Structure of the DNA-binding domain of NgTRF1 reveals unique features of plant telomere-binding proteins

Sunggeon Ko[1,3,†], Sung-Hoon Jun[2,†], Hansol Bae[2], Jung-Sue Byun[2,3], Woong Han[1,3], Heeyoung Park[1,3], Seong Wook Yang[2], Sam-Yong Park[4], Young Ho Jeon[5], Chaejoon Cheong[5], Woo Taek Kim[2,*], Weontae Lee[1,3,*] and Hyun-Soo Cho[2,3,*]

[1]Department of Biochemistry, [2]Department of Biology, [3]Protein Network Research Center, College of Life Sciences and Biotechnology, Yonsei University, Seoul 120-749, Korea, [4]Protein Design Laboratory, Yokohama City University, Suehiro 1-7-29, Tsurumi-ku, Yokohama 230-0045, Japan and [5]Magnetic Resonance Team, Korea Basic Science Institute (KBSI), Ochang, Chungbuk 363-883, Korea

ABSTRACT

Telomeres are protein–DNA elements that are located at the ends of linear eukaryotic chromosomes. In concert with various telomere-binding proteins, they play an essential role in genome stability. We determined the structure of the DNA-binding domain of NgTRF1, a double-stranded telomere-binding protein of tobacco, using multi-dimensional NMR spectroscopy and X-ray crystallography. The DNA-binding domain of NgTRF1 contained the Myb-like domain and C-terminal Myb-extension that is characteristic of plant double-stranded telomere-binding proteins. It encompassed amino acids 561–681 (NgTRF1$^{561-681}$), and was composed of 4 α-helices. We also determined the structure of NgTRF1$^{561-681}$ bound to plant telomeric DNA. We identified several amino acid residues that interacted directly with DNA, and confirmed their role in the binding of NgTRF1 to telomere using site-directed mutagenesis. Based on a structural comparison of the DNA-binding domains of NgTRF1 and human TRF1 (hTRF1), NgTRF1 has both common and unique DNA-binding properties. Interaction of Myb-like domain with telomeric sequences is almost identical in NgTRF1$^{561-681}$ with the DNA-binding domain of hTRF1. The interaction of Arg-638 with the telomeric DNA, which is unique in NgTRF1$^{561-681}$, may provide the structural explanation for the specificity of NgTRF1 to the plant telomere sequences, (TTTAGGG)$_n$.

INTRODUCTION

Telomeres are essential for eukaryotic genome stability (1). During the last decade, telomeres have been the subject of intense study because of the link between telomere function and cancer and aging (2,3). Telomeric DNA consists of tandem repeats of simple conserved sequences, and functions in maintaining the integrity of flanking chromosomal sequences during replication (1,4). Telomeric DNA that is shortened during replication is restored through the action of telomerase, a reverse-transcriptase that synthesizes telomeric DNA using its own RNA molecule as a template (5,6). The synthesis of telomeres by telomerase and telomere length is regulated by numerous telomere-binding proteins.

While the function of telomere-binding proteins in the regulation of telomere length is well characterized (7), other functions of them have also been described. The telomere-binding protein complex enables the DNA repair machinery to distinguish telomere ends from double-stranded DNA breaks. Defects in the telomere-binding protein complex trigger DNA damage response pathways that arrest the cell cycle and activate cell senescence or

apoptosis (8–10). Telomere-binding proteins also protect telomeres from inappropriate DNA repair reactions, such as end-to-end joining and exonucleolytic digestion (11). In humans, six telomere-specific proteins have been known to form a complex (12). Of them, three proteins, hTRF1, hTRF2 and hPOT1, directly bind to telomeric DNA sequences and they are interconnected by two additional proteins, hTIN2 and hTPP1. hTRF1 and hTRF2 are double-stranded DNA-binding proteins, while hPOT1 binds to single-stranded DNA. hTRF1 forms homodimers, and possesses a Myb-like domain through which it binds to specific DNA sequences (13–15). The role of hTRF1 in the regulation of telomere length has been demonstrated by gain-of-function studies (7,15) in which overexpression of wild-type allele caused telomere length to shorten and expression of a dominant negative allele resulted in progressive elongation of telomeres, until a new equilibrium was achieved. hTRF2 is a paralog of hTRF1 and its primary function is in telomere capping, which prevents end-to-end joining (11,16,17). hPOT1 has been proposed to function downstream of hTRF1 to relay the negative regulation to the telomere terminus (18). Several telomeric proteins have been identified in yeast. In the budding yeast *Saccharomyces cerevisiae*, RAP1 (scRAP1) is a double-stranded DNA-binding protein, and the primary telomere-binding protein (19–21). Excess scRAP1 bound at the telomere negatively regulates telomere elongation in *cis* through the inhibition of telomerase activity (22,23). However, while scRAP1 is functionally analogous to hTRF1, the two proteins are not homologous. In contrast, fission yeast contains an ortholog of hTRF1, TAZ1, which binds to telomeric DNA duplexes and negatively regulates telomere length (24).

The biological functions of telomeres and telomere-binding proteins have been studied extensively in humans and yeast. Double- or single-stranded telomere-binding proteins in plants have also been identified, which indicates that this class of proteins has been conserved throughout evolution (25–28). In *Arabidopsis*, there are at least 12 *TRF-like* (*TRFL*) genes that have a single Myb-like domain in their C-terminal region and they fall into two distinct gene families based on the presence or absence of the C-terminal Myb-extension (29). Recombinant TRFL family 1 proteins, which contain C-terminal Myb-extension form homo- and hetero-dimers and specifically interact with plant double-stranded telomeric DNA *in vitro*. TRFL family 2 proteins lack the C-terminal Myb-extension, similarly to nonplant telomere-binding proteins such as hTRF1 and hTRF2, but they cannot bind to telomeric DNA. Single myb histone (SMH) family proteins, which have a single Myb-like domain in their N-terminal region, also bind telomere DNA repeats *in vitro* and they are plant specific (30,31). The protein AtTRB1, a member of SMH family, interacts with the *Arabidopsis* homolog protein of hPOT1, AtPot1, suggesting its plant telomere-specific role (32). The physiological functions of telomere-binding proteins in plant have been studied recently. The expression of NgTRF1, a tobacco double-stranded telomere-binding protein, is regulated tightly in correlation with

cell division and the cell cycle (27). Overexpression of NgTRF1 resulted in a shorter telomere length compared to wild-type plants, whereas decreased expression of NgTRF1 resulted in a longer telomere length (33). Moreover, these perturbations of the expression of NgTRF1 caused apoptotic cell death. Recently, the *in vivo* function of rice telomere-binding protein, RTBP1, has been studied at the plant level (34). Loss-of-function (amorphic or hypomorphic) mutants of *RTBP1* exhibited defects in both vegetative and reproductive development, and these phenotypes correlated with the gradual acquisition of dysfunctional telomeres.

The structures of double-stranded telomere-binding proteins also have been studied mainly in human and yeast (35–38). The structures of full-length telomere-binding proteins have not been reported due to the presence of flexible linker regions within these proteins. Therefore, attention has focused on the structures of the Myb-like DNA-binding domains to understand telomere-binding mechanism. The DNA-binding domains of human double-stranded telomere-binding proteins, hTRF1 and hTRF2, are composed of three α-helices, and form a helix–turn–helix DNA-binding motif (35,37). The complex structure of the Myb-like domain of hTRF1 and telomeric DNA shows that the helix–turn–helix motif recognizes telomeric DNA sequences in the major groove and that the N-terminal region interacts with DNA in the minor groove (35,38). The crystal structure of hTRF2 DNA-binding domain in complex with telomeric DNA shows that it recognizes the same sequence as hTRF1 (35). RAP1 of budding yeast, scRAP1, contains two subdomains that are closely related in structure to Myb domains. The two subdomains are connected by a long linker, and they recognize and bind to two independent tandem telomeric repeats (36). The structure of the DNA-binding domain of *Arabidopsis* TRP1 (AtTRP1) was recently reported using NMR spectroscopy. It is composed of 4 α-helices suggesting that plant telomere-binding proteins have a unique DNA-binding domain compared to other organisms (39). Chemical shift perturbation assay suggested that helix 3 and the flexible loop connecting helix 3 and helix 4 are involved in the recognition of telomeric DNA sequence. Moreover, telomere DNA sequences have been identified as $(TTAGGG)_n$ in majority of eukaryotic organisms but $(TTTAGGG)_n$ in plants (40–42). However, the exact overall picture of how plant double-stranded telomere-binding proteins recognize plant telomeric sequence, $(TTTAGGG)_n$, and the role of plant-specific C-terminal Myb-extension in the telomeric sequence recognition is unclear. To address these questions, we solved the structure of the DNA-binding domain of NgTRF1 in complex with telomeric DNA. The molecular details of the interaction between the DNA-binding domain of NgTRF1 and plant telomeric DNA suggests that the interaction mode of plant and human double-stranded telomere-binding proteins are highly conserved during evolution. However, the plant-specific C-terminal Myb extension is required for the specific recognition by NgTRF1 of the sequence $(TTTAGGG)_n$, which is specific for plant telomeres.

