Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure and stability of designed TPR protein superhelices: unusual crystal packing and implications for natural TPR proteins

Tommi Kajander,[a]‡ Aitziber L. Cortajarena,[a] Simon Mochrie[b,c] and Lynne Regan[a,d]*

[a]Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, USA, [b]Department of Physics, Yale University, New Haven, USA, [c]Department of Applied Physics, Yale University, New Haven, USA, and [d]Department of Chemistry, Yale University, New Haven, USA

‡ Present address: Institute of Biotechnology, Research Program in Structural Biology and Biophysics, University of Helsinki, Finland.

Correspondence e-mail: lynne.regan@yale.edu

Received 14 March 2007
Accepted 17 May 2007

PDB References: TPR protein superhelices, 2f07, r2f07sf; 2avp, r2avpsf; 2hyz, r2hyzsf.

The structure and stability of repeat proteins has been little studied in comparison to the properties of the more familiar globular proteins. Here, the structure and stability of designed tetratricopeptide-repeat (TPR) proteins is described. The TPR is a 34-amino-acid motif which adopts a helix–turn–helix structure and occurs as tandem repeats. The design of a consensus TPR motif (CTPR) has previously been described. Here, the crystal structures and stabilities of proteins that contain eight or 20 identical tandem repeats of the CTPR motif (CTPR8 and CTPR20) are presented. Both CTPR8 and CTPR20 adopt a superhelical overall structure. The structures of the different-length CTPR proteins are compared with each other and with the structures of natural TPR domains. Also, the unusual and perhaps unique crystal-packing interactions resulting in pseudo-infinite crystalline superhelices observed in the different crystal forms of CTPR8 and CTPR20 are discussed. Finally, it is shown that the thermodynamic behavior of CTPR8 and CTPR20 can be predicted from the behavior of other TPRs in this series using an Ising model-based analysis. The designed protein series CTPR2–CTPR20 covers the natural size repertoire of TPR domains and as such is an excellent model system for natural TPR proteins.

1. Introduction

Repeat proteins play key roles in a variety of cellular processes. They are particularly prevalent in higher organisms, representing about 5% of annotated metazoan proteins in the PFAM database (Andrade *et al.*, 2001). Repeat proteins are constructed of linear strings of a small structural motif (typically 20–40 amino acids). When large numbers of the same type of repeat are present in tandem, they form elongated structures rather than the compact forms associated with the more familiar globular proteins. Several classes of repeat proteins have been shown to function by binding to other proteins or peptides, interactions that are facilitated by the extended binding surfaces that they present (D'Andrea & Regan, 2003; Kobe & Kajava, 2001).

The tetratricopeptide repeat (TPR), a 34-amino-acid motif, was first identified in yeast cell-division proteins (Sikorski *et al.*, 1990) and has since been found in a variety of proteins associated with diverse biological functions. TPR repeats are commonly found in tandem arrays, typically with 3–16 direct repeats (D'Andrea & Regan, 2003). The first structure of a three-TPR domain was solved by Barford and coworkers and revealed that the 34 amino acids of a single repeat formed a helix–turn–helix structure (Das *et al.*, 1998). The individual repeats were arranged relative to each other such that it was speculated that longer tandem arrays of TPRs, if this

arrangement were perpetuated, would form a superhelical structure.

We have previously described the design of a 'consensus TPR' (CTPR) whose sequence is derived from an alignment of individual TPR sequences. In this design, TPRs from all natural proteins are aligned, regardless of the number of TPRs in the protein or the position of a TPR within a tandem array (D'Andrea & Regan, 2003; Main *et al.*, 2003, 2005). Characterization of these CTPR proteins showed that the repeats adopt the TPR helix–turn–helix fold and that their stability increases as the number of tandem repeats increases. Moreover, we found that the relationship between the number of tandem TPR repeats and the thermodynamic behavior of the protein can be well described and predicted using a simple one-dimensional Ising model (Nelson, 2004; Zimm & Bragg, 1959), treating single helices as the folding units (Kajander *et al.*, 2005). Mello & Barrick (2004) have studied the thermodynamics of folding and unfolding of natural ankyrin-repeat domains and shown that their behavior can be described by a suitably scaled Ising model.

Here, we describe the crystal structures of consensus TPR proteins that contain eight and 20 direct tandem CTPR repeats (CTPR8 and CTPR20). We discuss the unusual and perhaps unique crystal-packing interactions and hypersymmetry that we observe in different crystal forms of these proteins. We compare the structures of CTPR proteins with different numbers of identical repeats. We also compare the structures of these designed proteins with natural TPR domains. Finally, we complement our previous studies on the length-dependence of the stability of the CTPR proteins and show that the denaturation behavior of CTPR20 can be predicted from the behavior of shorter proteins in this series using the one-dimensional Ising model.

2. Materials and methods

2.1. Molecular biology, cloning and protein purification

Proteins were designed and produced as described previously (Kajander *et al.*, 2005; Main *et al.*, 2003) and the gene encoding CTPR20 was constructed from two ten-repeat fragments by ligation (see below). Proteins were purified by affinity chromatography using either glutathione (Clonetech, Palo Alto, CA, USA) or nickel–NTA columns (Qiagen, Valencia, CA, USA) for GST or His-tag fusion proteins, respectively, and the tags were cleaved by either thrombin (Sigma-Aldrich, St Louis, MO, USA) or TEV-protease (Invitrogen, Carlsbad, CA, USA) digestion. Final protein preparations were obtained after a gel-filtration step through a HiLoad Superdex S200 column (Amersham Biosciences, Uppsala, Sweden). The gene for CTPR20 was generated using the pPROEX-HTb vector (Invitrogen, Carlsbad, CA, USA) containing the gene for CTPR10. A DNA fragment encoding ten TPR repeats was prepared by digestion with *Bam*HI and *Bgl*II restriction enzymes, which give the same sticky ends (New England Biolabs, Beverly, MA, USA). The gene encoding CTPR20 was created by ligation of the ten-repeat fragment into the pPROEX-HTb vector containing the CTPR10 gene singly digested with *Bam*HI. The orientation of the added ten-repeat fragment after ligation was verified by sequencing.

2.2. Mass-spectrometric analysis

Purified protein samples or single crystals were analysed by matrix-assisted laser desorption/ionization time-of-flight (MALDI–TOF) and nanoelectrospray ionization time-of-flight mass spectrometry (ESI–TOF) to determine the molecular weight of the material. Solutions were prepared for ESI–TOF analysis using the Millipore ZipTip system to remove salts and eluting into 50% acetonitrile, 0.2% formic acid. The mass spectra were acquired using a Micromass LCT electrospray (Micromass, Manchester, England). MALDI–TOF analysis was performed on a Voyager DE-PRO Biospectrometry workstation (AB Applied Biosystems, Foster City, CA, USA). The protein samples were mixed with nine volumes of the matrix solution (10 mg ml^{-1} sinapinic acid, 50% acetonitrile, 0.1% trifluoroacetic acid) and spotted onto the MALDI plate.

2.3. Crystallization and data collection

Purified CTPR8 protein was concentrated to 25–30 mg ml^{-1} in 20 m*M* Tris–HCl pH 7.5. Crystals were obtained readily at 295 K in several different screening conditions using the hanging-drop vapor-diffusion method. However, only those crystals initially obtained as tiny needles in 100 m*M* sodium acetate pH 4.6, 25% 2-methyl-2,4-pentanediol (MPD) and 20 m*M* $CaCl_2$ yielded well diffracting crystals that belonged to the trigonal space group $P3_121$ when conditions were refined to pH 5.0–5.5. At pH 5.0 crystals were still imperfectly ordered, but cryocooled crystals could be annealed using rapid (5 s) thawing and refreezing by briefly blocking the cold (100 K) nitrogen stream. Crystallization was attempted using solutions in which Ca^{2+} was replaced by various heavier cations in the hope that their anomalous signal could be used in phasing. For example, we obtained crystals by replacing the 20 m*M* $CaCl_2$ in the original conditions with 5–10 m*M* $CdCl_2$ or $CdSO_4$, 20–50 m*M* $BaCl_2$, 50 m*M* $SmCl_3$ or 50 m*M* $CoCl_2$. Well diffracting crystals were obtained from protein cocrystallized with either Cd^{2+} or Sm^{3+}, but the crystals obtained with $BaCl_2$ and $CoCl_2$ did not diffract. The crystals obtained with Cd^{2+} were found in both $P3_121$ and $P4_12_12$ space groups. Crystals obtained with $SmCl_3$ crystallized in space group $P2_12_12_1$. Thus, in total we obtained diffracting crystals of four different crystal forms belonging to three space groups. The CTPR8 crystals grown in the presence of Ca^{2+} or Cd^{2+} had the best diffraction quality. The Ca^{2+}-cocrystallized protein diffracted to 1.95 Å and the $P4_12_12$ crystal form with Cd^{2+} to 2.05 Å resolution. The other crystals diffracted to about 2.3 Å (Table 1).

