PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

The repeat domain of the type III effector protein PthA shows a TPR-like structure and undergoes conformational changes upon DNA interaction

Mário Tyago Murakami,* Mauricio Luis Sforça, Jorge Luiz Neves, Joice Helena Paiva, Mariane Noronha Domingues, André Luiz Araujo Pereira, Ana Carolina de Mattos Zeri, and Celso Eduardo Benedetti*

Laboratório Nacional de Biociências, Centro Nacional de Pesquisa em Energia e Materiais, 13083-970, Campinas, SP, Brazil

ABSTRACT

Many plant pathogenic bacteria rely on effector proteins to suppress defense and manipulate host cell mechanisms to cause disease. The effector protein PthA modulates the host transcriptome to promote citrus canker. PthA possesses unusual protein architecture with an internal region encompassing variable numbers of near-identical tandem repeats of 34 amino acids termed the repeat domain. This domain mediates protein–protein and protein–DNA interactions, and two polymorphic residues in each repeat unit determine DNA specificity. To gain insights into how the repeat domain promotes protein–protein and protein–DNA contacts, we have solved the structure of a peptide corresponding to 1.5 units of the PthA repeat domain by nuclear magnetic resonance (NMR) and carried out small-angle X-ray scattering (SAXS) and spectroscopic studies on the entire 15.5-repeat domain of PthA2 (RD2). Consistent with secondary structure predictions and circular dichroism data, the NMR structure of the 1.5-repeat peptide reveals three α-helices connected by two turns that fold into a tetratricopeptide repeat (TPR)-like domain. The NMR structure corroborates the theoretical TPR superhelix predicted for RD2, which is also in agreement with the elongated shape of RD2 determined by SAXS. Furthermore, RD2 undergoes conformational changes in a pH-dependent manner and upon DNA interaction, and shows sequence similarities to pentatricopeptide repeat (PPR), a nucleic acid-binding motif structurally related to TPR. The results point to a model in which the RD2 structure changes its compactness as it embraces the DNA with the polymorphic diresidues facing the interior of the superhelix oriented toward the nucleotide bases.

Proteins 2010; 78:3386–3395.

Key words: TAL effectors; tetratricopeptide repeat; pentatricopeptide repeat (PPR); *Xanthomonas citri.*

INTRODUCTION

Xanthomonas axonopodis pv. *citri* (Xac), the causal agent of citrus canker, induces the formation of raised pustules on the surface of the host plant. The pustules typically develop into larger corky and water-soaked lesions, which break the epidermis favoring bacterial dissemination.[1,2] Although it is known that canker lesions result from the intense division and expansion of the host cells at the site of infection, how exactly Xac induces cell division and growth is not yet clear. We have shown that Xac modulates the synthesis of cell growth regulators during the onset of infection[3] and that the common action of auxin and gibberellin is required for initial canker development.[4] Interestingly, these hormones alone or in combination had no apparent effect on citrus cell growth or division, suggesting that another factor is required for canker formation.[4] This additional factor is thought to be the Xac effector protein PthA, which is sufficient to promote cell hypertrophy when transiently expressed in citrus leaves.[5] Accordingly, expression of PthA proteins in citrus cells provokes transcriptional changes that overlap with those triggered by Xac infection associated with auxin and gibberellin action (Pereira and Benedetti, unpublished data).

PthA proteins differ from each other primarily by the number of near-identical repeats of 33–34 amino acids that are tandemly located in the central region of the protein. This domain confers host selectivity and is critical to determine pathogenicity.[6,7] PthA proteins are 95–97% identical to AvrBs3, the best known *Xanthomonas* type III effector protein. AvrBs3 is targeted to the nucleus of host cells where it modu-

Additional Supporting Information may be found in the online version of this article.
Grant sponsors: Fundação de Amparo à Pesquisa do Estado de São (FAPESP), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).
*Correspondence to: M.T. Murakami, Laboratório Nacional de Biociências, Centro Nacional de Energia e Materiais, R. Giuseppe Máximo Scolfaro, 10000, Campinas, SP, 13083-970, Brazil. E-mail: mario.murakami@lnbio.org.br or C.E. Benedetti, Laboratório Nacional de Biociências, Centro Nacional de Energia e Materiais, R. Giuseppe Máximo Scolfaro, 10000, Campinas, SP, 13083-970, Brazil. E-mail: celso.benedetti@lnbio.org.br.
Received 26 May 2010; Revised 16 July 2010; Accepted 24 July 2010
Published online 17 August 2010 in Wiley Online Library (wileyonlinelibrary.com).
DOI: 10.1002/prot.22846

lates transcription.[8] Notably, AvrBs3 recognizes different plant promoters and activates transcription of particular genes in both susceptible and resistant plants.[9,10] The interaction of AvrBs3 with its target DNA is mediated by its repeat domain,[9–12] which is also essential for AvrBs3 dimerization before its nuclear import.[13] Likewise, we found that the repeat domain of PthA proteins is critical for protein–protein interactions.[14] Therefore, it is becoming clear that the repeat domain of such effectors has a dual character as a protein scaffold that allows protein–protein and protein–DNA contacts with host targets.

An intriguing aspect of the repeat domain of such effectors is the variability found in every repeat unit at certain amino acid positions. In the case of the PthA variants, the polymorphism occurs preferentially at Positions 4 (D, E, Q, or A), 12 (N or H), 13 (I, D, G, or S), and 24 (R or A) of the repeat units.[14] Since most members of the AvrBs3/PthA protein family share high-sequence identity, it has been postulated that, besides the variation in the number of repeat units, the polymorphism within the repeat units would play a critical role in conferring the specificity required for the interactions with particular protein or DNA targets. Indeed, recent studies have demonstrated that the pair of residues at Positions 12–13 determines the DNA sequence specificity[15,16]; however, the structural basis for these interactions is presently unknown. Thus, structural information on the repeat region becomes highly relevant for understanding the molecular function of this unique protein domain. Although numerous homologues of AvrBs3/PthA proteins have been identified, no three-dimensional (3D) structure is yet available for any transcription activator-like (TAL) effector. Despite considerable efforts, we have not yet been able to crystallize any of the PthA variants or their repeat domains. Thus, we used a combination of circular dichroism (CD), dynamic light scattering (DLS), nuclear magnetic resonance (NMR), small-angle X-ray scattering (SAXS) and molecular modeling approaches to gain insights into the 3D structure of the PthA repeat domain. In this work, we present the first 3D structure of a peptide corresponding to 1.5-repeat units of PthA2 (PDB code: 2KQ5) and a low-resolution envelope for the entire repeat domain of PthA2 (RD2) that is in agreement with its theoretical tetratricopeptide repeat (TPR)-like structure. Spectroscopic studies also revealed that RD2 undergoes conformational changes in a pH-dependent manner and in the presence of DNA. Furthermore, we show that the repeat units of RD2 are similar to pentatricopeptide repeat (PPR), a 35 amino acid motif that is structurally related to TPR and have nucleic acid binding activity.[17,18]

