equation 6:

239 **Equation 6:** $q(n) = [0 \quad 1] \begin{bmatrix} \kappa_1\tau_{-1} & 1 \\ \kappa_1 & 1 \end{bmatrix} \cdots \begin{bmatrix} \kappa_n\tau_{n-1} & 1 \\ \kappa_n & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

240 This defines the fully folded state. The model allows for fitting of separate parameters (κ and

241 τ, thus ΔG_i, $\Delta G_{i-1,i}$ and *m*) to describe behaviour of the various repeat motif units by globally

242 fitting to data for degenerate CTPR protein compositions.

243 The fraction folded, λ_F is then simply defined as the sum of the subpartition functions

244 divided by the number of terms (repeat motifs) multiplied by the full partition function

245 (Equations 7-8):

246 **Equation 7:** $q(i) = [0 \quad 1] \cdots \begin{bmatrix} \kappa_i\tau_{i-1} & 0 \\ \kappa_i & 0 \end{bmatrix} \cdots \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

247 **Equation 8:** $\lambda_F = \frac{1}{nq(n)} \sum_{i=0}^{n} q(i)$

248 From the fitted variables the stability of any CTPR ensemble or part thereof ($\Delta G_{0\to j}^{H20}$) can be

249 calculated by adding energy terms (Equation 9):

250 **Equation 9:** $\Delta G_{0\to j}^{H20} = n\Delta G_i + (n-1)\Delta G_{i,j} = -RT \ln \kappa^n \tau^{(n-1)}$

251 where $\Delta G_{0\to j}^{H20}$ is the free energy of folding in water for a protein with *j* repeat motifs, *n* is the

252 number of folded repeat motifs in each protein, ΔG_i is the free energy of folding for the motif

253 at position i, and $\Delta G_{i-1,i}$ is the free energy for the interface between motifs at positions i-1 and
254 i.
255
256 **Stopped-flow fluorescence**
257 Aliquots of guanidinium hydrochloride (GdmHCl) were prepared by dispensing the
258 appropriate volume of stock solution of GdmHCl in sodium phosphate buffer (50 mM sodium
259 phosphate buffer pH 6.8, 150 mM NaCl) using a Hamilton Microlab ML510B dispenser. For
260 each protein, two aliquots (3 mL) were prepared to a final concentration of 10 μM of protein.
261 One aliquot was fully folded in sodium phosphate buffer (or low concentrations of GdmHCl)
262 and the other denatured in 6M GdmHCl. Samples were equilibrated at 10 °C or 25 °C for 2
263 hours. The proteins and the GdmHCl solutions were mixed at a 1:5 ratio. An excitation
264 wavelength of 280 nm was used, and the emission was measured using a 330 nm cut-off filter.
265 Unfolded protein was refolded by rapid mixing with increasing concentrations of GdmHCl up
266 to the denaturation mid-point as defined by equilibrium denaturation. Folded protein was
267 unfolded by rapid mixing with increasing concentrations of GdmHCl above the equilibrium
268 denaturation midpoint. Multiple traces were acquired at each GdmHCl concentration, averaged
269 and then fitted to a single exponential or a double exponential in GraphPad Prism.
270 Chevron plots that showed non-linear folding and/or unfolding arms were fitted using
271 a broad transition state barrier model originally described by Oliveberg and coworkers (ref).
272 Nevertheless, the fit was simply qualitative, as the refolding rates of these CTPR proteins are
273 faster than the limit of detection of our instrument:
274 **Equation 10:** $\ln k_{obs} = \ln\left(k_f^{H_2O} exp\,(-m_{k_f}[denaturant]\right) + \exp\,(-m_{k_f}^{*}[denaturant]^2) +$
275 $k_u^{H_2O} + \exp\,(m_{k_u}[denaturant]) + \exp\,(m_{k_u}^{*}[denaturant]^2))$
276
277

278 **Statistical analysis**

279 All measurements were performed in triplicate unless when indicated, and the errors for the
280 two-state fits are the standard errors of the mean. The errors from $\Delta G_{D-N}^{H_2O}$ calculation were
281 propagated from standard errors of the mean. Errors of the fitted variables by the 1-D
282 Heteropolymer Ising model were determined by calculating a covariance matrix from the
283 Jacobian matrix following a subsequent least-squares minimisation of the fit. Errors in $\Delta G_{0\to1}^{H20}$
284 were propagated from the errors obtained from the fitted variables.

285

286 **Data availability**

287 iPython Juypter notebooks of the Heteropolymer Ising model analysis are adjunted as
288 supplementary information. All data is available upon request. To create the figures in the
289 paper, the fitting results from PyFolding were exported as CSV files and plotted using the
290 program PRISM (GraphPad Software Inc, San Diego, USA).

291

292

Results

Design of consensus-repeat modules & loop extensions

In this study, we constructed 15 CTPR proteins that contain different numbers of repeats with two consensus repeat sequences differing by a point mutation and a single loop insertion of different lengths (shown schematically in Fig. 1). Comparison of the biophysical characteristics of all these different CTPR constructs enabled us to delineate the effects of loop insertion and of size of loop inserted versus the effects of point mutation. The 15 CTPR proteins consisted of: (i) a CTPR3 module (comprising three CTPR motifs), as studied previously by Regan and colleagues and referred to here as CTPR3-YD. In the third repeat there is a single point mutation, Y91D, relative to other published CTPR sequences (40). (ii) A six-repeat series built from two CTPR3-YD modules with either a native loop between the two modules (CTPR6-YD) or a poly-GS loop of differing length between 10 and 25 residues inserted between the two CTPR3-YD modules (CTPR6-YD-loop10, CTPR6-YD-loop15, etc.). The poly-GS loop contains a thrombin cleavage site that allowed us to demonstrate that the loop is solvent-accessible (see SI [Supplementary Information] – Fig S4). (iii) A series of four proteins (CTPR2, CTPR3, CTPR4 and CTPR6) comprising between two and four repeats of the original consensus sequence in (9). (iv) A 2-repeat series comprising CTPRa2 and CTPR2 with either a 10-residue or a 25-residue loop between the two repeats and versions of them with the Y-to-D point mutation. To simplify the analysis, all of the proteins lack the C-terminal so-called ‘solvating’ helix used in some previous studies. All expressed in *E. coli* in a soluble form and eluted as single monomer-sized peaks when subjected to size-exclusion chromatography (S.E.C).

317 **Comparison of the CTPR, CTPR-YD and CTPR-YD-loop constructs: loop extension**
318 **compromises the thermodynamic stability and cooperativity but not the overall native**
319 **structure.**

320 To determine whether loop insertion radically alters the secondary structure of the native state,
321 for example by unfolding repeats or decoupling sections of the CTPR array into independently
322 folding units, far-UV circular diochroism (CD) spectra were recorded and thermal/chemical
323 denaturations performed. Far-UV CD spectra show that the CTPR6-YD loop-extension
324 constructs have the same alpha-helical content as CTPR6-YD (Fig. 2a). Moreover, the CTPR6-
325 YD-loop series showed very high melting temperatures, similar to that of CTPR6-YD (Fig.
326 S2). Thus, a single loop extension of up to 25 residues does not compromise the native structure
327 of CTPR6-YD protein.