Unexpectedly, indexing of the diffraction data revealed that the unit cell in all the crystal forms ($P4_12_12$, $P3_121$ and $P2_12_12_1$) was too small to accommodate a whole CTPR8 molecule as the asymmetric unit (Table 1). Owing to the

Table 1
Data-collection and structure-solution statistics for CTPR8 and CTPR20.

Values in parentheses are for the highest resolution bin. FL, full-length protein containing the C-terminal A_{cap}-helix.

	CTPR8					CTPR20
Crystal form	$P3_121$ (Ca^{2+})	$P3_121$ (FL)†	$P3_121$ (Cd^{2+})	$P4_12_12$ (Cd^{2+})	$P2_12_12_1$ (Sm^{3+})	$P4_12_12$ (FL, Cd^{2+})
Wavelength (Å)	0.981	1.5418	1.60	1.60	1.50	1.5418
Unit-cell parameters (Å)	a = 68.23, c = 72.70	a = 68.76, c = 72.33	a = 68.55, c = 67.23	a = 54.24, c = 71.78	a = 36.16, b = 67.70, c = 70.79	a = 54.47, c = 71.80
Resolution range (Å)	50–1.95	50–2.5	50–2.3	50–2.05	50–2.3	50–2.8
Total No. of reflections	105466	10157	53603	106297	128603	41100
No. of unique reflections	18095‡	2366	15329‡	12784‡	14975‡	6377
Redundancy	5.8‡	—	3.5‡	8.3‡	8.6‡	6.4
$I/\sigma(I)$	31.2 (3.2)	2366	32.5 (2.3)	60.7 (7.7)	17.15 (4.9)	22.83 (3.65)
Completeness (%)	99.8 (99.9)	33.3 (26.6)	95.0 (61.1)	97.8 (94.1)	99.9 (99.5)	100.0 (100.0)
R_{merge}§ (%)	3.4 (35.5)	6.2 (30.5)	4.1 (32.5)	3.7 (23.7)	10.1 (20.8)	12.4 (45.91)
χ^2	0.900 (0.815)		1.127 (0.985)	1.136 (0.958)	1.61 (0.93)	1.121 (1.053)
FOM¶	—	—	0.32	0.38	0.34	—
FOM after solvent flattening	—	—	0.55	0.60	0.75	—

† Full data set was not collected. ‡ Friedel pairs measured as separate reflections. § $R_{merge} = \sum|I - \langle I\rangle|/\sum\langle I\rangle$, where I is the observed intensity and $\langle I\rangle$ is the average intensity. ¶ FOM is the figure of merit as reported by *SOLVE* and *RESOLVE* (Terwilliger & Berendzen, 1999).

Table 2
Mass spectrometry of CTPR samples.

Measured molecular weights are given in Da. The theoretical molecular weight for CTPR8* is that assumed for a construct proteolytically cleaved at DPR|S after the last repeat (consistent with hypothesis, this site mimics a thrombin cleavage site). FL, assumed full-length intact protein with capping helix. ND, not determined. Molecular-weight accuracy for MALDI–TOF is assumed to be around 100 Da.

	MALDI–TOF MW	ESI–TOF MW	Theoretical MW
CTPR8*	32700	32309 ± 9	32217.7
CTPR8 FL	34300†	ND	34230.8
CTPR20 FL	82700	82609 ± 13	82594.8

† No other major peak was observed apart from those at half mass/charge and quarter mass/charge ratio.

smaller than expected unit-cell size, we first suspected that degradation of the protein had occurred or that an incorrect space group had been assigned. It subsequently became evident from the Patterson self-rotation function and from the merging and scaling of data, phase calculation and subsequent refinement that the space groups had been correctly assigned. However, SDS–PAGE and mass-spectroscopic analysis of the crystals using MALDI–TOF and nanospray ESI–TOF revealed that the main component (>95% based on SDS–PAGE) was an eight-repeat fragment missing the last 15-residue capping helix (Table 2).

A closer inspection of the sequence revealed that the design of the longer (>3 repeats) consensus TPR proteins (Kajander *et al.*, 2005), which introduced two mutations at the last two positions of the consensus repeat (both in a turn), changing the sequence from –DPNN to –DPRS (Fig. 1), resulted in a sequence that resembles the protease cleavage site for thrombin (which minimally recognizes a Pro-Arg sequence in an exposed position, cutting after the Arg residue). Because we use thrombin to cleave the His tag in this construct, we hypothesized that we were also obtaining cleavage after the last TPR repeat before the capping helix; presumably, this was more accessible to thrombin than the other potential cleavage sites after each TPR repeat. We therefore recloned the CTPR constructs into a different vector (pPRO-EXHTb) which contains a His tag followed by a TEV cleavage site and no thrombin cleavage site. CTPR8 and CTPR20 have been successfully produced as intact proteins in this vector and crystallized and their size has been verified by mass spectroscopy (Table 2) and SDS–PAGE (data not shown).

CTPR20 crystallized under similar conditions to CTPR8, but at lower protein and precipitant concentrations and over a slightly wider pH range. CTPR20 also cocrystallized with both Ca^{2+} and Cd^{2+}. A trigonal ($P3_121$) crystal form was initially obtained with $CaCl_2$ using the same conditions as for CTPR8 and with 10 mg ml^{-1} protein; tetragonal ($P4_12_12$) crystals of CTPR20 with $CdCl_2$ were obtained with 100 m*M* sodium acetate pH 4.8, 10% MPD, 2.5 m*M* $CdCl_2$ at a protein concentration of 5 mg ml^{-1} and trigonal ($P3_121$) crystals were obtained with 100 m*M* sodium acetate pH 5.7, 2.5 m*M* $CdCl_2$, 2.5% MPD and 10 mg ml^{-1} protein. Diffraction data were collected in-house with a Rigaku Micromax-007 generator (Rigaku, The Woodlands, TX, USA) from cryocooled crystals at 100 K.

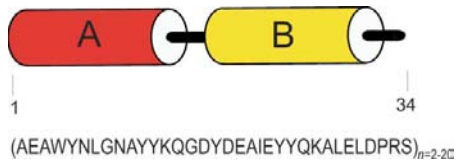

Figure 1
The CTPR motif. The designed consensus sequence is shown aligned with a schematic of the two helices of the TPR repeat. The structure of the CTPR proteins is always $(CTPR)_n$–capping helix. In CTPR8, for example, n = 8 and there are eight direct repeats of the CTPR sequence followed by AEA*KG*NLGNA*KQ*KQG, which is a modified version of the A-helix sequence designed to terminate the protein with a hydrophilic surface. Positions at which the sequence of the capping helix differs from that of an A-helix are in italics (Main *et al.*, 2003).

2.4. Structure solution and refinement

The structure of CTPR8 was solved by single-wavelength anomalous diffraction (SAD) in the three crystal forms described (see above) using the Cd^{2+} or Sm^{3+} anomalous signal with *SOLVE* and *RESOLVE* (Terwilliger & Berendzen, 1999), the output from which was used for further model building. The final 2.05 Å resolution tetragonal CTPR8 structure ($P4_12_12$) was first refined by rigid-body refinement in *CNS* (Brünger *et al.*, 1998) and then with *REFMAC* (Collaborative Computational Project, Number 4, 1994) with TLS refinement at 2.05 Å resolution (Table 1). The values of R and R_{free} were 20.5% and 24.1%, respectively. The trigonal $P3_121$ structure was refined with *CNS* at 2.3 Å resolution. In both cases an intermediate refinement step was performed using *ARP/wARP* (Perrakis *et al.*, 1999) and crystallographic waters were built in at this point. The R factors for this structure were higher (R = 26.6% and R_{free} = 30.4%), probably partly owing to the high Wilson B factor of the data (56 Å^2). The orthorhombic $P2_12_12_1$ $SmCl_3$-cocrystallized CTPR8 structure was first refined using *ARP/wARP*. Final refinement was performed using *REFMAC* to 2.3 Å resolution and the crystallographic waters built by *ARP/wARP* were edited using *Coot* (Emsley & Cowtan, 2004). The final R factors were R = 21.3% and R_{free} = 27.0%. All structures have excellent geometry and stereochemical properties (Table 3).

Table 3
CTPR structure-refinement statistics for various crystal forms.

