Characterization of two potentially universal turn motifs that shape the repeated five-residues fold—Crystal structure of a luminal pentapeptide repeat protein from *Cyanothece* 51142

GARRY W. BUCHKO,¹ SHUISONG NI,^{1,4} HOWARD ROBINSON,² ERIC A. WELSH,³ HIMADRI B. PAKRASI,³ AND MICHAEL A. KENNEDY^{1,4}

¹Biological Sciences Division, Pacific Northwest National Laboratory, Richland, Washington 99352, USA

²Biology Department, Brookhaven National Laboratory, Upton, New York 11973, USA

³Department of Biology, Washington University, St. Louis, Missouri 63130, USA

(RECEIVED June 19, 2006; FINAL REVISION August 21, 2006; ACCEPTED August 22, 2006)

Abstract

The genome of the diurnal cyanobacterium *Cyanothece* sp. PCC 51142 has recently been sequenced and observed to contain 35 pentapeptide repeat proteins (PRPs). These proteins, while present throughout the prokaryotic and eukaryotic kingdoms, are most abundant in cyanobacteria. The sheer number of PRPs in cyanobacteria coupled with their predicted location in every cellular compartment argues for important, yet unknown, physiological and biochemical functions. To gain biochemical insights, the crystal structure for Rfr32, a 167-residue PRP with an N-terminal 29-residue signal peptide, was determined at 2.1 Å resolution. The structure is dominated by 21 tandem pentapeptide repeats that fold into a right-handed quadrilateral β -helix, or Rfr-fold, as observed for the tandem pentapeptide repeats in the only other PRP structure, the mycobacterial fluoroquinolone resistance protein MfpA from *Mycobacterium tuberculosis*. Sitting on top of the Rfr-fold are two short, antiparallel α -helices, bridged with a disulfide bond, that perhaps prevent edge-to-edge aggregation at the C terminus. Analysis of the main-chain (Φ, Ψ) dihedral orientations for the pentapeptide repeats in Rfr32 and MfpA makes it possible to recognize the structural details for the two distinct types of four-residue turns adopted by the pentapeptide repeats in the Rfr-fold. These turns, labeled type II and type IV β -turns, may be universal motifs that shape the Rfr-fold in all PRPs.

Keywords: cyanobacteria; β -bridges; circular dichroism; thermal melt; right-handed parallel β -helix; single-bridge β -sheet; β -bulges

⁴Present address: Department of Chemistry and Biochemistry, Miami University, Oxford, OH 45056, USA.

Reprint requests to: Garry W. Buchko, Pacific Northwest National Laboratory, P.O. Box 999, EMSL Mail-Stop K8-98, Richland, WA 99352, USA; e-mail: garry.buchko@pnl.gov; fax: (509) 376-2303; or Michael A. Kennedy, Department of Chemistry and Biochemistry, Miami University, 160 Hughes Hall, Rm. 239, 701 East High St., Oxford, OH 45056, USA; e-mail: michael.kennedy@muohio.edu; fax: (513) 529-5715.

Article and publication are at <http://www.proteinscience.org/cgi/doi/10.1110/ps.062407506>.

As complete genome sequence information became available for many organisms, Bateman et al. (1998) discovered a novel family of proteins containing a tandem pentapeptide repeat that can be approximately described as A[D/N]LXX. Today, the Pfam database (Bateman et al. 2000) lists 2110 pentapeptide repeat proteins (PRPs) (Pfam00805). In roughly two-thirds of these proteins, the pentapeptide repeat is the only recognizable domain (Vetting et al. 2006). While the overwhelming majority of

PRPs have been identified in prokaryotes, they are also found in eukaryotes, including one protein in humans. The number of chromosomal PRPs is not evenly distributed in prokaryotic genomes. Photosynthetic cyanobacteria appear especially endowed with 16 PRPs identified in *Synechocystis* sp. strain PCC 6803 (Bateman et al. 1998) and 40 in *Nostoc punctiforme* (Vetting et al. 2006). The genome of the cyanobacterium *Cyanothece* sp. strain 51142 has recently been sequenced and 35 PRPs identified in its chromosome (E.A. Welsh, M. Liberton, J. Stockel, J.M. Jacobs, R.S. Fulton, S.W. Clifton, R.K. Wilson, R.D. Smith, L.A. Sherman, H.B. Pakrasi, et al., unpubl.).

The first protein observed to contain pentapeptide repeats dates back to 1995 with the discovery of the *hglK* gene in the cyanobacterium *Anabaena* sp. strain PCC 7120 (Black et al. 1995). The *hglK* gene encodes a 727-residue protein with an N terminus predicted to contain four membrane-spanning regions followed by a region containing 36 consecutive pentapeptide repeats of the consensus sequence ADLSG. Chemical mutagenesis was used to generate an *Anabaena* strain that introduced a stop codon just before the pentapeptide repeat domain in the *hglK* gene, producing mutants with a distinct morphology compared to the wild-type strain that were incapable of forming the thick glycolipid layer external to the cell wall. The conclusions were that the HglK protein was membrane-associated, and the pentapeptide repeat domain was necessary for glycolipid transport and/or localization during heterocyst formation. However, the precise biochemical function and the three-dimensional structure of the HglK protein remain unknown.

A 398-residue PRP, termed RfrA, with a motif organization similar to *Anabaena* 7120 HglK was later identified in the photosynthetic bacterium *Synechocystis* sp. strain 6803 (Chandler et al. 2003). Like HglK, RfrA contains four membrane-spanning regions at its N terminus followed by a shorter run of 12 consecutive pentapeptide repeats at its C terminus (Chandler et al. 2003). While RfrA appears to be involved in the regulation of a manganese transport system different from the more thoroughly characterized ABC-transporter system in *Synechocystis* 6803, the mechanism of regulation is unknown. Two hypotheses are that RfrA alters the expression of the second unknown manganese transporter (transcriptional) or it may reversibly modify the second transporter (post-translational).

The *hglK* and *rfrA* genes are present in the chromosomes of the cyanobacteria *Anabaena* and *Synechocystis*, respectively. In many species of nonphotosynthetic *Enterobacteriaceae*, plasmids encoding a protein containing tandem pentapeptide repeats have been identified (Tran and Jacoby 2002; Nordmann and Poirel 2005). Biochemical characterization of this protein, Qnr, shows that in vitro, it protects *Escherichia coli* DNA gyrase and *E. coli* DNA topoisomerase IV from the inhibitory effects

of the powerful fluoroquinolone antibiotics (Tran and Jacoby 2002; Tran et al. 2005). Fluoroquinolones exert their antibacterial properties by binding reversibly to normal DNA complexes formed between DNA gyrase and DNA topoisomerase IV (Drlica and Malik 2003). They stabilize a covalent tyrosyl-DNA phosphate ester that is normally a transient intermediate, preventing religation of the DNA. As a consequence, the phospho-phenolic linkage eventually is hydrolyzed and a DNA double-strand break is generated. The accumulation of double-strand breaks is lethal to the cell (van Gent et al. 2001). The Qnr protein was observed to compete with DNA for binding to DNA gyrase (Tran et al. 2005), suggesting that the antibiotic resistance provided by this PRP may be due to its interaction with DNA gyrase that prevents normal DNA binding. Structural evidence for such a mechanism was recently provided with the crystal structure of the first PRP, MfpA, from *Mycobacterium tuberculosis* (Hegde et al. 2005).

The structure of *M. tuberculosis* MfpA (mycobacterial fluoroquinolone resistance protein), a 183-residue protein that forms a dimer in solution, has recently been reported (Hegde et al. 2005). It was targeted for study because it was identified as a homolog (67% identical) to a newly discovered, 193-residue protein in *Mycobacterium smegmatis* that was shown to be responsible for fluoroquinolone resistance in this fast-growing *Mycobacterium* (Montero et al. 2001). *M. tuberculosis* MfpA contains 30 consecutive pentapeptide repeats, and the crystal structure revealed that they formed a novel type of right-handed quadrilateral β -helix with each pentapeptide repeat occupying one face of a nearly square repeating unit (Hegde et al. 2005). The tower-like motifs are aligned head to head in the MfpA dimer, and exhibit characteristics similar to B-form DNA, including size, shape, and predominately electronegative surface potential distribution. Indeed, the MfpA structure can be docked readily onto the crystal structure of an N-terminal construct of *E. coli* DNA gyrase A subunit (Morais Cabral et al. 1997), a protein with a large electropositive potential at the position where DNA is believed to bind, and act as a DNA mimic. This structural data suggesting a potential interaction between the MfpA dimer and DNA gyrase was supported by biochemical data showing that MfpA inhibits the supercoiling and relaxing activity of *E. coli* DNA gyrase (Hegde et al. 2005).

There are at least two other examples of bacterial plasmids encoding for proteins with pentapeptide repeats that offer antibiotic resistance. The *E. coli* McbG protein is responsible for resistance to the peptide antibiotic Microsin B17 (Garrido et al. 1988). Like fluoroquinolones, Microsin B17 generates DNA double-strand breaks through its interactions with DNA gyrase (Vizan et al. 1991), although details of the biochemical mechanism differ from fluoroquinolones with Microsin B17 trapping

a transient intermediate in the C-terminal domain of GyrB (Pierrat and Maxwell 2005). Another example is the oxetanocin A resistance factor discovered in the *oxrA* resistance locus in a plasmid from *Bacillus magisterium* (Morita et al. 1999). Oxetanocin A and its derivatives are potent inhibitors of viral DNA polymerases and HIV reverse transcriptase (Izuta et al. 1992). Given that McbG and OxrA contain 13 and nine tandem pentapeptide repeats, respectively, it has been suggested that this may be a sufficient number of consecutive repeats to provide resistance in a mechanism similar to that proposed for MfpA and fluoroquinolones, by acting as a DNA mimic for the antibiotic's target enzyme (Vetting et al. 2006).

It is clear that one biochemical function of PRPs expressed from bacterial plasmids is to provide resistance to fluoroquinolones and other antibiotics. Compelling evidence suggests that the mechanism of resistance is via DNA mimicry (Hegde et al. 2005; Vetting et al. 2006). The origins of antibiotic resistance genes on these plasmids are likely PRP genes present in chromosomal DNA of other organisms that have functions removed from antibiotic resistance. However, little is known about the biochemical function of chromosomal PRP genes. In order to gain a better understanding of the molecular function of PRPs, we have undertaken an effort to characterize the three-dimensional structure of proteins in this family from *Cyanothece* 51142, a diurnal cyanobacteria with 35 chromosomal PRP genes. Here, we discuss the general features of the amino acid sequences of these 35 PRPs, present the crystal structure for Rfr32, a 167-residue protein with 21 tandem pentapeptide repeats, and describe in detail the two types of turn motifs adopted by each pentapeptide repeat.