MATERIALS AND METHODS

Protein expression and purification

The 15.5-RD2 was amplified from *X. citri* plasmids and cloned into the *Bam*HI/*Sac*I sites of pET28a (Novagen). The construct was sequenced and used to transform *Escherichia coli* strain BL21 (DE3). Protein expression was induced at OD_{600} = 0.6 with 0.4 m*M* IPTG at 25°C. The cells were harvested after 3 h, centrifuged, and suspended in binding buffer (20 m*M* Tris–HCl pH 8.0, 200 m*M* NaCl, 15 m*M* imidazole, 5% (v/v) glycerol, 1 m*M* PMSF, 0.5% Nonidet P-40) and incubated on ice with lysozyme (1 mg mL^{-1}) and RNAse (10 μg mL^{-1}) for 30 min. Bacterial cells were disrupted by sonication and the soluble fraction was incubated with DNAse I (0.5 μg mL^{-1}) and $MgCl_2$ (2 m*M*). The suspension containing the 6xHis-RD2 was loaded on a 5-mL HiTrap chelating HP (GE Healthcare) column pre-equilibrated with binding buffer. The column was washed with 20 column volumes of the same buffer and the protein was eluted with imidazole gradient. After purification, dithiothreitol (DTT) was added to a final concentration of 1 m*M* and fractions displaying the highest purify were concentrated and loaded on a G75 16/60 Superdex (GE Healthcare) column, pre-equilibrated with 20 m*M* Tris–HCl pH 8.0, 200 m*M* NaCl, 5% (v/v) glycerol, and 1 m*M* DTT. RD2 was eluted as a single peak with a molecular mass corresponding to the monomer.

Dynamic light scattering

DLS experiments were carried out using a DynaPro 810 (Protein Solutions) apparatus equipped with a Peltier module for temperature control. The wavelength of the laser light and the output power were set to 830 nm and 30 mW, respectively. About 100 measurements were made at intervals of 20 s for each run. The DLS experiments were repeated several times with intervals of 30 min to check stability. Protein solutions of 1 mg mL^{-1} were prepared in 50 m*M* sodium acetate pH 5.0, 50 m*M* sodium phosphate pH 6.0 and 7.0, and 20 m*M* Tris–HCl pH 8.0. The complex of RD2:DNA was prepared in a molar ratio of 1.0:1.2, incubated for 12 h at 20°C, and centrifuged at 10,000*g* for 15 min at 4°C before analysis. Standard curves of bovine serum albumin were used for calibration and the experiments were conducted at 18°C. Hydrodynamic parameters were determined using the software DYNAMICS v.6.10.1.2.

CD measurements

CD spectroscopy experiments were conducted on a JASCO J-810 CD spectrophotometer equipped with a Peltier temperature control using 1-mm path quartz cuvettes. Spectra were acquired with a final protein concentration of 5.0–10 μ*M* in 20 m*M* Tris–HCl pH 8.0, containing 1.0 m*M* DTT and 100 m*M* NaCl. The RD2:DNA mixes were incubated for 3 h at 20°C and centrifuged (10,000*g*) for 15 min at 4°C before analyses. CD measurements were collected between 185 and 260 nm using a scanning rate of 50 nm min^{-1} with an average response time of 4 s.

Secondary structure variations were monitored as a function of changes in the initial CD spectrum upon addition of sodium dodecyl sulfate (SDS) for the 1.5-repeat unit. For all the spectra, an average of 10 scans was accumulated and the background spectrum of the buffer was subtracted.

NMR studies

NMR experiments were performed at 20°C using a Varian Inova 600 MHz spectrometer equipped with a cryogenic probe. Synthetic peptides (Proteimax, São Paulo, Brazil) corresponding to 1.0- or 1.5-repeat units of the repeat domain of PthA were dissolved in 20 m*M* phosphate buffer, pH 5.0, containing 70 m*M* SDS, 1.0 m*M* DTT, and 5% (v/v) D_2O, at a final concentration of ~1.0 m*M*. Peptide resonance peaks were assigned using standard methods including correlation spectroscopy,[19] total correlation spectroscopy (TOCSY),[20] and nuclear Overhauser enhancement spectroscopy (NOESY).[21] The TOCSY spectra were acquired using a DIPSI spin-lock sequence at a field strength of 10 kHz and a mixing time of 70 ms. NOESY spectra were recorded with mixing times of 150 and 300 ms. All 2D experiments were acquired in the phase-sensitive mode using the method of states.[22] Water suppression was achieved by low-power continuous wave irradiation during the relaxation delay or using the WATERGATE method.[23] Data were processed and analyzed using the NMRPipe/NMRVIEW software.[24] Before Fourier transformation, the time domain data were zero filled in both dimensions to yield a (4096 by 4096) data matrix. When necessary, a fifth-order polynomial baseline correction was applied after transformation and phasing. To obtain distance constraints, crosspeak volumes were estimated from the NOESY spectra.

The structure of the peptide corresponding to 1.5-repeat units of PthA was calculated in a semiautomated iterative manner with the program CYANA version 2.1,[25] using 100 starting conformers. CYANA 2.1 protocol was applied to calibrate and assign NOE crosspeaks. After the first few rounds of automatic calculations, the NOESY spectra were analyzed again to identify additional crosspeaks consistent with the structural model and to correct misidentified NOEs. The structures obtained were further refined by restrained minimization and molecular dynamic (MD) studies using the CNS software.[26] The 20 structures with the lowest target function were selected to represent the ensemble of peptide structures. The quality of the structures was analyzed with PROCHECK-NMR.[27] The NMR data were deposited in the biological magnetic resonance bank (BMRB) and protein data bank (PDB) under the entry codes 16589 and 2KQ5, respectively.

SAXS data collection and analysis

SAXS data were collected at the D11A-SAXS beamline at the Brazilian Synchrotron Light Laboratory (LNLS). The radiation wavelength was set to 1.488 Å and a charge-coupled device area detector (MARCCD 165 mm) was used to record the scattering patterns. The sample-to-detector distance was set to 1415.95 mm to give a scattering vector ranging from 0.10 to 3.5 nm^{-1}. The measurements were carried out with a sample concentration of 2 mg mL^{-1} at 18°C. Each protein sample was previously analyzed by DLS and only monodisperse solutions (polydispersity <20%) were used. Protein samples were centrifuged for 15 min at 10,000*g* to eliminate any existing aggregates immediately before each measurement. The scattering curves of the protein solutions and buffers were collected in frames of 300 s each to avoid radiation-induced protein damage. Each frame was carefully checked for possible bubbles or radiation-induced aggregation of the protein before calculating the average intensity and the associated experimental error. The experimental intensities were corrected for background, buffer contributions, detector inhomogeneities, and sample transmission.