	CTPR8			CTPR20
Crystal form	$P3_121$ (Cd^{2+})	$P4_12_12$ (Cd^{2+})	$P2_12_12_1$ (Sm^{3+})	$P4_12_12$ (Cd^{2+})†
Resolution range (Å)	50–2.3	50–2.05	50–2.3	50–2.8
R_{work}/R_{free}‡ (%)	26.6/30.4	20.1/24.1	21.3/27.0	29.0/31.2
No. of atoms	1055	578.5	1089.5	554
Protein atoms	1045	554	1057	554
Water molecules	8	23	31	—
Metal ions	2	1.5	1.5	—
B factor (Å^2) from Wilson plot	56.5	34.3	24.6	56.9
Average B factor from model (Å^2)	60.34	35.1	31.5	60.5
R.m.s.d. bond angles (°)	1.300	2.026	1.259	1.563
R.m.s.d. bond lengths (Å)	0.008	0.0027	0.016	0.011
Ramachandran plot§ (most favoured) (%)	91.8	93.3	93.3	95.0

† Identical to the CTPR8 $P4_12_12$ structure. ‡ $R = \sum |F_{obs} - F_{calc}| / \sum |F_{obs}|$, where F_{obs} are the observed structure factors and F_{calc} are the calculated structure factors. R_{work} is based on 90–95% of the data used in refinement and R_{free} is based on 10–5% of the data withheld for a cross-validation test. § No residues in the Ramachandran plots were in disallowed regions (the remainder were calculated to reside in additionally allowed regions).

The CTPR20 structure was solved directly by rigid-body refinement using the CTPR8 structure; surprisingly, the crystal forms of CTPR20 were almost identical to those of CTPR8 (Table 1) with the same unit-cell sizes, indicating the same hypersymmetric arrangement as for CTPR8. Indeed, the CTPR20 structure was shown to be identical to that of the CTPR8 $P4_12_12$ crystal form. The structure was refined to 2.8 Å resolution with R and R_{free} values of 29.0% and 31.2%, respectively, which are indicative of legitimate solution, and no significant conformational changes were observed in comparison to CTPR8 (see §3) based on difference Fourier OMIT maps and superpositioning after refinement (data not shown).

2.5. Small-angle X-ray scattering

Small-angle X-ray scattering (SAXS) measurements on CTPR8 were performed at concentrations of 3, 12, 25 and 60 mg ml^{-1} at NSLS beamline X21. At small wavevectors, the azimuthally averaged SAXS intensity of the highest concentration sample (data not shown) showed a weak peak at nonzero wavevector characteristic of weak interparticle correlations. In contrast, at the lowest two concentrations the azimuthally averaged SAXS intensity at small wavevectors was the same shape in both cases and decreased monotonically, consistent with the scattering from individual isolated scatterers. At larger wavevectors, the shape of the scattering is the same within error for all samples. Therefore, in order to present an experimental CTPR8 SAXS profile with as large a signal-to-noise rate as possible (Fig. 2), we scaled the scattering from the 60 mg ml^{-1} sample to match that of the 3 and 12 mg ml^{-1} samples at large and intermediate wavevectors. Fig. 2 shows a plot of the SAXS intensity from the 3 mg ml^{-1} sample for wavevectors less than 0.2 Å^{-1} and the scaled SAXS intensity from the 60 mg ml^{-1} sample for wavevectors greater than 0.2 Å^{-1}.

2.6. Circular-dichroism (CD) unfolding measurements

All CD experiments were performed using an Aviv CD spectrophotometer Model 215 (AVIV Instruments Inc.). Chemical denaturation, induced by guanidium hydrochloride (GuHCl), was performed using an automatic titrator (Microlab 500 series), monitoring the ellipticity at 222 nm. The equilibrium denaturation studies were performed by preparing two stock solutions at 3 µ*M* protein concentration in 50 m*M* phosphate pH 6.5, 150 m*M* NaCl and 0 or 7 *M* GuHCl for each protein. At each titration point the concentration of GuHCl was increased by 0.1 *M* by injecting denatured protein stock solution (7 *M* GuHCl), while the protein concentration was kept constant during the course of the experiment and the ellipticity at 222 nm was monitored after 45 min equilibration time. Measurements were performed in a 1 cm path-length cuvette with stirring at 298 K.

2.7. Application of the one-dimensional Ising model

The one-dimensional Ising model was originally derived to describe the behavior of a linear chain of interacting spins (s_i), which can take the value +1 or −1, in an external magnetic field (H). A similar model was subsequently applied to describe the polypeptide helix–coil transition (Zimm & Bragg, 1959; Nelson, 2004). In this case, the model describes the behavior of linear polymers composed of peptide units, each

of which is either in a helix or coil state, corresponding to spin +1 or −1, respectively. In our case, we have applied the Ising model to describe the stability of CTPRs. However, we treat entire helices in the CTPRs as individual spins, with +1 and −1 spins corresponding to folded and unfolded states, respectively (Kajander *et al.*, 2005).

All of the data within a given CTPR family is described in the model by three parameters, m_1, x_c and J, as described by Kajander *et al.* (2005). In brief, the free energy for folding according to the one-dimensional Ising model can be written as

$$\Delta G_{\rm fold} = 2RT \sum_{1}^{N} (-Js_i s_{i+1} - Hs_i),$$

where N is the number of helices, H is related to, for example, denaturant concentration (x) at a particular point on the unfolding curve by $H = (x_c - x)m_1$, with x_c corresponding to the transition midpoint in the limit of infinitely long CTPRs ($N \rightarrow \infty$), R is the thermodynamic gas constant and J specifies the coupling between helices. J, m_1 and x_c are obtained from fits to the Ising model form for the fraction folded (f). Within the model, all CTPRs in a series have the same J, m_1 and x_c and their unfolding behavior is distinguished only by N.

For the data shown in Fig. 3, $J = 1.83$, $x_c = 3.62$ and $m_1 = 1.29$. Experimentally, f is obtained, for example, from the ellipticity change measured as a function of denaturant as described previously (Kajander *et al.*, 2005). Within the model, the ΔG of folding for the whole repeat protein with N helices is given by the difference between completely folded and completely unfolded states, which is given by

$$\Delta G_{\rm fold} = RT(-2HN + 4J).$$

$4J$ is the penalty for terminal helices (*i.e.* in the model, fictitious −1 spins are added to the ends; Nelson, 2004). The values for H and J for our series of CTPRs obtained earlier were $J = 1.90$, $x_c = 3.82$ and $m_1 = 0.96$.

3. Results and discussion

3.1. The structure of CTPR8 and CTPR20

3.1.1. Crystal forms. We have solved structures of CTPR8 in three different crystal forms (in space groups $P4_12_12$, $P3_121$ and $P2_12_12_1$) and the structure of CTPR20 in one crystal form (in space group $P4_12_12$). Remarkably, the crystal forms and structures of CTPR8 and CTPR20 are very similar (Table 1). CTPR8 and CTPR20 crystallize as superhelices which stack head-to-tail with eightfold rotational symmetry (with a screw axis), with the eightfold symmetry generated by the eight TPR repeats. Both proteins crystallize with only two or four TPR repeats per asymmetric unit (Fig. 4). Thus, because there is only a half or a quarter of the whole molecule in the asymmetric unit for CTPR8 and only one tenth of the whole molecule for CTPR20 (it crystallizes in the $P4_12_12$ crystal form with the same unit-cell size as CTPR8), the structures of the full-length individual molecules are reconstructed by applying crystal symmetry and unit-cell translations. In the following sections, therefore, when we speak of the structures we are always referring to the structures reconstructed from the fragment present in the asymmetric unit of the particular crystal form.