Results and Discussion

Crystal growth and structure quality

SOSUisignal analysis (Gomi et al. 2004) of the native Rfr32 sequence identifies a 29-residue polypeptide starting at the N terminus that is postulated to direct the protein into the thylakoid lumen. Presuming the N-terminal 29 residues constitutes a signal polypeptide, it should be removed by a peptidase once it enters the thylakoid lumen, leaving a 138-residue protein (V30–Q167) that is the active cellular form. Efforts to express full-length recombinant Rfr32 in *E. coli* expression systems failed to generate significant levels of soluble protein, while a construct with the signal leader removed was reasonably successful (~15 mg/L medium). Consequently, the construct containing only Rfr32 residues V30–Q167 was used for our studies because this version was expressed in higher yields in a soluble form and it is likely the active cellular form of the protein. Crystals

were grown for truncated Rfr32 (V30–Q167) with and without a 43-residue, N-terminal tag containing an enterokinase cleavage site. Note that the amino acid sequence of the construct used for crystallization has been renumbered such that V30 in the native Rfr32 sequence corresponds to V2 in the crystal structure discussed here and the structures deposited in the Protein Data Bank.

Trigonal and tetragonal crystal forms of Rfr32 grew within a week using tagged Rfr32, whereas untagged Rfr32 crystallized only in the tetragonal form. X-ray diffraction data were collected on tagged Rfr32 (Se-Met labeled and unlabeled) in the trigonal crystal form and untagged Rfr32 (unlabeled) in the tetragonal crystal form. As listed in Table 1, both crystals were of comparable quality, with the unlabeled crystals of tagged Rfr32 diffracting to 2.1 Å resolution (PDB ID 2F3L) and Se-Met labeled crystals of untagged Rfr32 diffracting to 2.3 Å resolution (PDB ID 2G0Y). While tagged Rfr32 in the trigonal crystal form contained an N-terminal, 43-residue tag, no electron density was observed for this region. Instead, the first residue with reliable electron density was A7 (A35 in native, full-length Rfr32). SDS-PAGE analysis of protein stock solutions used for crystallization and crystals from unharvested crystallization drops suggest that the protein had not undergone proteolytic degradation before crystallization, indicating that the N-terminal tag was present but disordered in the crystals. Untagged Rfr32 only crystallized in the tetragonal crystal form, and the first residue with reliable electron density was S6 (S34 in native, full-length Rfr32). Evidently, the 43-residue tag on Rfr32 in the trigonal crystal form had little effect on the crystal structure, and this conclusion is corroborated by closer examination of the structures determined for both tagged and untagged Rfr32. The two crystal forms of Rfr32 contained only one molecule per asymmetric unit, and the structures of both molecules were essentially identical, with a backbone RMSD of 0.32 Å and an all heavy atom RMSD of 0.76 Å. Because the structure obtained for tagged Rfr32 in the trigonal crystal form diffracted to slightly higher resolution (PDB ID 2F3L), this structure is discussed in detail throughout the manuscript.

The quality of the crystal structure (2F3L) was assessed using PROCHECK (Laskowski et al. 1993) and MolProbity (Lovell et al. 2003). A Ramachandran plot of the coordinates using PROCHECK shows that all of the residues were in either the most favored regions (78%) or the additionally allowed (22%) regions. The average G-score was 0.30 (scores should be above −0.5). MolProbity analysis indicated that the overall protein geometry of the final model ranked in the 87th percentile (MER score of 1.87), where the 100th percentile is best among structures of comparable resolution. The clash score for “all-atom contacts” was 9.82, corresponding to

Table 1. Summary of data collection and structure refinement statistics for the two crystal forms of Rfr32

	Trigonal		Tetragonal
Data collection			
Data set	SAD	Native	Native
Space group	P3 ₂ 21	P3 ₂ 21	I4 ₁ 22
Unit-cell parameters (Å, °)	$a = b = 61.754$ $c = 83.708$ $\alpha = \beta = 90, \gamma = 120$	$a = b = 61.104$ $c = 83.391$ $\alpha = \beta = 90, \gamma = 120$	$a = b = 86.614$ $c = 89.745$ $\alpha = \beta = \gamma = 90$
Molecules/asymmetric unit	1.0	1.0	1.0
Matthews coefficient (Å ³ /Da)	3.1	3.0	2.8
Percent solvent (%)	60.5	59.6	56.5
X-ray source	X29	X29	X29
Temperature (K)	100	100	100
Resolution limits (Å)	30.0–2.3	50.0–2.1	50–2.3
Detector distance (mm)	300	200	300
Exposure time (sec)	1	5	3
Oscillation angle (°)	0.5	0.5	0.5
No. images	360	260	260
Total angle (°)	180	130	130
Mosaicity (°)	0.55	0.64	0.80
Wavelength (Å)	0.9790034	1.1	1.1
No. observations	140,848 (5072)	60,675 (2390)	45,488 (1986)
No. unique reflections	8322 (634)	9861 (613)	7582 (685)
Redundancy	17.0 (8.0)	6.2 (3.9)	6.1 (2.9)
R_{merge} (%) ^a	10.6 (28.5)	7.6 (21.8)	7.9 (20.5)
Completeness (%)	96.0 (74.5)	90.0 (57.1)	95.9 (89.4)
$I/\sigma(I)$	31.1 (4.2)	25.2 (3.6)	25.6 (3.3)
Structure refinement			
PDB ID		2F3L	2G0Y
R_{conv} (%) ^b		21.7	23.3
R_{free} (%) ^c		25.1	26.5
Protein atoms		989	995
Solvent atoms		25	16
Protein residues		131	132
Average B value, protein atoms (Å ²)		34.3	36.6
Average B value, solvent atoms (Å ²)		37.7	35.1
Average B value, total (Å ²)		34.4	36.5
RMSD from ideal bond lengths (Å)		0.006188	0.006069
RMSD from ideal bond angles (°)		1.00337	1.06414

Values in parentheses are for the highest resolution shell.

^a $R_{\text{merge}} = \sum_h \sum_j |I_{h,j} - \langle I_h \rangle| / \sum_h \sum_j I_{h,j}$, where I_h is shorthand for I_{hkl} and $I_{h,j}$ is the intensity of the j th observation of unique reflection h .

^b $R_{\text{conv}} = \sum_h |F_{\text{oh}}| - |F_{\text{ch}}| / \sum_h |F_{\text{oh}}|$, where F_{oh} and F are the observed and calculated structure factor amplitudes for reflection h .

^c R_{free} is equivalent to R_{conv} , but is calculated using a 5% disjoint set of reflections excluded from the maximum-likelihood refinement stages.

an 86th percentile ranking for structures of comparable resolution. While MolProbity analysis of the final model suggested that it contained three side chain rotamer outliers (residues I15, T25, and R45), closer inspection of the electron density for these residues did not justify a change in the rotamer orientations. Collectively, the PROCHECK and MolProbity assessments indicate that the final model is a high quality representation of the crystal structure of Rfr32.

Overall topology of the three-dimensional structure of Rfr32

Figure 1A is a Molscrip representation of Rfr32 (PDB ID 2F3L) using the Kabsch/Sander algorithm to identify

regions of secondary structure. The N terminus contains 21 consecutive pentapeptide repeats that form a right-handed parallel β -helix. While such helical β -sheet structures have been observed previously (Jenkins and Pickersgill 2001), PRPs are a unique subset of this group because four consecutive pentapeptide repeats form a nearly “square,” quadrilateral unit called a coil (Yoder et al. 1993; Jenkins and Pickersgill 2001; Vetting et al. 2006). These coils stack on top of one another to form a right-handed quadrilateral β -helix, or Rfr-fold. Consequently, the structure has four faces (Face 1 through Face 4) where each pentapeptide repeat on a single coil occupies one face of the tower. Rfr32 contains five complete, uninterrupted, stacked coils (labeled C1 through C5) giving the Rfr-fold a dimension of ~ 19 Å

Figure 1. (A) A Molscript representation of the Rfr32 (2F3L) crystal structure using the Kabsch/Sander algorithm to identify regions of secondary structure. α -Helices are colored red and β -strands blue. The side chains of C118 and C138 are highlighted to show the disulfide bond connecting α 1 and α 2. (B) Top view of A viewed from the C-terminal. (C) The $2F_o - F_c$ electron density map calculated using the structure factors and phases of the final model contoured at 2σ around the sulfur atoms of C118 and C138 identifying a disulfide bond (colored yellow).

in height and $\sim 13.4 \times 12.7 \times 12.4 \times 12.5$ Å in widths (average of Faces 1 through 4 as measured between backbone C- α carbons of the first and fifth residue of each pentapeptide repeat). The helix completes a revolution every 20 residues and travels ~ 4.8 Å along the helix axis, a distance similar to the separation between regular parallel β -strands. There is a very slight left-handed twist to the right-handed β -helix going from the N to C terminus, as can be seen in Figure 1B, a ribbon view of the tower's backbone from the top toward the N terminus. The coils of the tower are held together by short stretches of parallel β -sheets (Face 1) and β -bridges (Faces 2–4), both discussed in more detail below, which are integral to the quadrilateral shape of the Rfr-fold. At the C terminus of the Rfr-fold are two antiparallel α -helices (V111–C118 and T132–S135). While the first α -helix projects upwards from the fifth coil (C5) at an $\sim 60^\circ$ angle, the second shorter α -helix rests on top of Face 4 of C5, and this is more clearly seen in Figure 1B. Hydrogen bonds between the backbone atoms of G101 and F104 with the side chain groups of T132 and S135, respectively, help stabilize α 2 on top of Face 4 of C5. An 11-residue loop between the two α -helices folds over the side of Face 4 where it is stabilized by three hydrogen bonds between N81, G100, and T103 with N125, N125, and T128, respectively. The two α -helices are linked by a disulfide bond between C118 and the penultimate residue in the protein, C138. An electron density map supporting the assignment of the disulfide bond is shown in Figure 1C. Interestingly, this disulfide bond is observed despite growing the crystals in

the presence of 1.0 mM dithiothreitol, suggesting that it is protected from reduction.

Figure 2 shows the amino acid sequence of Rfr32 with the residues aligned according to position in the coils and faces of the Rfr-fold. The center residue of each pentapeptide repeat is designated i with the preceding residues labeled $i - 2$ and $i - 1$ and the following residues labeled $i + 1$ and $i + 2$. Figure 3A illustrates that the side chains of the $i - 2$ and i residues all point toward the interior of the tower and pack the middle of the Rfr-fold. In Rfr32, the i th residues are almost exclusively Leu or Phe (17/21), resulting in a stacked column of phenylalanine side chains interspersed with leucine side chains. The aromatic residues are gray in Figure 2 and indicated in various colors in Figure 3A to show that, except for F104 (magenta), they all stack in groups in Face 1 (black; Y9, F29, and F49), Face 2 (red; F19 and F39), and Face 3 (blue; F74 and F94). The side chains of the aromatic residues all have similar χ_1 and χ_2 torsion angles ($\chi_1 = -65 \pm 6^\circ$, $\chi_2 = 85 \pm 5^\circ$). While the side chain of the i th residue is predominately a large hydrophobic group, the side chain of the $i - 2$ residue is predominately a small hydrophobic group (13/21 are Ala). As illustrated in Figure 3B, these $i - 2$ residues also are aligned in columns. The net result is that the interior of the Rfr-fold is predominantly hydrophobic, alternating between columns of large and small side chains, and devoid of water. Indeed, the crystal structure revealed no water molecules in the interior.