MATERIALS AND METHODS

Cloning, protein expression and purification and DNA preparation

DNA fragments encoding the deletion mutants of NgTRF1 were cloned into pPROEX (Invitrogen) for hexa-histidine tagging proteins expression and the plasmid pGEX4T-1 (Amersham Biosciences) for glutathione-S-transferase (GST) fusion proteins expression. For structural studies, hexa-histidine tagging proteins (pPROEX vector) were used. The DNA-binding domain of NgTRF1, $NgTRF1^{561-681}$, was overexpressed in *Escherichia coli* BL21 (DE3)-CodonPlus strain (Stratagene). The cells were grown at 37°C to an optical density at 600 nm (OD_{600}) of 0.5, then 1 mM isopropyl-1-thio-β-D-galactopyranoside (IPTG) was added to induce protein overexpression at 25°C. After an additional 8 h of growth, the cells were harvested and subjected to centrifugation.

The cell pellets were suspended in 25 mM sodium phosphate (pH 7.0), 100 mM NaCl and sonicated. Recombinant proteins were initially purified by Ni–NTA column (Amersham Biosciences), followed by TEV protease digestion and a second round of Ni–NTA chromatography to remove the fusion tags. Further purification of the protein was carried out using a Superdex 75 gel-filtration column (Amersham Biosciences).

For DNA interaction studies using EMSA assay, GST-fusion proteins (pGEX4T-1 vector) were used. Several NgTRF1 mutants (Figure 1) were overexpressed in *E. coli* BL21 (DE3)-CodonPlus strain (Stratagene). For overexpression of NgTRF1 1 mM IPTG was used when the cell growth reached optical density at 600 nm (OD_{600}) of 0.6 in 37°C in shaking incubator. After IPTG induction, the cells were incubated at 25°C with gently shaking. After 8–10-h incubation at 25°C, the cells were harvested and subjected to centrifugation. When the cells were sonicated in 25 mM sodium phosphate (pH 7.0), 100 mM NaCl and centrifuged to remove cell debris, supernatants was loaded on glutathione-Sepharose™ high performance resin (Amersham Biosciences) and GST-fusion NgTRF1 was purified. For removing GST-tag, thrombin protease (Amersham Biosciences) was used. Superdex 75 gel-filtration column (Amersham Biosciences) was applied to further purification.

Complementary strands of a 14mer consisting of two repeats of the telomeric DNA sequence (5′-TTTAGGG TTTAGGG-3′) were chemically synthesized and solved in distilled water. For annealing, each oligomers of same molar ratio were mixed and put in the 94°C Dry-Bath (Barnstead Co., Ltd) for 5 min. The Dry-Bath containing DNA mixture was cooled down slowly in room temperature. To remove additive chemicals from DNA synthesis and do NMR experiments, DNA solution was dialyzed to protein buffer (pH 7.0 in 25 mM sodium phosphate buffer with 100 mM NaCl) using dialysis membrane, MWCO 5 KDa (Spectra/Por® dialysis Co., Ltd) for 12 h. 1D NMR experiment in Bruker DRX 500 MHz confirmed double-strand DNA. DNA concentrations were determined by measuring absorbance at 260 nm.

NMR spectroscopy

The telomere DNA-binding domain of NgTRF1 was prepared in a solution of 90% H_2O, 10% D_2O or 99.9% D_2O, pH 7.0 in 25 mM sodium phosphate buffer, 100 mM NaCl. The purified DNA-binding domain of NgTRF1 was concentrated to 1 mM by centrifugation using an Amicon filter unit (Milipore). All NMR spectra were recorded at 298 K using Bruker DRX500 MHz and DRX600 MHz spectrometers equipped with a tripleresonance inversed probe with x, y, z gradient shielding. A cryoprobe was also used. ¹H chemical shifts were referenced directly to internal 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS). ¹⁵N and ¹³C shifts were referenced indirectly. The strong solvent resonance was suppressed by water-gated pulse sequence combined with pulsed-field-gradient (PFG) pulses. Backbone and Cβ resonances were assigned using the following techniques in succession: two-dimensional (2D) ¹H-¹⁵N HSQC, constant-time-¹H-¹³C HSQC and 3D HNCO, HNCACB, CBCA(CO)NH and HNCA spectra. In some experiments, HN(CA)CO and HN(CO)CA were also collected. Side-chain and Hα assignments were obtained using HBHANH, H(CC)(CO)NH-TOCSY, ¹⁵N-edited NOESY, ¹³C-edited NOESY, HCCH-TOCSY and HNHA spectra. As a final step, HCCH-TOCSY was collected after solvent exchange to D_2O. Distance restraints for the DNA-binding domain of NgTRF1 were obtained using ¹⁵N-edited NOESY and ¹³C-edited NOESY spectra, with mixing times of 100–150 ms in 90% H_2O, 10% D_2O or 99% D_2O to extract NOE information. Slowly exchanging amide protons were identified by lyophilizing a fully protonated sample in H_2O to dryness, re-dissolving it in a solution of 99.99% D_2O, and then acquiring the 2D ¹H-¹⁵N HSQC spectrum immediately, or after 1 day.

¹⁵N-¹H NOE values were calculated as the ratio of the intensities of paired ¹⁵N-¹H correlation peaks from interleaved spectra acquired with and without ¹H presaturation during a recycle time of 5 s.

All NMR data were processed using Bruker XWIN-NMR (Bruker Instruments) and NMRPipe/NMRDraw software (43) on a Linux-operating PC workstation and analyzed (resonance assignments and cross-peak picking/integration, etc.) using Sparky 3.60 software. In the acquisition dimension the small residual water resonance was removed by a solvent-suppression time domain filter, zero-filled to twice the size and Fourier-transformed. All indirect dimensions were processed using a linear prediction (LP) to enhance resolution. The size of the ¹⁵N time domain was doubled by mirror image LP. Forward–backward LP was applied to the ¹³C and ¹H domains. HNCO was used to resolve overlap in ¹H-¹⁵N HSQC spectra.