The radius of gyration (R_g) was evaluated using Guinier approximation[28] as implemented in the program PRIMUS.[29] The indirect Fourier transform package GNOM[30] was used to evaluate the pair-distance distribution function $p(r)$. The low-resolution envelope of the RD2 protein was determined using *ab initio* modeling implemented in DAMMIN.[31] An averaged model was generated from several runs using the DAMAVER suite of programs.[32] The SAXS model and the NMR structures were superimposed with SUPCOMB.[33]

Homology molecular modeling of the RD2

The crystal structures of importin β-1 from *Saccharomyces cerevisiae* (2BPT)[34] and TIP120 from *Homo sapiens* (1U6G)[35] were used as 3D templates for restraint-based modeling as implemented in the MODELLER program.[36] The overall model was improved enforcing the proper stereochemistry using spatial restraints and CHARMM energy terms, followed by conjugate gradient simulation based on the variable target function method.[36] Ten models were built for the RD2 sequence based on the (m)GenThreader alignment. All models were evaluated with the DOPE potential and the one with the lower global score was selected for explicit solvent MD simulation using GROMACS[37] to check its stability and consistency. The overall and local quality analyses of the final model were assessed by VERIFY3D,[38] PROSA,[39] and VADAR.[40] Three-dimensional structures were displayed, analyzed, and compared using the program COOT.[41]

RESULTS AND DISCUSSION

The 1.5-repeat peptide of PthA shows a TPR-like fold

Secondary structure algorithms predict only α-helices and turns for the repeat domains of PthA proteins. Not

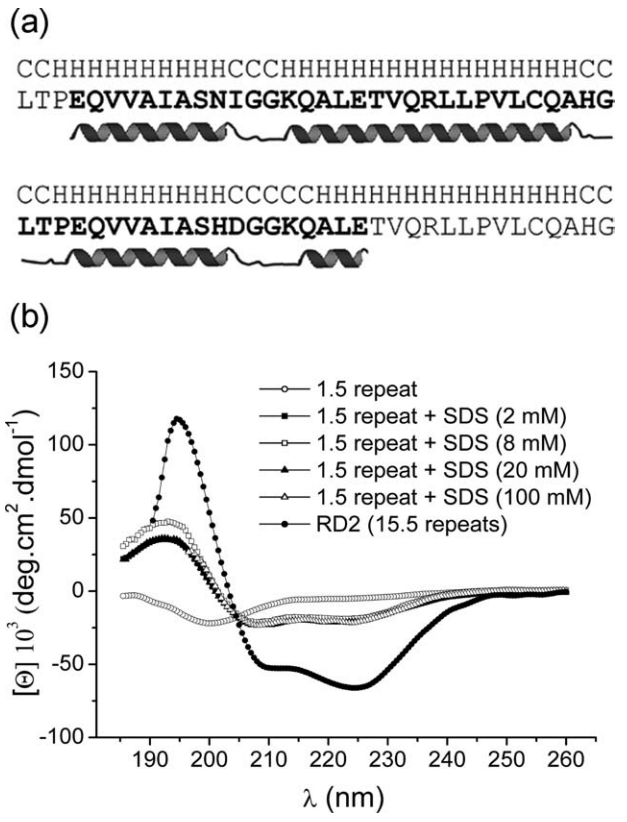

Figure 1

Structural features of the repetitive units of RD2. (**a**) PSIPRED secondary structure prediction of two consecutive repeat units of RD2 showing coils (C) and helices (H). The 1.5-repeat unit peptide used for structural resolution is shown in bold and the secondary structure elements determined by NMR are represented underneath the sequence. (**b**) CD curves of the purified 15.5-repeat domain (5 μ*M*) of PthA2 (RD2) in comparison with that of the 1.5-repeat unit peptide (10 μ*M*) in the presence and absence of different SDS concentrations.

surprisingly, the predictions are almost identical within each of the repeat units, as shown for two consecutive repeats found in the PthA variants [Fig. 1(a)]. The predictions are consistent with the far-UV CD spectrum of the entire 15.5-RD2 showing two minimum peaks at 208 and 222 nm and a positive peak at 193 nm, which indicate that RD2 has high contents of α-helices [Fig. 1(b)].

CD data analysis has indicated that the 34-amino acid peptide corresponding to one repeat unit of PthA is not structured in solution (not shown). Thus, we have used *in silico* 3D modeling algorithms to find a minimal peptide size that possesses the stereochemical requirements for a structured motif, and we observed that the 1.5-repeat peptide is the minimal structural unit of the repeat region of PthA. CD measurements indicated that the 1.5-repeat peptide is partially unfolded in solution; however, addition of 2 m*M* SDS was sufficient to increase the contents of secondary structural elements, particularly α-helices [Fig. 1(b)]. Although SDS has been successfully used to investigate the structure of membrane peptides, it has also been shown to stabilize conformations in peptides with propensity to form α-helices via electrostatic interactions.[42] We therefore used NMR techniques to solve the structure of the 1.5-repeat peptide in the presence of SDS since no experimental 3D model of the repeat unit was yet available.

The 1.5-repeat structure was solved using 982 distance constraints derived from the NOESY spectra. The inter-residue NOEs correlate with the chemical shift index and show a dense pattern of $d\alpha N(i,i+3)$, $d\alpha\beta(i,i+3)$, and $d\alpha N(i,i+4)$ NOEs involving residues V3-I10, Q14-Q20, P24-A29, and V37-G45, indicating that these regions possess an helical fold (Supporting Information Figure 1). Statistics of an ensemble of 20 lower energy structures [Fig. 2(a)] show no significant violations of the molecular geometry parameters (Table I). In addition, the Ramachandran plot analysis shows that all *phi–psi* angles are in allowed regions and the root mean square deviations (RMSD) for the backbone and side chains are within expected values, indicating a good stereochemical quality of the ensemble.

The peptide folds into a helical-bundle structure that is very similar to the TPR topology [Fig. 2(a)], thus corroborating the secondary structure predictions and CD data. In fact, the CD curve of RD2 is quite similar to those of TPR domains with a 222/208 nm ratio of ~1.3, which indicates coiled-coil-like conformations.[43–46] The specific interactions responsible for these conformational states are the presence of a series of hydrogen bonds between the backbone carbonyl (C=O) of residues "*i*" and the backbone amide (NH) of residues "*i* + 4" spanning the helical regions. The secondary structure is also stabilized by hydrophobic interactions that keep the helical regions close together. Preferential spatial arrangement is confirmed by NOE interactions between side chains of the hydrophobic residues V4-S8 with T18-L22 and L16-C27 with L32-H43 [Fig. 2(b)].

Interestingly, the two hypervariable diresidues NI and HD, known to specify preferential interactions with adenine and cytosine in DNA target sites, respectively,[15,16] are structurally close in the NMR structures [Fig. 2(c)]. This would be expected considering that each pair of the variable residues recognizes one nucleotide in a linear fashion, that is, one variable diresidue to one nucleotide in the DNA target sequence.[15,16] We also noticed that the K residues, which typically play a central role in protein–DNA interactions,[47] are located adjacently to the polymorphic diresidues [Fig. 2(c)].