Figure 3B illustrates that the side chains of the $i - 1$, $i + 1$, and $i + 2$ residues all point away from the interior of the tower and form the exterior, solvent-exposed surface of the Rfr-fold. While the interiorly directed side chains are primarily hydrophobic, the exteriorly directed side chains are typically hydrophilic. However, there are a few hydrophobic solvent-exposed side chains, colored black in Figure 3B with the corresponding residues underlined in Figure 2, and these form three small hydrophobic “islands” on the protein's surface on Faces 1, 2, and 4 that may be important sites for binding interactions with

	Face1	Face2	Face3	Face4		Coil
107	<u>AVLDV</u>				111	<u>C6</u>
87	<u>AIMMR</u>	TKFDD	AKITG	ADFSL	106	<u>C5</u>
67	<u>GLAYL</u>	TSFKG	ADLTN	AYLTE	86	<u>C4</u>
47	<u>AVFNG</u>	SALIG	ADLHG	ADLTN	66	<u>C3</u>
27	<u>AQFTN</u>	ADLTD	SNFSE	ADLRG	46	<u>C2</u>
7	<u>ASVED</u>	VKLIG	EDFSG	KSLTY	26	<u>C1</u>
	-2-1 +1+2	-2-1 +1+2	-2-1 +1+2	-2-1 +1+2		

Figure 2. Structure-based sequence alignment of the 21 tandem pentapeptide repeats in the Rfr-fold (A7–V111) of Rfr32. The residue position in the pentapeptide repeat, relative to the central residue i , is labeled on the bottom. The side chains of the $i - 2$ and i residues are in the interior of the Rfr-fold, while the side chains of the $i - 1$, $i + 1$, and $i + 2$ residues form the exterior. Aromatic residues in the i th position are colored gray; hydrophobic residues on the surface of the Rfr-fold are underlined.

Figure 3. Stick illustration of the regular alignment of the side chains in the Rfr-domain (A7–V111) of Rfr32. The main chain backbone of each coil is traced in a different color with every C- α carbon shown as a sphere. Except where noted below, the side chain atoms are colored by element type: (white) hydrogen, (green) carbon, (blue) nitrogen, (red) oxygen. (A) Only the side chains of the $i - 2$ and i residue for each pentapeptide repeat are displayed, and are observed to stack regularly inside the tower. The four groups of aromatic side chains are highlighted: (black) Y9, F29, F49; (red) F19, F39; (blue) F74, F94; (magenta) F104. (B) Only the side chains of the $i - 1$, $i + 1$, and $i + 2$ residue for each pentapeptide repeat are displayed and are observed orientated outside the tower. The hydrophobic side chains are highlighted in black.

another luminal protein or with the thylakoid membrane. One consequence of the small hydrophobic islands is that there is no large, contiguous, negatively charged surface on the protein (data not shown), as observed for MfpA. Aside from a small, contiguous, negatively charged region through Faces 3 and 4, the four-sided structure lacks a uniform charge distribution, having distinctly charged surfaces on each face.

Detailed description of the Rfr-fold

The general structural features of the Rfr-fold in Rfr32 are similar to those observed in the only other solved structure of a protein with an Rfr-fold, MfpA (Hegde et al. 2005; Vetting et al. 2006). Such similarities were

predicted by Vetting et al. (2006) because the repeating nature of the pentapeptide sequence in PRPs suggests that they should also adopt similar repeating conformations throughout the structure. However, with a second structure of an Rfr-fold to analyze, it is possible to characterize in more detail the structural properties that may be universal to all Rfr-folds.

Figure 4 is a plot of the main chain (Φ, Ψ) dihedral orientations for Rfr32 residues A7–V111 that make up the Rfr-fold. Clearly, two distinct patterns are observed for the five residues constituting each coil on Face 1 and for the five residues constituting each coil on Faces 2, 3, and 4. Figure 5 is a Ramachandran plot of the data in Figure 4 coded based on position in the pentapeptide repeat with residues in Face 1 colored blue and residues in Faces 2, 3,

Figure 4. Plot of the main chain (Φ, Ψ) dihedral torsion angles for the 21 consecutive pentapeptide repeats (residues A7–V111) that make up the Rfr-fold of Rfr32. The Φ torsion angles are connected with a dashed line and labeled with open black squares for residues in Face 1 and closed blue squares for residues in Faces 2–4. The Ψ torsion angles are connected with a solid line and labeled with open green circles for residues in Face 1 and closed red circles for residues in Faces 2–4.

Figure 5. Ramachandran plot of the main chain (Φ, Ψ) dihedral torsion angle pairs for the 21 consecutive pentapeptide repeats (residues A7–V111) of Rfr32 color coded on the basis of position in the repeat. Residues in Face 1 are open and colored blue; residues in Faces 2–4 are solid and colored red. Symbols: (circles) $i - 2$; (squares) $i - 1$; ($\times$) i ; (+) $i + 1$; (diamonds) $i + 2$. The major differences between the (Φ, Ψ) pairs of residues in Face 1 and Faces 2–4 are highlighted by enclosure in red (i) and blue ($i + 1$) circles or ovals. The blue arrows identify the Ramachandran nomenclature for the two turn motifs defined from the $i + 1$ to $i + 2$ residue: (solid) type II; (dashed) type IV.

and 4 colored red. The first observation is that only 76% of the residues in the Rfr-fold lie in the most favored region while the remaining 24% lie in the additionally allowed region. This is somewhat surprising given the high quality and resolution of the X-ray diffraction data, and suggests that there is something unique to the Rfr-fold. Closer inspection of the Ramachandran plot reveals that the (Φ, Ψ) pairs for the $i - 2$ (circles), $i - 1$ (squares), and $i + 2$ (diamonds) residues are clustered into regions of the Ramachandran plot based on their position in the pentapeptide sequence regardless of their Face position in the Rfr-fold. For the i th ($\times$) and $i + 1$ (+) residues, the (Φ, Ψ) pairs from Face 1 are clustered into a different region than the pairs from Faces 2, 3, and 4. These four

regions, circled red (i) and blue ($i + 1$) in Figure 5, differ by ~ 90 – 110° in Φ and Ψ , indicating that the two distinct conformations of the pentapeptide repeat differ by an $\sim 90^\circ$ rotation of the peptide unit between residue i and $i + 1$.

A summary of the general main chain (Φ, Ψ) dihedral orientations of the residues in the two pentapeptide conformations is shown in Table 2. The first entries are the general (Φ, Ψ) dihedral orientations proposed by Vetting et al. (2006) based on their single structure, and the second value is a refinement of the general orientations based on our second structure (for specifics, see Table 3). Two features common in all the pentapeptide repeats are a β -bridge in the $i - 1$ position of the repeat and, as illustrated in Figure 1A, an $\sim 90^\circ$, right-handed, four-residue turn between each pentapeptide repeat. A turn in a protein is defined if the carbonyl of residue i hydrogen bonds with the amide of residue $i + n$ (Kabsch and Sander 1983). The β -turn is a very common four-residue turn ($n = 3$) between residues that are not in an α -helix with a $C_\alpha(i)$ to $C_\alpha(i + 3)$ distance of $< 7 \text{ \AA}$ (Richardson 1981; Shepherd et al. 1999). β -Turns effect a reversal in the direction of the protein backbone and are typically subclassified into nine different types on the basis of the main-chain (Φ, Ψ) dihedral values (Wilmot and Thornton 1988). In Rfr32 the mean $C_\alpha(i)$ (residue i) to the following $C_\alpha(i + 3)$ (residue $i - 2$ of the following pentapeptide repeat) distance is always $< 7.0 \text{ \AA}$ in all four turns of each coil. Hence, tandem pentapeptide repeats (1) contain no α -helices, (2) contain at least one β -bridge, (3) effect a change in the direction of the protein backbone, and (4) have a $C_\alpha(i)$ to $C_\alpha(i + 3)$ distance $< 7.0 \text{ \AA}$, features characteristic of β -turns (Richardson 1981; Shepherd et al. 1999). Therefore, one way of describing the Rfr-fold is as a collection of two types of secondary structure elements, β -turns, involving residues $i, i + 1, i + 2$, of one pentapeptide repeat and the first residue of the following pentapeptide repeat, ($i - 2$), connected by isolated β -bridges involving the $i - 1$ residue (Vetting et al. 2006). These two β -turns fall into types II and IV

Table 2. Summary of the general main chain (Φ, Ψ) dihedral torsion angle pairs for each residue in the pentapeptide repeat

Pentapeptide repeat position	Type II β -turn	DSSP motif (II)	Type IV β -turn	DSSP motif (IV)
$i - 1$	120/120 (–105/115)	β -bridge ()	–120/120 (–125/125)	β -bridge ()
i	–120/20 (–115/25)	turn	–120/120 (–125/115)	β -bridge ()
$i + 1$	–60/120 (–55/130)	coil/bend	–120/120 (–120/130)	β -bridge ()
$i + 2$	60/20 (65/10)	bend	60/20 (55/40)	bend
$i - 2$	–80/160 (–85/150)	coil	–80/160 (–80/150)	coil

The pentapeptide repeat position is relative to the central residue in the five residue repeat (See Fig. 2). The first values are from Vetting et al. (2006). The values in parentheses are refined values based on the averages observed in the crystal structures of Rfr32 and MfpA. The values in bold highlight the major difference between the two turns and reflect an $\sim 90^\circ$ rotation of the peptide unit between residue i and $i + 1$.