NMR structural calculations

Distance restraints were derived from cross-peaks in ¹⁵N-edited NOESY ($\tau m = 100$ and 150 ms) and ¹³C-edited NOESY ($\tau m = 150$ ms) spectra. Slowly exchanging amide protons were identified by acquiring a series of ¹H-¹⁵N HSQC spectra after dissolving lyophilized protein into 100% D_2O. Angle constraints were obtained
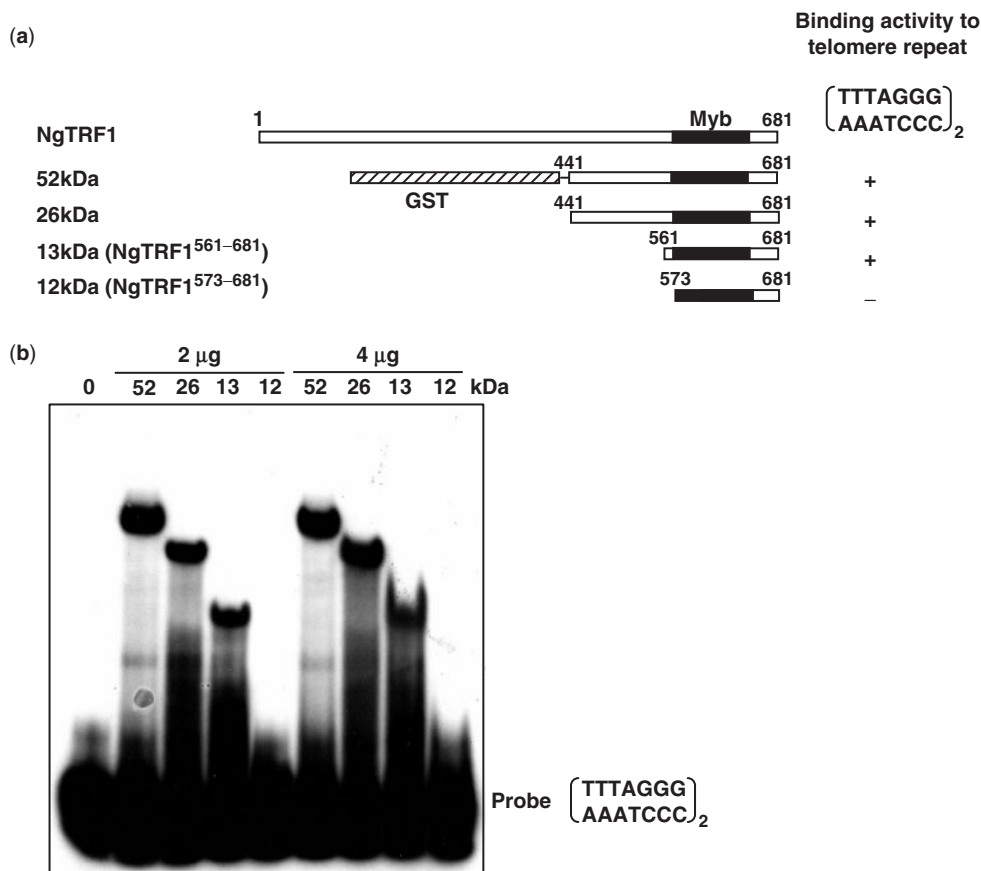
Figure 1. Mapping of the DNA-binding domain of NgTRF1. (**a**) Schematic representation of full-length and deletion mutants of NgTRF1. GST is represented by the shaded box; black boxes represent the Myb-like domain of NgTRF1; regions outside the Myb-like domain are represented by open boxes. The molecular mass of each mutant polypeptide is indicated in the left column. The binding activity of each mutant protein is presented in the right column. (**b**) Gel retardation assays. Total of 2 or 4 µg of the indicated purified protein were incubated with 0.25 ng of radiolabeled double-stranded telomeric DNA (TTTAGGG)₂, and then subjected to electrophoresis on a nondenaturing polyacrylamide gel. Protein–DNA complexes were visualized by autoradiography.

from the TALOS prediction (44). Protein structure was calculating using the CYANA program version 2.1, which combines automated assignment of NOE cross-peaks and structural calculations. Chemical shift tolerances were set at 0.02 p.p.m. for protons, and 0.3 p.p.m. for nitrogen and carbon. Additional tolerances were set at 0.03 p.p.m. and 0.04 p.p.m. for protons and heavy atoms, respectively. NMR-derived experimental restraints contained 1168 unambiguous NOEs [175 intraresidue, 296 sequential, 206 medium-range $(2 < = |i - j| < 5)$ and 490 long-range $(|i - j| > = 5)$ NOE constraints], 86 distance restraints for 43 backbone hydrogen bonds and 168 dihedral angle restraints, which were used for the structural calculations. Seven cycles of the CYANA routine were performed, and each cycle consisted of 10 000 steps of torsion-angle dynamics with a simulated annealing protocol. 100 structures were calculated and 20 structures with the lowest target function values were chosen for analysis. Structures with the lowest NOE energies were retained and validated using PROCHECK (45). Structures were analyzed and visualized using PyMOL (DeLano Scientific LLC, San Carlos, CA) and MOLMOL (46).

Crystallization and X-ray data collection

Purified NgTRF1[561–681] was concentrated to 10 mg ml^{-1} and crystallized using the hanging-drop vapor diffusion method at 290 K. Crystals were grown in a solution of 0.1 M Tris pH 8.5 and 25% (w/v) PEG 3350. The crystals were soaked with 1 mM Pt(NO₃)₄ for 2 weeks to prepare platinum-derivative crystals and then transferred to a cryoprotective solution containing 20% (v/v) glycerol. A diffraction data set at 1.9 Å was collected on beam line 5A of the Photon Factory (PF) in Japan. The NgTRF1[561–681]–DNA complex was prepared by mixing purified protein and DNA at a molar ratio of 2:1 (protein:DNA). After 2 h incubation at 4°C, the complex was purified by Superdex 75 gel-filtration chromatography (Amersham Biosciences) and concentrated to 24 mg ml^{-1}. The NgTRF1[561–681]–DNA complexes were crystallized in a solution of 100 mM Bis–Tris pH6.5, 50 mM CaCl₂ and 28% PEG 550 MME and then cryoprotected in 20% glycerol. A diffraction data set at 2.4 Å was collected on beam line 4A at the Pohang Accelerator Laboratory, Korea. Data sets were processed and scaled using DENZO and SCALEPACK (47) from the HKL package.

Crystal structure determination

The crystal of NgTRF1[561–681] belonged to space group $P2_12_12_1$ and had unit-cell dimensions of $a = 40.12$ Å, $b = 48.26$ Å and $c = 52.01$ Å. Assuming one molecule per asymmetric unit, the Matthews coefficient V_M was calculated as 2.1 Å^3/Da, which corresponded to a solvent content of 40.8%. The structure was determined using the multi-wavelength anomalous dispersion (MAD) method with platinum-derivative crystals. Phasing and automatic tracing were performed using the program of Shelx and Sharp programs (48). The expected platinum sites were identified using the program Shelx, and density modification and automated model building were carried out using Sharp. Tracings of the C$_\alpha$ atoms were done using the O program (49) until most of the residues were fitted into the electron density map. Using the CNS program suite (50), several cycles of simulated annealing, minimization and B group refinement followed by manual model rebuilding were carried out. Final R values for data in the resolution range of 20.0–1.9 Å were as follows: $R_{factor} = 21.0\%$ and $R_{free} = 23.3\%$. The final model did not include the following residues: residues 561–577 and 661–681. These amino acid residues constituted flexible loops whose electron densities were very weak or undetectable.

The crystal of NgTRF1[561–681] in complex with telomeric DNA belonged to space group P3$_2$ with unit-cell dimensions of $a = b = 70.84$ Å and $c = 68.71$ Å. It contained two complexes per asymmetric unit. The structure was determined by molecular replacement using the CCP4 version of MOLREP (51) and the structure of NgTRF1[561–681] as a search model. The final model does not include the following residues: residues 561–573 and 661–681. These amino acid residues constitute flexible loops whose electron densities were very weak or undetectable.

Site-directed mutagenesis and gel retardation assay

Site-directed mutagenesis was carried out using the QuikChange[TM] site-directed mutagenesis system (Stratagene). Preparation of proteins for the gel mobility shift assay was described previously (27).

Isothermal titration calorimetry (ITC)

ITC was performed using a VP-ITC system (Microcal Inc.) at 25°C in a solution of 25 mM sodium phosphate (pH 7.0), 100 mM NaCl. NgTRF1[561–681] and telomeric DNA were dialyzed extensively against buffer prior to analysis. The final concentrations of NgTRF1[561–681] and DNA were 40 μM and 210 μM, respectively. The DNA solution was injected 17 times into a 1.8-ml sample cell containing NgTRF1[561–681] and analysis was carried out using Microcal Origin software. Individual injections were integrated following manual adjustment of the baseline. Heats of dilution and mixing were determined from a separate control experiment or from the end-point of the titration in the same way described above, and the values were subtracted prior to curve fitting using a one-site model.