The RD2 has an elongated shape

The *ab initio* molecular shape of RD2 was determined by SAXS analysis and the corresponding profiles of the scattering curve and distance distribution function $p(r)$ are shown in Figure 3(a). SAXS calculations revealed a maximum dimension of 129 Å from the $p(r)$ curve and a gyra-

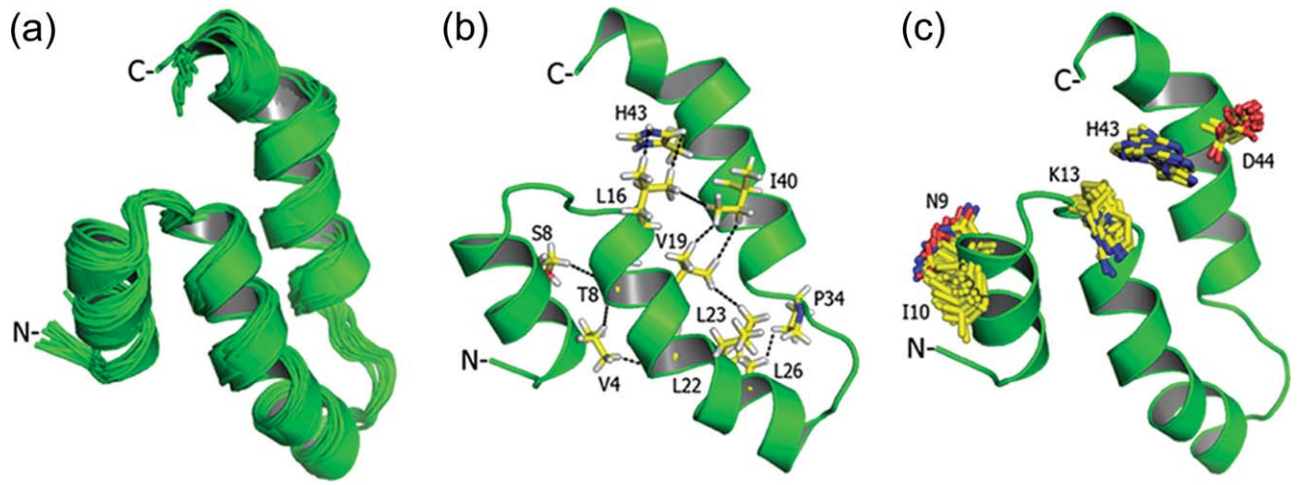

Figure 2

The NMR structure of the 1.5-repeat peptide of PthA. (**a**) Ensemble of the 20 lowest energy structures showing three helices connected by two turns. (**b**) The lowest energy structure showing the hydrophobic interactions between the side chains of residues VVAIA and LLPVL, which stabilize the structure. (**c**) Orientation of the hypervariable diresidues NI and HD relative to the K residue in the peptide structure.

tion radius of 40.82 ± 0.01 Å for the RD2 molecule. These values are in agreement with those obtained from the *in silico* modeling (see below) and indicate that RD2 is a monomer in solution. Analytical size-exclusion chromatography and DLS data also suggest that RD2 is a monomer in solution with a hydrodynamic radius of around 4.5 nm for a sample with low polydispersity index (15%).

The low resolution envelope of RD2 was derived without imposing any constraints [Fig. 3(b)]. Structural alignments of each individual molecular model, obtained by *ab initio* shape reconstruction, showed that they are very similar as indicated by the normalized spatial discrepancy (NSD) values, which are around 0.8. The RD2 model reveals an elongated or extended shape consistent with the notion that the repeat units are structurally arranged in tandem along the repeat region [Fig. 3(b)]. The elongated shape of the RD2 molecule can also be noted from the $p(r)$ curve profile [Fig. 3(a)].

Molecular modeling predicts a TPR-like superhelical structure for RD2

Fold assignment searches for RD2 using PSIPRED[48] returned 10 protein structures with significant scores, all displaying TPR domains. The crystal structures of the yeast importin β1 (PBD code: 2BPT) and human TIP120 (PDB code: 1U6G) were used as templates to generate a restraint-based 3D model of RD2 [Fig. 4(a)]. The model shows good local and global stereochemical properties with a *Z*-score of 6.8. Analyses of the Ramachandran plot indicate that 93% of the RD2 residues are in most favorable regions, 5% are in additional allowed regions and only 2% are in disallowed regions. In addition, local quality analysis assessed by plotting the energies as a function of the amino acid positions shows no positive values thus highlighting the good stereochemical quality of the model and its suitability for structural analysis and comparisons. The NMR structure of the 1.5-repeat peptide superimposes adequately with the RD2 model with an RMSD of 2.7 Å [Fig. 4(b)], thus corroborating the *in silico* model.

The RD2 3D model conserves all the structural features found in TPR-containing proteins encompassing 31 antiparallel α-helices that fold into a superhelical structure [Fig. 4(a)] and is in agreement with previous molecular modeling studies.[49] Furthermore, the superhelix model is consistent with the RD2 CD data showing an average 222/208 nm ratio >1.1, which is typical of coiled-coil regions. The repeat units of RD2 have the same amino acid segment length of TPRs and their superposition indicates that

Table I

Structural Statistics of the 1.5-Repeat Unit Peptide

NOEs	
Total number	982
Short range $\|i-j\| \leq 1$	650
Medium range $1 < \|i-j\| < 5$	156
Long range $\|i-j\| \geq 5$	176
CYANA	
Target function	1.01 ± 0.15
Distance violation >0.20	0
Dihedral angles >5°	0
RMSD	
Backbone (total)	0.72 ± 0.13
Side chain (total)	1.54 ± 0.17
PROCHECK	
Most favorable	84.1%
Additional allowed	15.3%
Generously allowed	0.6%
Not allowed	0.0%

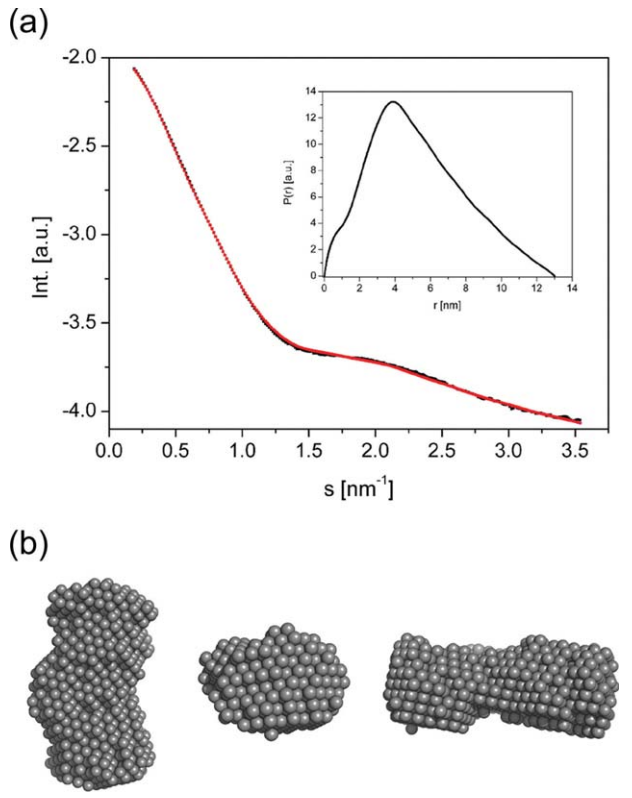

Figure 3

Ab initio shape determination of RD2. (**a**) Experimental SAXS curve and the theoretical fitting of the data (black and red, respectively) using the program GNOM. The inlet shows the pair distance distribution function $p(r)$. (**b**) Different views of the *ab initio* low resolution DAMMIN envelop model of RD2 in solution. The NSD values for a set of 15 DAMMIN and GASBOR models ranged from 0.85 to 0.95 in both calculations.

they are quite similar to one another (RMSD $\leq$ 1.5 Å). As observed in the structure of the 1.5-repeat peptide [Fig. 2(b)], the superhelix is stabilized by contacts of hydrophobic residues across the α-helical faces resulting in a continuous hydrophobic core. Additional polar interactions between close side chains contribute to the protein packing and stabilization. Notably, the *in silico* model fits remarkably well on the SAXS envelope and displays a NSD of 1.25, indicating high-shape complementarity between the *ab initio* envelope and the surface of the theoretical superhelix model [Fig. 4(c)].