328 Next, chemical denaturation experiments were performed by monitoring both
329 tryptophan fluorescence (there is a tryptophan residue in each repeat) and CD (monitored at
330 222 nm). Initially, all curves were fitted to a two-state equation to give the midpoints of
331 unfolding ($D_{50\%}$), *m*-values and free energies of unfolding (Table 1). Figure 2b & c shows a
332 comparison of the denaturation curves of the CTPR6-YD-loop proteins with those of the CTPR
333 series and CTPR-YD series, from which a number of features and trends are apparent.

334 First, each chemical denaturation curve, whether monitored by CD or fluorescence,
335 showed a single unfolding transition. Moreover, there is good agreement between denaturation
336 curves monitored by CD and by fluorescence. This result indicates that denaturation occurs via
337 concurrent loss of native secondary and tertiary structure. Importantly, the native pre-transition
338 baselines of the CD-monitored denaturations were essentially flat. Thus, the single loop and
339 single point mutation-containing proteins do not partially unfold before the major transition.

340 Second, the chemical denaturations of the four loop variants overlay and give the same
341 $D_{50\%}$ and *m*-values when fitted to a two-state equation. Significantly, these values are lower

342 than those of the parent protein, CTPR6-YD, yet higher than "half" of it (CTPR3-YD). The
343 inserted loop, therefore, appears to cause a loss in stability and cooperativity, and this effect is
344 independent of the length of the loop. However, since the CTPR6-YD-loop variants have
345 significantly higher $D_{50\%}$ and m-values than those of CTPR3-Y91D, the repeats must be folded
346 as a CTPR6 unit rather than exist as two fully uncoupled CTPR3-YD halves. The two-state
347 fits of the data indicate an apparent loss in stability of 7.5 kcal mol^{-1} upon loop extension (Table
348 1).

349

350 Table 1. Parameters obtained by fitting the equilibrium denaturation data to a two-state model
351 for the CTPR, CTPR-YD and the CTPR6-YD-loop proteins series.

Experiment /Protein	$D_{50\%}$ (M)	m-value (kcal mol^{-1} M^{-1})	$\Delta G_{D-N}^{H_2O}$ (kcal mol^{-1})
Equilibrium denaturation monitored by Fluorescence			
CTPR2	3.53 ± 0.01	2.09 ± 0.04	-7.4 ± 0.1
CTPR3	4.30 ± 0.02	2.8 ± 0.2	-12.0 ± 0.9
CTPR4	4.80 ± 0.01	4.0 ± 0.3	-19.2 ± 0.9
CTPR6	5.30 ± 0.03	4.8 ± 0.2	-25.5 ± 1.1
CTPR2-YD	2.28 ± 0.01	2.1 ± 0.04	-4.8 ± 0.1
CTPR3-YD	3.93 ± 0.02	2.9 ± 0.2	-11.4 ± 0.8
CTPR6-YD-loop10	4.35 ± 0.01	3.4 ± 0.1	-14.9 ± 0.6
CTPR6-YD-loop15	4.32 ± 0.02	3.1 ± 0.3	-13.4 ± 1.1
CTPR6-YD-loop20	4.37 ± 0.02	3.4 ± 0.3	-14.9 ± 1.3
CTPR6-YD-loop25	4.38 ± 0.02	3.1 ± 0.2	-13.6 ± 0.9
CTPR6-YD	4.99 ± 0.03	4.5 ± 0.5	-22.5 ± 2.3
Equilibrium denaturation monitored by CD			
CTPR2	3.50 ± 0.01	2.2 ± 0.03	-7.6 ± 0.1
CTPR3	4.46 ± 0.01	3.3 ± 0.04	-10.3 ± 0.1
CTPR4	4.85 ± 0.01	4.8 ± 0.1	-23.3 ± 0.5
CTPR6	5.41 ± 0.01	4.9 ± 0.1	-26.5 ± 0.5
CTPR2-YD	2.32 ± 0.02	1.8 ± 0.1	-4.2 ± 0.1
CTPR3-YD	3.96 ± 0.01	2.31 ± 0.03	-9.6 ± 0.1
CTPR6-YD-loop10	4.19 ± 0.01	2.69 ± 0.06	-11.3 ± 0.3
CTPR6-YD-loop15	4.24 ± 0.01	2.64 ± 0.04	-11.2 ± 0.2
CTPR6-YD-loop20	4.21 ± 0.01	3.08 ± 0.04	-12.9 ± 0.2
CTPR6-YD-loop25	4.20 ± 0.01	2.88 ± 0.04	-12.1 ± 0.2
CTPR6-YD	4.97 ± 0.01	4.3 ±0.1	-21.3 ± 0.4

352 All measurements were performed in triplicate, and the errors listed are the standard errors of
353 the mean. The $\Delta G_{D-N}^{H_2O}$ for the loop-extension proteins are apparent values only, as their low m-
354 values indicate that the unfolding transitions are not fully cooperative.
355
356
357 **Un/folding kinetics of the loop-extension constructs show that loss of thermodynamic**
358 **stability is mainly through increased rates of unfolding & TPR motifs are not uncoupled.**
359 The unfolding and refolding kinetics of the proteins was measured using stopped-flow
360 fluorescence. The refolding traces for all proteins were fitted to the sum of two exponential
361 phases, the faster of which constituted ~80-95% of the overall amplitude (Fig. S3). The smaller,
362 slower phase could be the result of proline isomerization, as there is a proline residue in each
363 CTPR module (at the end of the second helix). The refolding traces at GdmHCl concentrations
364 below 2.5 M were too fast to be fitted accurately. The unfolding traces were fitted to a single
365 exponential phase (Fig. S3).

366 Both unfolding and refolding kinetics are shown in Fig. 2d as chevron plots. These
367 show that all proteins exhibit curvature in both the refolding arm and the unfolding arm.
368 Therefore, although the kinetics is more complex than a simple two-state transition, two effects
369 of loop extension are readily apparent. First, the loop-extension proteins have rate constants
370 for unfolding that lie between those of the 3-repeat and 6-repeat arrays, CTPR3-Y91D and
371 CTPR6-Y91D. Second, loop extensions have only a small effect on the refolding rates. Thus,
372 the kinetics show that the major effect of the loop extension is to destabilise the native state via
373 increased unfolding rates. Moreover, the intermediate nature of the loop constructs' chevron
374 plots corroborates the equilibrium finding that the loops do not completely uncouple the 6-
375 repeat protein into two CTPR3-YD units.
376
377 **Delineating the effects of loop extension on stability & cooperativity using 1-D**
378 **heteropolymer Ising model analysis**

