

Crystallization Notes

Insights into the structural variation between pentapeptide repeat proteins—Crystal structure of Rfr23 from *Cyanothece* 51142

Garry W. Buchko^{a,*}, Howard Robinson^b, Himadri B. Pakrasi^c, Michael A. Kennedy^{a,*,1}

^a Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA 99352, USA

^b Biology Department, Brookhaven National Laboratory, Upton, NY 11973-5000, USA

^c Department of Biology, Washington University, St. Louis, MI 63130-4899, USA

Received 23 July 2007; received in revised form 1 November 2007; accepted 9 November 2007

Available online 26 December 2007

Abstract

Cyanothece sp. PCC 51142 contains 35 pentapeptide repeat proteins (PRPs), proteins that contain a minimum of eight tandem repeated five-residues (Rfr) of the general consensus sequence A[N/D]LXX. Published crystal structures of PRPs show that the tandem pentapeptide repeats adopt a type of right-handed quadrilateral β -helix called an Rfr-fold. To characterize how structural features of Rfr-folds might vary with different amino acid sequences, the crystal structure of *Cyanothece* Rfr23 (174 residues) was determined at 2.4 Å resolution. The structure is dominated by an Rfr-fold capped at the N-terminus with a nine-residue α -helix (M26*–E34). The Rfr-fold of Rfr23 contains four structural features previously unobserved in Rfr-folds. First, Rfr23 is composed entirely of type II β -turns. Second, the pentapeptide repeats are not consecutive in the primary amino acid sequence. Instead, Rfr23 contains 24-residues protruding outside one corner of the first complete N-terminal coil of the Rfr-fold (L56–P79) (24-residue insertion). Third, a disulfide bond between C39 and C42 bridges the β -turn between the first and second pentapeptide repeats in the first coil (disulfide bracket). NMR spectroscopy indicates that the reduction of the disulfide bracket with the addition of DTT destroys the entire Rfr-fold. Fourth, a single-residue perturbs the Rfr-fold slightly in the last coil between the C-terminal two pentapeptide repeats (single-residue bulge). © 2007 Elsevier Inc. All rights reserved.

Keywords: Cyanobacteria; Disulfide bond; Right-handed parallel β -helix; Edge-to-edge aggregation; β -Turns; Repeated five-residues

1. Introduction

Since Bateman et al. (1998) first reported a novel family of proteins containing consecutive pentapeptide repeats the number of pentapeptide repeat proteins (PRPs) listed in the Pfam data bank (Pfam00805) (Bateman et al., 2000) has escalated from a little over 1000 in late 2005 to nearly 3500 today. PRPs are identified by tandem repeated five-residues (Rfr) of the general consensus sequence A[D/N]LXX. While present in eukaryotes, PRPs are found pri-

marily in prokaryotes and especially in cyanobacteria (Vetting et al., 2006). Thirty-five PRPs have been identified in *Cyanothece* sp. PCC 51142 and they are predicted to be located in the lumen/periplasm, plasma membrane, and the cytosol (Buchko et al., 2006a). The combination of sheer numbers and diverse cellular locations argues that PRPs have important, as yet unknown, biochemical and physiological functions in cyanobacteria (Kieslebach et al., 1998).

Despite the vast number of PRPs listed in the Pfam database, only three crystal structures have been published. The first was MfpA (Rv3361c), a 183-residue protein associated with mycobacterial fluoroquinolone resistance in *Mycobacterium tuberculosis* (Hegde et al., 2005; Vetting et al., 2006). The second was Rfr32, a 167-residue protein in *Cyanothece* 51142 (Buchko et al., 2006a). The third was Np275, a 98-residue protein from *Nostoc punctiforme*

* Corresponding authors. Fax: +1 509 376 2303.

E-mail addresses: garry.buchko@pnl.gov (G.W. Buchko), michael.kennedy@muohio.edu (M.A. Kennedy).

¹ Present address: Department of Chemistry and Biochemistry, Miami University, Oxford, OH 45056, USA.

(Vetting et al., 2007). MfpA, Rfr32, and Np275 contain 30, 21, and 17 consecutive pentapeptide repeats, respectively, and their crystal structures reveal that these repeated five-residues adopt a novel type of right-handed quadrilateral β -helix, an Rfr-fold, reminiscent of a square tower with four distinct faces (Buchko et al., 2006a; Hegde et al., 2005; Vetting et al., 2006). Each pentapeptide repeat occupies one face of the Rfr-fold with four consecutive repeats completing a 20-residue coil that rises ~ 4.8 Å with every revolution. The tower-like feature is produced by a stacking of the coils and is stabilized by a network of inter-coil and intra-coil hydrogen bonds and hydrophobic side chain interaction in the interior of the tower. The coils themselves are composed of two distinct, four-residue type II and type IV β -turns that may be universal motifs of the Rfr-fold in all PRPs (Buchko et al., 2006a).

The biochemical function of *MsMfpA* and other PRPs expressed from bacterial plasmids may be to provide resistance to fluoroquinolones and other antibiotics by acting as a DNA mimic (Drlica and Malik, 2003; Hegde et al., 2005). Fluoroquinolones kill cells by generating double-strand DNA break after binding to the complex formed between DNA and DNA gyrase or DNA topoisomerase IV (Drlica and Malik, 2003). MfpA has a predominately electronegative surface potential, similar in size and shape to B-form DNA, that can be modeled onto the electropositive DNA-binding surface of DNA gyrase (Hegde et al., 2005). By binding to DNA gyrase or DNA topoisomerase IV, MfpA prevents the normal DNA substrate from binding, an event that would be lethal in the presence of fluoroquinolones. While there is compelling evidence that antibiotic resistance is a key biochemical function of PRP expressed from bacterial plasmids, the biochemical function of PRPs expressed from chromosomal DNA is not known. *Cyanothece* Rfr32 was originally targeted for study to gain further structure-based insights into the functions of PRPs in cyanobacteria. Given the repetitive nature of the pentapeptide repeat sequence, it was not surprising to observe that the overall architecture adopted by the tandem five-residue repeats was similar in *MsMfpA* and Rfr32, as predicted (Bateman et al., 1998; Hegde et al., 2005). However, differences were also observed between both Rfr-folds, likely related to the sequential ordering of type II and type IV β -turns, which result in different twists to the Rfr-fold and different surfaces exposed to the solvent. Most importantly perhaps, no uniform electronegative charge distribution on the solvent exposed surface was observed suggesting that perhaps the major role of the Rfr-fold in Rfr32 was not to act as a DNA mimic. Indeed, the third PRP structure, Np275, adds support to the latter conclusion (Vetting et al., 2007). To further test the DNA mimic hypothesis, to gather additional information on the effect of sequence variation on structural features of the Rfr-fold, and to gain insights into the molecular function of PRPs in cyanobacteria, we have determined the crystal structure of a second PRP from *Cyanothece* 51142, Rfr23.