3.1.2. Overall structures of the proteins. Both CTPR8 and CTPR20 form ordered superhelical structures with regular geometry and exactly eight repeats forming one superhelical turn (Figs. 4 and 5). Thus, CTPR8 comprises one superhelical

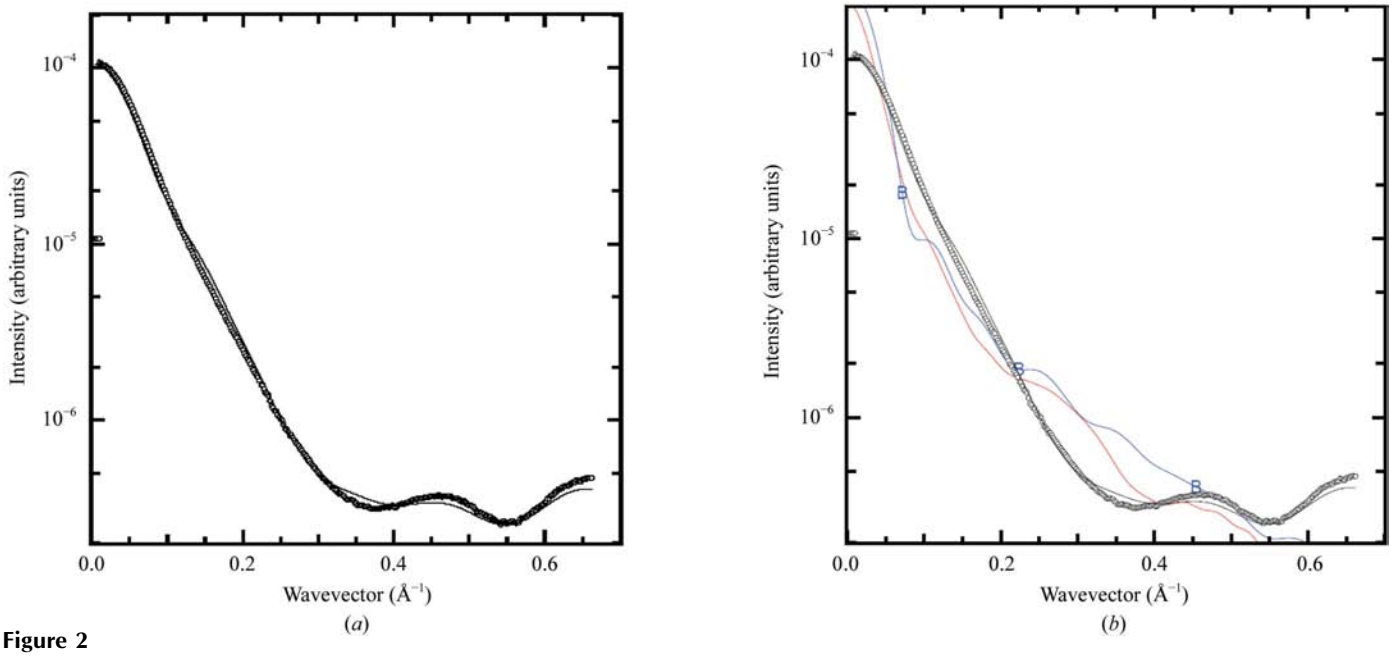

Figure 2
SAXS data. (*a*) The experimental SAXS profile (symbols) overlaid with the calculated profile (black line) for monomeric CTPR8. (*b*) The experimental and calculated data shown in (*a*) overlaid with calculated SAXS profiles for models of side-by-side (blue) and head-to-tail dimers (red) of CTPR8.

turn with overall molecular dimensions of approximately 80 × 38 Å (Fig. 4) and CTPR20 consists of 2.5 superhelical turns with dimensions of approximately 200 × 38 Å (Fig. 4). The exact pitch and superhelical characteristics of both proteins are defined by the particular crystal symmetry and unit-cell size. In all three crystal forms analyzed, the superhelical axis of the TPR 'fiber' runs along one or more crystallographic axis and there is one superhelical turn per superhelix in the unit cell. Thus, the unit-cell axis that the superhelical axis coincides with also corresponds to the exact value of the superhelical pitch. In crystals with space groups $P4_12_12$ and $P2_12_12_1$ this is the length of the c axis and in space group $P3_121$ this is the length of a and b (which are equal by symmetry; Table 1 and Fig. 5). The exact eightfold symmetry dictates that the superhelical twist per repeat for all the structures is 45°.

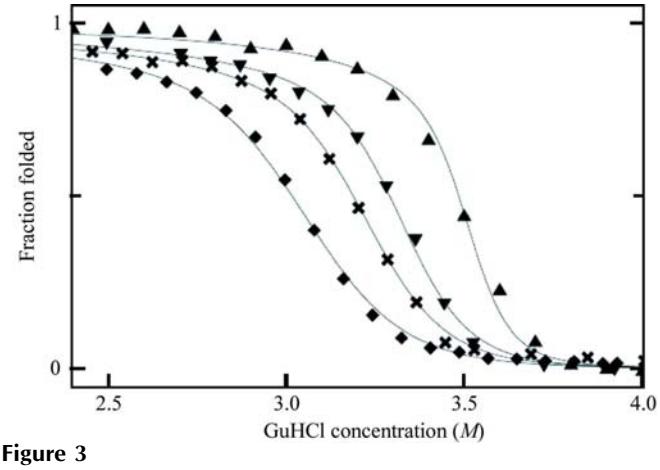

Figure 3
The CD unfolding of the CTPR6, CTPR8, CTPR10 and CTPR20 proteins as a function of GuHCl (dots) and fit to the Ising model (lines). The one-dimensional Ising-model parameters for the fit are $J = 1.83$, $x_c = 3.62$ and $m_1 = 1.29$.

Each of the individual TPR repeats in CTPR8 and CTPR20 adopts the typical helix–turn–helix structure of this motif (Blatch & Lassle, 1999; Das *et al.*, 1998; Main *et al.*, 2003). In addition, alignment of individual repeats from the different CTPR8 and CTPR20 structures shows that their main-chain atoms are superimposable, as are those of the side chains of the key hydrophobic consensus residues (Trp4, Leu7, Gly8, Tyr11, Gly15, Ala20, Tyr24, Ala27 and Pro32; Figs. 1 and 6). Significant differences between repeats are only observed in the side-chain conformations of the nonconserved solvent-exposed polar residues on the B-helix of the TPR repeat (Fig. 6). In fact, the structure of a single TPR repeat is highly conserved in all the different CTPR structures: CTPR2, CTPR3, CTPR8 and CTPR20 (see below; Main *et al.*, 2003).

The overall backbone conformational variation between the different crystal forms of CTPR8 ranges from an r.m.s.d. of