379 The above two-state fitting of the equlibrium denaturation data is only of limited, qualitative
380 use, given that there is clearly evidence of deviation of the loop-extended protein from this
381 simple model. Global Ising model analysis of repeat-protein denaturation curves has been
382 shown to be an effective means of quantifying repeat-protein energetics, as it enables us to
383 dissect the contribution that individual repeat units make (inter-repeat interfacial energy and
384 intrinisic repeat stability) to the overall stability and cooperativity (11, 42–44). Here we use a
385 heteropolymer ising model, as our TPR arrays are composed of non-identical repeat motifs
386 (Fig. 1 & S.I.). We therefore globally fitted 27 denaturation curves of the following eleven
387 proteins (the majority of which were performed in triplicate) - the CTPR series (CTPR2,
388 CTPR3, CTPR4 and CTPR6), the loop series (CTPR6-YD-loop10, CTPR6-YD-loop15,
389 CTPR6-YD-loop20 and CTPR6-YD-loop25), and the mutant series (CTPR2-YD, CTPR3-YD
390 and CTPR6-YD) (Fig. 3) - thereby determining the energetics of all three types of repeat units
391 used (CTPR, CTPR-YD and CTPR-YD-loop), where the unit of repetition was defined as the
392 whole TPR motif i.e. helix-turn-helix-loop. As there was no significant length dependence of
393 the stability of the CTPR6-YD-loop series, we fitted all of them with the same energetic terms.
394 The denaturation curves were first converted to fraction unfolded (using Equation 3), as the
395 CD data showed that there was no pre-transition unfolding of the proteins. The heteropolymer
396 model was able to describe the equilibrium denaturation curves of these eleven proteins with a
397 total of nine globally-fitted parameters. These parameters are the intrinsic stability (ΔG_i), the
398 interfacial stability (ΔG_{ij}) and the m-value (m_i) for each of the three types of repeat units (CTPR,
399 CTPR-YD and CTPR-YD-loop). Fig 3 show the high quality of the fits, and Table 2
400 summarises the results.

401 The ising model confirms the two-state analysis showing that the CTPR-YD loop-
402 containing repeat is the least stable, followed by the point mutation-containing CTPR-YD
403 repeat, with the CTPR repeat being the most stable. Futhermore, the ising model analysis shows

404 that the destabilising effect of the point mutation is mostly localised to the intrinsic energy
405 term, whereas the effect of loop extension was mostly localised to the interfacial energy term
406 with little effect on the intrinsic energy term. Thus the energetic effect of the loop insertion
407 relative to that of the point mutation can be calculated as $\Delta\Delta G = \Delta G_{0\rightarrow1}^{H20}$ (CTPR variant 1) -
408 $\Delta G_{0\rightarrow1}^{H20}$ (CTPR variant 2), where $\Delta G_{0\rightarrow1}^{H20} = \Delta G_{i}^{H20} + \Delta G_{i-1,i}^{H20}$. Table 3 summarises the results and
409 shows the effect of the point mutation (3.3 ± 0.3 kcal mol^{-1}) compared with the loop (4.3 ± 0.4
410 kcal mol^{-1}). This means that the loop value is four times the energetic cost of a 10-residue loop
411 extension observed previously for globular proteins (1.1 kcal mol^{-1}) as calculated by the Ising
412 model and seven times as calculated by the two-state model (45). The difference between the
413 two may be a result of partially folded intermediate states being taken into account in the Ising
414 model.
415
416 Table 2. Values of intrinsic (ΔG_{i}) and interfacial (ΔG_{ij}) stabilities for the three different repeat
417 units analysed using the heteropolymer model. Only the intrinsic stability term has a
418 denaturant dependence (m_{i}).
419

Repeat type	[a]ΔG_{i} (kcal mol^{-1})	[a]ΔG_{ij} (kcal mol^{-1})	[a]m_{i} (kcal mol^{-1} M^{-1})	[b]$\Delta G_{0\rightarrow1}^{H20}$ (kcal mol^{-1})
CTPR	-0.59 ± 0.12	-6.08 ± 0.08	1.1 ± 0.7	-6.7 ± 0.2
CTPR-YD	2.53 ± 0.07	-5.60 ± 0.01	0.6 ± 0.3	-3.1 ± 0.2
CTPR-YD-loop	2.59 ± 0.03	-1.36 ± 0.03	0.38 ± 0.12	1.2 ± 0.1

[a]Errors of the fitted variables were determined by calculating a covariance matrix from the Jacobian matrix following a subsequent least-squares minimisation of the fit.

[b]$\Delta G_{0\rightarrow1}^{H20} = \Delta G_{i}^{H20} + \Delta G_{i-1,i}^{H20}$, i.e. the stability gained when a single repeat is added to a folded TPR ensemble. Errors in $\Delta G_{0\rightarrow1}^{H20}$ were propagated from the errors obtained from the fitted variables.

420
421
422 Table 3. Energetic costs of the YD mutation and the loop extension, calculated as the changes
423 in the free energy of unfolding ($\Delta\Delta G$) from a heteropolymer Ising model fit.
424

	[a]$\Delta\Delta G$ (kcal mol^{-1})
CTPR to CTPR-YD	+ 3.6 ± 0.3
CTPR to CTPR-YD-loop	+ 7.9 ± 0.2
CTPR to CTPR-loop	+ 4.3 ± 0.2

[a] $\Delta\Delta G = \Delta G_{0\to1}^{H20}$ (CTPR variant 1) - $\Delta G_{0\to1}^{H20}$ (CTPR variant 2), where $\Delta G_{0\to1}^{H20} = \Delta G_{i}^{H20} + \Delta G_{i-1,i}^{H20}$. Errors in $\Delta\Delta G$ were propagated from the errors in $\Delta G_{0\to1}^{H20}$.

425 **Loop extension incurs only a small energetic cost in the context of a two-repeat array**

426 The additivity rule of the Ising model allows us to predict the stability of a protein comprising
427 any combination of CTPR, CTPR-YD and CTPR-loop units. The large stability loss of loop
428 extension observed for the 6-repeat protein would be predicted to render a 2-repeat protein with
429 the YD mutant (CTPR2-YD-loop) to be mostly unfolded and a 2-repeat protein (CTPR2) to be
430 very destabilised (see predicted denaturation curve for CTPR2-YD-loop in Fig. 4b). To test
431 this prediction, we made four two-repeats proteins with and without the Y-to-D mutation and
432 with loop extensions of 10 residues and 25 residues: CTPR2-YD-loop10 and CTPR2-YD-
433 loop25, CTPR2-loop10 and CTPR2-loop25. Previous reports on CTPR proteins have
434 demonstrated how changing the amino-acid composition of the short loop between repeats has
435 a small but significant effect on the interfacial stability. Specifically, changing the sequence
436 from NN to RS results in a loss in stability of 1 kcal mol^{-1} due to differences in side-chain
437 interactions upon mutation. This effect was found to follow the additivity rule of the Ising
438 model (32). We do not know how the loop extension would affect these loop interactions, and
439 therefore, we made an additional CTPR2 variant with the DPRS sequence (CTPRa2) for
440 comparative purposes.

441
442
443 Figure 4a shows a comparison of the CD spectra of CTPR2-loop25 and CTPRa2. As CTPRs
444 are all-helical proteins, they should show a double minimum in the CD spectrum at 208 nm
445 and 222 nm. However, CTPR proteins do not have a pronounced 208 nm minimum (9, 23, 39).
446 Interestingly, the spectrum of CTPR2-loop25 did show the double minimum expected for an
447 α-helical protein. The similar 222 nm ellipticities of CTPR2-loop25 and CTPR2 indicate that
448 loop extension does not compromise the overall structure of the protein.