2. Structure determination and refinement

Details of the methods used to express, purify, and crystallize selenomethione labeled Rfr23 (T27-D174) were previously reported (Buchko et al., 2006b). For NMR spectroscopy, a similar approach was used to obtain uniformly ^{15}N - and ^{15}N -, ^{13}C -labeled protein using minimal medium (Miller) containing $^{15}\text{NH}_4\text{Cl}$ (1 mg/mL) and $^{15}\text{NH}_4\text{Cl}$ (1 mg/mL)/[$^{13}\text{C}_6$]-D-glucose (2 mg/mL), respectively. Dithiothreitol (DTT) was omitted from the size exclusion chromatography buffer in purifying Rfr23 for NMR spectroscopy.

All NMR data were collected on 1–2 mM samples in buffer with and without 2 mM DTT containing 50 mM NaCl, 20 mM Tris-HCl, pH 7.4 (25 °C) using Varian Inova-900, -800, and -600 spectrometers equipped with triple resonance probes and pulse field gradients. Pulse sequences for the two-dimensional HSQC and all three-dimensional experiments (HSQC-NOESY, HNCACB) were from Protein Pack. To directly show the effect that DTT had on Rfr23, a ^1H - ^{15}N HSQC spectra was collected on a Rfr23 sample prepared in the absence of DTT (–DTT) and then recollected immediately after making the sample 2 mM in DTT (+DTT) directly in the same NMR tube. Overall rotational correlation times (τ_c) were estimated for Rfr23 (\pm DTT) from backbone amide ^{15}N $T_{1\rho}/T_1$ ratios (Szyperski et al., 2002).

X-ray diffraction data for “blade-like” Rfr23 crystals were collected at the National Synchrotron Light Source (NSLS) at Brookhaven National Laboratory at the X29A beamline using an ADSC Q315 CCD detector (Buchko et al., 2006b). A nearly complete SAD diffraction data sets was collected to 2.4 Å resolution, with the details of the data collection and processing reported previously (Buchko et al., 2006b). Using the SAD data set, approximately 90% of the structure was built automatically using the program RESOLVE (Terwilliger, 1999, 2000, 2001a,b) with the remainder built manually using XtalView/Xfit (McRee, 1999). The model was iteratively refined using the refine.inp algorithm in CNS (Brünger et al., 1998) employing the maximum likelihood target using amplitudes. As a final check on the model the stereochemical quality was assessed using the program PROCHECK (Laskowski et al., 1996) and MolProbity (Lovell et al., 2003) and any conflicts addressed. MolProbity analysis indicated that the overall protein geometry of the final model ranked in the 96th percentile (MolProbity score of 1.90) where the 100th percentile is best among structures of comparable resolution. The clash score for “all-atoms” was 12.49 corresponding to an 89th percentile ranking for structures of comparable resolution. Overall, the MolProbity assessment indicated that the final model was a high quality representation of the crystal structure of Rfr23, a conclusion corroborated by PROCHECK analyses. Note that the large percentage of residues in additionally allowed regions (26%) reflects properties of the Rfr-fold (Buchko et al., 2006a; Hegde et al., 2005; Vetting et al., 2007, 2006). The atomic coordi-

Table 1
Summary of data collection and structure refinement statistics for Rfr23^a

<i>Data collection</i>	
Space group	I4 ₁
Unit-cell parameters (Å, °)	$a = 106.42$, $b = 106.42$, $c = 52.43$ $\alpha = \beta = \gamma = 90$
Molecules/asym. unit	1.0
Matthews coeff. (Å ³ /kDa)	3.9
Percent solvent (%)	68.5
X-ray source	X29
Temperature (K)	100
Resolution limits (Å)	30.0–2.4 (2.49–2.40)
Detector distance (mm)	300
Exposure time (s)	3
Oscillation angle (°)	0.5
No. images	720
Total angle (°)	360
Mosaicity (°)	0.31–0.61
Wavelength (Å)	0.9792
No. unique reflections	24422 (2235)
Redundancy	6.7 (5.3)
R _{sym} I (%) [†]	7.9 (21.2)
Completeness (%)	96.9 (83.3)
I/σ(I)	29.6 (6.0)
<i>Structure refinement</i>	
PDB ID	2O6W
R _{conv} (%) [‡]	19.7
R _{free} (%) [§]	22.8
Protein atoms	1128
Solvent molecules	62
Ligands	1 (As)
Protein residues	123
Average B value, main chain atoms (Å ²)	41.3
Average B value, side chain atoms (Å ²)	45.4
Average B value, total protein atoms (Å ²)	43.3
Average B value, solvent atoms (Å ²)	48.8
Rms deviations from ideal bond lengths (Å)	0.006
Rms deviations from ideal bond angles (°)	1.04
<i>Ramachandran statistics</i>	
Most favored	73%
Additionally allowed	26%
Generously allowed	1%

^a Values in parenthesis are for the highest resolution shell (2.49–2.40 Å).

[†] $R_{\text{sym}}I = \frac{\sum_{hkl} \sum_j |I_j(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_j I_j(hkl)}$, where $I_j(hkl)$ is the intensity of the j th symmetry-related reflection of $I(hkl)$ and $\langle I(hkl) \rangle$ is its average.

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

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

nates for the crystal structure of Rfr23 (A25*-D174) have been deposited in the RCSB Protein Data Bank (2O6W) and a summary of the structure refinement statistics are given in Table 1.