NMR titration of DNA binding

Based on imino proton resonances (10–14 p.p.m.), the NgTRF1[561–681]–DNA complex formation was monitored by recording 1D spectrum with molar ratio of DNA and protein of 1:0.5, 1:1, 1:1.5, 1:2, 1:2.5, respectively.

RESULTS

Identification of the DNA-binding domain of NgTRF1 for structural analysis

NgTRF1 was identified as a double-stranded telomeric DNA-binding protein in tobacco (27). Similar to the double-stranded telomeric DNA-binding proteins of other species, NgTRF1 contains a Myb-like domain in its C-terminal region (Figure 1a), which has a high level of homology to that of the double-stranded telomere-binding proteins of other organisms. In contrast to hTRF1, however, plant telomere-binding proteins, including NgTRF1, contain an additional C-terminal extension of the Myb-like domain (C-terminal Myb-extension) that is highly conserved (27,29,52). Previously, it was shown that the Myb-like domain and the C-terminal Myb-extension mediate binding to telomeric DNA (27,29,52). To map the DNA-binding domain of NgTRF1, we generated a set of deletion mutants, and examined their ability to bind to plant telomeric DNA using an electrophoretic mobility shift assay (Figure 1). NgTRF1[573–681], which consisted of the Myb-like domain and the C-terminal Myb-extension, was not sufficient for binding to the telomeric sequence (TTTAGGG)$_2$. In contrast, NgTRF1[561–681], which contained 12 additional N-terminal amino acids, exhibited DNA-binding activity. To confirm the DNA-binding properties of NgTRF1[561–681], we performed a titration experiment using one-dimensional NMR and increasing molar ratios of DNA:protein [1:0.5, 1:1, 1:1.5, 1:2, 1:2.5; (TTTAGGG)$_2$: NgTRF1[561–681]]. We observed significant chemical shift perturbations (in the range of 10–14 p.p.m.) of the imino hydrogen atoms of the DNA bases upon the addition of protein (data not shown), which provided strong evidence that NgTRF1[561–681] binds to DNA. Based on these results, we used NgTRF1[561–681] for subsequent structural study.

Structure determination

We first analyzed NgTRF1[561–681] using NMR spectroscopy. All peptide backbone resonance assignments of NgTRF1[561–681] were completed with the exception of the 3 N-terminal residues, which were not resolved due to resonance overlap (Figure 2a). Residues 582–660 adopt a well-defined tertiary structure and it was refined to a root mean square deviation (RMSD) of 0.86 Å for backbone atoms. Most of the Φ, Ψ angles of the final structures were appropriately distributed in the Ramachandran plot and the structural statistics are presented in Table 1.

We also determined the crystal structure of a platinum-derivative of NgTRF1[561–681] using MAD phasing. The crystal belonged to space group of $P2_12_12_1$ with one molecule per asymmetric unit. The electron density maps of the 17 N-terminal residues and the 21 C-terminal
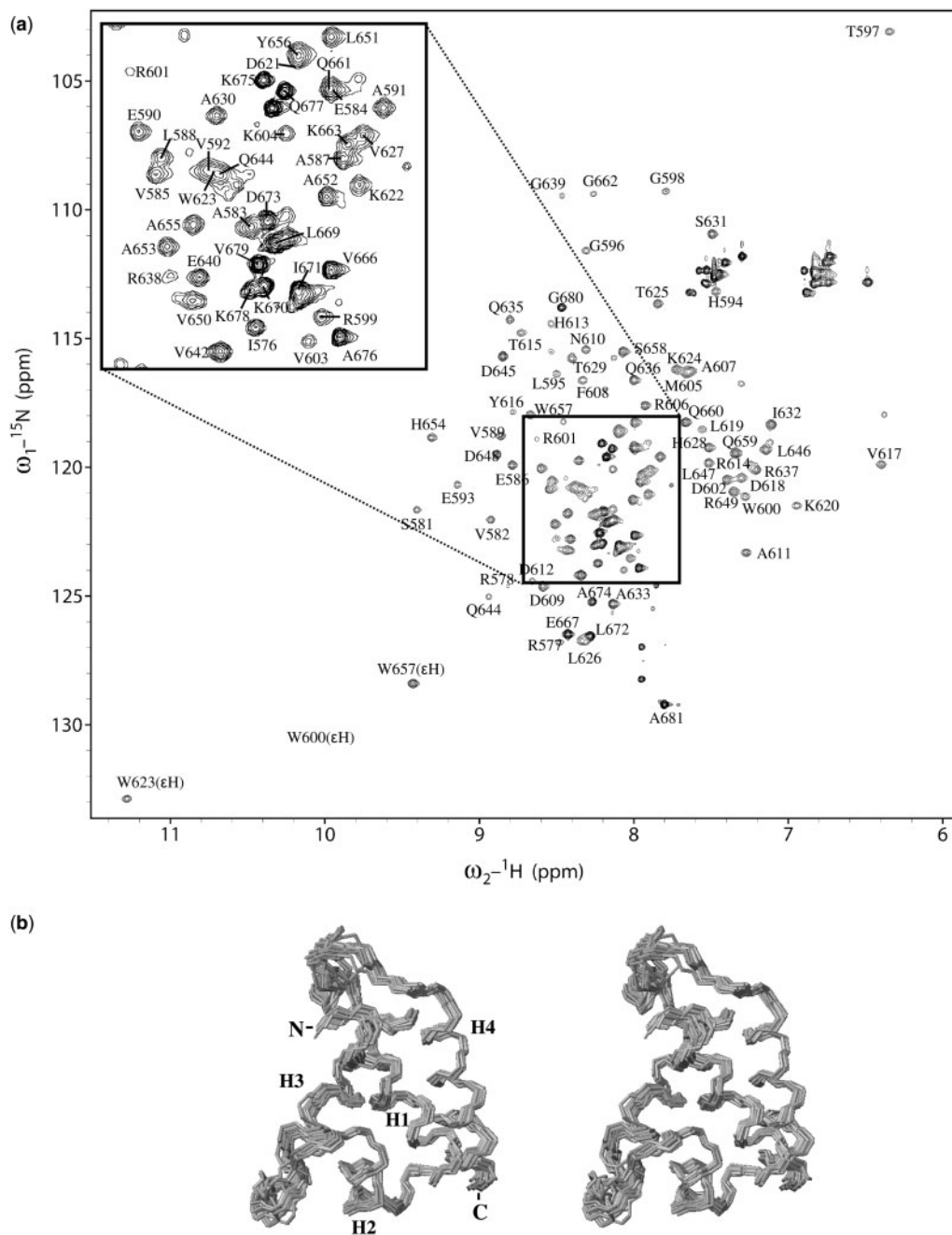
Figure 2. NMR spectrum and solution structure of NgTRF1[561–681]. (**a**) ^{1}H-^{15}N HSQC spectrum of ^{13}C/^{15}N labeled NgTRF1[561–681]. The spectrum was acquired at pH 7.0 and 298 K using a Bruker 600 MHz spectrometer. (**b**) The stereo plots of the backbone atoms of the 20 lowest-energy structures of NgTRF1[561–681] are superimposed with respect to the average structure of the defined residues of 582–659 (N, C_α, C′, O).

residues, which were predicted to form flexible loops based on the solution structure, were disordered.

The structure of NgTRF1[561–681] bound to telomeric DNA is determined using molecular replacement with the structure of NgTRF1[561–681] as a search model. The complex crystal belonged to space group P3$_2$ with two complex molecules per asymmetric unit. The electron density map of the telomeric DNA was evident after map drawing and the telomeric DNA structure of hTRF1–DNA complex was used for tracing. The electron density maps for the 13 N-terminal residues and the

21 C-terminal residues, which formed flexible loops in the structure of apo NgTRF1[561–681], were also disordered. Data collection, phasing and refinement statistics are presented in Table 1.