449 Figure 4b shows a comparison between the experimentally observed denaturation curves of all
450 the CTPR2 variant proteins (CTPR2, CTPRa2, loops and YD series) with the Ising-predicted
451 denaturation curve based on the energetic terms obtained from the CTPR6 variants, as
452 discussed above. As can be seen, all of the two-repeat proteins had the same *m*-value within
453 error, indicating that folding cooperativity is not perturbed by loop extension (Table 4). The
454 stability loss due to the DPNN-to-DPRS mutation was 1 kcal mol^{-1}, the same as the value
455 obtained from the six-repeat data (and consistent with previous measurements (32)). However,
456 the energetic cost of the loop extension in the 2-repeat protein was ~2.5-fold smaller than the
457 value of 4.3 kcal mol^{-1} obtained from the heteropolymer model for the CTPR-loop in the 6-
458 repeat protein. It is also noteworthy that this energetic cost is length-dependent, unlike the
459 length-independent effect of loop extension observed for the 6-repeat array. Fersht and
460 colleagues used the following polymer model to predict the entropic cost of a loop extension
461 in a globular protein (46):

462 **Equation 11:** $$\Delta\Delta G = -\mathrm{T}\,\Delta\Delta S_{config.} = -\mathrm{T}\left(-\frac{3}{2}\right)\mathrm{R}ln\left(\frac{n+\delta n}{n}\right)$$

463 where *n* is the loop length, and *δn* is the length of the extension. Accordingly, the entropic cost
464 should be 1.1 kcal mol^{-1} for a 10-residue loop extension and 1.75 kcal mol^{-1} for a 25-residue
465 loop extension. These values are much closer to those observed for the loop extensions in the
466 2-repeat array (Table 5). As would be expected, globally fitting the CTPR2-YD-loop proteins
467 together with the other series to the heteropolymer Ising model produced values that were not
468 thermodynamically consistent with the data, further underlining the observation that loop
469 extension in a CTPR2 array is not energetically equivalent to loop extension in a CTPR6 array.

470

471 Table 4. Fit of the equilibrium denaturation data to a two-state model for the CTPR2 proteins.
472 $\Delta\Delta G_{D-N}^{H_2O}$ values are calculated as the difference in $\Delta G_{D-N}^{H_2O}$ relative to CTPR2a.

Protein	$D_{50\%}$ (M)	*m*-value (kcal mol^{-1} M^{-1})	$\Delta G_{D-N}^{H_2O}$ (kcal mol^{-1} M^{-1})

CTPR2a	3.07 ± 0.02	2.09 ± 0.04	-6.4 ± 0.1
CTPR2n	3.53 ± 0.01	2.09 ± 0.04	-7.4 ± 0.1
CTPR2-YD	2.28 ± 0.01	2.09 ± 0.04	-4.8 ± 0.1
CTPR2-YD-loop10	1.43 ± 0.02	2.09 ± 0.04	-3.0 ± 0.1
CTPR2-YD-loop25	1.25 ± 0.02	2.09 ± 0.04	-2.6 ± 0.1
CTPR2n-loop10	2.71 ± 0.02	2.09 ± 0.04	-5.7 ± 0.1
CTPR2n-loop25	2.45 ± 0.02	2.09 ± 0.04	-5.1 ± 0.1

All measurements were performed in triplicate. [a]Errors are the standard errors of the mean. [b]Errors were propagated from the errors obtained from the fitted variables.

473 Table 5. Energetic cost of the point mutation and loop extensions in a two-repeats array of
474 CTPRs. $\Delta\Delta G_{D-N}^{H_2O}$ values are calculated as the difference in $\Delta G_{D-N}^{H_20}$ between the specified
475 proteins

	$\Delta\Delta G_{D-N}^{H_2O}$ (kcal mol^{-1})
Cost of YD mutation in CTPR2n	2.6 ± 0.1
Cost of RS loop instead of NN loop	1.0 ± 0.1
Cost of loop10 in CTPR2-YD	1.8 ± 0.1
Cost of loop25 in CTPR2-YD	2.2 ± 0.1
Cost of loop10 in CTPR2	1.7 ± 0.1
Cost of loop25 in CTPRn	2.3 ± 0.1
Theoretical entropic cost of a loop10*	1.1
Theoretical entropic cost of a loop25*	1.7

*The theoretical entropic cost of both loop lengths. [b]Errors were propagated from the errors obtained from the fitted variables.

476

477 # Discussion

478 Here we have asked whether TPR proteins can be functionalised by extending the loops
479 between repeats and how these structural alterations affect their folding. It is interesting to
480 compare TPRs to ANK-repeat proteins in this respect (7, 13, 23, 44), as the major differences
481 between them are the lengths of the helices and of the inter-repeat loops. ANK proteins have
482 shorter helices that contribute less to stability than the longer TPR ones. However, the long
483 semi-structured inter-repeat loops in ANKs contribute to high overall stability through forming

484 network of stabilising hydrogen bonds. This creates a large mismatch between intrinsic and
485 interfacial stabilities, thereby resulting in highly cooperative folding (19).

486 The mismatch of intra- and inter-repeat stabilities is smaller in the TPRs (13, 39). The
487 interfacial stability of the CTPRs is provided mainly by the hydrophobic packing between
488 alpha-helical residues in adjacent repeats with a smaller contribution from specific interactions
489 of residues in the inter-repeat loop. Disruption of the loop contacts upon mutation of the NN
490 sequence to RS decreases the overall stability of the repeat (~1 kcal mol^{-1} per loop) (32).
491 According to polymer theory a loop extension of 10 residues, should have a similar sized
492 energetic cost as the NN to RS mutation, with longer loops having greater entropic penalty
493 (45). Consequently, we would expect that a loop-extended CTPR array should still be highly
494 stable. However, what we observe is different from this prediction: a single loop extension
495 introduced into the two middle repeats of a six-repeat array causes a much larger than expected
496 and length-independent decrease in both stability and cooperativity. Strikingly, when the same
497 loop is inserted into a two-repeat array only a small and length-dependent loss of stability is
498 observed, similar to that predicted by polymer theory. Moreover, there was no significant effect
499 on the *m*-value, indicating that cooperativity of the two-repeat array is not compromised by the
500 loop extension. In contrast, loop insertion in a six-repeat array lowered the both the *m*-value
501 and $D_{50\%}$ and brought these values close to, but importantly, still larger than that of a three-
502 repeat array.

503 The folding behaviour of CTPR proteins is dependent on the number of repeats: CTPR2
504 has been described as the most two-state like, resembling a four-helix bundle (i.e. a globular
505 protein) as much as a tandem-repeat array. Increasing the number of repeats in the array results
506 in an increase in the overall stability of the protein because of the nearest-neighbour
507 cooperativity between repeats and the mismatch between intrinsic and interfacial stability. The
508 central repeats have been shown by hydrogen-deuterium exchange experiments to be the most

509 highly protected from solvent and therefore the least likely to explore unfolded conformations
510 (11, 47, 48). Moreover, the degree of protection increases with increasing number of repeats in
511 the array, the trend breaking down only when the number of repeats in the array is sufficiently
512 large for intermediates to be populated. We have shown that loop extension weakens the
513 unfolding cooperativity of the array. We would therefore expect the loop-extended repeat to be
514 much less protected from hydrogen-deuterium exchange than the consensus counterpart. TPRs
515 (and ankyrin repeat proteins) have been shown to exhibit dynamic spring-like behaviour in
516 solution, whereby a spring constant can be used to define the frequency of the protein
517 "breathing" (16, 49–52) – thus, the loss of nearest neighbour cooperativity and stability induced
518 by loop extension should manifest as an increase in dynamic properties at the loop-extended
519 interface.