TPR domains are a well-known protein scaffolds that mediate interprotein associations and assembly of multiprotein complexes in numerous organisms. The domain was first identified in yeast proteins required for mitosis and RNA synthesis as a degenerated 34-amino acid sequence arranged in tandem repeats.[50,51] Since then, studies have highlighted its diversity of arrangements and functions.[52,53] More recently, it was found that TPR domains can self-interact and promote protein oligomerization.[43,54,55] Thus, the TPR-superhelical structure of RD2 is consistent with the fact that RD2 also self-interacts in two-hybrid assays and associates with a number of citrus proteins, including the TPR domain of TDX.[14]

Conformational changes of RD2 upon interaction with DNA

Although TPR domains have not been reported to bind nucleic acids directly, they are found in numerous DNA and RNA-binding proteins.[51,53,56,57] To gain insights into how the RD2 superhelix could bind DNA with

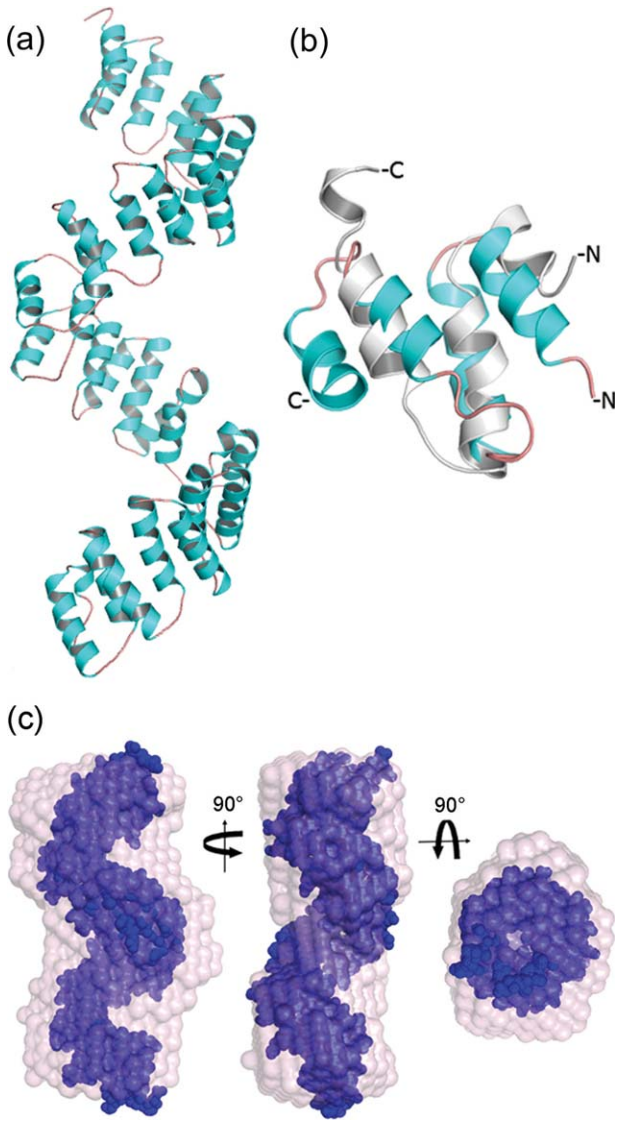

Figure 4

The structural model of RD2. (**a**) A cartoon representation of the TPR-like superhelical structure of RD2 depicting helices in cyan and loops in magenta. (**b**) Superposition of the NMR (gray) and modeled structures (cyan) of the PthA 1.5-repeat unit. (**c**) Orthogonal views illustrating the superposition of the TPR-like superhelical structure of RD2 onto its SAXS envelope.

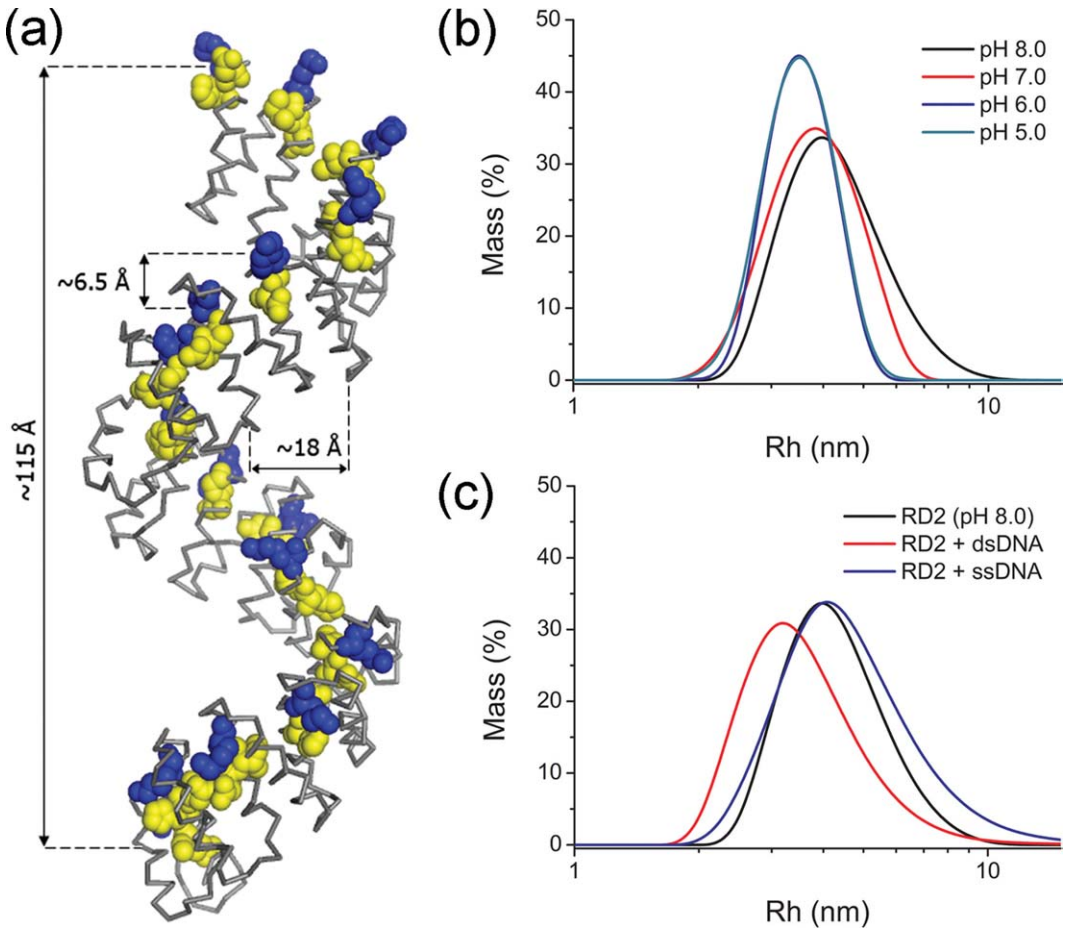

Figure 5

Conformational changes of RD2. (**a**) A ribbon representation of the RD2 TPR-like structure highlighting the relative location of the hypervariable diresidues (yellow spheres) and the lysines (blue spheres). The approximate dimensions of the height, cleft and spacing between two consecutive hypervariable diresidues are shown. DLS curves of RD2 at different pHs (**b**) and after incubation with the single-strand (ss) or double-strand (ds) DNA sequence ACACATTCTAAAATTTATATAAACCCTCATCCATTTCC derived from a citrus PthA-induced promoter (**c**).