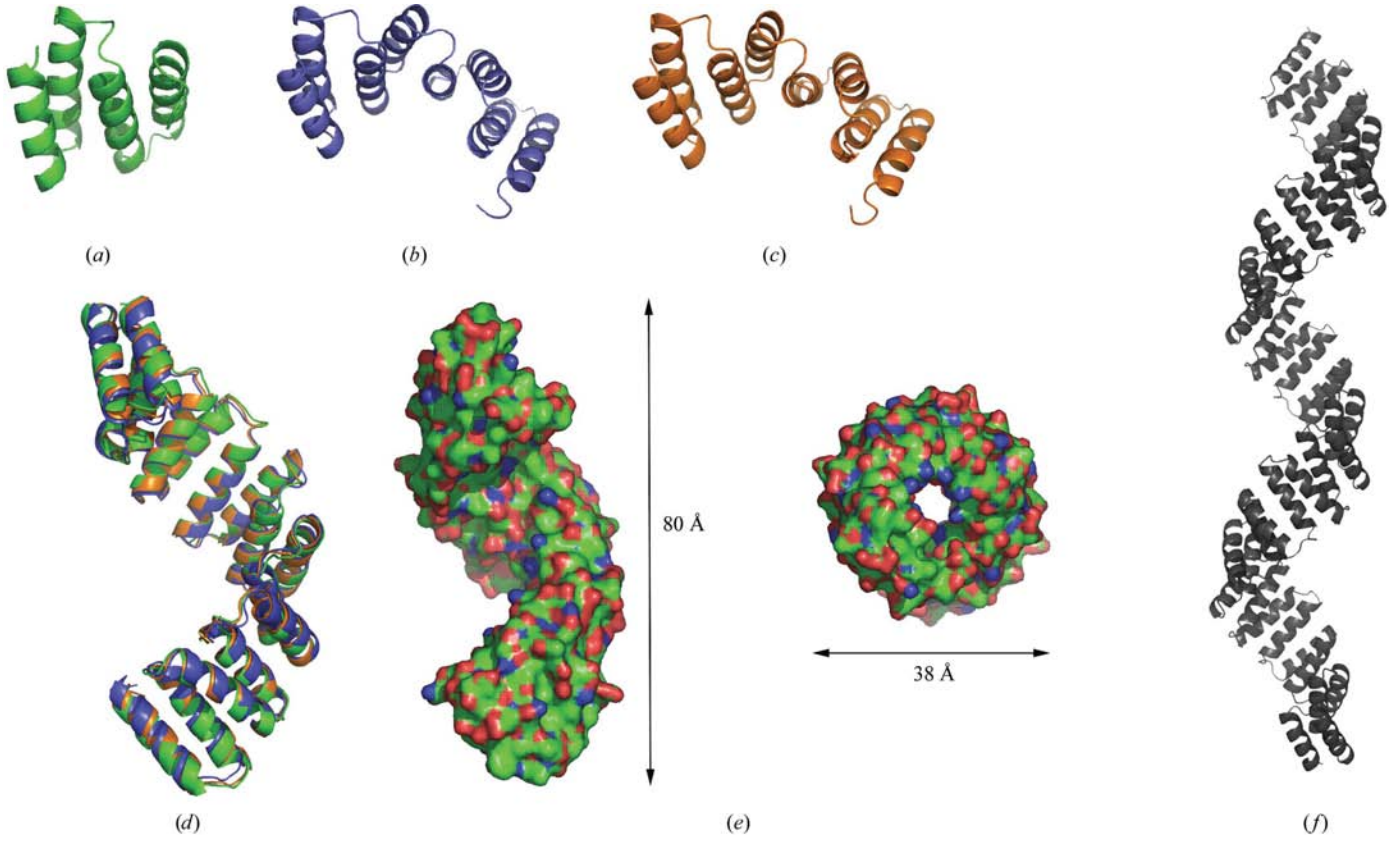

Figure 4
CTPR crystal structures. The asymmetric units of the different crystal forms of CTPR8 are shown in ribbon representation. (*a*) $P4_12_12$, (*b*) $P3_121$, (*c*) $P2_12_12_1$. (*d*) Alignment of the CTPR8 eight-repeat fragment from different crystal forms, colored as in (*a*)–(*c*). (*e*) A molecular-surface representation of CTPR8 in two different views, with the dimensions shown (C atoms green, N atoms blue and O atoms red). (*f*) Ribbon representation of the structure of CTPR20.

1.102 Å for the $P4_12_12$ and $P3_121$ structures to 0.643 Å for the $P3_121$ and $P2_12_12_1$ and 0.563 Å for the $P2_12_12_1$ and $P4_12_12$ structures (all for four repeats and alignment of 136 C^{α} atoms). This variation appears to derive from differences in the superhelical conformation and from typical variation between different crystal structures of the same protein: the C^{α}-atom r.m.s.d. of 25 different crystal forms of wild-type and closely related variants of lysozyme has been reported to be around 0.2–0.5 Å (Zhang *et al.*, 1995), while a PDB-wide pairwise study of proteins whose structures were solved in multiple crystal forms found an r.m.s.d. between pairs of 0.5–0.6 Å for organized secondary structure (Eyal *et al.*, 2005).

The superhelical pitch varies between about 68 and 72 Å in the various crystal forms. The diameter of the superhelix, perpendicular to its axis, varies with the pitch. In $P4_12_12$ and $P2_12_12_1$, where the pitches are 71.8 and 70.8 Å, respectively, the diameters are also similar, at 38.35 and 38.20 Å, respectively, whereas for the $P3_121$ structure the pitch is 68.6 Å and the diameter is 42.1 Å. Thus, the superhelix is slightly stretched from the $P3_121$ conformation in the other crystal forms. Superposition of individual repeats (*i.e.* the A–B helix pairs; Fig. 1) and the inter-repeat helix pairs (B–A, where A is the A-helix from the next repeat) gives C^{α}-atom pair r.m.s.d.s of between 0.16 and 0.83 Å; a slightly higher variation is observed for interaction between repeats [aligned individual repeats have r.m.s.d.s from 0.16 to 0.45 Å and inter-repeat helix pairs (B–A′) from 0.14 to 0.83 Å]. However, conformational variation appears to be as large for inter-repeat interactions in one structure as between structures (as judged by variation of helical packing angles and the pairwise r.m.s.d. values above). Thus, it appears that the overall differences in the superhelical pitch are a combination of small variations in packing distributed over several helix interactions and interactions are the same overall in different crystal forms.

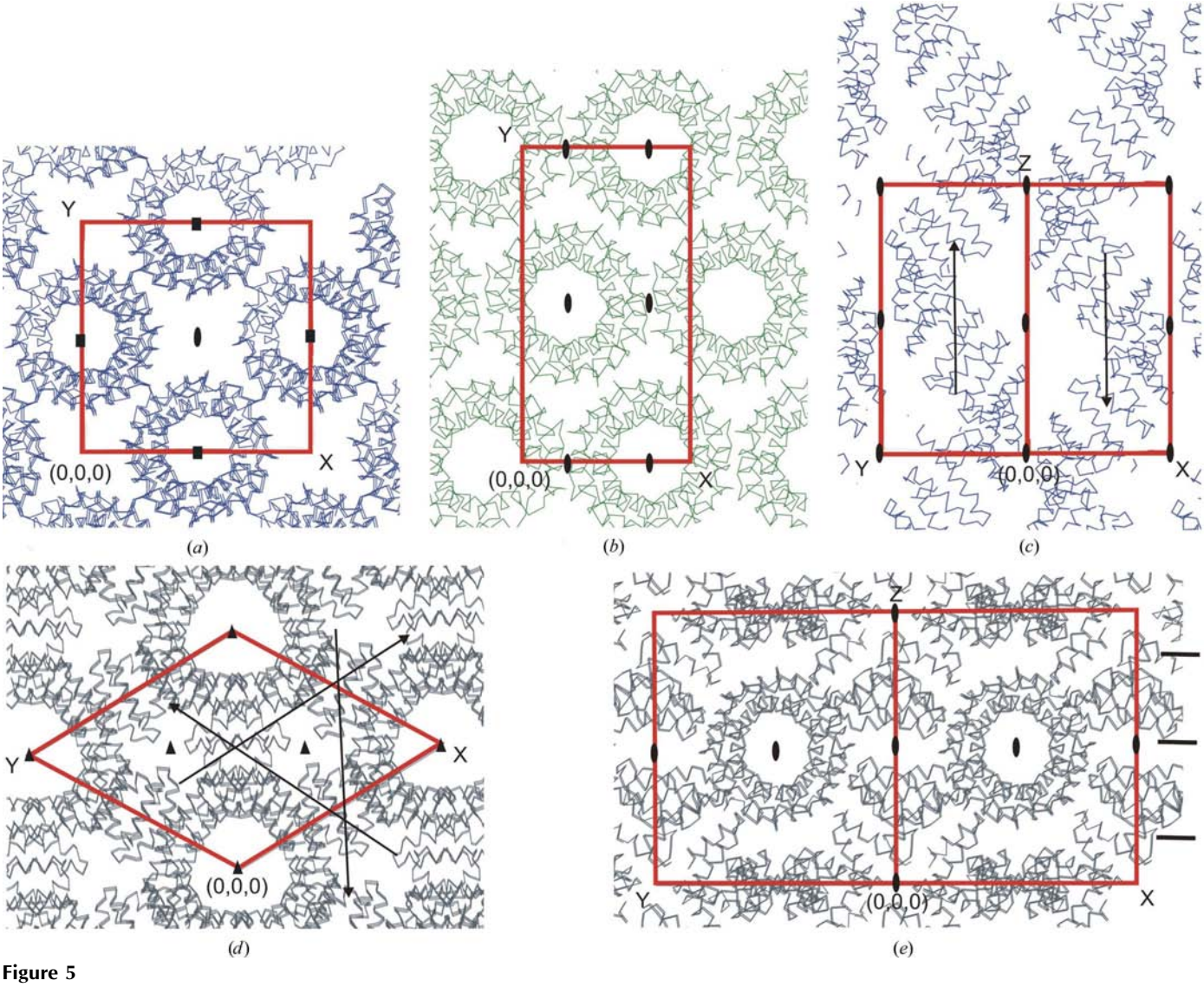

Figure 5
Crystal packing. Different views of the crystal lattice in $P4_12_12$ (*a*, *c*), $P2_12_12_1$ (*b*) and $P3_121$ (*d*, *e*) space groups. The crystal axes (*x*, *y*, *z*) and positions of different twofold (black ellipses), threefold (black triangles) and fourfold (black squares) symmetry operators are indicated. The red boxes represent the unit cells in each crystal form. Arrows in (*c*) and (*d*) indicate the long axis of the crystalline superhelices. The thick bands on the right in (*e*) indicate the positions of layers of crystalline superhelices.

The conformation of the longer CTPR8 and CTPR20 superhelices is very similar to the smaller CTPR3 structure. Alignment of the CTPR3 structure (Main *et al.*, 2003) with fragments of the superhelical CTPR structures gives r.m.s.d. values varying from 0.848 to 1.250 Å against the 3.5 repeats of CTPR3. Thus, it appears that inter-helical interactions are conserved as the chain length increases from the small TPR-repeat domains to elongated superhelical structures.