520 In conclusion, our study shows that the introduction of loops into CTPR arrays is context
521 dependent and can lead to a more dynamic and a less stable CTPR protein array than expected.
522 TPR proteins function as molecular scaffolds (53–56), and long loops of 10 or more residues
523 are commonly observed (30). The break in cooperativity, the population of intermediates and
524 the dynamic and mechanical consequences of a weakened inter-repeat interface may be
525 important for their mechanism of action and/or regulation of binding partners. Importantly, we
526 have shown that large inter-repeat loop extensions can nevertheless produce very stable and
527 natively folded CTPR arrays. Although folding cooperativity is weakened, it is not completely
528 destroyed. Thus, our study demonstrates that CTPR arrays are amenable to both large and small
529 loop insertions ready to be exploited in various biotechnology and biomedical applications.

530 ## Author contributions

531 AP and LSI conceived and designed the experiments. AP carried out the experiments, AP,
532 ERM and AL performed the data analysis, and AP, LSI, ERM and AL wrote the manuscript.
533

534 ## Acknowledgements

535 LSI acknowledges the support of a Senior Fellowship from the UK Medical Research
536 Foundation. AP was supported by a BBSRC Doctoral Training Programme scholarship and an
537 Oliver Gatty Studentship. LSI and ERM acknowledge support from a Leverhulme Trust Project
538 Grant.
539

540 ## Competing financial interests

541 The authors declare no competing financial interests.

542 # References

543 1. Kajava, A. V. 2001. Review: proteins with repeated sequence structural prediction and
544 modeling. J. Struct. Biol. 134: 132–44.

545 2. Kajava, A. V. 2012. Tandem repeats in proteins: from sequence to structure. J. Struct.
546 Biol. 179: 279–88.

547 3. Allan, R.K., and T. Ratajczak. 2011. Versatile TPR domains accommodate different
548 modes of target protein recognition and function. Cell Stress Chaperones. 16: 353–67.

549 4. Andrade, M.A., C. Petosa, S.I. O'Donoghue, C.W. Müller, and P. Bork. 2001.
550 Comparison of ARM and HEAT protein repeats. J. Mol. Biol. 309: 1–18.

551 5. Andrade, M.A., C. Perez-Iratxeta, and C.P. Ponting. 2001. Protein repeats: structures,
552 functions, and evolution. J. Struct. Biol. 134: 117–31.

553 6. Tewari, R., E. Bailes, K.A. Bunting, and J.C. Coates. 2010. Armadillo-repeat protein
554 functions: questions for little creatures. Trends Cell Biol. 20: 470–81.

555 7. Aksel, T., and D. Barrick. 2014. Direct observation of parallel folding pathways
556 revealed using a symmetric repeat protein system. Biophys. J. 107: 220–32.

557 8. Binz, H.K., M.T. Stumpp, P. Forrer, P. Amstutz, and A. Plückthun. 2003. Designing
558 Repeat Proteins: Well-expressed, Soluble and Stable Proteins from Combinatorial
559 Libraries of Consensus Ankyrin Repeat Proteins. J. Mol. Biol. 332: 489–503.

560 9. Main, E.R.G., Y. Xiong, M.J. Cocco, L. D'Andrea, and L. Regan. 2003. Design of
561 stable alpha-helical arrays from an idealized TPR motif. Structure. 11: 497–508.

562 10. Parmeggiani, F., R. Pellarin, A.P. Larsen, G. Varadamsetty, M.T. Stumpp, O. Zerbe,
563 A. Caflisch, and A. Plückthun. 2008. Designed armadillo repeat proteins as general
564 peptide-binding scaffolds: consensus design and computational optimization of the
565 hydrophobic core. J. Mol. Biol. 376: 1282–304.

566 11. Wetzel, S.K., G. Settanni, M. Kenig, H.K. Binz, and A. Plückthun. 2008. Folding and
567 Unfolding Mechanism of Highly Stable Full-Consensus Ankyrin Repeat Proteins. J.
568 Mol. Biol. 376: 241–257.

569 12. Aksel, T., and D. Barrick. 2009. Biothermodynamics, Part A. Elsevier.

570 13. Kajander, T., A.L. Cortajarena, E.R.G. Main, S.G.J. Mochrie, and L. Regan. 2005. A
571 new folding paradigm for repeat proteins. J. Am. Chem. Soc. 127: 10188–90.

572 14. Tang, K.S., B.J. Guralnick, W.K. Wang, A.R. Fersht, and L.S. Itzhaki. 1999. Stability
573 and folding of the tumour suppressor protein p16. J. Mol. Biol. 285: 1869–1886.

574 15. Lowe, A.R., and L.S. Itzhaki. 2007. Rational redesign of the folding pathway of a
575 modular protein. Proc. Natl. Acad. Sci. U. S. A. 104: 2679–84.

576 16. Serquera, D., W. Lee, G. Settanni, P.E. Marszalek, E. Paci, and L.S. Itzhaki. 2010.
577 Mechanical unfolding of an ankyrin repeat protein. Biophys. J. 98: 1294–301.

578 17. Hutton, R.D., J. Wilkinson, M. Faccin, E.M. Sivertsson, A. Pelizzola, A.R. Lowe, P.
579 Bruscolini, and L.S. Itzhaki. 2015. Mapping the Topography of a Protein Energy
580 Landscape. J. Am. Chem. Soc. 137: 14610–25.

581 18. Zweifel, M.E., D.J. Leahy, F.M. Hughson, and D. Barrick. 2003. Structure and
582 stability of the ankyrin domain of the Drosophila Notch receptor. Protein Sci. 12:
583 2622–32.

584 19. Barrick, D., D.U. Ferreiro, and E.A. Komives. 2008. Folding landscapes of ankyrin

585 repeat proteins: experiments meet theory. Curr. Opin. Struct. Biol. 18: 27–34.

586 20. Croy, C.H., S. Bergqvist, T. Huxford, G. Ghosh, and E.A. Komives. 2004. Biophysical
587 characterization of the free IkappaBalpha ankyrin repeat domain in solution. Protein
588 Sci. 13: 1767–77.

589 21. Ferreiro, D.U., S.S. Cho, E.A. Komives, and P.G. Wolynes. 2005. The Energy
590 Landscape of Modular Repeat Proteins: Topology Determines Folding Mechanism in
591 the Ankyrin Family. J. Mol. Biol. 354: 679–692.

592 22. Parra, R.G., R. Espada, N. Verstraete, and D.U. Ferreiro. 2015. Structural and
593 Energetic Characterization of the Ankyrin Repeat Protein Family. PLOS Comput.
594 Biol. 11: e1004659.

595 23. Javadi, Y., and E.R.G. Main. 2009. Exploring the folding energy landscape of a series
596 of designed consensus tetratricopeptide repeat proteins. Proc. Natl. Acad. Sci. 106:
597 17383–17388.