sequence specificity, we looked at the spatial distribution of the polymorphic residues at Positions 12 (N or H) and 13 (I, D, or G). As observed in the 1.5-repeat peptide structure [Fig. 2(c)], the NI, HD, and NG pairs are located at the same side of the molecule at the tip of helix1, in close proximity to the K residues [Fig. 5(a)]. Furthermore, it is evident from the RD2 model that the polymorphic diresidues are mostly oriented to the inner face of the superhelix [Fig. 5(a)]. Thus, the only way the polymorphic residues could contact DNA would be the protein embracing the double strand, which is consistent with the fact that ligand binding in TPR proteins predominantly involves the concave surface of the TPR domain.[53] The atomic distances of the superhelix clefts (18 Å wide on average) supports this notion [Fig. 5(a)]. However, both the modeled structure and the SAXS envelop of RD2 have a molecular length of ~115 Å [Fig. 5(a)], which would cover ~34 nucleotides, that is, twice as many nucleotides as predicted to bind RD2.[15] Therefore, to be able to bind a 17-nucleotide target sequence, as predicted by the linear recognition model of one nucleotide per hypervariable diresidues,[15,16] the RD2 protein would have to undergo some degree of compaction, like a coil spring being compressed. Surprisingly, when the hydrodynamic behavior of RD2 was investigated, substantial changes in its hydrodynamic radius (R_h) due to pH changes were found [Fig. 5(b)]. At pH 8.0, the RD2 molecule displays an R_h of 4.5 nm that is consistent with the SAXS measurements done at similar pH. However, at more acidic pH, which might mimic the DNA surface, the R_h values are around 3.5 nm, indicating that RD2 can assume a more compact structure [Fig. 5(b)]. Similarly, RD2 significantly increased its compactness (R_h of 3.6 nm) in the presence of a DNA fragment derived from a citrus promoter upregulated by PthA2 [Fig. 5(c)]. Changes in the hydrodynamic radius was not clearly observed when the corresponding single-strand sense DNA was used [Fig. 5(c)], indicating that compaction preferentially occurs upon double-strand DNA interaction.

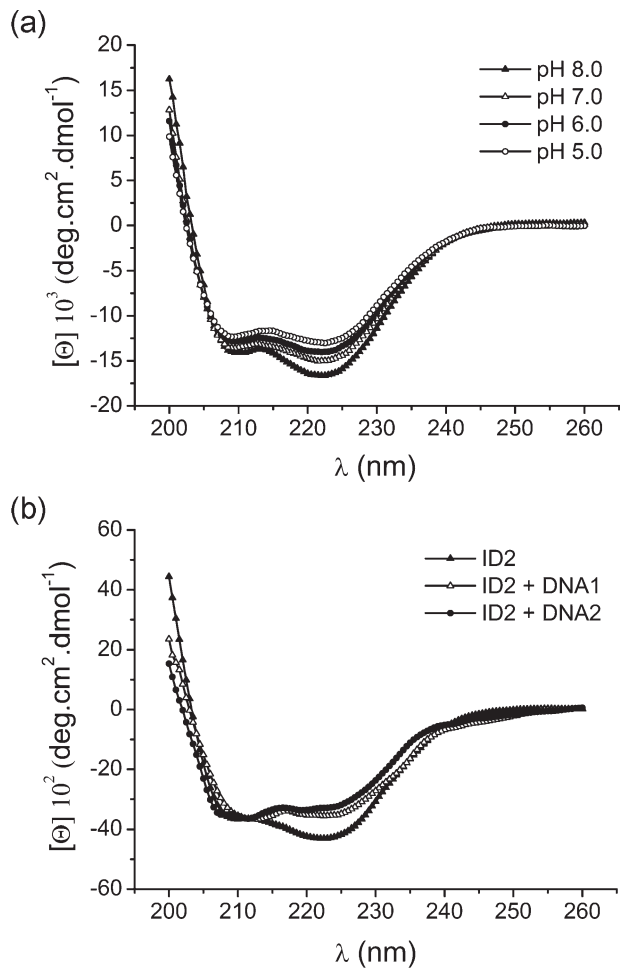

Figure 6

CD spectra of RD2 at different pHs and in the presence of DNA. CD curves of RD2 protein taken at different pHs (**a**) and after incubation with two DNA fragments (**c**), showing major changes in the 222/208 nm ratios at acidic pH and in the presence of DNA. The DNA1 sequence (ACACATTCTAAAATTTAT ATAAACCCTCATCCATTTCC) is derived from a citrus promoter upregulated by PthA2 whereas DNA2 contains the predicted PthA2 binding site[15] (ACAC ATTCTAAAATTTACACACCTCTTTTAATATTTCC).

To further investigate these conformational changes, the secondary structure of RD2 at different pHs and in the presence of DNA was analyzed by CD spectroscopy. At acidic pH, RD2 apparently looses some of its α-helical contents, as judged by the lower intensities of the 208 and 222 nm negative signals at pHs 7.0, 6.0, and 5.0, relative to pH 8.0 [Fig. 6(a)]. However, the 222/208 nm ratio shifted from 1.3 at pH 8.0 to nearly 1.0 at pH 5.0, indicating a decrease in coiled-coil-like conformations at acidic pHs. The 222/208 nm ratio of RD2 also diminished in the presence of DNA. This was observed not only with the citrus promoter fragment but also with a DNA fragment containing the predicted PthA2 binding site[15] [Fig. 6(b)]. Interestingly, it has been shown that a disulfide bond within the TPR region of the barley Sgt1 protein is thought to promote a compaction of the helical bundles.[43] The reduced Sgt1 has a CD 222/208 nm ratio of ~1.2, whereas oxidized Sgt1 has a ratio of ~0.9,[43] indicating that compaction of the TPR helical bundles is associated with a decrease in its coiled-coil-like conformation.

The repeat units of RD2 are similar to PPR motifs

As TPR domains show no obvious sequence similarities to the repeat units of RD2 and so far have not been reported to bind DNA directly, we searched for TPR-related motifs with known nucleic acid binding properties. Surprisingly, we found that the consensus sequence of PPR motifs[58,59] is very similar to that of the repeat units of RD2 (see Fig. 7). In addition, BLAST searches using the RD2 sequence as query also identified a number of Arabidopsis (Q9MA95.2, NP_187175.1) and rice (Os02g0555100, OsJ_07125, OsJ_12287) proteins containing PPR motifs.