3. Overall topology of the three-dimensional structure of Rfr23

Fig. 1A is a cartoon representation of the crystal structure of Rfr23 with a surface representation in Fig. 1B showing an electrostatic projection of the solvent accessible

surface. The structure is dominated by 23 five-residue repeats that adopt an Rfr-fold (Buchko et al., 2006a; Hegde et al., 2005; Vetting et al., 2006), a unique type of right-handed parallel β -helix (Jenkins and Pickersgill, 2001). The Rfr-fold consists of four consecutive pentapeptide repeats that form a nearly “square”, quadrilateral unit called a coil (Jenkins and Pickersgill, 2001; Vetting et al., 2006; Yoder et al., 1993) that stack on top of one another to form a four-faced “tower” (Face 1 through Face 4) where each pentapeptide repeat on a single coil occupies one face of the tower. The β -helix of Rfr23 contains five and three-quarter coils (labeled C1 through C6) (Buchko et al., 2006b). Resting on the N-terminal of the Rfr-fold of Rfr23 is a nine-residue α -helix (M26*-E34) connected to the first pentapeptide repeat by two residues. The α -helix is amphipathic (De Grado et al., 1981) with a predominantly hydrophilic side exposed to the solvent and a hydrophobic side that interacts with the hydrophobic side chains in the interior of the Rfr-fold. Termini at both ends of a “naked” Rfr-fold contain exposed hydrogen donors and acceptors in position to form “edge-to-edge” β -bridges and β -sheets with another molecule, leading to edge-to-edge aggregation of the protein (Richardson and Richardson, 2002). A role of the N-terminal α -helix may be to prevent such edge-to-edge aggregation from occurring (Buchko et al., 2006a; Richardson and Richardson, 2002).

The amino acid sequence of Rfr23 is shown in Fig. 2 with the residues aligned according to position in the coils and faces of the Rfr-fold as determined from the crystal structure. Reading the sequence from the N- towards the C-terminal, the center residue of each pentapeptide repeat is designated i with the preceding residues labeled $i - 1$ and $i - 2$ and the following residues labeled $i + 1$ and $i + 2$. As observed with the other PRP crystal structures (Buchko et al., 2006a; Hegde et al., 2005), the side chains of the $i - 2$ and i residues in Rfr23 all point towards the interior of the tower and pack the middle of the Rfr-fold. In Rfr23 the i th residues are almost exclusively Leu (19/23) with the exceptions observed in the first and last coil. 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 smaller, often hydrophobic group (10/23 are Ala in Rfr23). As observed in the crystal structures of MfpA, Rfr32, and Np275, no water molecules are observed in the interior of the Rfr-fold of Rfr23.

Also, as observed in the Rfr-fold of the three other reported PRP structures, the side chains of the $i - 1$, $i + 1$, and $i + 2$ residues of Rfr23 all point away from the interior of the tower and form the exterior, solvent exposed surface of the Rfr-fold. Fig. 2 shows that the majority of these exteriorly directed residues are hydrophilic, in contrast to the hydrophobic side chains in the interior. However, as illustrated in Fig. 2 (grey and underlined), a few hydrophobic solvent exposed side chains are present on Faces 3, 2, and especially 1 that form small hydrophobic “islands” on the hydrophilic surfaces. One consequence of these dispersed, small hydrophobic islands is that there

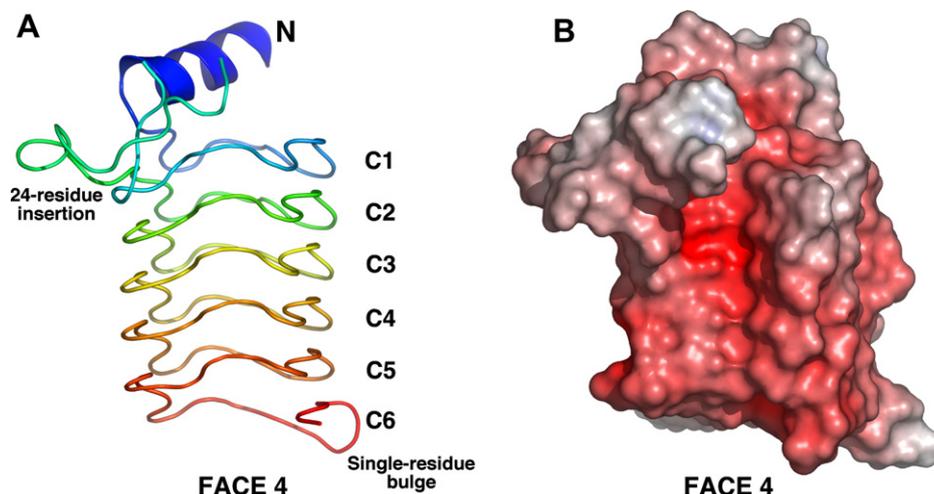


Fig. 1. (A) Cartoon representation of the structure of Rfr23 (2O6W) colored spectrally (red to violet) from the C-terminus. The structure is dominated by an Rfr-fold composed of nearly six complete coils, C1–C6. At the corner of Face 4 and Face 1, between coils C1 and C2, is a 24-residue insertion. In the terminal coil, C6, about the corner of Face2 and Face 3, is a small perturbation to the Rfr-fold labeled a “single-residue bulge”. The single-residue bulge is more visible in the top view of the structure shown in Fig. 4. (B) Approximately the same orientation as in (A) displaying the solvent accessible electrostatic surface on Face 4. Electrostatic projections are drawn (Pymol) at a level of +11 (blue) and –11 (red) kT/e , where k is the Boltzmann’s constant, T is the absolute temperature, and e is the magnitude of the electron charge.