3.2. Crystal packing and data analysis

3.2.1. Head-to-tail packing and hypersymmetry. In all crystal forms, the individual CTPR molecules stack head to tail, forming an apparently continuous pseudo-infinite crystalline helical fiber which can associate with a variety of symmetry arrangements, including tetragonal, trigonal and orthorhombic lattices (Fig. 5). CTPR8 and CTPR20 crystallize in the same tetragonal and trigonal crystal forms and make the same crystal contacts, with metal ions mediating contacts between adjacent superhelical fibers (Figs. 4 and 5, Table 1).

The orthorhombic ($P2_12_12_1$) crystal form has very similar crystal packing to the tetragonal ($P4_12_12$) crystal form, with superhelical fibers running next to each other in an antiparallel arrangement (Fig. 5). The significant differences are that instead of superhelical fourfold symmetry in $P4_12_12$, the crystalline intrasuperhelical symmetry breaks into a twofold symmetry in $P2_12_12_1$; as a result, while the *a* and *b* axes in $P4_12_12$ are necessarily equal, in $P2_12_12_1$ the equivalent axis lengths are different ($a = b = 54.24$ Å in $P4_12_12$ and $a = 36.16$, $b = 67.70$ Å in $P2_12_12_1$; Table 1). The $P2_12_12_1$ *c* axis is equivalent to the tetragonal *c* axis (*i.e.* parallel to the superhelical axis; Fig. 5) and these are approximately the same length, this being the helical pitch of the eight-repeat unit.

Secondly, in $P4_12_12$ the superhelices are equally spaced in all directions perpendicular to the fourfold axis along the *c* axis, while in the $P2_12_12_1$ cell the antiparallel superhelices pack closer to each other in one direction (along the *a* axis) and are as a result 'squeezed' further apart in distance along the *b* axis (Fig. 5).

The asymmetric unit in the trigonal ($P3_121$) and orthorhombic ($P2_12_12_1$) crystal forms is comprised of four CTPR repeats (Fig. 4), whereas in the tetragonal ($P4_12_12$) crystal form it is comprised of only two repeats (Fig. 4). Because all the repeats are identical and the molecules stack upon each other in the crystal, forming a pseudo-infinite superhelical structure, the smallest crystallographic asymmetric unit does not need to cover the whole molecule (indeed, in the case of CTPR20 this would be impossible as the whole molecule would not fit in the unit cell even in the lowest symmetry space group *P*1). Consequently, we could call this structure 'hypersymmetric', with static disorder at the repeat level (Fig. 7), which in this case means that the ends of the molecules cannot be located in electron density. This is because of register disorder between adjacent crystalline superhelices, which results in the break between the ends of two stacked molecules in one crystalline superhelix (Fig. 7) being located between each repeat in the asymmetric unit, with 1/8 occupancy for CTPR8, and thus density within a superhelix is continuous throughout the unit cell. To our knowledge, this is the first example for a protein of there being less than one complete molecule in the asymmetric unit.

Similar head-to-tail stacking of helices has been observed for DNA and RNA double-stranded oligonucleotides (see, for example, Brennen *et al.*, 1986; Shah & Brünger, 1999). In these examples, the dominant crystal contacts between adjacent double helices are made by the external sugar-phosphate backbone. When such helices stack end to end, but the sequence is out of phase with the length of the unit cell, heterogeneity in the composition of the asymmetric unit can result (Brennen *et al.*, 1986; Shah & Brunger, 1999; Klosterman *et al.*, 1999; Mueller *et al.*, 1999). In the worst case, complete static disorder for each base-pair position will be observed. Shah & Brunger (1999) suggested that this could have similar effects on intensities as translational parallel noncrystallographic symmetry (Rogers & Wilson, 1953) and hypothesized that the intensity statistics might deviate from normal in a predictable manner. For their RNA, with fourfold static disorder, the values of the $\langle I\rangle^2/\langle I^2\rangle$ and $\langle F\rangle^2/\langle F^2\rangle$ ratios were found to deviate in the fashion predicted for static disorder based on results from calculated structure-factor amplitudes (four different conformations averaged in the asymmetric unit), with a value of 4.1 for the $\langle I\rangle^2/\langle I^2\rangle$ ratio.

In contrast, for CTPR8 and CTPR20 no disorder in electron density is apparent because each CTPR repeat is identical in sequence and static disorder in the unit cell only occurs at the level of a whole repeat. Thus, the structure appears to be appropriately described by application of crystallographic space-group symmetry operators and unit-cell translations to the fragment present in the asymmetric unit, as one molecule (polypeptide chain) extends beyond an asymmetric unit in the crystal. Analysis of the intensity statistics for the crystals of

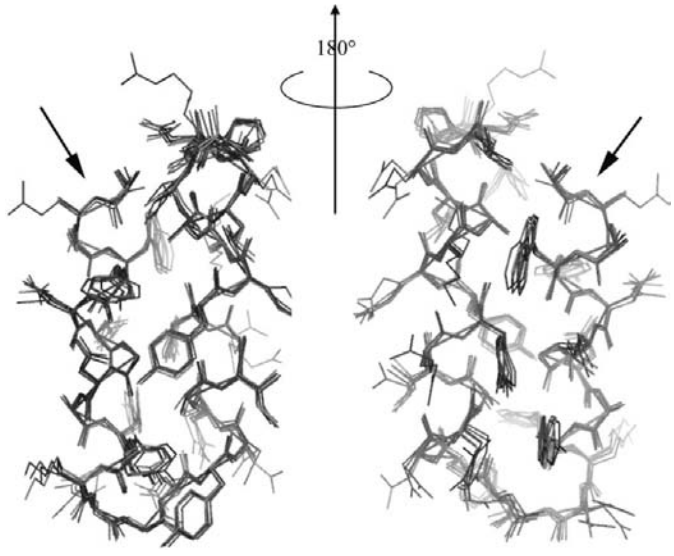

Figure 6
Alignment of the individual repeats of CTPR8 structures solved in different crystal forms. Arrows indicate the A-helix of the repeat, with all ten repeats from the $P3_121$, $P4_12_12$ and $P2_12_12_1$ crystal structures aligned.

CTPR8 and CTPR20 showed no systematic deviations from typical intensity statistics, although for the CTPR8 trigonal form crystallized with Cd^{2+}, the CTPR20 tetragonal crystal form and the CTPR20 trigonal form with $CaCl_2$ the $\langle I\rangle^2/\langle I^2\rangle$ values were higher than the expected value of 2.0 for the normal untwinned acentric case (2.44, 2.27 and 2.60). However, for the other four crystal forms of CTPR8 and CTPR20 the values are very close to 2.0 (between 1.99 and 2.09) and thus there is no consistent systematic trend to be observed.

In the case of the CTPR structures, intramolecular symmetry is fulfilled by crystallographic symmetry operators and there is static disorder on the repeat level (density appears as continous and individual molecules in different crystalline superhelices can be out of register with respect to each other by an integral number of repeats). However, for the CTPRs the disorder does not appear to be significant on the crystal level, in contrast to the RNA-duplex examples (Mueller *et al.*, 1999; Shah & Brunger, 1999), where the density was interpreted to actually consist of several different conformations. Therefore, the positional disorder observed here does not result in any systematic deviation in the intensity statistics.

While for proteins such cases of intramolecular symmetry in which the complete covalent molecule is generated by crystallographic symmetry have not been observed previously, they are common in other types of crystals. For example, approximately 14% of the small-molecule structures in the Cambridge Structural Database (http://www.ccdc.cam.ac.uk) have less than one molecule in the asymmetric unit. Rare cases of static disorder have been reported for proteins; for example, in a complex of trypsin with mung-bean inhibitor the crystal-packing contacts are fully defined by the trypsin molecule alone and the inhibitor is found in two alternative conformations in the asymmetric unit (Lin *et al.*, 1993).

3.2.2. Metal ion-mediated inter-superhelical crystal contacts. Apart from the head-to-tail stacking of the superhelical monomers, crystal contacts between monomers are mediated almost entirely by divalent or trivalent cations. In both the $P3_121$ and $P4_12_12$ crystal forms of CTPR8, crystal contacts between superhelices are mediated by Cd^{2+} ions interacting with acidic groups on the convex outside of the TPR superhelix. In the $P4_12_12$ crystal form, all inter-superhelix interactions are mediated by one unique interface, with a Cd^{2+} occupying a special position exactly on the crystallographic twofold axis (and two other symmetry-related Cd^{2+} ions on each side; *i.e.* there are 1.5 Cd^{2+} ions in the asymmetric unit). The Cd^{2+} on the twofold axis is coordinated by Glu50 from symmetry-related superhelices, with water molecules occupying the vacant coordination positions. The other Cd^{2+} is coordinated by Asp18 and Glu19 from one superhelix and Glu63 from another, again with water molecules filling the vacant coordination positions.