Table 3. Mean main chain (Φ, Ψ) dihedral torsion angle pairs for each residue in the pentapeptide repeat of Rfr32 and MfpA

Pentapeptide repeat position	Rfr32 Type II β -turn (15)	MfpA Type II β -turn (17)	Rfr32 Type IV β -turn (5)	MfpA Type IV β -turn (9)
$i - 1$	$-107 \pm 13/113 \pm 11$	$-100 \pm 9/113 \pm 10$	$-130 \pm 18/124 \pm 10$	$-125 \pm 8/128 \pm 5$
i	$-114 \pm 7/30 \pm 15$	$-114 \pm 9/21 \pm 8$	$-127 \pm 7/117 \pm 20$	$-124 \pm 6/117 \pm 20$
$i + 1$	$-57 \pm 10/128 \pm 6$	$-54 \pm 8/132 \pm 6$	$-124 \pm 18/131 \pm 12$	$-122 \pm 16/126 \pm 5$
$i + 2$	$66 \pm 8/12 \pm 9$	$67 \pm 7/8 \pm 9$	$55 \pm 4/39 \pm 10$	$55 \pm 7/38 \pm 12$
$i - 2$	$-89 \pm 20/151 \pm 9$	$-84 \pm 18/148 \pm 10$	$-70 \pm 8/152 \pm 11$	$-87 \pm 16/155 \pm 10$

The pentapeptide repeat position is relative to the central residue in the repeat (see Fig. 2). Values for MfpA taken from the PDB structure 2BM5 (Hegde et al. 2005). Number in parentheses indicates the number of turns used to determine the average with only values for complete coils used: C1–C5 for Rfr32, C1–C7 for MfpA. Furthermore, for MfpA, pentapeptide repeats containing, and following, a proline residue were not included in the average (colored yellow in Fig. 7).

because the main chain (Φ, Ψ) dihedral values of the central two residues, $i + 1$ and $i + 2$, in the four-residue turn are within 30° of the “ideal” values for these types (Wilmot and Thornton 1988). Note that while the type II β -turn is common (Hutchinson and Thornton 1994; Pabasik et al. 2005), the type IV β -turn is a miscellaneous bin for β -turns that do not fall into any of the other categories (Richardson 1981) and is actually the most populated type (Hutchinson and Thornton 1994). As illustrated in Figure 5 and highlighted in bold in Table 2, the major difference between the two types of β -turns is the Ψ and Φ torsion angles of the i and $i + 1$ residues due to an $\sim 90^\circ$ rotation of the peptide unit between these two residues. The consequence of this single peptide unit rotation is an altered network of intercoil and intracoil hydrogen bonding as illustrated in two examples for Rfr32 in Figure 6.

Figure 6A illustrates the hydrogen bonding network for three coils (C2, C3, and C4) on Face 2 viewed from the front and the top. The pentapeptide repeat on Face 2 of each coil forms a type II β -turn with the $i - 2$ residue of the following pentapeptide. Only the $i - 1$ β -bridge residue contributes both an amide proton and carbonyl oxygen to intercoil hydrogen bonding. The main chain amide of the i th residue and the main chain carbonyl of the $i + 1$ residue are roughly orthogonal to the plane of the other main chain atoms of the $i - 2$ through $i + 1$ residues, and consequently, they cannot form intercoil hydrogen bonds. However, in this orthogonal orientation, the carbonyl of the i th residue is near the amide of the $i - 2$ residue where it forms an intracoil hydrogen bond as shown in the top view in Figure 6A. Such a hydrogen bond is a characteristic of a DSSP defined turn (Kabsch and Sander 1983). As illustrated by a solid blue arrow in Figure 5, the Ramachandran nomenclature (Wilmot and Thornton 1990) for the type II β -turn in the Rfr-fold is $\beta_P\gamma$.

Figure 6B illustrates the hydrogen bonding network for three coils (C2, C3, and C4) on Face 1 viewed from the front and the top. In this example, the pentapeptide repeat on each coil forms a type IV β -turn with the $i - 2$ residue

of the following pentapeptide. The approximate 90° rotation of the peptide unit between the i th and $i + 1$ residue places the main chain amide of the i th residue and the main chain carbonyl of the $i + 1$ residue into the plane of the other main chain atoms of the $i - 2$ through $i + 1$ residues. Now this amide and carbonyl group can fully form intercoil hydrogen bonds with the main chain atoms of the pentapeptide above and below it if these pentapeptide repeats are also in the same type IV β -turn position, as is the case in Figure 6B. However, one consequence of this $\sim 90^\circ$ rotation of the peptide unit between the i and $i + 1$ residue is that the main chain carbonyl of the i th residue is no longer near the amide proton of the $i - 2$ residue, as shown in the top view in Figure 6B, and now these groups cannot form an intracoil hydrogen bond. As a result of the loss of intracoil hydrogen bonds for intercoil hydrogen bonds, the pentapeptide that forms a type IV β -turn is more extended (~ 0.9 Å) than the pentapeptide that forms a type II β -turn. Furthermore, because of the absence of a hydrogen bond between the carbonyl of residue i with the amide of residue $i + n$, this is not a “turn” as defined by the DSSP convention (Kabsch and Sander 1983). As illustrated by a blue dashed arrow in Figure 5, the Ramachandran nomenclature (Wilmot and Thornton 1990) for the type IV β -turns in the Rfr-fold is β_{α_L} , which is different from the $\beta_P\gamma$ observed for the type II β -turns. This small difference is reflected in the Ramachandran nomenclature arrows for both turns shown in Figure 5, they cross but are not coincident.

One consequence of such a collection of turns in an Rfr-fold is that a type IV β -turn may not exist in isolation on the face of an Rfr-fold (e.g., a type IV β -turn enveloped by type II β -turns above and below it) because there will be no new intracoil hydrogen bond between two stacked $i + 1$ residues (Fig. 6B, front) to compensate for the intercoil hydrogen bond between sequential i and $i - 2$ residues (Fig. 6A, top) that is lost in a type IV β -turn. In the only two solved PRP structures containing an Rfr-fold, type IV β -turns in isolation are not observed (see Fig. 8 below).

However, the situation is not that simple because the side chain of the $i - 2$ residue is oriented inside the core of the Rfr-fold (Fig. 3A). In type IV β -turns these residues are often serine and threonine (Hegde et al. 2005). The hydroxyl groups of these side chains are observed to hydrogen bond with their own backbone amide or the backbone carbonyl group of the i th residue on the coil directly below it (Hegde et al. 2005), acting a bit like a backbone mimic (Eswar and Ramakrishnan 1999) and providing some added stabilization to the turn. Note that if type IV β -turns must, at a minimum, be present in pairs on a face of an Rfr-fold, then they would also always meet one of the major requirements to be called a β -bulge. Bulges are believed to play an important biological role in proteins, affecting the direction of the β -sheet and the positioning of important residues for

function (Chan et al. 1993). While the classical definition of a β -bulge is two residues in the bulged β -strand opposite one residue in the adjacent β -strand (Richardson et al. 1978), a β -bulge can be any irregularity in a β -sheet involving two strands (Chan et al. 1993), including irregularities directly opposite each other. In the latter example, called P-bent, the displaced residues on both strands occupy the α_R region of Ramachandran space and bends the parallel β -sheet $\sim 45^\circ$ (Chan et al. 1993). In Rfr32, the displaced residues, $i + 2$ in the pentapeptide repeat, occupies the α_L region of Ramachandran space and bends the β -sheet $\sim 90^\circ$. β -Bulges in parallel sheets are very uncommon with the majority, $\sim 90\%$, observed in antiparallel β -sheets (Richardson et al. 1978; Chan et al. 1993).

As mentioned, one of the common features of the Rfr-fold is the β -bridge at residue $i - 1$. Only when two type

Figure 6. Examples of the two types of β -turns adopted by the pentapeptide repeat and the network of inter- and intracoil main chain hydrogen bonds. Shown are the front and top views of the main chain backbone atoms of three adjacent pentapeptide repeats on coils C2 through C4 of Rfr32. (A) Three adjacent type II β -turns on Face 2. The main chain carbonyl of the i th residue and the amide of the $i + 1$ residue are orientated $\sim 90^\circ$ out of the plane of the other main chain atoms of the $i - 2$ through $i + 1$ residues and cannot form intercoil hydrogen bonds (front). Instead, the carbonyl of the i th residue is near the amide of the $i - 2$ residue, and forms an intracoil hydrogen bond with the next pentapeptide repeat (top). (B) Three adjacent type IV β -turns on Face 1. The main chain carbonyl of the i th residue and the amide of the $i + 1$ residue are orientated in the plane of the other main chain atoms of the $i - 2$ through $i + 1$ residues and form intercoil hydrogen bonds (front). The carbonyl of the i th residue is too distant from the amide of the $i - 2$ residue to form an intracoil hydrogen bond with the next pentapeptide repeat (top).

IV β -turns exist on the same face of an Rfr-fold are there two adjacent β -bridges present to form a DSSP defined β -ladder, that is, at the same time, a DSSP defined β -sheet (Kabsch and Sander 1983). As illustrated in Figure 1A, all the type IV β -turns occur on Face 1 of Rfr32 to form one continuous parallel β -sheet. On the three remaining faces of Rfr32 parallel β -bridges are aligned to form a long, single-bridge β -sheet on each face. While isolated β -bridges are occasionally observed in protein structures (Richardson and Richardson 2002), to the best of our knowledge the stacked parallel β -bridges observed in the faces of the Rfr-fold are unique in protein fold “space.”

Comparison with MfpA

Figures 7 and 8 are side-by-side comparisons of the crystal structures of Rfr32 (2F3L) and the only other solved structure of a protein with an Rfr-fold, MfpA (2BM5) (Hegde et al. 2005), using the structures refined to the highest resolution (2.1 and 2.0 Å, respectively). To

simplify the figures, only one MfpA molecule in the C-terminal, head-to-head dimer is shown (monomer A). As illustrated in Figures 7 and 8, MfpA and Rfr32 share a similar overall architecture—an N-terminal, right-handed quadrilateral β -helix (Rfr-fold) capped with a pair of α -helices. The pentapeptide repeats adopt a similar registration in both molecules, with four tandem repeats defining a coil of the tower and each coil rising ~ 4.8 Å along the helix axis. Indeed, a Dali search (Holm and Sander 1998) using residues A7–L106 of Rfr32 returns a Z-score of 19.2 and an RMSD of 1.0 Å with residues Q2–G101 of MfpA indicating that the Rfr-fold is very similar in both molecules. MfpA also contains two α -helices toward the C-terminal with one helix incorporated into the last coil on Face 3 and the second one sitting over the top of the last pentapeptide repeat on Face 2.

The pentapeptide repeats in the Rfr-fold of Rfr32 and MfpA all adopt one of two general conformations, a type II or type IV β -turn. This is evident from the analysis of the data in Table 3 that lists the mean main chain Φ and Ψ torsion angles for the two types of turns in Rfr32 and MfpA. The means for 17 out of 20 of the listed torsion angles are $<5^\circ$ apart. Of the three listed torsion angles that differ by $>5^\circ$, the means still fall within the standard deviations. In Figure 7 the type II and type IV β -turns adopted by the pentapeptide repeats are colored blue and cyan, respectively, for MfpA and Rfr32. The face on the left-hand side of the cyan colored type IV β -turn also contains a DSSP defined β -sheet while the face on the left-hand side of the blue colored β -turn contains a single-bridge β -sheet. As mentioned earlier, the type IV β -turns always appear, at a minimum, in pairs on a face. One obvious difference between Rfr32 and MfpA is that the two types of turns are clustered on individual faces of Rfr32 while they are mixed in the N-terminal half of MfpA. Perhaps the different arrangement of the turns is a mechanism for generating different surfaces on the faces of an Rfr-fold.