The overall structure of NgTRF1[561–681]

Superimposition of the solution and crystal structures of NgTRF1[561–681] (residues 578–660) showed that the two structures are almost identical, with a RMSD of 1.078 Å for the C_α atoms. The most prominent deviation

Table 1. Structural statistics

	NgTRF1[561–681]			Complex DNA
	Peak	Edge	Remote	
Data collection				
Beam		PF		PLS 4A
Space group		P2$_1$2$_1$2$_1$		P3$_2$
Resolution	50–1.9	50–1.9	50–1.9	50–2.4
Wavelength (Å)	1.0715	1.0719	1.0332	0.97929
Total reflections	218.186	110.694	115.470	118.466
Unique reflection	9258	9240	9268	15094
Completeness, %	98.2 (96.9)	97.3 (94.0)	96.7 (92.5)	95.8 (85.7)
R_{sym}, %[a]	7.4 (32.4)	6.8 (25.9)	5.9 (25.4)	9.5 (23.1)
Average I/σ (I)	55.8 (5.7)	34.9 (2.9)	34.7 (2.5)	13.2 (2.8)
Structure refinement				
Resolution (Å)		20–1.9		20–2.4
Reflections				
R_{cyst}, %[b]		21.0		24.6
R_{free}, %[c]		23.3		25.6
Protein		678		1409
Water		99		123
Rms deviations				
Bonds (Å)		0.007		0.007
Angles (°)		1.217		1.169
Ramachandran plot, %[d]				
Most favored		94.6		88.2
Additional allowed		5.4		11.1
Generally allowed		0		0.7
disallowed		0		0
NMR				
NOE distance restraints				
All			1317	
short range ($|i-j|$ <=1)			702	
Medium range (1 ≤<$|i-j|$ <5)			337	
Long range ($|i-j|$ >5)			278	
Hydrogen bonds distance restraints[a] (no.)			28	
Mean RMS deviations from the average coordinate				
Backbone atoms (N,C$_\alpha$,C′,O)			Helix only[e]: 0.54±0.16 Å (582–659: 0.86±0.19 Å)	
Heavy atoms			Helix only: 1.32±0.18 Å (582–659: 1.74±0.25 Å)	
Ramachandran plot (%)				
Most favored regions			78.2	
Additional allowed regions			21.5	
Generously allowed regions			0.3	
Disallowed regions			0.0	

[a]$R_{sym} = \Sigma|I_{obs}-I_{avg}|/I_{obs}$, where I_{obs} is the observed intensity of individual reflection and I_{avg} is average over symmetry equivalents.
[b]$R_{cyst} = \Sigma||F_{obs}|-|F_{calc}||/\Sigma|F_{obs}| \times 100$ for 95% of recorded data.
[c]R_{free} is the R-factor calculated by using 5% of the reflection data chosen randomly and omitted from the start of refinement.
[d]Calculated with program PROCHECK.
[e]Helix regions indicate helix 1 (A.A. 582–595), helix 2 (A.A. 600–607), helix 3 (A.A. 616–631), and helix 4 (A.A. 644–659).

was in the loop between helix 2 and helix 3 of NgTRF1[561–681]. When we restricted the comparison to the 4 helices of NgTRF1[561–681], the two structures were highly similar, with an RMSD of 0.713 Å for the C$_\alpha$ atoms. Especially, both N- and C-terminal regions are determined to be very flexible based on NOEs and dynamics data. This finding is well agreed with X-ray crystallographic data, showing that electron densities of both regions are not observed. Therefore, we conclude that the structure and topology of both solution and crystal structure are almost same.

NgTRF1[561–681] consisted of four α-helices and connecting loops, with helix 1 (residues 582–595), helix 2 (residues 600–607), helix 3 (residues 616–631) and helix 4 (residues 644–659) (Figures 2b and 3a). The N-terminal (residues 561–581) and C-terminal (residues 660–681) regions formed long flexible loops in the solution structure. In the crystal structure, shorter regions of the N-terminal (residues 578–581) and C-terminal (residue 660) loops were resolved in the electron density map. The canonical Myb-like domain of NgTRF1, which encompasses residues 574–629, extended from the N-terminal loop to helix 3 (Figure 3a). The end of helix 3 (residues 630–631), the loop between helix 3 and helix 4, helix 4 and C-terminal loop comprised the C-terminal Myb-extension that is unique feature of plant double-stranded telomere-binding proteins. The packing of 4 helices dictates the overall folding of NgTRF1[561–681]. Helix 3 was nearly perpendicular to and interacted with helices 1, 2 and 4 (Figure 3a). Helix 4, which is present
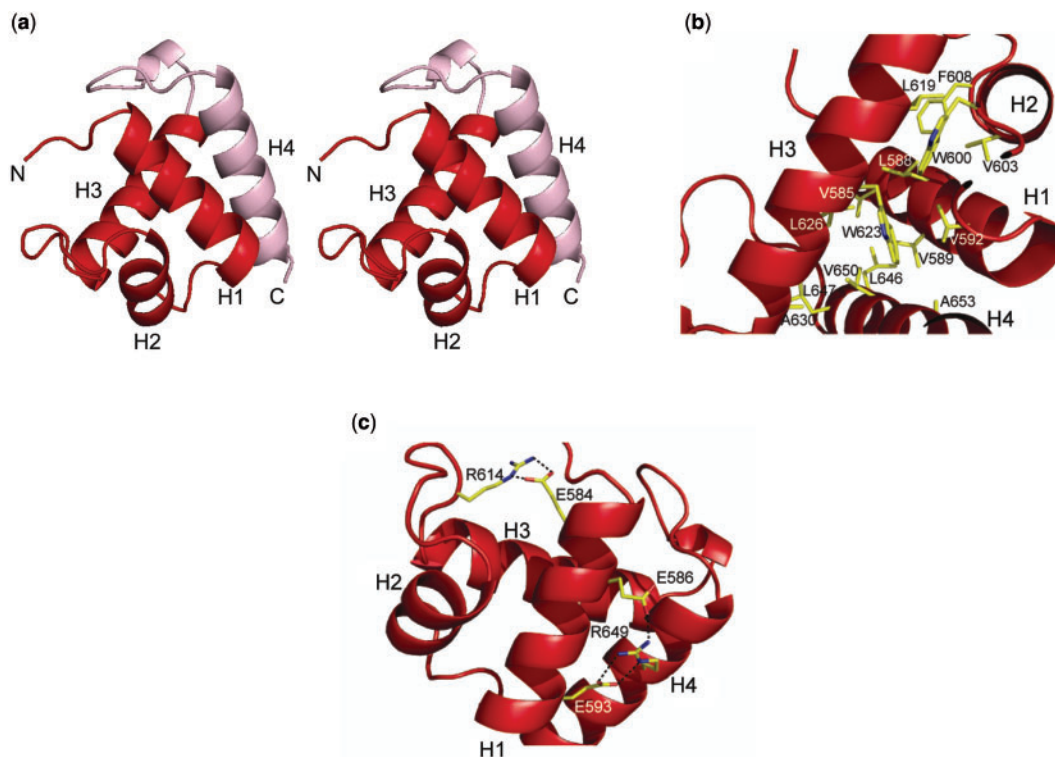
Figure 3. Crystal structure of NgTRF1$^{561-681}$. (**a**) The overall structure of NgTRF1$^{561-681}$ in ribbon representation. The canonical Myb-like domain is colored in red, and the C-terminal Myb-extension is colored in light pink. The helices are numbered in sequence from N- to C-terminal. (**b**) View of the hydrophobic core formed by the side-chains of residues from all four helices (depicted in yellow). (**c**) Two salt bridges stabilizing the overall structure of NgTRF1$^{561-681}$. Ionic interactions are drawn as dashed lines.