	Face1	Face2	Face3	Face4	Coil
25*			AMTNNLYRLELE		36
37	REC <u>V</u> G	CNLEG	VNLPR	ENFG*	55 C1
80	VDLTR	ANLSN	ANLYQ	SDLSS	99 C2
100	I <u>I</u> LEN	A <u>I</u> LVE	TNLS	TDLEN	119 C3
120	A <u>I</u> LIG	ANLQ	ANLEN	ANLQ	139 C4
140	ANLEN	ANLRG	A <u>I</u> LTG	VNLEE	159 C5
160	THLKG	IETDKNTVWD			174 C6
	-2-1 +1+2	-2-1 +1+2	-2-1 +1+2	-2-1 +1+2	

Fig. 2. Structure-based sequence alignment of the tandem pentapeptide repeats in the Rfr-fold of Rfr23. The first two residues (A25*M26*) are non-native and remain after enterokinase removal of 41 of the 43 residues of the N-terminal tag. The residue position in the pentapeptide repeat, relative to the central residue i , is labeled on the bottom. The side chains of the $i-1$, $i+1$, and $i+2$ residues form the exterior of the Rfr-fold and the side chains of the $i-2$ and i residues occupy the interior. Consecutive interior cysteine side chains, C39 (i) and C42 ($i-2$), form a disulfide bond that stabilizes a type II turn involved two tandem pentapeptide repeats. The aromatic residue in the i th position, F52 and W173, are underlined and italicized, the hydrophobic residues on the surface of the Rfr-fold are underlined and colored grey, and the α -helix is enclosed in a box. The asterisk is a 24-residue insertion, L56–P79, that lies outside the Rfr-fold.

is no large, contiguous, negatively charged surface on the protein that may mimic the negatively charged phosphate backbone of DNA except, as shown in Fig. 1B, for Face 4. The latter Face of Rfr23 contains a contiguous negatively charged surface. However, this surface is linear unlike the helical phosphodiester backbone of DNA.

Detailed analyses of the Rfr-fold in the previously published PRP structures revealed that the Rfr-fold in these proteins were constructed by two distinct types of four-residue β -turns, type II and type IV, involving residues i , $i+1$, and $i+2$ of one pentapeptide repeat and the first residue, $i-2$, of the following pentapeptide repeat. Type II and type IV β -turns may be recognized by the pattern of inter-coil and intra-coil backbone hydrogen bonding (see

caption Fig. 4) and via an analysis of the main chain (Φ, Ψ) dihedral values. For MfpA and Rfr32 a narrow range of main chain (Φ, Ψ) dihedral values was observed dependant on the position in the pentapeptide repeat and the type of β -turn (Buchko et al., 2006a). In Rfr23 the main chain (Φ, Ψ) dihedral values of all 23 pentapeptide repeats fall near the previously reported approximate values for a type II β -turn (Buchko et al., 2006a) and all display a network of inter-coil and intra-coil backbone hydrogen bonding typical of type II β -turns. Because the overall length of the pentapeptide repeat in a type II β -turn is approximately 0.9 Å shorter than in a type IV β -turn, (Buchko et al., 2006a; Hegde et al., 2005) the Rfr-fold in Rfr23, composed entirely of one type of β -turn, is more “square” than an Rfr-fold containing a mixture of type II and type IV

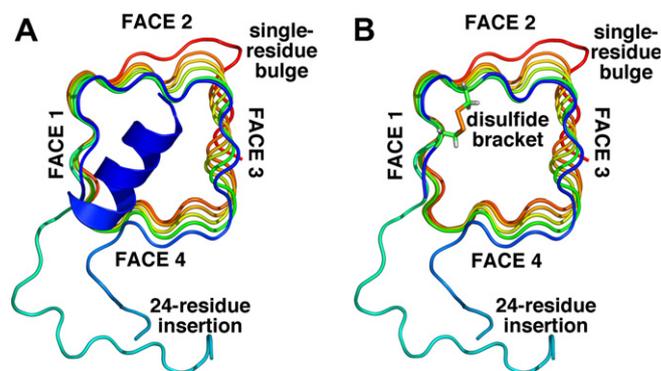


Fig. 3. Cartoon representation of the structure of Rfr23 viewed looking down the N-terminal. The position of the 24-residue insertion at the corner of Face 4 and Face 1 in the Rfr-fold is apparent as is the perturbation to the regular pattern of the Rfr-fold due to the single-residue bulge in C6. (A) Entire protein shown. (B) The N-terminal A-helix (M26*–E34) has been removed and the side chains of C39 and C42 included to show the position of the C39–C42 disulfide bond (disulfide bracket).