The Rfr-fold of MfpA has eight consecutive, complete helical turns with a prominent kink after the fourth helical turn that was attributed to a *cis*-proline between C4 and C5 on Face 4 (the turn before and after this residue is colored yellow in Fig. 7). The kink induces a 12° change in the helical axis of coils C1–C4 and C5–C8, which may be essential to its function as a DNA mimic, causing a sigmoidal shape across the head-to-head dimer (Hegde et al. 2005; Vetting et al. 2006). On the other hand, the Rfr-fold of Rfr32 contains no proline residues and no such kink, and therefore, Rfr32 may be unable to function as a DNA mimic.

Another small difference between the two structures is that the Rfr-fold in MfpA is more twisted than in Rfr32 with the twist most prominent before the kink (C1–C4). Vetting et al. (2006) suggested that the twist may be

Figure 7. Stylized stereoview of the two crystal structures of MfpA (2BM5) (Hegde et al. 2005) and Rfr32 (2F3L) highlighting the position of the type II and type IV β -turns in blue and cyan, respectively. The structures are drawn with the N-terminal on the *bottom*, α -helices colored red, and Face 1 aligned in the *front*.

Figure 8. Comparison of the position of the aromatic residues (black) in the Rfr-fold in the crystal structures of MfpA (2BM5) (Hegde et al. 2005) and Rfr32 (2F3L). The structures are drawn with the N-terminal on the bottom and Face 1 aligned on the right-hand side.

driven by stacking interactions of the interior aromatic side chains that minimize the negative interaction between the π -electron clouds (Hunter et al. 1991). While such stacking interactions may contribute to the twist in the Rfr-fold, it likely is not the major contributor to the twist. Figure 8 highlights the position of the aromatic residues in the structures of Rfr32 and MfpA. The latter Rfr-fold contains two stacks, one of three (Face 2) and one of four (Face 3) aromatic residues, while Rfr32 contains three stacks, one of three (Face 1) and two of two (Faces 2 and 3) aromatic residues. Granted, the more extended aromatic stacks in MfpA may generate more twist, but, Rfr32 and the lower coils of MfpA both contain the same overall number of stacked aromatic residues, seven. A more likely reason for the twist is the difference in overall length of the pentapeptide repeat in a type IV β -turn versus a type II β -turn, ~ 0.9 Å. In Figure 7 the two types of turns are colored cyan (type IV) and blue (type II), and clearly, there is a mixing of the turns in the lower coils of MfpA while in Rfr32 all the type IV β -turns are on Face 1. Consequently, the lower coils of MfpA will be more twisted than the coils of Rfr32.

Related to the difference in the length of the two β -turns and aromatic side chain stacking, Vetting et al. (2006) observed that phenylalanine residues predominated the *i* position of pentapeptide repeats adopting type IV β -turns. They suggested that the more extended type IV β -turn conformation better accommodated the bulky aromatic group than the type II β -turn conformation, as seven out of 10 aromatic residues were observed in type

IV β -turns in the Rfr-fold of MfpA. However, the correlation may have been serendipitous, as only three out of the eight aromatic residues in the Rfr-fold of Rfr32 are observed in type IV β -turns.

The second α -helix at the C-terminal of MfpA interacts in an antiparallel fashion with the identical helix in a second molecule to form an intermolecular head-to-head dimer. Similar head-to-head dimers were observed in the four crystal forms that were characterized, and dimers were also observed to form in solution (Hegde et al. 2005; Vetting et al. 2006). In contrast, Rfr32 is a monomer in solution, as shown by size-exclusion chromatography and nuclear magnetic resonance spectroscopy (data not shown), and the packing in the two crystals of Rfr32 differed from MfpA such that the C-terminal α -helices of two molecules do not contact each other (data not shown). Therefore, if dimer formation is essential for MfpA to function as a DNA mimic, the absence of a similar dimer in Rfr32 may hint toward a distinct function in the luminal space of cyanobacteria.

Circular dichroism profile and thermal stability of cleaved Rfr32

Circular dichroism spectroscopy is a powerful tool to probe the conformation of proteins in solution (Woody 1974; Smith and Pease 1980) because small changes in the backbone conformation can cause strong changes in the CD spectrum (Manning et al. 1988). While the CD spectra of α -helices and β -sheets are well characterized, there remains some ambiguity regarding the pure components CD spectra of different types of β -turns (Perczel et al. 1992). Understanding the component contribution of β -turns to the CD spectrum is important because β -turns are a common structural motif, comprising up to 25% of the structure of all folded proteins and peptides (Kabsch and Sander 1983; Wilmot and Thornton 1988). One of the reasons for the ambiguity in the contribution of β -turns to CD spectra is the scarcity of proteins and model compounds that are purely one type of β -turn. The crystal structure of Rfr32 shows $\sim 75\%$ of the protein forms a right-handed quadrilateral β -helix structure with 80% of the residues participating in two types of β -turns (75% type II and 25% type IV) with one face of the fold adopting a canonical parallel β -sheet structure and three faces of the fold forming single-bridge β -sheets (Fig. 1A). Consequently, the CD spectrum of Rfr32 should be dominated by β -turn and parallel β -sheet features.

A far-UV CD spectrum of Rfr32, obtained using a sample with the N-terminal 43-residues removed so that the spectrum was free of contributions from this section, is shown in Figure 9A. The spectrum is dominated by one feature, a minimum band at ~ 216 nm with no distinct maximum band. As might be expected, this

Figure 9. (A) Circular dichroism spectrum of untagged Rfr32 (260 μ M) at 25°C in buffer containing 100 mM NaCl, 20 mM potassium phosphate, 1 mM DTT (pH 7.4). (B) CD thermal melt for untagged Rfr32 (20 μ M) in buffer containing 100 mM NaCl, 20 mM potassium phosphate, 1.0 mM DTT (pH 7.4). The data were collected at 216 nm in 2.5°C intervals between 10°C and 80°C.

spectrum most closely resembles the pure component CD spectra for β -turns and parallel β -sheets (Perczel et al. 1992). Despite having two short α -helices at the C-terminal, the characteristic double minimum at 222 nm and 208–210 nm and maximum between 190 nm and 195 nm (Holzwarth and Doty 1965) is buried under the major band. Once the correlation between pentapeptide sequence and β -turn type is better understood, by genetically modifying Rfr32 to remove the C-terminal α -helices and convert the type IV β -turns into type II β -turns, it may be possible to construct an Rfr-fold that is composed entirely of type II β -turns and obtain a CD spectrum even more dominated by this component. Alternatively, some of the other *Cyanothece* PRPs may natively be free of extraneous secondary structure and contain an Rfr-fold dominated by a single type of turn.

To assay the thermal stability of Rfr32, the ellipticity at 216 nm was measured as a function of temperature between 10°C and 80°C. Typically, a phase transition is observed as the folded protein becomes denatured and the ellipticity concomitantly decreases with heating (Buchko et al. 2000; Chang et al. 2003; Kwok and Hodges 2003). Such a transition occurs for Rfr32, as shown in Figure 9B. There is a very gentle decrease in the ellipticity at 216 nm from 10°C to $\sim$ 40°C, upon which a step decrease occurs up to 55°C, at which point a plateau is reached. The inflection point for the transition, which is non-reversible, is $\sim$ 48°C, and likely reflects the unraveling of the Rfr-fold.

Analysis of the PRP family in *Cyanothece* 51142

Table 4 lists the 35 PRPs identified in the genome of *Cyanothece* 51142. SOSUI signal analysis (Gomi et al. 2004) of these PRP sequences predicts that seven will be located in the lumen/periplasm, nine in the plasma membrane, and the rest in the cytosol. While the proteins vary in size from 105 to 930 residues, 80% of them

contain <400 amino acids. Analysis of their sequences predicts that they contain as few as 14 (Rfr33) and as many as 61 (Rfr01) pentapeptide repeats. Except for Rfr08, Rfr02, Rfr17, and Rfr16, all the pentapeptide repeats are tandem. Figure 10 graphically illustrates the predicted composition of the 35 *Cyanothece* 51142 PRPs in terms of pentapeptide repeat (red), N-terminal (dark blue), C-terminal (light blue), and other regions (white). For proteins with <400 residues, a predicted Rfr-domain constitutes >50% of the residues in all but three PRPs. For proteins with >400 residues, the predicted RFR-domain constitutes less than one-third of the protein, and in two of these larger proteins the pentapeptide repeats are not all tandem. As observed in the other cyanobacteria, the predicted Rfr-domain is located toward the C-terminal in

Table 4. Summary of features of the 35 PRPs from *Cyanothece*

Name	Length	Number of repeats ^a	Predicted location	BLAST result
Rfr29	105	17 (81%)	Cytoplasmic	
Rfr34	141	13 (47%)	Cytoplasmic	
Rfr28	145	20 (69%)	Lumenal/periplasmic	
Rfr30	146	16 (55%)	Lumenal/periplasmic	
Rfr12	153	26 (85%)	Cytoplasmic	
Rfr19	166	19 (57%)	Membrane	
Rfr35	166	18 (55%)	Cytoplasmic	
Rfr27	167	16 (50%)	Membrane	
Rfr32	167	21 (63%)	Lumenal/periplasmic	
Rfr23	172	17 (49%)	Membrane	
Rfr11	176	27 (77%)	Cytoplasmic	
Rfr24	182	26 (72%)	Cytoplasmic	
Rfr10	183	33 (90%)	Cytoplasmic	
Rfr26	194	26 (67%)	Cytoplasmic	
Rfr21	221	22 (50%)	Cytoplasmic	DNAJ domain
Rfr13	236	28 (59%)	Cytoplasmic	
Rfr08	240	42 (88%) ^b	Cytoplasmic	
Rfr22	261	18 (34%)	Lumenal/periplasmic	
Rfr05	268	47 (88%)	Cytoplasmic	
Rfr14	272	26 (48%)	Cytoplasmic	
Rfr06	273	41 (75%)	Membrane	
Rfr04	295	49 (83%)	Cytoplasmic	
Rfr15	319	24 (38%)	Cytoplasmic	
Rfr03	320	60 (94%)	Cytoplasmic	
Rfr07	325	34 (52%)	Cytoplasmic	
Rfr01	330	62 (94%)	Cytoplasmic	
Rfr20	367	22 (30%)	Cytoplasmic	
Rfr02	399	62 (78%) ^b	Lumenal/periplasmic	
Rfr33	439	12 (14%)	Membrane	RfrA (Mg uptake)
Rfr09	472	29 (33%)	Membrane	Ser/Thr protein kinase
Rfr18	554	22 (20%)	Cytoplasmic	Ser/Thr protein kinase
Rfr31	637	23 (18%)	Membrane	
Rfr17	682	27 (20%) ^b	Membrane	
Rfr25	820	18 (11%)	Cytoplasmic	UvrD
Rfr16	930	29 (16%) ^b	Membrane	

^aNumber in parentheses is the percentage of residues in the PRP predicted to be located in pentapeptide repeats.