additionally in the DNA-binding domains of plant double-stranded telomere-binding proteins, interacted predominantly with helix 1 and helix 3. Helix 4 and helix 1 were packed together with an angle of ~30° between their helical axes. The overall structure of NgTRF1$^{561-681}$ is stabilized by a hydrophobic core composed of residues from all four helices: Val-585, Leu-588, Val-589 and Val-592 of helix 1; Trp-600, and Val-603 of helix 2; Leu-619, Trp-623, Leu-626 and Ala-630 of helix 3; Leu-646, Leu-647, Val-650 and Ala-653 of helix 4 (Figure 3b). Phe-580 in the N-terminal loop and Phe-608 in the loop between helix 2 and helix 3 also take part in the formation of hydrophobic core. The indole rings of Trp-600 (helix 2) and Trp-623 (helix 3) contribute to the formation of the hydrophobic core and appear to stabilize the overall structure. These residues are highly conserved in plant double-stranded telomere-binding proteins, which indicates that these hydrophobic interactions play a major role in stabilizing the overall folding of NgTRF1$^{561-681}$ (28). There were two charge–charge interactions that stabilize the structure of NgTRF1$^{561-681}$ in long distance (Figure 3c). The salt bridge between Glu-584 and Arg-614 made a connection between helix 1 and the loop between helix 2 and helix 3. The second ionic interaction was between Arg-649 of helix 4 and Glu-586 and Glu-593 of helix 1. The sequences of helix2 and helix 3 of NgTRF1$^{561-681}$ are similar to the prototype sequence of the helix–turn–helix motif (53),

thus, it is likely that they form the structure of a helix–turn–helix motif. By comparison with the double-stranded telomere-binding proteins of other organisms, helix 3 of NgTRF1$^{561-681}$ most likely functions as a DNA recognition helix. The relative configuration of helix 4, and the presence of both hydrophobic and ionic interactions that involve this helix and contribute to the overall structure of the DNA-binding domain, suggest that helix 4 does not interact directly with the DNA, but rather, plays a pivotal role in the overall stabilization of this protein.

Binding of NgTRF1$^{561-681}$ to telomeric DNA

To quantitate the binding of NgTRF1$^{561-681}$ to telomeric DNA, we carried out a series of ITC experiments (Figure 4a). The binding isotherms were fitted to a one-site binding model and a protein:DNA ratio of 2. The K_d value of NgTRF1$^{561-681}$ bound to telomeric DNA was 4.15×10^{-8} M, and the enthalpy of complex formation was -1.462×10^4 cal/mol at 25°C in ITC buffer. These results indicated that NgTRF1$^{561-681}$ binds to double-stranded telomeric DNA (nucleotide sequences: TTTAGGGTTTAGGG) at a molar ratio of 2:1 (protein:DNA) with high affinity.

We also probed the DNA-binding sites of NgTRF1$^{561-681}$ by monitoring changes in the 2D-HSQC spectra of NgTRF1$^{561-681}$ by titrating telomeric DNA. Based on assignment of the backbone resonances of
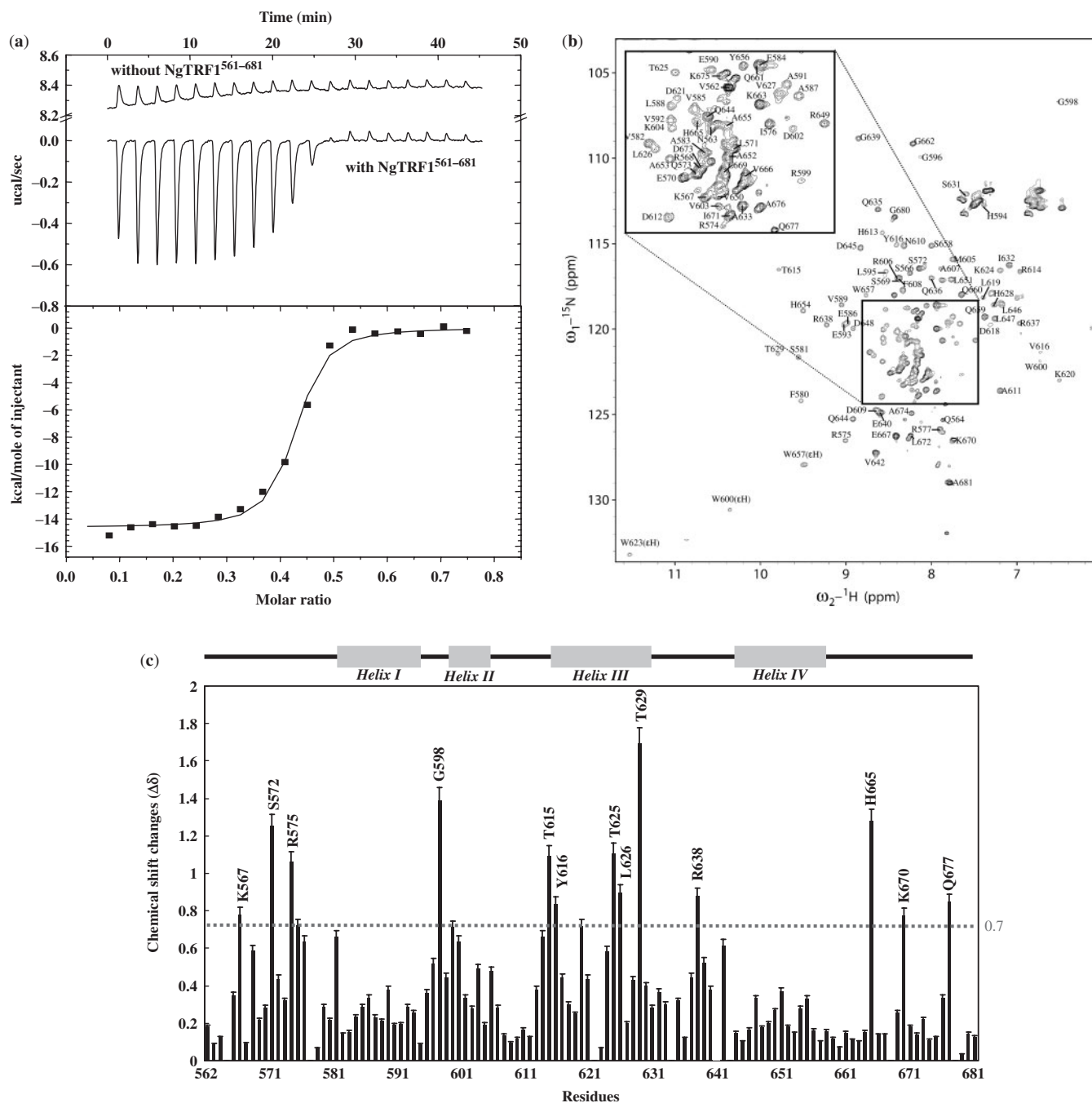
Figure 4. Characterization of binding of NgTRF1[561–681] to telomeric DNA. (**a**) ITC analysis of the binding of the NgTRF1[561–681] to the telomeric DNA sequence (TTTAGGG)$_2$ showing that binding is exothermic. Upper panel: 210 uM telomeric DNA was injected into a 1.8-ml sample cell containing 40 uM NgTRF1[561–681]. As a control, the DNA solution was injected into a sample cell that did not contain NgTRF1[561–681] (upper trace). Lower panel: Nonlinear least squares fit of the data varying the stoichiometry (n), the enthalpy of the reaction (ΔH) and the association constant (K_a). (**b**) The ^{1}H-^{15}N HSQC spectra of NgTRF1[561–681] with an equimolar amount of telomeric DNA. (**c**) Chemical shift perturbation of NgTRF1[561–681] upon DNA binding, with a cutoff at 0.7 p.p.m. Deviations in the chemical shifts of the indicated residues upon DNA binding are displayed. Chemical shift changes were calculated as $(\Delta\delta_{tot} = [(\delta_{HN}W_{HN})^2 + (\delta_N W_N)^2 + (\delta_{CO}W_{CO})^2 + (\delta_{Ca}W_{Ca})^2]^{1/2} \geqslant 0.7$ p.p.m., where δ_i is the chemical shift of nucleus i, and W_i denotes its weight factor, $W_{HN} = 1$, $W_N = 0.154$, $W_{ca} = 0.276$ and $W_{CO} = 0.341$). Schematic of the secondary structure of NgTRF1[561–681] is shown above the panel. (**d**) A plot of the peptide backbone dynamics of NgTRF1[561–681] upon binding to DNA. Heteronuclear NOEs (XNOE) were plotted for each residue and the corresponding secondary structures (depicted above the panel). (**e**) The locations of residues identified by chemical shift perturbation analysis are shown on the surface of NgTRF1[561–681].