PPRs are formed by 2–26 tandem arrays of a degenerate 35 amino acid motif with an average of 9–12 motifs per protein.[17,58] PPR proteins have been associated with the transcription and translational machineries, playing roles in mRNA stabilization and RNA editing.[17,18,60] Interestingly, some of the citrus proteins that were isolated in a two-hybrid screening as interactors of PthAs are associated with RNA stabilization.[14] Although no PPR structures are known, tandem PPR motifs are predicted to fold into a superhelical structure just like the TPR superhelix.[18,58] Accordingly, the CD spectrum of maize PPR5 resembles that of RD2[17] and a structural model of PPRs superposes well to the RD2 superhelix (results not shown). Thus, the structure of RD2 proposed here is consistent with the notion that nucleic acid-binding PPR motifs display a TPR-like fold. Although PPR proteins are almost absent in prokaryotes, the fact that they are predominantly found in plant mitochondria and chloroplasts[61] raises the possibility that PthAs and related type III effectors might have evolved from a common PPR ancestor protein.

```
ETVQRLLPVLCQAHGLTPEQVVAIASHDG-GKQAL RD2 rep2
ETVQRLLPVLCQAHGLTPDQVVAIASHDG-GKQAL RD2 rep6
ETVQRLLPVLCQAHGLTPQQVVAIASNGG-GKQAL RD2 rep8
EEA..LY..M....G..PN..TYNALINAYAK.G. PPR cons.
*  .  *   :    *  *:  .  *  .. .* .
```

Figure 7

The repeat units of RD2 are similar to PPR motifs. Protein sequence alignment between three polymorphic repeat units of RD2 and the consensus sequence from 1303 pentatricopeptide (PPR) motifs[57] performed by ClustalW2 (EMBL-EBI). Identical residues are shaded and indicated by asterisks, whereas similar residues are indicated by single or double dots. Dots in the PPR consensus sequence represent the most variable residues in PPR motifs.

In conclusion, this work provides the first experimental structural data of the PthA repeat domain. Both the NMR structure of the 1.5-repeat peptide and the SAXS envelope of RD2 corroborate the theoretical model that predicts a TPR-superhelical structure for RD2. Moreover, DLS and CD studies show that RD2 undergoes conformational rearrangements upon DNA interaction thus supporting the idea that the protein embraces the DNA with differential compactness as to cover a variable number of nucleotides.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the Laboratório Nacional de Biociências (LNBio) and the Laboratório Nacional de Luz Síncrotron (LNLS) for providing the NMR facilities and SAXS beamline time, respectively. They also thank Renata Rocha de Oliveira for technical assistance on the CD measurements.

REFERENCES

1. Brunings AM, Gabriel DW. *Xanthomonas citri*: breaking the surface. Mol Plant Pathol 2003;4:141–157.
2. Wichmann G, Bergelson J. Effector genes of *Xanthomonas axonopodis pv. vesicatoria* promote transmission and enhance other fitness traits in the field. Genetics 2004;166:693–706.
3. Cernadas RA, Camillo LR, Benedetti CE. Transcriptional analysis of the sweet orange interaction with the citrus canker pathogens *Xanthomonas axonopodis* pv. *citri* and *Xanthomonas axonopodis* pv. *aurantifolii*. Mol Plant Pathol 2008;9:609–631.
4. Cernadas RA, Benedetti CE. Role of Auxin and Gibberellin in citrus canker development and in the transcriptional control of cell-wall remodeling genes modulated by *Xanthomonas axonopodis* pv *citri*. Plant Sci 2009;177:190–195.
5. Duan YP, Castañeda GZ, Erdos G, Gabriel DW. Expression of a single, host-specific, bacterial pathogenicity gene in plant cells elicits division, enlargement, and cell death. Mol Plant Microbe Interact 1999;12:556–560.
6. Swarup S, Yang Y, Kingsley MT, Gabriel DW. An *Xanthomonas citri pathogenicity gene. pth A*, pleiotropically encodes gratuitous avirulence on nonhosts. Mol Plant Microbe Interact 1992;5:204–213.
7. Al-Saadi A, Reddy JD, Duan YP, Brunings AM, Yuan Q, Gabriel DW. All five host-range variants of *Xanthomonas citri* carry one PthA homolog with 17.5 repeats that determines pathogenicity on citrus, but none determine host-range variation. Mol Plant Microbe Interact 2007;20:934–943.
8. Kay S, Bonas U. How *Xanthomonas* type III effectors manipulate the host plant. Curr Opin Microbiol 2009;12:37–43.
9. Kay S, Hahn S, Marois E, Hause G, Bonas U. A bacterial effector acts as a plant transcription factor and induces a cell size regulator. Science 2007;318:648–651.
10. Römer P, Hahn S, Jordan T, Strauss T, Bonas U, Lahaye T. Plant pathogen recognition mediated by promoter activation of the pepper Bs3 resistance gene. Science 2007;318:645–648.
11. Kay S, Hahn S, Marois E, Wieduwild R, Bonas U. Detailed analysis of the DNA recognition motifs of the *Xanthomonas* type III effectors AvrBs3 and AvrBs3Deltarep16. Plant J 2009;59:859–871.
12. Römer P, Strauss T, Hahn S, Scholze H, Morbitzer R, Grau J, Bonas U, Lahaye T. Recognition of AvrBs3-like proteins is mediated by specific binding to promoters of matching pepper Bs3 alleles. Plant Physiol 2009;150:1697–1712.
13. Gürlebeck D, Szurek B, Bonas U. Dimerization of the bacterial effector protein AvrBs3 in the plant cell cytoplasm prior to nuclear import. Plant J 2005;42:175–187.
14. Domingues MN, Souza TA, Cernadas RA, de Oliveira MLP, Docena C, Farah CS, Benedetti CE. The *Xanthomonas citri* effector protein PthA interacts with citrus proteins involved in nuclear transport, protein folding and ubiquitination associated with DNA repair. Mol Plant Pathol 2010;11:663–675.
15. Boch J, Scholze H, Schornack S, Landgraf A, Hahn S, Kay S, Lahaye T, Nickstadt A, Bonas U. Breaking the code of DNA-binding specificity of TAL-Type III effectors. Science 2009;326:1509–1512.
16. Moscou MJ, Bogdanove AJ. A simple cipher governs DNA recognition by TAL Effectors. Science 2009;326:1501.
17. Williams-Carrier R, Kroeger T, Barkan A. Sequence-specific binding of a chloroplast pentatricopeptide repeat protein to its native group II intron ligand. RNA 2008;14:1930–1941.
18. Pfalz J, Bayraktar OA, Prikryl J, Barkan A. Site-specific binding of a PPR protein defines and stabilizes 5′ and 3′ mRNA termini in chloroplasts. EMBO J 2009;28:2042–2052.
19. Aue WP, Bartholdi E. Ernst RR2-dimensional spectroscopy—application to nuclear magnetic-resonance. J Chem Phys 1976;64:2229–2246.
20. Braunschweiler L, Erns RR. Coherence transfer by isotropic mixing: application to proton correlation spectroscopy. J Magn Reson 1983;53:521–528.
21. Kumar A, Ernst RR, Wüthrich K. A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton–proton crossrelaxation networks in biological macromolecules. Biochem Biophys Res Commun 1980;95:1–6.
22. States DJ, Haberkorn RA, Ruben DJ. A two-dimensional nuclear overhauser experiment with pure absorption phase in 4 quadrants. J Magn Reson 1982;48:286–292.
23. Piotto M, Saudek V, Sklenar V. Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. J Biomol NMR 1992;2:661–665.
24. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR 1995;6:277–293.
25. Güntert P. Automated nmr structure calculation with cyana. Methods Mol Biol 2004;278:353–378.
26. Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang J-S, Kuszewski J, Nilges M, Panuu NS, Read RJ, Rice LM, Simonson T, Warren LG. Crystallography and NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr 1998;54:905–921.
27. Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR 1996;8:477–486.
28. Guinier A, Fournet G. Small angle scattering of X-rays. New York: Wiley; 1955. pp17–19.
29. Konarev PV, Volkov VV, Sokolova AV, Koch MHJ, Svergun DI. PRIMUS: a windows PC-based system for small-angle scattering data analysis. J Appl Crystallogr 2003;36:1277–1282.
30. Svergun DI. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. J Appl Crystallogr 1992;25:495–503.
31. Svergun DI. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. Biophys J 1999;76:2879–2886.
32. Volkov VV, Svergun DI. Uniqueness of *ab initio* shape determination in small-angle scattering. J Appl Crystallogr 2003;36:860–864.
33. Kozin MB, Svergun DI. Automated matching of high- and low-resolution structural models. J Appl Crystallogr 2001;34:33–41.
34. Liu SM, Stewart M. Structural basis for the high-affinity binding of nucleoporin Nup1p to the *Saccharomyces cerevisiae* importin-beta homologue, Kap95p. J Mol Biol 2005;349:515–525.