598 24. Main, E.R.G., S.E. Jackson, and L. Regan. 2003. The folding and design of repeat
599 proteins: reaching a consensus. Curr. Opin. Struct. Biol. 13: 482–9.

600 25. Cheng, C.Y., V.A. Jarymowycz, A.L. Cortajarena, L. Regan, and M.J. Stone. 2006.
601 Repeat motions and backbone flexibility in designed proteins with different numbers
602 of identical consensus tetratricopeptide repeats. Biochemistry. 45: 12175–83.

603 26. Tamaskovic, R., M. Simon, N. Stefan, M. Schwill, and A. Plückthun. 2012. Designed
604 ankyrin repeat proteins (DARPins) from research to therapy. Methods Enzymol. 503:
605 101–34.

606 27. Cortajarena, A.L., F. Yi, and L. Regan. 2008. Designed TPR modules as novel
607 anticancer agents. ACS Chem. Biol. 3: 161–6.

608 28. Grove, T.Z., M. Hands, and L. Regan. 2010. Creating novel proteins by combining
609 design and selection. Protein Eng. Des. Sel. 23: 449–55.

610 29. Phillips, J.J., C. Millership, and E.R.G. Main. 2012. Fibrous nanostructures from the
611 self-assembly of designed repeat protein modules. Angew. Chem. Int. Ed. Engl. 51:
612 13132–5.

613 30. Bakkum, A.L., and R.B. Hill. 2017. Removal of a consensus proline is not sufficient to
614 allow tetratricopeptide repeat oligomerization. Protein Sci. .

615 31. Grove, T.Z., and L. Regan. 2012. New materials from proteins and peptides. Curr.
616 Opin. Struct. Biol. 22: 451–6.

617 32. Phillips, J.J., Y. Javadi, C. Millership, and E.R.G. Main. 2012. Modulation of the
618 multistate folding of designed TPR proteins through intrinsic and extrinsic factors.
619 Protein Sci. 21: 327–338.

620 33. Garibyan, L., and N. Avashia. 2013. Polymerase chain reaction. J. Invest. Dermatol.
621 133: e6.

622 34. Gasteiger, E., C. Hoogland, A. Gattiker, S. Duvaud, M.R. Wilkins, R.D. Appel, and A.
623 Bairoch. 2005. Protein Analysis Tools on the ExPASy Server 571 571 Protein
624 Identification and Analysis Tools on the ExPASy Server. .

625 35. Perez-Riba, A., and L.S. Itzhaki. 2017. A method for rapid high-throughput
626 biophysical analysis of proteins. Sci. Rep. 7: 9071.

627 36. Jackson, S.E., and A.R. Fersht. 1991. Folding of chymotrypsin inhibitor 2. 1. Evidence
628 for a two-state transition. Biochemistry. 30: 10428–35.

629 37. Myers, J.K., C.N. Pace, and J.M. Scholtz. 1995. Denaturant m values and heat capacity
630 changes: relation to changes in accessible surface areas of protein unfolding. Protein
631 Sci. 4: 2138–48.

632 38. Lowe, A.R., A. Perez-Riba, L.S. Itzhaki, and E. Main. 2017. PyFolding: An open-
633 source software package for graphing, analysis and simulation of thermodynamic and
634 kinetic models of protein folding. doi.org. : 191593.

635 39. Millership, C., J.J. Phillips, and E.R.G. Main. 2016. Ising Model Reprogramming of a
636 Repeat Protein's Equilibrium Unfolding Pathway. J. Mol. Biol. 428: 1804–17.

637 40. Grove, T.Z., J. Forster, G. Pimienta, E. Dufresne, and L. Regan. 2012. A modular
638 approach to the design of protein-based smart gels. Biopolymers. 97: 508–517.

639 41. Kajander, T., A.L. Cortajarena, S. Mochrie, and L. Regan. 2007. Structure and stability
640 of designed TPR protein superhelices: unusual crystal packing and implications for
641 natural TPR proteins. Acta Crystallogr. Sect. D Biol. Crystallogr. 63: 800–811.

642 42. Aksel, T., A. Majumdar, and D. Barrick. 2011. The contribution of entropy, enthalpy,
643 and hydrophobic desolvation to cooperativity in repeat-protein folding. Structure. 19:
644 349–60.

645 43. Millership, C., J.J. Phillips, and E.R.G. Main. 2016. Ising Model Reprogramming of a
646 Repeat Protein's Equilibrium Unfolding Pathway. J. Mol. Biol. 428: 1804–1817.

647 44. Kloss, E., N. Courtemanche, and D. Barrick. 2008. Repeat-protein folding: new
648 insights into origins of cooperativity, stability, and topology. Arch. Biochem. Biophys.
649 469: 83–99.

650 45. Nagi, A.D., and L. Regan. 1997. An inverse correlation between loop length and
651 stability in a four-helix-bundle protein. Fold. Des. 2: 67–75.

652 46. Ladurner, A.G., and A.R. Fersht. 1997. Glutamine, alanine or glycine repeats inserted
653 into the loop of a protein have minimal effects on stability and folding rates. J. Mol.
654 Biol. 273: 330–7.

655 47. Cortajarena, A.L., S.G.J. Mochrie, and L. Regan. 2008. Mapping the energy landscape
656 of repeat proteins using NMR-detected hydrogen exchange. J. Mol. Biol. 379: 617–26.

657 48. Main, E.R.G., K. Stott, S.E. Jackson, and L. Regan. 2005. Local and long-range
658 stability in tandemly arrayed tetratricopeptide repeats. Proc. Natl. Acad. Sci. U. S. A.
659 102: 5721–6.

660 49. Lamboy, J.A., H. Kim, K. Suk, T. Ha, and E.A. Komives. 2011. Visualization of the
661 nanospring dynamics of the I κ B α ankyrin repeat domain in real time. Proc Natl Acad
662 Sci U S A. 108: 10178–10183.

663 50. Lee, G., K. Abdi, Y. Jiang, P. Michaely, V. Bennett, and P.E. Marszalek. 2006.
664 Nanospring behaviour of ankyrin repeats. Nature. 440: 246–249.

665 51. Li, L., S. Wetzel, A. Plückthun, and J.M. Fernandez. 2006. Stepwise unfolding of
666 ankyrin repeats in a single protein revealed by atomic force microscopy. Biophys. J.
667 90: L30-2.

668 52. Cohen, S.S., I. Riven, A.L. Cortajarena, L. De Rosa, L.D. D'Andrea, L. Regan, and G.
669 Haran. 2015. Probing the Molecular Origin of Native-State Flexibility in Repeat
670 Proteins. J. Am. Chem. Soc. 137: 10367–10373.

671 53. D'Andrea, L.D., and L. Regan. 2003. TPR proteins: the versatile helix. Trends
672 Biochem. Sci. 28: 655–62.

673 54. Schmid, A.B., S. Lagleder, M.A. Gräwert, A. Röhl, F. Hagn, S.K. Wandinger, M.B.
674 Cox, O. Demmer, K. Richter, M. Groll, H. Kessler, and J. Buchner. 2012. The
675 architecture of functional modules in the Hsp90 co-chaperone Sti1/Hop. EMBO J. 31:
676 1506–17.