In the $P3_121$ crystal form of CTPR8 there are two interfaces between adjacent superhelices related by the $P3_121$ twofold-symmetry operators along the *ab* plane. One Cd^{2+} ion is coordinated in each interface by Glu56 and Glu63 from each molecule. Additionally, there is a salt bridge between Glu90 and Lys94 from adjacent molecules (there are actually two Cd^{2+} ions on this interface related by a crystallographic twofold). The other interface contains Cd^{2+} coordinated by Asp16 from one molecule and Glu121 and Glu120 from another. All the Cd^{2+} ions in both crystal forms are tetrahedrally coordinated by carboxylate side chains and water molecules. Similarly, the $P2_12_12_1$

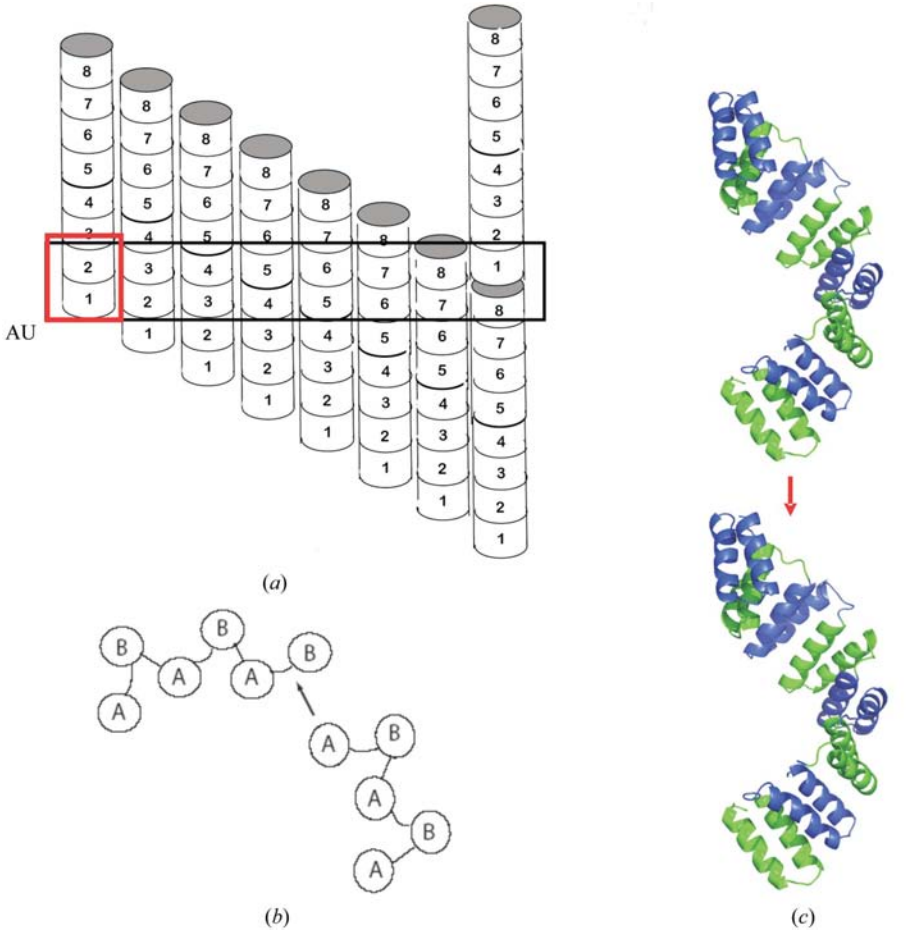

Figure 7
Schematic representation of the crystal-packing interactions between superhelical molecules. (*a*) As an example, in $P4_12_12$ there are two repeats (numbered 1–8) within the asymmetric unit (indicated as AU; red box). For this arrangement there are eight equally possible two-repeat arrangements for the asymmetric unit. (*b*) Schematic illustration of the stacking of helices between repeats of individual molecules. The first A-helix of the next molecule must always pack against the last repeat of the previous molecule and therefore the C-terminal capping helix must be displaced. (*c*) Ribbon representation of the superhelical stacking with each single repeat coloured yellow or blue.

inter-superhelical contacts are mediated by the Sm^{3+} ions, which are mainly coordinated by carboxylate side chains and water molecules. There are three Sm^{3+} ions per asymmetric unit in the $P2_12_12_1$ structure (all octahedrally coordinated). One is coordinated by Glu19 and Asp18 and three water molecules and the backbone carbonyl group of Leu98 of a symmetry-related superhelix. Around this site, two Arg residues (at position 33 of a TPR repeat) from the middle repeats (repeats two and three in the asymmetric unit) are involved in crystal contacts. Arg101 is hydrogen bonded to Gln14 and Arg67 is 'stacked' against the side chain of Glu22 over the interface. Also in this interface, Glu50 mediates an inter-superhelical interaction through a water molecule to Glu97. The other two Sm^{3+} ions can be found on another interface. One is coordinated by Gln16 and Asp86 from one superhelix and Glu52 and Asp53 from another. The other Sm^{3+} ion is coordinated by Glu56 and Asp120 and Glu121 from a symmetry-related molecule, with the remaining coordination sites being occupied by water molecules.

3.2.3. Head-to-tail stacking of CTPR molecules in crystals but not in solution. We initially observed 'head-to-tail' packing in crystals containing CTPR8 molecules that were lacking the C-terminal 'capping' A-helix (A_{cap}). It was natural that AB-AB-AB-AB-AB-AB-AB-AB units could pack head to tail by continuing the same B–A interactions between molecules (Fig. 7). We were therefore surprised that the same crystal packing was observed for versions of CTPR8 and CTPR20 that retained the C-terminal 'solvating' A_{cap} helix. The only way for AB-AB-AB-AB-AB-AB-AB-AB-A_{cap} molecules to pack head to tail is if the C-terminal A_{cap}-helix is displaced to allow B–A packing (it should be also noted that if the A_{cap} is packed against the B-helix of the last repeat as in other TPR structures, the current crystal packing and crystal form would not be possible; Fig. 7*b*). The presence of a single band on an SDS–PAGE gel and the observation of a single mass by mass-spectrometric analysis confirms that for both CTPR8 and CTPR20 the C-terminal A_{cap}-helix is present in the material crystallized (Table 2) and there is no protein that lacks the C-terminal solvating helix present in the crystals. The C-terminal A_{cap}-helix was designed to replace solvent-exposed hydrophobic residues with a more hydrophilic sequence than the consensus A-helix (through the mutations W4K, Y5G, Y11K and Y12Q, where the numbering refers to repeat positions in the A-helix). No electron density corresponding to the C-terminal helix is visible for either the CTPR8 or CTPR20 crystals. We must therefore conclude that in the crystal environment the energy of intermolecular B–A interactions is more favorable than that of intramolecular B–A_{cap} interactions.

Because of the intermolecular interactions that we observed in the crystals, we considered it important to examine whether there was any indication of intermolecular interactions in solution. We saw no evidence of higher molecular-weight species by gel filtration (data not shown). We also collected small-angle X-ray scattering (SAXS) data for CTPR8, which confirmed the elongated monomeric state of the protein in solution. In Fig. 2, we plot the SAXS intensity from the 3 mg ml^{-1} sample for wavevectors less than 0.2 Å^{-1} and the scaled SAXS intensity from the 60 mg ml^{-1} sample for wavevectors greater than 0.2 Å^{-1}. Also shown in Fig. 2, as a solid line, is the predicted SAXS profile of a single monomeric CTPR8 determined from the CTPR8 $P4_12_12$ crystal structure using the *CRYSOL*25 program. The good agreement between the model and these data supports the idea that CTPR8 is indeed a monomer in solution. To further check this idea, in Fig. 2 we compare the experimental SAXS profile with that of two hypothetical CTPR8 dimers: one a head-to-tail dimer (red) and the other a side-by-side dimer (blue). In neither case is there a match to the observed scattering.