35. Goldenberg SJ, Cascio TC, Shumway SD, Garbutt KC, Liu J, Xiong Y, Zheng N. Structure of the Cand1-Cul1-Roc1 complex reveals regulatory mechanisms for the assembly of the multisubunit cullin-dependent ubiquitin ligases. Cell 2004;119:517–528.
36. Fiser A, Sali A. Modeller: generation and refinement of homology-based protein structure models. Methods Enzymol 2003;374:461–491.
37. Lindahl E, Hess B, van der Spoe DL. Gromacs 3.0: a package for molecular simulation and trajectory analysis. J Mol Model 2001;7:306–317.
38. Eisenberg D, Luthy R, Bowie JU. VERIFY3D: assessment of protein models with three-dimensional profiles. Methods Enzymol 1997;277:396–404.
39. Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res 2007;5:407–410.
40. Willard L, Ranjan A, Zhang H, Monzavi H, Robert FB, Syakes BD, Wishart DS. VADAR: a web server for quantitative evaluation of protein structure quality. Nucleic Acids Res 2003;31:3316–3319.
41. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr 2004;60:2126–2132.
42. Montserret R, McLeish MJ, Böckmann A, Geourjon C, Penin F. Involvement of electrostatic interactions in the mechanism of peptide folding induced by sodium dodecyl sulfate binding. Biochemistry 2000;39:8362–8373.
43. Nyarko A, Mosbahi K, Rowe AJ, Leech A, Boter M, Shirasu K, Kleanthous C. TPR-mediated self-association of plant SGT1. Biochemistry 2007;46:11331–11341.
44. Cortajarena AL, Regan L. Ligand binding by TPR domains. Protein Sci 2006;15:1193–1198.
45. Carrigan PE, Sikkink LA, Smith DF, Ramirez-Alvarado M. Domain:domain interactions within Hop, the Hsp70/Hsp90 organizing protein, are required for protein stability and structure. Protein Sci 2006;15:522–532.
46. Mirus O, Bionda T, von Haeseler A, Schleiff E. Evolutionary evolved discriminators in the 3-TPR domain of the Toc64 family involved in protein translocation at the outer membrane of chloroplasts and mitochondria. J Mol Model 2009;15:971–982.
47. Luscombe NM, Laskowski RA, Thornton JM. Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. Nucleic Acids Res 2001;29:2860–2874.
48. McGuffin LJ, Bryson K, Jones DT. The PSIPRED protein structure prediction server. Bioinformatics 2000;16:404–405.
49. Schornack S, Meyer A, Römer P, Jordan T, Lahaye T. Gene-for-gene-mediated recognition of nuclear-targeted AvrBs3-like bacterial effector proteins. J Plant Physiol 2006;163:256–272.
50. Sikorski RS, Boguski MS, Goebl M, Hieter P. A repeating amino acid motif in CDC23 defines a family of proteins and a new relationship among genes required for mitosis and RNA synthesis. Cell 1990;60:307–317.
51. Hirano T, Kinoshita N, Morikawa K, Yanagida M. Snap helix with knob and hole: essential repeats in *S. pombe* nuclear protein nuc2+. Cell 1990;60:319–328.
52. Lamb JR, Tugendreich S, Hieter P. Tetratrico peptide repeat interactions: to TPR or not to TPR? Trends Biochem Sci 1995;20:257–259.
53. D'Andrea LD, Regan L. TPR proteins: the versatile helix. Trends Biochem Sci 2003;28:655–662.
54. Bansal PK, Nourse A, Abdulle R, Kitagawa K. Sgt1 dimerization is required for yeast kinetochore assembly. J Biol Chem 2009;284:3586–3592.
55. Krachler AM, Sharma A, Kleanthous C. Self-association of TPR domains: lessons learned from a designed, consensus-based TPR oligomer. Proteins 2010;78:2131–2143.
56. Weber P, Fulgosi H, Piven I, Müller L, Krupinska K, Duong VH, Herrmann RG, Sokolenko A. TCP34, a nuclear-encoded response regulator-like TPR protein of higher plant chloroplasts. J Mol Biol 2006;357:535–549.
57. Tanaka A, Takano Y, Ohnishi Y, Horinouchi S. AfsR recruits RNA polymerase to the afsS promoter: a model for transcriptional activation by SARPs. J Mol Biol 2007;369:322–333.
58. Small ID, Peeters N. The PPR motif—a TPR-related motif prevalent in plant organellar proteins. Trends Biochem Sci 2000;25:46–47.
59. Bentolila S, Alfonso AA, Hanson MR. A pentatricopeptide repeat-containing gene restores fertility to cytoplasmic male-sterile plants. Proc Natl Acad Sci USA 2002;99:10887–10892.
60. Pfalz J, Liere K, Kandlbinder A, Dietz KJ, Oelmüller R. pTAC2, -6, and -12 are components of the transcriptionally active plastid chromosome that are required for plastid gene expression. Plant Cell 2006;18:176–197.
61. O'Toole N, Hattori M, Andres C, Iida K, Lurin C, Schmitz-Linneweber C, Sugita M, Small I. On the expansion of the pentatricopeptide repeat gene family in plants. Mol Biol Evol 2008;25:1120–1128.