677 55. Zhang, Z., K. Kulkarni, S.J. Hanrahan, A.J. Thompson, and D. Barford. 2010. The
678 APC/C subunit Cdc16/Cut9 is a contiguous tetratricopeptide repeat superhelix with a
679 homo-dimer interface similar to Cdc27. EMBO J. 29: 3733–3744.

680 56. Haslbeck, V., J.M. Eckl, C.J.O. Kaiser, K. Papsdorf, M. Hessling, and K. Richter.
681 2013. Chaperone-interacting TPR proteins in Caenorhabditis elegans. J. Mol. Biol.
682 425: 2922–39.

683

684

685

686 Figure 1. Sequences, topologies & modelled structures of the 15 proteins used in this study.
687 (A) The repeating TPR motif sequences used (each repeat contains two alpha-helices): CTPRn
688 (red), CTPR-YD (grey) and CTPR-YD-loop (green).
689 (B) Topology of the CTPR series of four proteins containing only the "CTPR" motif (CTPR2,
690 CTPR3, CTPR4 and CTPR6) (9). Repeats are coloured as per panel A to show that all proteins
691 in this series contain only the CTPR sequence.
692 (C) Topology of the CTPR series containing "CTPR", "CTPR-Y91D" and "CTPR-Y91D-
693 loop" motifs (CTPR3-YD, CTPR6-YD, CTPR6-YD-loop10, CTPR6-YD-loop15, etc.).
694 Repeats are coloured as per panel A to show where the loop-containing and YD-containing
695 repeats occur.
696 (D) Topology of the CTPR2 series containing "CTPRn", "CTPR-Y91D" and "CTPR-Y91D-
697 loop" motifs (CTPRa2 and CTPR2 with either a 10-residue or a 25-residue loop between the
698 two repeats and a version of them with the Y-to-D point mutation). Repeats are coloured as per
699 panel A to show where the loop-containing and YD-containing repeats occur.
700 (E) Ribbon repesentatation of the atomic structures of CTPR2, CTPR3 and CTPR6 based on
701 the crystal structure 2HYI (41). The dots represent the fact that this series also includes CTPR4
702 (not shown). Repeats are coloured as per panel A to show that all proteins contain only the
703 CTPRn sequence.
704 (F) Ribbon repesentatation of the atomic structures of CTPR3-YD, CTPR6-YD and CTPR6-
705 YD-loop proteins based on crystal structure 2HYI (41). Repeats are coloured as per panel A to
706 show that, for example, CTPR3-YD is composed of two CTPR repeats and a C-terminal CTPR-
707 YD repeat. In the representation of the CTPR6-YD-loop proteins, the CTPR-YD-loop motif is
708 located in repeat 3 (green). The loops were inserted after the third repeat (green) and before the
709 fourth repeat (red). Sequences for all proteins are found in Table S1.
710
711 Figure 2. Biophysical analysis and comparison of the CTPR, CTPR-Y91D and CTPR-Y91D-
712 loop series of proteins. (a) Far-UV CD spectra, (b & c) averaged equilibrium denaturation
713 curves monitored by (b) CD at 222 nm and converted to Molar Ellipticity and (c) fluorescence
714 converted to fraction unfolded for ease of comparison and (d) chevron plots. The denaturation
715 curves are fitted to a two-state model. The chevron plots are fitted to a two-state model in which
716 folding and unfolding reaction proceed via a broad transition-state model. Measurements were
717 performed at 25 ºC in 50 mM sodium phosphate buffer pH 6.8, 150 mM NaCl.
718
719 Figure 3. Equilibrium denaturation curves for the CTPR, CTPR-YD and CTPR-YD-loop
720 proteins fitted globally to a 1-D heteropolymer Ising model. (a) Topologies used for each
721 protein when fit to the Heteropolymer Ising model - CTPR repeat (red), the CTPR-YD repeat
722 (black) and the CTPR-YD-loop repeat (green). The minimum unit of repetition was set as an
723 individual helix-turn-helix-loop repeat.

724 Figure 4. Effects of loop extension on the two-repeat CTPR array. (a) CD spectra of CTPRa2,
725 CTPR2 and CTPR2-loop25. (b) Equilibrium denaturation curves monitored by fluorescence
726 (converted to fraction unfolded for comparison) for all CTPR2 variants (CTPR2, CTPRa2,
727 loops and YD series) and CTPR2-YD-loop predicted according to Ising behaviour. The data
728 are fitted to a two-state model. All measurements were performed at 25 ºC in 50 mM sodium
729 phosphate buffer pH 6.8, 150 mM NaCl.
730