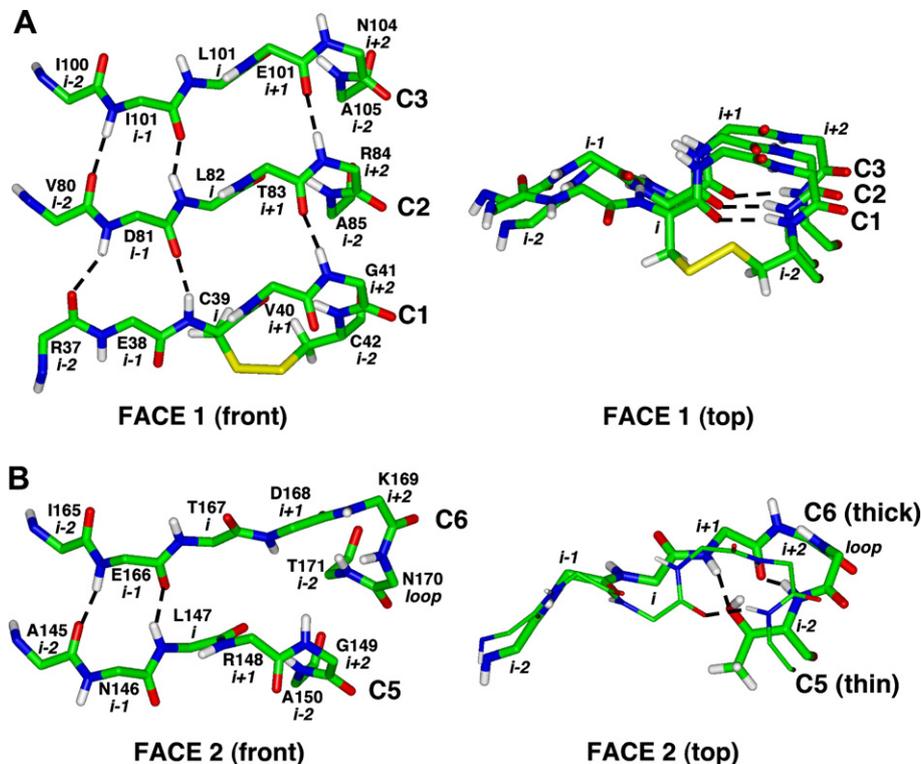


Fig. 4. (A) The front and N-terminal top views of the main chain backbone atoms (red, oxygen; green, carbon; blue, nitrogen; white, proton; yellow, sulfur) of three adjacent pentapeptide repeats on coils C1 through C3 on Face 1 of Rfr23 highlight the C39–C42 disulfide bond and the network of inter- and intra-coil main chain hydrogen bonds in the stacked type II β -turns. Note that the hydrogen atoms were not obtained directly from the crystal structure but estimated based on a 1.0 Å N–H bond length. 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 inter-coil hydrogen bonds (front). Instead, the carbonyl of the i th residue is near the amide of the $i - 2$ residue and forms an intra-coil hydrogen bond with the next pentapeptide repeat (top) that is typical of type II β -turn. In the Rfr-fold the side chain of all i and $i - 2$ residues are directed towards the center of the structure and evidently, when they are cysteine residues, can oxidize to form an intramolecular disulfide bond. The position of the disulfide bond at the corner of a coil between adjacent pentapeptide repeats “brackets” the type II β -turn and may provide it with extra stability. (B) The front and N-terminal top views of the main chain backbone atoms of two adjacent pentapeptide repeats on coils C5 and C6 on Face 2 of Rfr23 highlighting the effect of the single-residue bulge on the network of inter- and intra-coil main chain hydrogen bonds between the two coils. In the top view C5 is the thinner coil. To accommodate the single-residue insertion (N170) in C6 the two major structural deviations from a classical type II β -turn observed in Rfr-folds are (1) the main chain carbonyl of the i th residue and the main chain amide of the $i + 1$ residue are no longer approximately orthogonal to the plane of the inter-coil hydrogen bonding network and (2) the main chain carbonyl of the $i + 1$ residue and the main chain amide of the $i + 2$ residue are now approximately orthogonal to plane of the inter-coil hydrogen bonding network.

β -turns. Indeed, Rfr23 has average widths of $\sim 12.5 \times 12.4 \times 12.5 \times 12.3$ Å (average of Faces 1 through 4 as measured between backbone C- α carbons of the first and fifth residue of each pentapeptide repeat). Another consequence of an Rfr-fold containing only one type of β -turn is that there is less twist to the Rfr-fold and this is somewhat evident in Fig. 3, a view of Rfr23 down the long axis of the molecule. Similar observation regarding the shape of an Rfr-fold composed entirely of one type of β -turn (type II) was made for Np275 (Vetting et al., 2007).

4. Structural variation in the Rfr-fold

The general features of the Rfr-fold of Rfr23 are similar to those observed for the Rfr-fold in the other published PRP crystal structures, MfpA (Hegde et al., 2005; Vetting et al., 2006), Rfr32 (Buchko et al., 2006a), and Np275 (Vetting et al., 2007). Indeed, a DALI search (Holm and Sander, 1998) using residues R37–G55 and V80–D174 (the Rfr-fold) of Rfr23 returned *Z*-scores of 16.5, 15.7, and

14.8 against the Rfr-fold of Np275, Rfr32, and MfpA, respectively. Given the repeating nature of the pentapeptide repeat sequence in PRPs, similarities in the structure of all PRPs were predicted (Bateman et al., 1998; Vetting et al., 2006). However, a closer inspection of the structure of Rfr23 reveals three differences that show there are sequence-dependent structural variations to the Rfr-fold. The first is a rather large, 24-residue interruption in the Rfr-fold between the first and second coils (the 24-residue insertion). The second is an intraresidue disulfide bond between C39 and C42 in the first coil (the disulfide bracket). The third is a single-residue bulge between two Faces in the last coil (the single-residue bulge).

4.1. The 24-residue insertion

The most significant difference between the structure of Rfr23 and the other published PRP structures is that the 23 pentapeptide repeats that compose the Rfr-fold in Rfr23 are not all tandem in the primary amino acid sequence.

As illustrated in Fig. 3 the 24-residue insertion hangs out of the corner of Face 1 and Face 4 at the transition between coils C1 and C2. No electron density is observed for three out of the 24 residues in the insertion, suggesting that part of this region may be natively disordered (Stogios et al., 2007). The 24-residue insertion is positioned near the N-terminus and is accommodated with a one-residue gap in the Rfr-fold, the $i + 2$ residue of the fourth pentapeptide repeat. The sequence of the insertion, LKYR-IPRSSSPLSVTPFGMDKAKP, has a net positive charge at neutral pH, and this may explain its predisposition to sit over the negatively charged surface of Face 4 (Fig. 1B).

Previously, the program HMMER (Eddy, 1998) v2.3.2 was used to identify 35 PRPs in the genome of *Cyanothece* 51142 (Buchko et al., 2006a). Such an analysis predicted that 49% of Rfr23 consisted of an Rfr-fold with 17 tandem pentapeptide repeats. The crystal structure reported here indicates that approximately two-thirds of Rfr23 consists of an Rfr-fold (assuming the absent N-terminal signal region does not adopt an Rfr-fold) formed by 23 pentapeptide repeats. One reason for the discrepancy is because a criterion of the cyanobacteria-specific HMM model used to search for PRPs was 12 consecutive pentapeptide repeats in the center of the alignment. The four consecutive pentapeptide repeats prior to the 24-residue insertion was not long enough to be identified by the program. Another reason for the discrepancy is the program was not able to make allowances for a single-residue bulge that allows an additional 4/5 pentapeptide repeat to settle into a Rfr-fold. Given that the HMMER program predicts that 15 out of the 35 PRPs identified in *Cyanothece* are 50% or less Rfr-fold, the Rfr-fold in some of these PRPs may be larger than predicted due to insertions that interrupt the contiguity of pentapeptide repeats. Furthermore, it is possible that some PRPs were not identified in the *Cyanothece* genome because they did not contain at least 12 tandem pentapeptide repeats, but, would contain

enough pentapeptide repeats to establish a Rfr-fold if insertions are present and accounted for. The determination of more PRP structures is necessary in order to fine tune programs used to predict Rfr-motifs from primary amino acid sequences.