^bInterruption in the RFR periodicity.

Figure 10. Predicted residue composition of the 35 PRPs in *Cyanothece* 51142. The PRPs are drawn sequentially following the order in Table 4 (increasing number of amino acid residues) using the following coloring pattern that is proportional to predicted composition: (red) pentapeptide repeat; (dark blue) N-terminal; (light blue) C-terminal; (white) other regions. The PRP with an asterisk is Rfr32.

the majority of the *Cyanothece* 51142 PRPs, especially in proteins that likely contain multiple domains (Vetting et al. 2006).

As discussed earlier, the Pfam database (Bateman et al. 2000) currently identifies 2110 proteins with pentapeptide repeat domains. A bioinformatics study of a smaller Pfam list of PRPs (1061) by Vetting et al. (2006) revealed that approximately half of these proteins were currently known, unique proteins. Out of these PRPs, the vast majority were in prokaryotes, and ~40% had an additional, non-Rfr, domain. The additional domains could be grouped into 20 categories, with only seven containing more than two members. The top three most populated domains were the WD40 β -transducin repeat (56), Ser/Thr protein kinase (11), and tetratricopeptide_1 repeat (11) (Vetting et al. 2006). Table 4 indicates that 15 out of the 35 *Cyanothece* PRPs are >50% nonpentapeptide repeat, and consequently, could also potentially contain an additional domain. A BLAST study (Altschul et al. 1990) of the 35 *Cyanothece* PRPs indicates that only four contain an additional, identifiable domain as listed in Table 4. Two are Ser/Thr kinases, one a DnaJ domain, and the fourth a UvrD/REP helicase. The latter domain catalyzes ATP dependent unwinding of double-stranded DNA to single-stranded DNA, and was the only domain not identified in the study by Vetting et al. (2006). Given that very little is known about the biological function of PRPs, it may be that the diversity of protein sequences attached to some of the other Rfr-folds may have novel, uncharacterized folds and functions. For example, only 14% of Rfr33 is composed of pentapeptide repeats, and this protein is homologous to RfrA, a protein with an Rfr-domain that has been associated with manganese uptake in *Synechocystis* (Chandler et al. 2003).

In addition to perhaps performing a unique biological function, protein sequences that often straddle an Rfr-fold may also serve an additional function as stabilizers of the

Rfr-fold. The top and bottom of the “naked” Rfr-fold contains exposed hydrogen donors and acceptors in position to form “edge-to-edge” β -bridges and β -sheets with another molecule. If the Rfr-fold really was naked, these ends could lead to edge-to-edge aggregation of the protein (Richardson and Richardson 2002). However, as shown in Figure 1B, the C-terminal of the Rfr-fold is capped by a pair of α -helices connected by a loop that sits on top of the edge of the last coil, nicely protecting three out of the four faces on the C-terminal edge of the Rfr-fold from forming edge-to-edge aggregates. At the N-terminal, the Rfr32 construct used for crystallization contains a large ~5-kDa polypeptide tag that could prevent the N-terminal from forming edge-to-edge aggregates. Untagged Rfr32 contained seven residues prior to the first pentapeptide repeat that could perform the same function, and interestingly, these molecules packed N-terminal-to-N-terminal in the crystal. Without this tag, native Rfr32 contains a 29-residue signal sequence that may form a similar function at the N-terminal before the protein reaches its destination in the thylakoid lumen. Richardson and Richardson (2002) observed that β -sheet edge protection was common in the structures of other β -helical proteins. Perhaps the C-terminal α -helices in Rfr32 have no biochemical function except to prevent the C-terminal of the Rfr-fold from aggregating. MfpA also contains a pair of α -helices at the C-terminal that may perform a similar role, especially when they associate with another MfpA molecule to form a head-to-head dimer. Note that 10 of the 35 PRPs identified in *Cyanothece* have few, if any, predicted non-Rfr-fold residues at the C terminus (Fig. 10). It will be interesting to see if these PRPs without a C-terminal sequence are monomers, dimers, or higher order aggregates in solution.

Conclusions

The Rfr-fold is a special subset in the right-handed parallel β -helix family of protein structures, with at least 16 right-handed parallel β -helices having been observed and listed in the SCOP structure database (Murzin et al. 1995). Like the Rfr-fold, these β -helices also have coils with spacing of ~4.8 Å (Jenkins and Pickersgill 2001). However, unlike the coils in the Rfr-fold where the sequence length of each coil is 20 residues, the sequence lengths of the coils in the other β -helices range from 30+ (Badger et al. 2005) to 12 (Liou et al. 2000) residues. Furthermore, even within the same β -helix, the length of the coil may vary in contrast to the consistent 20-residue length observed in each coil in the Rfr-fold. At least three different types of stacking occur in the interior of β -helices—aliphatic stacks, aromatic stacks, and polar stacks (Jenkins and Pickersgill 2001). Aromatic and aliphatic stacks are observed at the *i*th residue position

of the pentapeptide repeat in the Rfr-fold, while aliphatic stacks are observed in the $i - 2$ position. A major difference between the Rfr-fold and most of the other right-handed parallel β -helices is the number and length of the “faces” in the β -helix. The Rfr-fold contains four faces that vary by <1 Å in length, while the right-handed parallel β -helices have three or four faces with cross-sections that are triangular (Graether et al. 2000), square (Liou et al. 2000), rectangular (Badger et al. 2005), or even L-shaped (Emsley et al. 1996). While all right-handed parallel β -helices, including the Rfr-fold, contain parallel β -sheets, only the Rfr-fold contains linear stacked arrays of β -bridges aligned to form single-residue β -sheets. Interestingly, parallel β -sheets are more rigid than antiparallel β -sheets (Emberly et al. 2004), suggesting that the architecture of the Rfr-fold and all right-handed parallel β -helices may be especially sturdy, although the relatively low melting temperature determined for Rfr32 by CD spectroscopy contradicts this hypothesis. Amidst the diversity of shapes adopted by right-handed parallel β -helices, the Rfr-fold appears to be the only group of right-handed β -helices that may be readily predicted from the amino acid sequence. Analysis of the two available crystal structures of proteins containing pentapeptide repeats suggests that the structure of this sequence-identifiable right-handed β -helix, the Rfr-fold, is shaped by individual pentapeptide repeats adopting one of two turn motifs. These two turn motifs may be universal to the Rfr-fold in all 2110 PRPs in the Pfam database.

While the small family of right-handed parallel β -helix structures share many common features, there is also a lot of variation in the size and shape of these structures (Jenkins and Pickersgill 2001). Likewise, there is also variation in the known functions of these right-handed parallel β -helices, ranging from pectate lyases (Yoder et al. 1993) to hyperactive antifreeze proteins (Graether et al. 2000). On the other hand, because of the repetitive nature of the pentapeptide repeat sequence, it was predicted that the structures adopted by such tandem sequences would all be very similar (Bateman et al. 1998; Hegde et al. 2005). The second structure of a protein containing an Rfr-fold reported here, Rfr32, supports these predictions as there are striking similarities in overall architecture of the Rfr-fold in Rfr32 and MfpA. However, there are also differences, likely related to the sequential ordering of type II and type IV β -turns, which result in different twists to the β -helix and different surfaces exposed to the solvent. Differences on the solvent-exposed surface may also be introduced with “exceptions” to the general pentapeptide repeat motif. The Pfam definition of the pentapeptide repeat is A[D,N]LXX (Bateman et al. 2000); however, a more precise definition of the consensus sequence by Vetting

et al. (2006) is [S,T,A,V][D,N][L,F][S,T,R][G]. The latter definition is not exclusive, since the Rfr-fold can tolerate exceptions, especially with regard to solvent-exposed residues in the $i - 1$, $i + 1$, and $i + 2$ position in the pentapeptide repeat (see Fig. 2). Overall, such differences between Rfr-folds may result in variations to the biochemical functions of the Rfr-fold. Indeed, these differences are pronounced enough to suggest that Rfr32 may have a function different from the one proposed for MfpA.

There is convincing evidence that one biochemical function of PRPs expressed from bacterial plasmids is to provide resistance to fluoroquinolones and other antibiotics via a mechanism that involves DNA mimicry (Hegde et al. 2005; Vetting et al. 2006). These plasmid genes that convey antibiotic resistance likely developed from PRP genes present in chromosomal DNA as a special niche that originated secondarily to its primary role or function in cyanobacteria. However, little is known about the biochemical function of the pentapeptide repeat domain in these chromosomal PRP gene products, and nothing is known about their mechanism of action. Cyanobacteria are unique in the sheer number of proteins with predicted Rfr-domains. As observed in other cyanobacteria (Kieslebach et al. 1998), the 35 PRPs in *Cyanothece* 51142 are predicted to be located in all of the cellular compartments (Table 4), and only one of these compartments, the cytosol, contains DNA. These observations, sheer numbers, and disparate cellular locations argue for an important physiological function for PRPs (Kieslebach et al. 1998) in cyanobacteria that likely does not involve DNA mimicry. Further studies of the structure and biochemical function of proteins containing Rfr-folds are necessary in order to refine our emerging understanding of this intriguing family of proteins.

Material and methods

Annotation and analysis of the PRP genes from Cyanothece 51142

The PRP genes in *Cyanothece* were identified using the program HMMER (Eddy 1998) v2.3.2. The process involved searching the genomes of 18 cyanobacteria for pentapeptide repeats as defined in the Pfam database (Bateman et al. 2000) and then, through an iterative process, generating a cyanobacteria-specific HMM model for the pentapeptide repeat. This cyanobacteria-specific HMM model, that consisted of 12 pentapeptide repeats in the center of the alignment, was then used to search the newly sequenced genome of *Cyanothece*. Thirty-five PRP genes were identified.

The acronym PRP is used throughout the manuscript to describe proteins containing pentapeptide repeats. However, the PPR acronym for the pentapeptide repeat motif conflicts with PPR nomenclature already used to define pentatracheptide

repeat motifs in a large family of proteins in *Arabidopsis thaliana* (Small and Peteers 2000). Consequently, we are adopting the nomenclature first used by Chandler et al. (2003) to annotate the PRP genes from *Synechocystis* 6803, to annotate the PRP genes from *Cyanothece* 51142, repeated five-residues (Rfr). As a result, the 35 PRP genes from *Cyanothece* 51142 are annotated *Rfr1* through *Rfr35*, based on their sequential position in the chromosome.