DNA-bound NgTRF1[561–681], we found that there were significant perturbations in the NMR resonances in the presence of DNA (Figure 4b). The majority of the residues that underwent large changes in chemical shift were located in three regions of NgTRF1[561–681]: the N-terminal loop, helix 3 and the C-terminal loop (Figure 4c). XNOE data show that both N- and C-terminal regions of the protein become more rigid in

the presence of telomere DNA (Figure 4d). It has been reported that the N-terminal loop and helix 3 of other organisms have also been shown to be involved in DNA binding (Nishikawa *et al.*, 2001). The majority of the residues that showed a significant chemical shift perturbation (>0.7 p.p.m.) upon DNA binding were distributed on one surface of the molecule (Figure 4e).

The overall structure of DNA-bound NgTRF1$^{561-681}$

The Myb-like domain of NgTRF1$^{561-681}$ contacted the DNA directly (Figure 5a). A helix–turn–helix motif composed of helix 2, the loop between helix 2 and helix 3 and helix 3 interacted with the major groove of the DNA, with helix 3 as a DNA recognition helix, while the short N-terminal arm (residues 574–581) interacted with the minor groove of the DNA. In addition, C-terminal Myb-extension of NgTRF1$^{561-681}$ also interacted with the DNA. The loop between helix 3 and helix 4 interacted with the minor groove of the DNA (Figure 5a). Overall, the structures of DNA-bound and unbound NgTRF1$^{561-681}$ were similar with an RMSD of 0.581 Å for the C_α atoms (for the residues 578–660). The hydrophobic and ionic interactions that stabilized the structure of NgTRF1$^{561-681}$ were sustained in the

DNA-bound form of the molecule. The major difference between DNA-bound and unbound NgTRF1$^{561-681}$ was in the N-terminal arm, and the loop between helix 3 and helix 4 (Figure 5b). Residues 574–577 were clearly evident from the electron density map of the DNA-bound form of NgTRF1$^{561-681}$, but not the free form of the molecule, which suggested that in the absence of DNA, this region is flexible, and that binding to DNA make it more rigid (Figure 5b and c). Based on a structural comparison of DNA-bound and unbound NgTRF1$^{561-681}$, it could be speculated that upon binding to DNA, the loop between helix 3 and helix 4 changes its conformation in order to fit into and interact with the minor groove of the DNA (Figure 5b and d).

Structural comparison of NgTRF1$^{561-681}$ and the DNA-binding domain of hTRF1

NgTRF1 is homologous to hTRF1. The full-length ORFs are 17% homologous, and there is 27% homology in the Myb-like domain alone (27). Although the DNA-binding domain of NgTRF1 has an additional C-terminal Myb-extension, the structures of the DNA-binding domains of the two proteins in complex with DNA are very similar, with an RMSD of 0.666 Å for the C_α atoms (Table 2,
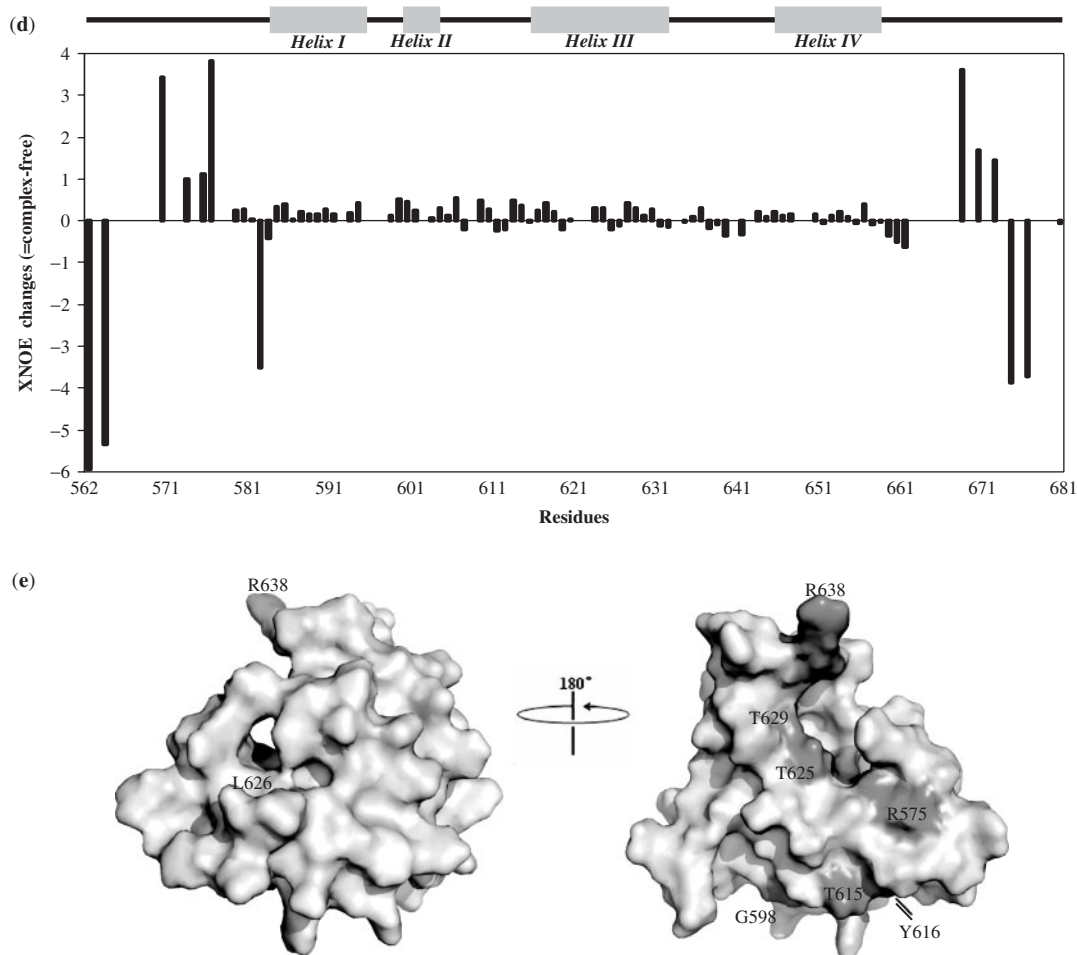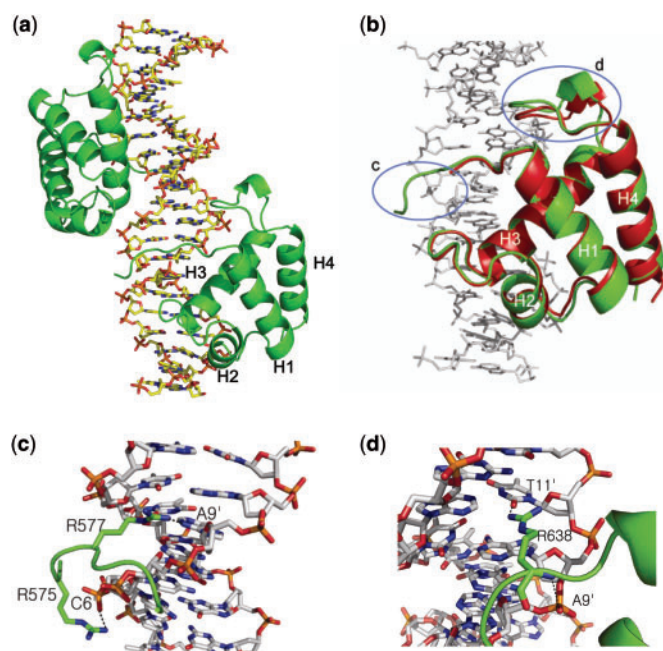

(d)

(e)

Figure 4. Continued

Figure 6). A hydrophobic core formed by helices 1, 2 and 3 is conserved in NgTRF1$^{561-681}$ and hTRF1. In addition, helix 4 of NgTRF1$^{561-681}$ also took part in the formation of the hydrophobic core. The hydrophobic interaction