731

A

CTPR motif

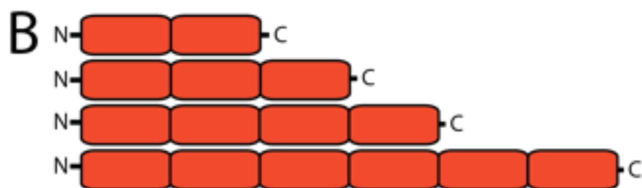
AEAWYNLGNAYYKQGDYQKAIEYYQKALELDPNN

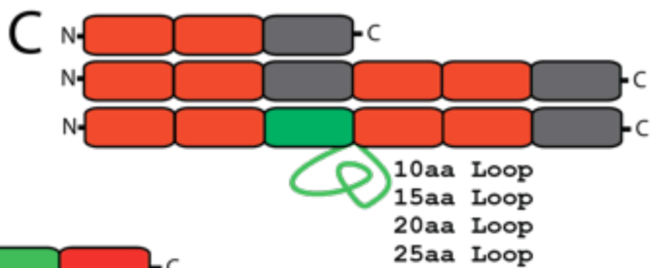
CTPR-Y91D motif

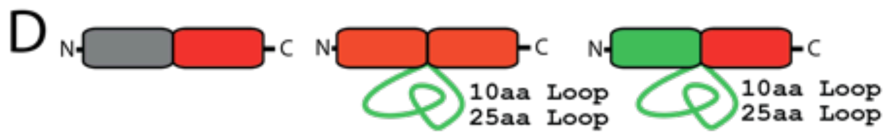
AEAWYNLGNAYYKQGDYQKAIEDYQKALELDPNN

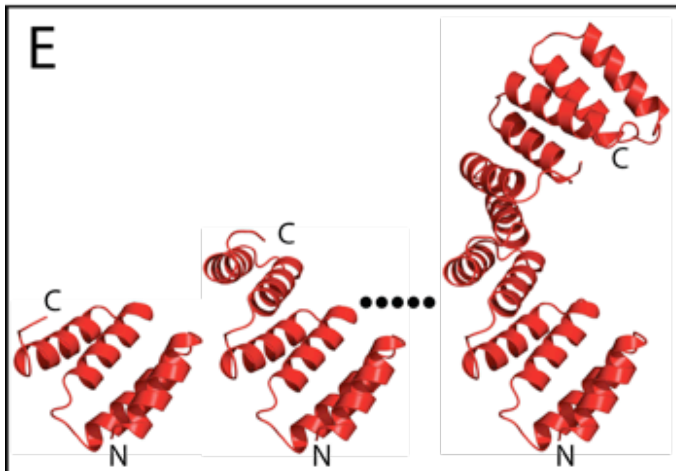
CTPR-Y91D-loop motif

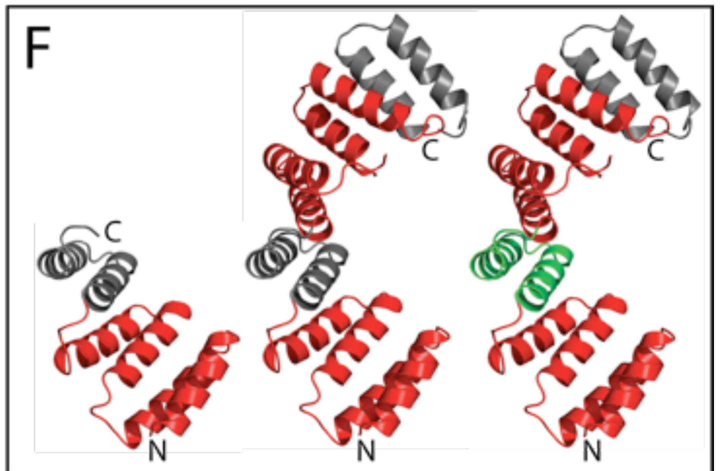
AEAWYNLGNAYYKQGDYQKAIEDYQKALELDPNN LOOPs:10-25aa

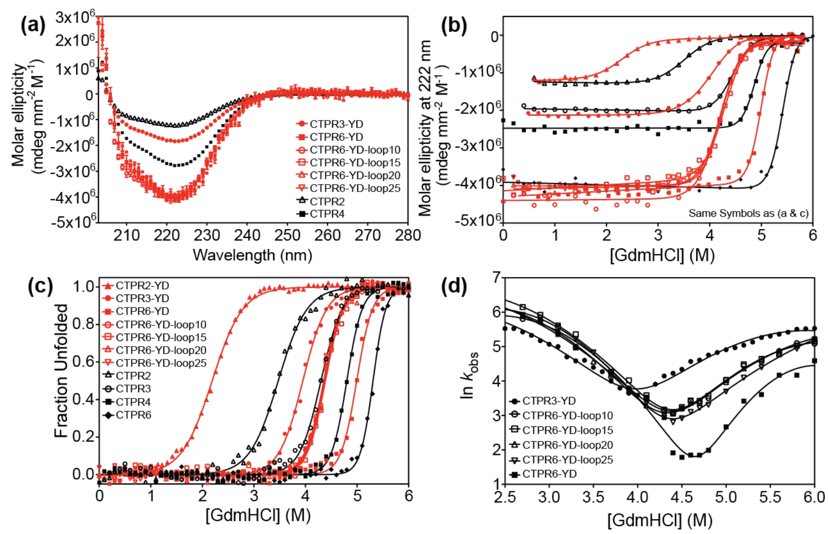




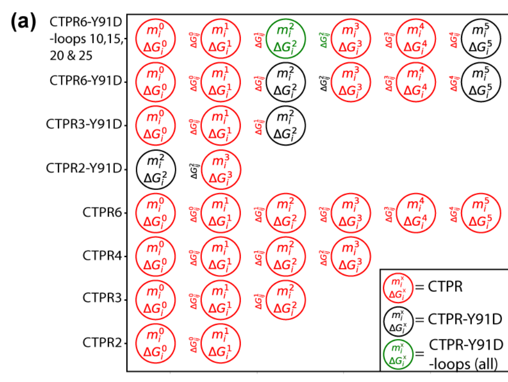




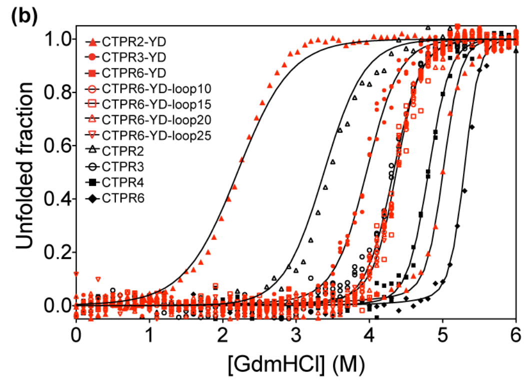




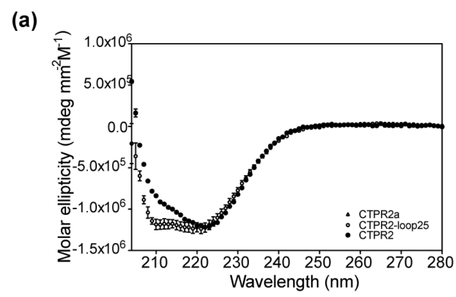
(a)
Molar ellipticity (mdeg mm^{-2} M^{-1})
Wavelength (nm)
CTPR3-YD
CTPR6-YD
CTPR6-YD-loop10
CTPR6-YD-loop15
CTPR6-YD-loop20
CTPR6-YD-loop25
CTPR2
CTPR4
(b)
Molar ellipticity at 222 nm (mdeg mm^{-2} M^{-1})
[GdmHCl] (M)
Same Symbols as (a & c)
(c)
Fraction Unfolded
[GdmHCl] (M)
CTPR2-YD
CTPR3-YD
CTPR6-YD
CTPR6-YD-loop10
CTPR6-YD-loop15
CTPR6-YD-loop20
CTPR6-YD-loop25
CTPR2
CTPR3
CTPR4
CTPR6
(d)
ln k_{obs}
[GdmHCl] (M)
CTPR3-YD
CTPR6-YD-loop10
CTPR6-YD-loop15
CTPR6-YD-loop20
CTPR6-YD-loop25
CTPR6-YD

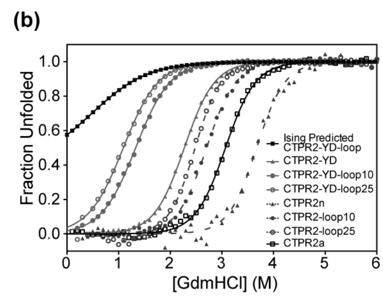


(a)
CTPR6-Y91D -loops 10,15, 20 & 25
CTPR6-Y91D
CTPR3-Y91D
CTPR2-Y91D
CTPR6
CTPR4
CTPR3
CTPR2
= CTPR
= CTPR-Y91D
= CTPR-Y91D -loops (all)



(b)
CTPR2-YD
CTPR3-YD
CTPR6-YD
CTPR6-YD-loop10
CTPR6-YD-loop15
CTPR6-YD-loop20
CTPR6-YD-loop25
CTPR2
CTPR3
CTPR4
CTPR6
Unfolded fraction
[GdmHCl] (M)


(a)
Molar ellipticity (mdeg mm²M⁻¹)
1.0x10⁶
5.0x10⁵
0.0
-5.0x10⁵
-1.0x10⁶
-1.5x10⁶
210
220
230
240
250
260
270
280
Wavelength (nm)
CTPR2a
CTPR2-loop25
CTPR2


(b)
Fraction Unfolded
1.0
0.8
0.6
0.4
0.2
0.0
0
1
2
3
4
5
6
[GdmHCl] (M)
Ising Predicted
CTPR2-YD-loop
CTPR2-YD
CTPR2-YD-loop10
CTPR2-YD-loop25
CTPR2n
CTPR2-loop10
CTPR2-loop25
CTPR2a