Cloning, expression, and purification

The *Rfr32* gene minus the N-terminal 29 residues containing the signal peptide was amplified using the genomic DNA of *Cyanothece* sp. ATCC 51142 and the oligonucleotide primers 5'-ATCGAGGTCTCACATGGTCACTGGCTCCAGTGC-3' and 5'-TGACTGGTCTCCGAGCTATTGACATCGTAAGGACTCACGG-3' (Midland). The amplified *Rfr32* gene, corresponding to Rfr32 residues V30–Q167, was inserted into the NcoI/XhoI-digested expression vector pET30b (Novagen) such that a 43-residue tag containing six consecutive histidine residues was added to the N terminus of the gene product. The recombinant plasmid was transformed into *E. coli* BL21 (DE3) and methionine-auxotrophic B834 (DE3) cells (Novagen). The SeMet-substituted protein was expressed in the B834 (DE3) cells following an autoinduction protocol using minimal medium supplemented with 34 μ g/mL kanamycin, 30 μ g/mL chloramphenicol, and 200 μ g/mL of each individual amino acid except for the inclusion of 10 μ g/mL methionine and 125 μ g/mL selenomethionine. After autoinduction at 25°C, the cells were harvested by centrifugation and frozen at 193 K. Thawed cells were resuspended in 32 mL lysis buffer (0.3 M NaCl, 50 mM sodium phosphate, 10 mM imidazole at pH 8.0) and brought to 0.2 μ M in PMSF prior to three passes through a French press (SLM Instruments). Following sonication for 30 sec, the cell debris was spun at 25,000g for 1.5 h. After passage through a 0.45- μ m syringe filter, the supernatant was loaded onto a 20 mL Ni-NTA affinity column (Qiagen) and washed stepwise with 50 mL of buffer (0.3 M NaCl, 50 mM sodium phosphate at pH ~8.0) containing increasing concentrations of imidazole (5, 10, 20, 50, and 250 mM). The fraction containing Rfr32 eluted primarily with the 250 mM imidazole fraction (~15 mg/L medium). If the N-terminal 43-residue tag was to be removed, the protein fraction was dialyzed (8-kDa molecular-weight cutoff) overnight at 277 K in 4 L of enterokinase buffer (50 mM NaCl, 20 mM Tris-HCl at pH 7.4). Following volume reduction to ~1 mL (Amicon Centriprep-10), the N-terminal tag was removed at room temperature with ~1 unit of enterokinase (GenScript Corp.) per 2 mg of protein and 1 μ L of 1.0 M CaCl₂. Protein with and without the N-terminal tag was then purified on a Superdex75 HiLoad 26/60 column (Amersham Pharmacia Biotech) that simultaneously exchanged it into the buffer used for crystallization (100 mM NaCl, 20 mM Tris, 1.0 mM dithiothreitol at pH 7.1). Using a flow rate of 2.5 mL/min, tagged and untagged Rfr32 eluted with retention times of 70 and 80 min, respectively, values characteristic of protein with calculated monomeric molecular weights of 19,563 Da and 14,848 Da, respectively (enterokinase cleaved Rfr32 contains an alanine prior to the starting methionine). SDS-PAGE showed the proteins to be >98% pure.

Circular dichroism spectroscopy

Circular dichroism data were obtained on an Aviv Model 62DS spectropolarimeter calibrated with an aqueous solution of

ammonium d-(+)-camphorsulfonate. Measurements were made on two solutions of untagged Rfr32 in buffer containing 100 mM NaCl, 20 mM potassium phosphate, 1 mM DTT (pH 7.4). A thermal denaturation curve for Rfr32 was obtained on a 20 μ M solution in a quartz cell of 0.1-cm path length by recording the ellipticity at 216 nm in 2.5°C intervals from 10°C to 80°C. A far-UV wavelength spectrum between 190 nm and 250 nm was recorded on a 260 μ M solution of Rfr32 in a quartz cell of 0.02-cm path length at 25°C. The spectrum was the result of averaging three consecutive scans with a bandwidth of 1.0 nm and a time constant of 1.0 sec. The wavelength spectrum was processed by first subtracting a blank spectrum followed by baseline correction and noise reduction.

Crystallization methods

Vapor-diffusion crystallization trials using hanging drops were set up on tagged and untagged Rfr32 at room temperature (~22°C) using screens from Hampton Research. Two differently shaped diffraction quality crystals were grown under two different conditions. Long, hollow tubes with a hexagonal face that adopted a trigonal crystalline lattice were harvested ~6 d after mixing 2 μ L of tagged protein (~3 mg/mL) with 2 μ L of buffer containing 30% PEG 1500 and 10% glycerol. Both native and SeMet-substituted crystals were grown under these conditions. Bipyramidal-shaped crystals that adopted a tetragonal crystalline lattice were harvested 2–3 d after mixing 2 μ L of untagged protein (~3 mg/mL) with 2 μ L of buffer containing 30% PEG 4000, 0.1 M Tris-HCl, 0.2 M MgCl₂, 10% glycerol (pH 8.5). Under these latter conditions only SeMet-substituted crystals were grown.

Structure determination and refinement

X-ray diffraction data for the Rfr32 crystals were collected at the National Synchrotron Light Source (NSLS) at Brookhaven National Laboratory. Data were collected at the X29A beamline using an ADSC Q315 CCD detector. A three-wavelength MAD data set at 2.3 Å resolution (SeMet-labeled) and a native data set at 2.1 Å resolution (unlabeled) were collected on trigonal crystals of tagged Rfr32. A native data set at 2.3 Å resolution (SeMet-labeled) was collected on a tetragonal crystal of untagged Rfr32. All data collection statistics are summarized in Table 1. The images were integrated and scaled with HKL2000 (Otwinowski and Minor 1997). The heavy atom sites in the selenium-labeled Rfr32 in the trigonal crystal lattice were determined using the SHELX program suite (Schneider and Sheldrick 2002; Sheldrick 2002, 2003) and HKL2MAP (Pape and Schneider 2004). The peak wavelength data of the MAD data set, which contained useful anomalous signal out to 2.8 Å resolution, was used together with the program SOLVE (Terwilliger and Berendzen 1999) (www.solve.lanl.gov) to produce an electron density map at 2.8 Å resolution. Fragments of the structure were built automatically into the 2.8 Å resolution map using RESOLVE (<http://www.csb.yale.edu/userguides/datamanip/solve/html/>) (Terwilliger 1999, 2000, 2001a,b), and the remainder of the structure was built manually using XtalView/Xfit (McRee 1999). The model was further iteratively refined using a native data set extending out to 2.1 Å resolution and the refine.inp algorithm in CNS (<http://cns.csb.yale.edu/v1.0/>) (Brünger et al. 1998) employing the maximum likelihood target using amplitudes. The stereochemical quality of the final model was assessed using the program PROCHECK (Laskowski et al.

1993) and MolProbity (<http://molprobity.biochem.duke.edu>) (Lovell et al. 2003). Once the structure in the trigonal crystal form of tagged Rfr32 was refined, it was used as a search model for molecular replacement using the program MOLREP (Vagin and Teplyakov 1997) to solve the structure in the tetragonal crystal lattice of untagged Rfr32 using the native data set at 2.3 Å resolution. The structure refinement statistics are also given in Table 1. The Rfr32 structures refined in the trigonal and tetragonal crystal forms have been submitted to the Protein Data Bank (PDB ID 2F3L and 2G0Y, respectively).

Acknowledgments

This work is part of an EMSL Membrane Biology Scientific Grand Challenge project at the W.R. Wiley Environmental Molecular Sciences Laboratory, a national scientific user facility sponsored by U.S. Department of Energy's Office of Biological and Environmental Research (BER) program located at Pacific Northwest National Laboratory (PNNL). PNNL is operated for the U.S. Department of Energy by Battelle. We thank PNNL scientists Drs. Theresa A. Ramelot and John R. Cort for their assistance with computational software, and the X29A beam line scientists at the National Synchrotron Light Source at Brookhaven National Laboratory for their assistance. Support for beamline X29A at the National Synchrotron Light Source comes principally from the Offices of Biological and Environmental Research and of Basic Energy Sciences of the U.S. Department of Energy, and from the National Center for Research Resources of the National Institutes of Health.