4.2. The disulfide bracket

Rfr23 is composed entirely of type II β -turns and consequently, the carbonyl of the i th residue is always near the amide of the next sequential $i - 2$ residue where it can form an intra-coil hydrogen bond as shown in the top view of Fig. 4A. Supplementing one intra-coil hydrogen bond in Rfr23 is an intra-coil disulfide bond between C39 and C42. Cysteine-39 and -42 are i th and $i - 2$ residues, respectively, pentapeptide positions where the side chains are directed into the interior of the Rfr-fold. As illustrated in Fig. 4A, the disulfide bond does not perturb the geometry of the turn in C1 relative to the turns shown for C2 and C3 (top). Instead, the disulfide bond appears to bracket the turn between pentapeptide repeats, hence the term ‘disulfide bracket’. There is some slight perturbation to the main chain conformations N-terminal to the turn, primarily for the $i + 1$ residue, however, this may be because the pentapeptide repeat is the first repeat of the coil.

Disulfide bonds are an important protein component, often contributing to protein stability, activity, and folding (Kadokura, 2006; Thornton, 1981). To assess the structural importance of the lone disulfide bond in Rfr23, the reducing agent DTT was added to Rfr23 and the structural consequence probed by NMR spectroscopy. Such experiments show that the reduction of the disulfide bracket results in the complete collapse of the structured protein into an unfolded protein. This unfolding is illustrated most convincingly in Fig. 5A and B, ^1H - ^{15}N HSQC spectra of Rfr23 in the absence and presence of 2 mM DTT, respec-

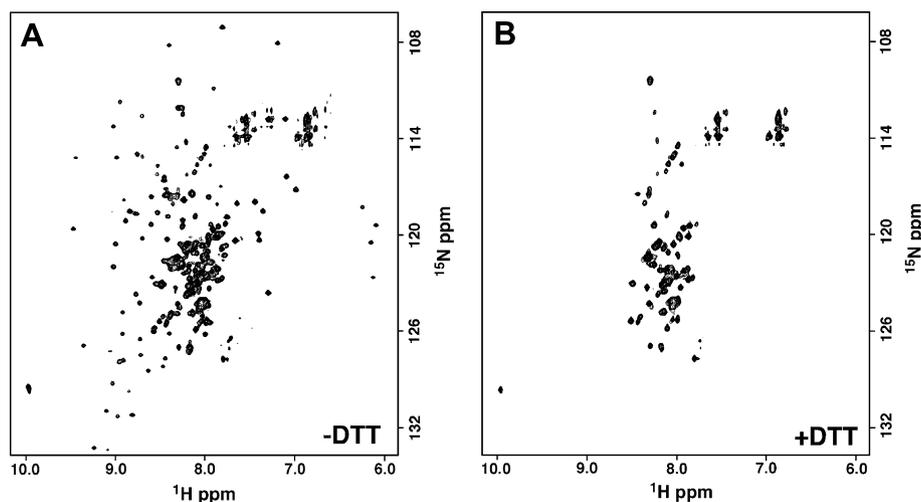


Fig. 5. Comparison of the ^1H - ^{15}N TROSY HSQC spectra of untagged Rfr23 without DTT (A) and after making the buffer (50 mM NaCl, 20 mM Tris-HCl, pH 7.4) 2 mM in DTT (B). Approximately three-times more amide cross peaks are observed in the absence of DTT. Spectra collected under identical conditions at 25 °C and a ^1H resonance frequency of 800 MHz.

tively. In Fig. 5A approximately 160 backbone amide cross peaks appear and they are mostly narrow and widely dispersed in both the nitrogen and proton dimension, features characteristic of a structured, monomeric protein. The estimated isotropic overall rotational correlation time (τ_c) for the protein generating the spectrum in Fig. 5A is 7.9 ns as inferred from ^{15}N spin relaxation times (Szyperski et al., 2002). Such a correlation time is typical of a folded 16.4 kDa protein (Palczewska et al., 2001) and corroborates with the retention time observed by size exclusion chromatography (data not shown). Fig. 5B is the spectrum of the same sample of Rfr23 after making the solution 2 mM in DTT. Now only approximately 60 backbone amide cross peaks are observed and the proton chemical shifts largely fall within a small window as is typical of a protein lacking significant secondary structure (Buchko et al., 2007). Analysis of three-dimensional HNCACB data collected on a ^{13}C -, ^{15}N -labeled Rfr23 sample in the presence of DTT confirm such a conclusion, there is little chemical shift dispersion in the $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ carbons of amino acid residues of the same type. The estimated isotropic overall rotational correlation time for the protein generating the spectrum in Fig. 5B is 4.9 ns, typical of a folded ~ 10 kDa protein. Because SDS-PAGE showed no change in the molecular weight for Rfr23 due to the addition of DTT, the reduction of the disulfide bracket must be effecting the unfolding of the protein. Note that all the cross peaks observed in the ^1H - ^{15}N HSQC spectrum of Rfr23 in the presence of DTT exist as a subset of cross peaks in the ^1H - ^{15}N HSQC spectrum of Rfr23 collected in the absence of DTT. This observation suggests that the Rfr23 sample without DTT either contains a mixture of structured and unstructured protein, or, Rfr23 contains a common “structural” feature that is not dependent on the presence or absence of DTT. The latter conclusion is unlikely because the only unstructured region of the protein is the 24-residue insertion and four of these residues are prolines. The 60 cross peaks in the “unstructured” spectrum (Fig. 5B) is three times greater than the 20 cross peaks expected for the disordered 24-residue insertion.