3.3. Implications for natural TPR domain structures

The designed CTPR unit adopts the same structure in the context of the different-length proteins for which we have solved high-resolution crystal structures: CTPR2, CTPR3, CTPR8 and CTPR20. Moreover, the observed and predicted superhelical structures are the same for all these proteins. This result confirms the prediction of regular superhelices for extended arrays of TPR units, originally proposed by Barford

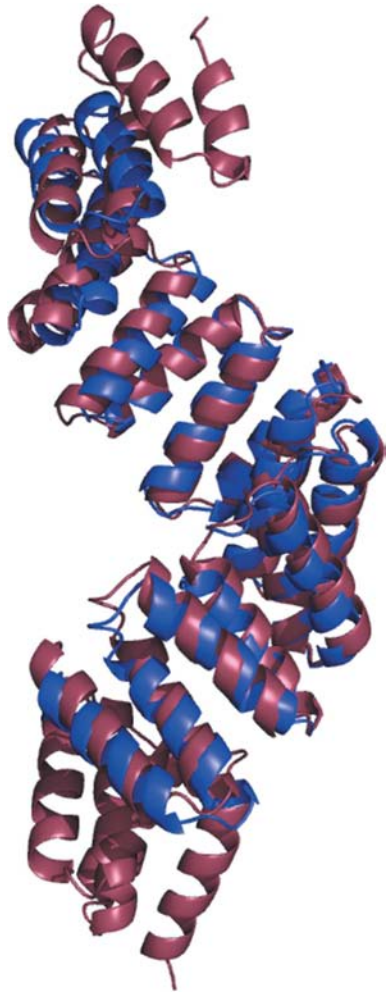

Figure 8
Alignment of repeats 3–10 of the TPR domain of OGT (red) with the tetragonal crystal structure of CTPR8 (blue).

and coworkers based on the structure of the 3TPR domain of PP5 (Das *et al.*, 1998). Proteins with up to 16 direct TPR repeats have been predicted in the sequence databases (D'Andrea & Regan, 2003) and some ORFs with significantly more are predicted by automated methods (*e.g.* PFAM and SMART). The structure of the 11.5 TPR-repeat fragment of the 13 TPR-repeat domain of the enzyme O-linked GlcNAc transferase (OGT; Jinek *et al.*, 2004) reveals a similar superhelical structure to that of the CTPR proteins. The CTPR proteins and the TPR domain of OGT all have eight repeats per superhelical turn. The superhelical pitch of the TPR domain of CTPR8 and CTPR20 are similar; however, in OGT the superhelical pitch varies along the molecule, from 58–64 Å at the C-terminus to 66–67 Å at the N-terminus. The TPR domain of OGT aligned with CTPR8 gives an r.m.s.d. of 1.94 Å for 245 aligned C^{α} atoms (Fig. 8).

We propose that when natural TPRs are present in uninterrupted tandem arrays, they all adopt a similar superhelical pitch with eight repeats per superhelical turn. The TPR motif has sufficient common features to specify not only the individual TPR helix–turn–helix structure but also interactions between repeats. Extended models generated from natural TPR domains (TPR2A, TPR1 of HOP and the TPR domain of PP5) support this hypothesis (data not shown; Main *et al.*, 2003; Das *et al.*, 1998).

Not all proteins with multiple TPRs will form such extended superhelical arrays. The peroxisomal import receptor (PEX-5), for example, has multiple TPR repeats, but these are separated by an unstructured linker into two typical 3TPR regions which fold over into a closed 'clamshell' structure around the ligand (Gatto *et al.*, 2000). The yeast Tom70 protein also has multiple TPRs. In this structure, separate 3TPR and 8TPR domains stack together tail to tail in a somewhat analogous fashion to the superhelical structures that we describe (Wu & Sha, 2006). In Tom70, however, neither the sequences of the individual repeats nor the superhelical structure is as regular as those observed in the OGT and CTPR proteins.

3.4. Stability of tandem arrays of TPRs

We have previously described the unfolding behavior of a series of CTPR proteins with different numbers of identical repeats, varying from two to ten, and demonstrated that the unfolding of this series can be described minimally using the one-dimensional Ising model with only three parameters (Kajander *et al.*, 2005). Here, we show that the denaturation behavior of CTPR20 can also be described and predicted using this simple treatment. The unfolding of CTPR6, CTPR8, CTPR10 and CTPR20 as a function of GuHCl (solid symbols) is shown in Fig. 3, together with the best-fit Ising-model description (solid lines), and CTPR20 unfolding fits to the one-dimensional Ising model as predicted.

4. Conclusions

Here, we present the structures of extended arrays of identical TPR repeats. We show that these proteins adopt a superhelical fold with eight TPR repeats per superhelical turn. In the crystals, the superhelices pack head to tail, generating extremely long superhelices that extend throughout the crystal. This arrangement represents a special case of crystal packing in which intramolecular hypersymmetry in a protein is described purely by crystallographic symmetry in the absence of NCS or static conformational disorder. We compare the structure of these designed proteins with that of a natural 11.5-repeat TPR domain of OGT that also contains an uninterrupted array of TPR units. This domain adopts the same superhelical structure as that of designed proteins. Finally, we demonstrate that the denaturation behavior of CTPR20 is as predicted from the behavior of shorter proteins in this series.

We are grateful to Joana Tala for technical assistance with CTPR8 protein expression and purification, Jimin Wang for insightful discussions on crystallographic matters, and Matt Calabrese and Jeff Knight for help with ESI–MS. Diffraction data were collected at BNL NSLS beamline X6A and at SAXS on BNL beamline X21. We gratefully acknowledge postdoctoral fellowships from the Spanish Ministry of Culture, Education and Sports (to ALC) and the Helsingin Sanomat Centennial Foundation, Finland (to TK). Use of the National Synchrotron Light Source, Brookhaven National Laboratory was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences under contract No. DE-AC02-98CH10886.

References

Andrade, M. A., Perez-Iratxeta, C. & Ponting, C. P. (2001). *J. Struct. Biol.* **134**, 117–131.
Blatch, G. L. & Lassle, M. (1999). *Bioessays*, **21**, 932–939.
Brennen, R. G., Westhof, E. & Sundaralingam, M. (1986). *J. Biomol. Struct. Dynam.* **3**, 649–665.
Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
D'Andrea, L. D. & Regan, L. (2003). *Trends Biochem. Sci.* **28**, 655–662.
Das, A. K., Cohen, P. W. & Barford, D. (1998). *EMBO J.* **17**, 1192–1199.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Eyal, E., Gerzon, S., Potapov, V., Edelman, M. & Sobolev, V. (2005). *J. Mol. Biol.* **351**, 431–442.
Gatto, G. J. Jr, Geisbrecht, B. V., Gould, S. J. & Berg, J. M. (2000). *Nature Struct. Biol.* **7**, 1091–1095.
Jinek, M., Rehwinkel, J., Lazarus, B. D., Izaurralde, E., Hanover, J. A. & Conti, E. (2004). *Nature Struct. Mol. Biol.* **11**, 1001–1007.
Kajander, T., Cortajarena, A. L., Main, E. R., Mochrie, S. G. & Regan, L. (2005). *J. Am. Chem. Soc.* **127**, 10188–10190.
Klosterman, P. S., Shah, S. A. & Steitz, T. A. (1999). *Biochemistry*, **38**, 14784–14792.
Kobe, B. & Kajava, A. V. (2001). *Curr. Opin. Struct. Biol.* **11**, 725–732.

Lin, G., Bode, W., Huber, R., Chi, C. & Engh, R. A. (1993). *Eur. J. Biochem.* **212**, 549–555.
Main, E. R., Stott, K., Jackson, S. E. & Regan, L. (2005). *Proc. Natl Acad. Sci. USA*, **102**, 5721–5726.
Main, E. R., Xiong, Y., Cocco, M. J., D'Andrea, L. & Regan, L. (2003). *Structure*, **11**, 497–508.
Mello, C. C. & Barrick, D. (2004). *Proc. Natl Acad. Sci. USA*, **101**, 14102–14107.
Mueller, U., Muller, Y. A., Herbst-Irmer, R., Sprinzl, M. & Heinemann, U. (1999). *Acta Cryst.* D**55**, 1405–1413.
Nelson, P. (2004). *Biological Physics: Energy, Information, Life.* New York: W. H. Freeman & Co.
Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.
Rogers, D. & Wilson, A. J. C. (1953). *Acta Cryst.* **6**, 439–449.
Shah, S. A. & Brunger, A. T. (1999). *J. Mol. Biol.* **285**, 1577–1588.
Sikorski, R. S., Boguski, M. S., Goebl, M. & Hieter, P. (1990). *Cell*, **60**, 307–317.
Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* D**55**, 849–861.
Wu, Y. & Sha, B. (2006). *Nature Struct. Mol. Biol.* **13**, 589–593.
Zhang, X. J., Wozniak, J. A. & Matthews, B. W. (1995). *J. Mol. Biol.* **250**, 527–552.
Zimm, B. H. & Bragg, J. K. (1959). *J. Chem. Phys.* **31**, 526–535.