References

- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. 1990. Basic alignment research tool. *J. Mol. Biol.* **215**: 403–410.
- Badger, J., Sauder, J.M., Adams, J.M., Antonysamy, S., Bain, K., Bergseid, M.G., Buchanan, S.G., Buchanan, M.D., Batiyenko, Y., Christopher, J.A., et al. 2005. Structural analysis of a set of proteins resulting from a bacterial genomics project. *Proteins* **60**: 787–796.
- Bateman, A., Murzin, A.G., and Teichmann, S.A. 1998. Structure and distribution of pentapeptide repeats in bacteria. *Protein Sci.* **7**: 1477–1480.
- Bateman, A., Birney, E., Durbin, R., Eddy, S.R., Howe, K.L., and Sonnhammer, E.L. 2000. The Pfam protein families database. *Nucleic Acids Res.* **28**: 263–266.
- Black, K., Buikema, W.J., and Haselkorn, R. 1995. The *hglK* gene is required for localization of heterocyst-specific glycolipids in cyanobacterium *Anabaena* sp. strain PCC 7120. *J. Bacteriol.* **177**: 6440–6448.
- 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., et al. 1998. Crystallography and NMR System (CNS): A new software suite for macromolecular structure determination. *Acta Crystallogr.* **D54**: 905–921.
- Buchko, G.W., Hess, N.J., Bandaru, V., Wallace, S.S., and Kennedy, M.A. 2000. Spectroscopic studies of zinc(II)- and cobalt(II)-associated *Escherichia coli* formamidopyrimidine-DNA glycosylase: Extended X-ray absorption fine structure evidence for a metal-binding domain. *Biochemistry* **40**: 12441–12449.
- Chan, A.W., Hutchinson, E.G., Harris, D., and Thornton, J.M. 1993. Identification, classification, and analysis of β -bulges in proteins. *Protein Sci.* **2**: 1574–1590.
- Chandler, L.E., Bartsevich, V.V., and Pakrasi, H.B. 2003. Regulation of manganese uptake in *Synechocystis* 6803 by RfrA, a member of a novel family of proteins containing a repeated five-residue domain. *Biochemistry* **42**: 5508–5514.
- Chang, J.-F., Hall, B.E., Tanny, J.C., Moazed, D., Filman, D., and Ellenberger, T. 2003. Structure of the coiled-coil dimerization motif of Sir4 and its interaction with Sir3. *Structure* **11**: 637–649.
- Drlica, K. and Malik, M. 2003. Fluoroquinolones: Action and resistance. *Curr. Top. Med. Chem.* **3**: 249–282.
- Eddy, S.R. 1998. Profile hidden Markov models. *Bioinformatics* **14**: 755–763.
- Emberly, E.G., Mukhopadhyay, R., Tang, C., and Wingreen, N.S. 2004. Flexibility of β -sheets: Principal component analysis of database protein structures. *Proteins* **55**: 91–98.
- Emsley, P., Charles, I.G., Fairweather, N.F., and Isaacs, N.W. 1996. Structure of *Bordetella pertussis* virulence factor P69 pertactin. *Nature* **381**: 90–92.
- Eswar, N. and Ramakrishnan, C. 1999. Secondary structures without backbone: An analysis of backbone mimicry by polar side chains in protein structure. *Protein Eng.* **12**: 447–455.
- Garrido, M.C., Herrero, R., Kolter, R., and Moreno, F. 1988. The export of the DNA-replication inhibitor Microsin B17 provides immunity for the host-cell. *EMBO J.* **7**: 1853–1862.
- Gomi, M., Sonoyama, M., and Mitaku, S. 2004. High performance system for signal peptide prediction: SOSUlsignal. *Chem-Bio Info. J.* **4**: 142–147.
- Graether, S.P., Kuiper, M.J., Gagne, S.M., Walker, V.K., Jia, Z., Sykes, B.D., and Davies, P.L. 2000. β -Helix structure and ice-binding properties of a hyperactive antifreeze protein from an insect. *Nature* **406**: 325–328.
- Hegde, S.S., Vetting, M.W., Roderick, S.L., Mitchenall, L.A., Maxwell, A., Takiff, H.E., and Blanchard, J.S. 2005. A fluoroquinolone resistance protein from *Mycobacterium tuberculosis* that mimics DNA. *Science* **308**: 1480–1483.
- Holm, L. and Sander, C. 1998. Touring protein fold space with Dali/FSSP. *Nucleic Acids Res.* **26**: 316–319.
- Holzwarth, G.M. and Doty, P. 1965. The ultraviolet circular dichroism of polypeptides. *J. Am. Chem. Soc.* **87**: 218–228.
- Hunter, C.A., Singh, J., and Thornton, J.M. 1991. π - π Interactions—The geometry and energetics of phenylalanine phenylalanine interactions in proteins. *J. Mol. Biol.* **218**: 837–846.
- Hutchinson, E.G. and Thornton, J.M. 1994. A revised set of potentials for β -turn formation in proteins. *Protein Sci.* **3**: 2207–2216.
- Izuta, S., Shimada, N., Kitigawa, M., Suzuki, M., Kojima, K., and Yoshida, S. 1992. Inhibitory effects of triphosphate derivatives of oxetanocin G and related compounds on eukaryotic and viral DNA polymerases and human immunodeficiency virus reverse transcriptase. *J. Biochem.* **112**: 81–87.
- Jenkins, J. and Pickersgill, R. 2001. The architecture of parallel β -helices and related folds. *Prog. Biophys. Mol. Biol.* **77**: 111–175.
- Kabsch, W. and Sander, C. 1983. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonding and geometrical features. *Bio-polymers* **22**: 2577–2637.
- Kieslebach, T., Mant, A., Robinson, C., and Schroder, W.P. 1998. Characterization of an *Arabidopsis* cDNA encoding a thylakoid lumen protein related to a novel “pentapeptide repeat” family of proteins. *FEBS Lett.* **428**: 241–244.
- Kwok, S.C. and Hodges, R.S. 2003. Clustering of large hydrophobes in the hydrophobic core of two-stranded α -helical coiled-coils controls protein folding and stability. *J. Biol. Chem.* **278**: 35248–35254.
- Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. 1993. PROCHECK: A program for checking the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**: 283–291.
- Liou, Y.C., Tocilj, A., Davies, P.L., and Jia, Z. 2000. Mimicry of ice structure by surface hydroxyls and water of a β -helix antifreeze protein. *Nature* **406**: 322–324.
- Lovell, S.C., Davis, I.W., Arendall III, W.B., de Bakker, P.I.W., Word, J.M., Prisant, M.G., Richardson, J.S., and Richardson, D.C. 2003. Structure validation by $C\alpha$ geometry: ϕ , ψ , and $C\beta$ deviation. *Proteins* **50**: 437–450.
- Manning, C.M., Illangasek, M., and Woody, R.W. 1988. Circular dichroism studies of distorted α -helices, twisted β -sheets and β -turns. *Biophys. Chem.* **31**: 77–86.
- McRee, D.E. 1999. Xtalview/Xfit—A versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.* **125**: 156–165.
- Montero, C., Mateu, G., Rodriguez, R., and Takiff, H.E. 2001. Intrinsic resistance of *Mycobacterium smegmatis* to fluoroquinolones may be influenced by new pentapeptide protein MfpA. *Antimicrob. Agents Chemother.* **45**: 3387–3392.
- Morais Cabral, J.H., Jackson, A.P., Smith, C.V., Shikotra, N., Maxwell, A., and Liddington, R.C. 1997. Crystal structure of the breakage-reunion domain of DNA gyrase. *Nature* **388**: 903–906.
- Morita, M., Tomita, K., Ishizawa, M., Tagaki, K., Kawamura, F., Takahashi, H., and Morino, T. 1999. Cloning of oxetanocin A biosynthetic and resistance genes that reside on a plasmid of *Bacillus magisterium* strain NK84-0128. *Biosci. Biotechnol. Biochem.* **63**: 563–566.
- Murzin, A.G., Brenner, S.E., Hubbard, T., and Chothia, C. 1995. SCOP: A structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* **247**: 536–540.
- Nordmann, P. and Poirel, L. 2005. Emergence of plasmid-mediated resistance to quinolones in *Enterobacteriaceae*. *J. Antimicrob. Chemother.* **56**: 463–469.

- Otwinowski, Z. and Minor, W. 1997. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**: 307–326.
- Pabasik Jr., N., Fleming, P.J., and Rose, G.D. 2005. Hydrogen-bonded turns in proteins: The case for a recount. *Protein Sci.* **14**: 2910–2914.
- Pape, T. and Schneider, T.R. 2004. HKL2MAP: A graphical user interface for phasing with SHELX programs. *J. Appl. Crystallogr.* **37**: 843–844.
- Perczel, A., Park, K., and Fasman, G.D. 1992. Deconvolution of the circular dichroism spectra of proteins: The circular dichroism spectra of the antiparallel β -sheet in proteins. *Proteins* **13**: 57–69.
- Pierrat, O.A. and Maxwell, A. 2005. Evidence of the role of DNA strand passage in the mechanism of action of Microcin B17 on DNA gyrase. *Biochemistry* **44**: 4204–4215.
- Richardson, J.S. 1981. The anatomy and taxonomy of protein structure. *Adv. Protein Chem.* **34**: 167–339.
- Richardson, J.S. and Richardson, D.C. 2002. Natural β -sheet proteins use negative design to avoid edge-to-edge aggregation. *Proc. Natl. Acad. Sci.* **99**: 2754–2759.
- Richardson, J.S., Getzoff, E.D., and Richardson, D.C. 1978. The β -bulge: A common small unit of nonrepetitive protein structure. *Proc. Natl. Acad. Sci.* **75**: 2574–2578.
- Schneider, T.R. and Sheldrick, G.M. 2002. Substructure solution with SHELXD. *Acta Crystallogr. D Biol. Crystallogr.* **58**: 1772–1779.
- Sheldrick, G.M. 2002. Macromolecular phasing with SHELXE. *Z. Kristallogr.* **217**: 644–650.
- Sheldrick, G.M. 2003. *SHELXC*. Göttingen University, Germany.
- Shepherd, A.J., Gorse, D., and Thornton, J.M. 1999. Prediction of the location and type of β -turns in proteins using neural networks. *Protein Sci.* **8**: 1045–1055.
- Small, I.D. and Peteers, N. 2000. The PPR motif—A TPR-related motif prevalent in plant organellar proteins. *Trends Biochem. Sci.* **25**: 46–47.
- Smith, J.A. and Pease, L.G. 1980. Reverse turns in peptides and proteins. *CRC Crit. Rev. Biochem.* **8**: 315–399.
- Terwilliger, T.C. 1999. Reciprocal-space solvent flattening. *Acta Crystallogr. D Biol. Crystallogr.* **55**: 1863–1871.
- Terwilliger, T.C. 2000. Maximum-likelihood density modification. *Acta Crystallogr. D Biol. Crystallogr.* **56**: 965–972.
- Terwilliger, T.C. 2001a. Map-likelihood phasing. *Acta Crystallogr. D Biol. Crystallogr.* **57**: 1763–1775.
- Terwilliger, T.C. 2001b. Maximum-likelihood density modification using pattern recognition of structural motifs. *Acta Crystallogr. D Biol. Crystallogr.* **57**: 1763–1775.
- Terwilliger, T.C. and Berendzen, J. 1999. Automated MAD and MIR structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **55**: 849–861.
- Tran, J.H. and Jacoby, G.A. 2002. Mechanism of plasmid-mediated quinolone resistance. *Proc. Natl. Acad. Sci.* **99**: 5638–5642.
- Tran, J.H., Jacoby, G.A., and Hooper, D.C. 2005. Interaction of the plasmid-encoded quinolone resistance protein QnrA with *Escherichia coli* topoisomerase IV. *Antimicrob. Agents Chemother.* **49**: 3050–3052.
- Vagin, A. and Teplyakov, A. 1997. MOLREP: An automated program for molecular replacement. *J. Appl. Crystallogr.* **30**: 1022–1025.
- van Gent, D.C., Hoeijmakers, J.H.J., and Kanaar, R. 2001. Chromosomal stability and the DNA double-stranded break connection. *Nat. Rev. Genet.* **2**: 196–206.
- Vetting, M.W., Hegde, S.S., Fajardo, J.E., Fiser, A., Roderick, S.L., Takiff, H.E., and Blanchard, J.S. 2006. Pentapeptide repeat proteins. *Biochemistry* **45**: 1–10.
- Vizan, J.L., Hernandez-Chico, C., del Castillo, I., and Moreno, F. 1991. The peptide antibiotic Microcin B17 induces double-strand cleavage of DNA mediated by *E. coli* DNA gyrase. *EMBO J.* **10**: 467–476.
- Wilmot, C.M. and Thornton, J.M. 1988. Analysis and prediction of the different types of β -turn in proteins. *J. Mol. Biol.* **203**: 221–232.
- Wilmot, C.M. and Thornton, J.M. 1990. β -Turns and their distortions: A proposed new nomenclature. *Protein Eng.* **3**: 479–493.
- Woody, R.W. 1974. *Studies of theoretical circular dichroism of polypeptides: Contributions of β -turns*, pp. 338–360. John Wiley & Sons, New York.
- Yoder, M.D., Keen, N.T., and Jurnak, F. 1993. New domain motif: The structure of pectate lyase C, a secreted plant virulence factor. *Science* **260**: 1503–1507.