Disulfide bonds may contribute up to 6 kcal/mol to the stability of a protein at optimal temperatures (Betz, 1993; Darby and Creighton, 1995; Matsumura and Matthews, 1991). Perhaps more robust Rfr-folds may be designed *de novo* by introducing disulfide brackets into Rfr-folds. While the cytoplasm harbors many mechanisms to keep cysteines in the reduced state, oxidized cysteines may remain protected from reduction in extracytoplasmic compartments if they are not secreted outside the cell. SignalP analysis (Bendtsen et al., 2004) of the native Rfr23 sequence predicts an N-terminal signal sequence that could direct the protein into extracytoplasmic compartments or outside the cell. Because disulfide bond reduction is occasionally observed as a mechanism to regulate the activity of proteins (Hogg, 2003), perhaps the disulfide bracket plays some role in regulating the (unknown) activity of Rfr23.

4.3. The single-residue bulge

As illustrated schematically in Fig. 2, N170 is ‘looped out’ of the sequential sequence of pentapeptide repeats in order to stack W173 into the interior of the Rfr-fold in an *i*th-type position. To accommodate the single-residue bulge, shown in detail in Fig. 4B with coils C5 and C6 on Face 2, the main chain atoms adopt (Φ, Ψ) dihedral values that deviate from the average type II β -turn values $(-140/180)$, $(-159/-177)$, $(-62/-21)$, $(-82/-18)$, and $(-67/126)$ for *i*, *i* + 1, *i* + 2, bulge, and *i* - 2, respectively). In C6, the main chain carbonyl of the *i*th residue and the main chain amide of the *i* + 1 residue are no longer approximately orthogonal to the plane of the inter-coil hydrogen bonding network, but are now about half way between being orthogonal and in the inter-coil hydrogen bonding plane. As a consequence, the carbonyl of the *i*th residue, T167, cannot form an intra-coil hydrogen bond with the amide of the sequential *i* - 2 residue as shown in the top view in Fig. 4B. Instead, the amide of D168 and the side chain hydroxyl oxygen of T171 (*i* - 2) form an intra-coil hydrogen bond. Another significant difference in the single-residue bulge is the main chain carbonyl of the *i* + 1 residue and the main chain amide of the *i* + 2 residue are now approximately orthogonal to plane of the inter-coil hydrogen bonding network, and consequently, an inter-coil backbone hydrogen bond is lost (between G149 amide and D168 carbonyl). However, the carbonyl of the *i* + 2 residue (D168) is now near the amide of the sequential *i* - 2 residue (T171) and an intra-coil hydrogen bond can form. The single-residue bulge in Rfr23 may be a special situation, since it is situated in the C-terminal coil of an Rfr-fold. It will be interesting to observe if single-residue bulges in other Rfr-folds are accommodated similarly, especially in the interior of Rfr-folds. Perhaps the role of the distortions to the regular Rfr-fold produced by the single-residue bulge is to prevent edge-to-edge aggregation from occurring at the C-terminus.

5. Concluding remarks

The sheer number of PRPs observed in the sequenced genomes of photosynthetic cyanobacteria coupled with their predicted location in all the cyanobacteria cellular compartments argues for an important physiological function for these proteins in cyanobacteria (Buchko et al., 2006a; Kieslebach et al., 1998). However, the biological role of the Rfr-fold is not known. Persuasive evidence suggests that the 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). Based on this observation, one hypothesis is that Rfr-folds in PRPs expressed from chromosomal DNA, by mimicking DNA, may target the catalytic domains to DNA-binding proteins and exert some post-translational control over them (Vetting et al., 2006). This hypothesis is not supported with the PRP struc-

ture from *N. punctiforme* (Np275) and the two PRP structures from *Cyanotheca* (Rfr32 and Rfr23). While the surface of Np275 is even more electronegative than MfpA, it failed to show any significant ability to interact with DNA gyrase (Vetting et al., 2007). Rfr32 did not contain a helical electronegative surface that mimicked the phosphodiester backbone stairway of double-stranded DNA (Buchko et al., 2006a). While one face of Rfr23 contains a well-defined electronegative surface (Fig. 1B) it does not contain a contiguous electronegative surface with similarities in size, shape, and electrostatics to double-stranded DNA. Hence, DNA-mimicry may not be the biological role of the Rfr-fold in PRPs from cyanobacteria and the antibiotic resistant properties of plasmid PRP gene products likely developed as a secondary property from PRP gene products in chromosomal DNA.

Structures of more PRPs with Rfr-folds are necessary to enable a more accurate prediction of Rfr-folds based on primary amino acid sequence. In the four deposited PRP structures, the side chains of the $i - 2$ and i residues are always directed towards the interior of the Rfr-fold while the side chains of the $i - 1$, $i + 1$, and $i + 2$ residues are always directed towards the exterior of the Rfr-fold. Hence, due to the repeating nature of the pentapeptide repeats and the regular structure of the Rfr-fold, much information about the interior and exterior of the Rfr-fold can be obtained from the primary amino acid sequence alone if the registration of the pentapeptide repeats is known. However, if the Rfr-fold fails to begin at the predicted first pentapeptide repeat, then such alignment is incorrect and so is the predicted external and internal alignment of the side chains. Indeed, the use of molecular replacement methods to determine the crystal structure of Np275 was hampered because of the lack of information regarding the alignment of the Rfr-fold (Vetting et al., 2007). Alignment is also incorrect if there are unpredicted interruptions in the Rfr-fold, as noted here with the 24-residue insertion between the first and second coils of the Rfr-fold of Rfr23. Consequently, additional structures of PRPs in different sequence contexts are required in order to generate more robust algorithms to more accurately predict Rfr-folds from primary amino acid sequences. At the same time additional PRP structures may reveal additional structural variations to the Rfr-fold and refine our emerging understanding of this intriguing family of proteins.

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 US Department of Energy's Office of Biological and Environmental Research (BER) program located at Pacific Northwest National Laboratory (PNNL). PNNL is operated for the US Department of Energy by Battelle. We thank Dr. Eric A. Welsh (Washington University—St. Louis) for useful dis-

cussions and Dr. Farhad Forouhar (Miami University—Ohio) for assistance in scaling the data and building the 24-residue insertion. The assistance of the X29A beam line scientists at the National Synchrotron Light Source at Brookhaven National Laboratory is appreciated. 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 US Department of Energy, and from the National Center for Research Resources of the National Institutes of Health. This manuscript has been authored by Battelle Memorial Institute, Pacific Northwest Division, under Contract No. DE-AC05-76RL0 1830 with the US Department of Energy.

References

- Bateman, A., Murzin, A.G., 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., Sonnhammer, E.L., 2000. The Pfam protein families database. *Nucleic Acids Res.* 28, 263–266.
- Bendtsen, J.D., Nielsen, H., von Heijne, G., Brunak, S., 2004. Improved prediction of a signal peptides: SignalP 3.0. *J. Mol. Biol.* 340, 783–795.
- Betz, S.F., 1993. Disulfide bonds and the stability of globular proteins. *Protein Sci.* 2, 1551–1558.
- 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., 1998. Crystallography and NMR System (CNS): a new software suite for macromolecular structure determination. *Acta Crystallogr. D* 54, 905–921.
- Buchko, G.W., Ni, S., Robinson, H., Welsh, E.A., Pakrasi, H.B., Kennedy, M.A., 2006a. Characterization of two potentially universal turn motifs that shape the repeated five-residues fold—crystal structure of a luminal pentapeptide repeat protein from *Cyanotheca* 51142. *Protein Sci.* 15, 2579–2595.
- Buchko, G.W., Robinson, H., Ni, S., Pakrasi, H.B., Kennedy, M.A., 2006b. Cloning, expression, crystallization and preliminary crystallographic analysis of a pentapeptide-repeat protein (Rfr23) from the bacterium *Cyanotheca* 51142. *Acta Cryst.* F62, 1251–1254.
- Buchko, G.W., Ni, S., Lourette, N.M., Reeves, R., Kennedy, M.A., 2007. NMR resonance assignments of the high mobility group protein HMGA1. *J. Biomol. NMR* 38, 185.
- Darby, N., Creighton, T.E., 1995. Disulfide bonds in protein folding and stability. *Methods Mol. Biol.* 40, 219–252.
- De Grado, W.F., Kezdy, F.J., Kaiser, E.T., 1981. Design, synthesis and characterization of a cytotoxic peptide with melittin-like activity. *J. Am. Chem. Soc.* 103, 679–681.
- Drlica, K., 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.
- Hegde, S.S., Vetting, M.W., Roderick, S.L., Mitchenall, L.A., Maxwell, A., Takiff, H.E., Blanchard, J.S., 2005. A fluoroquinoline resistance protein from *Mycobacterium tuberculosis* that mimics DNA. *Science* 308, 1480–1483.
- Hogg, P.J., 2003. Disulfide bonds as switches for protein function. *Trends Biochem. Sci.* 28, 210–214.
- Holm, L., Sander, C., 1998. Touring protein fold space with Dali/FSSP. *Nucleic Acids Res.* 26, 316–319.
- Jenkins, J., Pickersgill, R., 2001. The architecture of parallel β -helices and related folds. *Prog. Biophys. Mol. Biol.* 77, 111–175.
- Kadokura, H., 2006. Oxidative protein folding: many different ways to introduce disulfide bonds. *Antioxid. Redox Signal.* 8, 731–733.

- Kieslebach, T., Mant, A., Robinson, C., 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.
- Laskowski, R.A., Rullmann, J.A.C., MacArthur, M.W., Kaptein, R., Thornton, J.M., 1996. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* 8, 477–486.
- Lovell, S.C., Davis, I.W., Arendall III, W.B., de Bakker, P.I.W., Word, J.M., Prisant, M.G., Richardson, J.S., Richardson, D.C., 2003. Structure validation by $C\alpha$ geometry: π , σ , and $C\beta$ deviation. *Proteins* 50, 437–450.
- Matsumura, M., Matthews, B.W., 1991. Stabilization of functional proteins by introduction of multiple disulfide bonds. *Methods Enzymol.* 202, 336–356.
- McRee, D.E., 1999. A versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.* 125, 156–165.
- Palczewska, M., Groves, P., Ambrus, A., Kaleta, A., Kover, K., Batta, G., Kuznicki, J., 2001. Structural and biochemical characterization of neuronal calretinin domain I–II (residues 1–100). *Eur. J. Biochem.* 268, 6229–6237.
- Richardson, J.S., Richardson, D.C., 2002. Natural β -sheet proteins use negative design to avoid edge-to-edge aggregation. *Proc. Natl. Acad. Sci. USA* 99, 2754–2759.
- Stogios, P.J., Chen, L., Prive, G.G., 2007. Crystal structure of the BTB domain from the LRF/ZBTB7 transcriptional regulator. *Protein Sci.* 16, 336–342.
- Szyperski, T., Yeh, D.C., Sukumaran, D.K., Moseley, H.N.B., Montelione, G.T., 2002. Reduced-dimensionality NMR spectroscopy for high-throughput protein resonance assignment. *Proc. Natl. Acad. Sci. USA* 99, 8009–8014.
- Terwilliger, T.C., 1999. Reciprocal-space solvent flattening. *Acta Crystallogr. D* 55, 1863–1871.
- Terwilliger, T.C., 2000. Maximum-likelihood density modification. *Acta Crystallogr. D* 56, 965–972.
- Terwilliger, T.C., 2001a. Maximum-likelihood density modification using pattern recognition of structural motifs. *Acta Crystallogr. D* 57, 1763–1775.
- Terwilliger, T.C., 2001b. Map-likelihood phasing. *Acta Crystallogr. D* 57, 1763–1775.
- Thornton, J.M., 1981. Disulphide bridges in globular proteins. *J. Mol. Biol.* 151, 261–287.
- Vetting, M.W., Subray, S., Hegde, S.S., Hazleton, K.Z., Blanchard, J.S., 2007. Structural characterization of the fusion of two pentapeptide repeat proteins, Np275 and Np276, from *Nostoc Punctiforme*: resurrection of an ancestral protein. *Protein Sci.* 16, 755–760.
- Vetting, M.W., Hegde, S.S., Fajardo, J.E., Fiser, A., Roderick, S.L., Takiff, H.E., Blanchard, J.S., 2006. Pentapeptide repeat proteins. *Biochemistry* 45, 1–10.
- Yoder, M.D., Keen, N.T., Journak, F., 1993. New domain motif: the structure of pectate lyase C, a secreted plant virulence factor. *Science* 260, 1503–1507.