Figure 5. Structure of NgTRF1$^{561-681}$ in complex with telomeric DNA. (**a**) Overall structure of NgTRF1$^{561-681}$ in complex with telomeric DNA. Two molecules of NgTRF1$^{561-681}$ bind to one telomeric DNA. The helices are numbered in sequence from N- to C-terminal. (**b**) Structural comparison of NgTRF1$^{561-681}$ in the DNA-bound (green) and -unbound (red) states. Blue circles, I and II, indicate the major conformational differences between the two structures. Close-up view of circle I and II were presented in (**c**) and (**d**), respectively. (c) Detailed view of the interactions of the N-terminal arm with DNA. Hydrogen bond between side chain of Arg-577 and base of A9′, and ionic interaction between side chain of Arg-575 and phosphate of C6′. Hydrogen bond and ionic interaction are drawn as dashed lines. (d) Detailed view of the interaction of the loop between helix 3 and helix 4 with DNA. Hydrogen bond between side chain of Arg-638 and base of T11′, and hydrogen bond between the amino nitrogen of Arg-638 and phosphate of A9′.

between helix 3 and helix 4 of NgTRF1$^{561-681}$ made helix 3 longer than that of the DNA-binding domain of hTRF1 (Table 2). NgTRF1$^{561-681}$ had four salt bridges. There were two unique ionic interactions in NgTRF1$^{561-681}$, Arg-649 (in helix 4) with Glu-586 and with Glu-593 (in helix 1), further stabilizing the structure. The salt bridge between Arg-575 and Asp-618 links the N-terminal arm and helix 3, resulting in a broader area of interaction with the DNA. Arg-575 was not detected in the electron density map of DNA-unbound NgTRF1$^{561-681}$, which suggested that the conformation of DNA-bound NgTRF1$^{561-681}$ was stabilized further by this ionic interaction.

The orientation of the DNA-binding domains of DNA-binding proteins is influenced primarily by non-specific interactions with the phosphate-sugar DNA backbone (38). We identified several positively charged surface areas in the structure of NgTRF1$^{561-681}$ that may mediate its interaction with the DNA backbone using electrostatic surface potentials (Figure 7a and b). Positively charged residues from helix 3 comprised one broad patch (Figure 7b, I). NgTRF1$^{561-681}$ interacted with the major groove of the DNA through helix 3, similar to other double-stranded telomere-binding proteins (Figures 7a). In particular, Tyr-616, Lys-622 and Thr-629 of NgTRF1$^{561-681}$ interacted with the phosphate groups of nucleotides T2, A8′ and A9′, respectively (Figure 8). Tyr-616, Lys-622 and Thr-629 are equivalent to Ser-417, Arg-423 and Leu-430 of hTRF1, respectively. Ser-417 of hTRF1 also interacted with the phosphate group of nucleotide T2, albeit indirectly, through a water molecule (35). Arg-423 of hTRF1 interacted with the phosphate group of nucleotide A7′, while Leu-430 of hTRF1 did not interact with the DNA. Patch 1 extended to Arg-601 of helix 2 (Figure 7b, II), which interacted with the phosphate group at nucleotide T2 (Figure 8). The equivalent position in hTRF1, Ser-404, also interacted with the phosphate group of the DNA backbone at nucleotide T2. These results indicated that the mechanism of interaction of NgTRF1$^{561-681}$ with the DNA backbone of the major groove is conserved in plant and human double-stranded telomere-binding proteins. A prominent positively charged surface patch

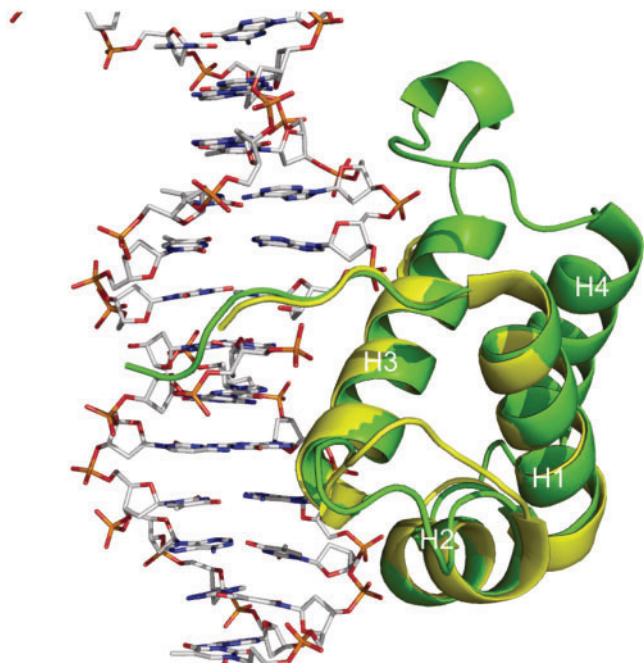**Table 2.** Structural comparison of NgTRF$^{561-681}$ and the DNA binding domain of hTRF1

RMSD† (0.666 Å)	NgTRF1$^{561-681}$	hTRF1 (DNA-binding domain)	Structure conservation
Hydrophobic Interactions			
Helix I, II and III	Phe580, Leu588, Val592,	Trp383, Leu391, Val395,	O^{a}
	Trp600, Val603, Phe608,	Trp403, Ile406, Phe412,	O
	Leu619, Trp623, Leu626	Leu420, Trp424, Met427	O
Helix I, III and IV	Val583, Val589, Ala630,	–	X^{b}
	Leu646, Leu647, Val650, Ala653	–	X
Salt bridges			
	Glu584–R614	Glu387–Arg415	O
	–	Asp388–Arg423	X
	Glu586–Arg649	–	X
	Glu593–Arg649	–	X
	Arg575–Asp618	–	X

RMSD† is for C$_{\alpha}$ atoms between 576–629 of NgTRF$^{561-581}$ and the DNA-binding domain of hTRF1.
^{a}O indicates the structural conservation in the two structures.
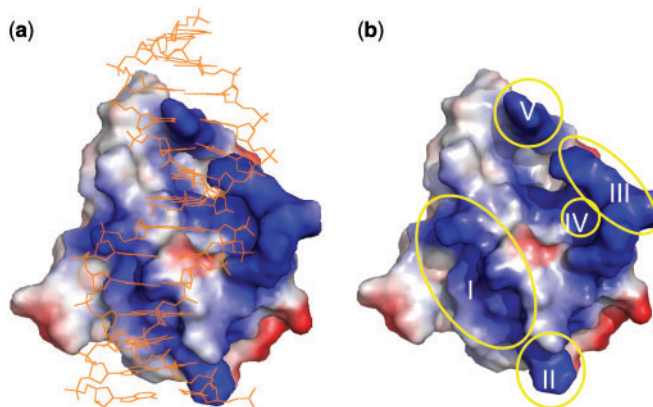^{b}X indicates the structural difference in the two structures.

composed of three residues of the N-terminal arm of NgTRF1$^{561-681}$ mediated the interaction of NgTRF1$^{561-681}$ with the minor groove of the DNA (Figure 7a and b, III). Arg-575, Arg-578 and Phe-580 of NgTRF1$^{561-681}$ interacted with the phosphate groups of the DNA backbone at nucleotides C6′, C7′ and A8′, respectively (Figure 8). Gln-381 and Trp-383 of hTRF1, which are equivalent to Arg-578 and Phe-580 of NgTRF1$^{561-681}$, also interacted with the phosphate groups of the DNA backbone at nucleotides C6′ and A7′, respectively (35). The equivalent residue of Arg-575 of NgTRF1$^{561-681}$ in hTRF1 was absent from the reported structure of hTRF1. Patch 3 extended to Arg-614 in the loop between helix 2 and helix 3 (Figure 7b, IV), which interacted with the phosphate group of the DNA backbone at nucleotide C7′ (Figure 8). The equivalent residue in hTRF1, Arg-415, also interacted with the phosphate group at nucleotide C6′ (35). Especially, Arg-638 in the loop between helix 3 and helix 4 created a surface-positive

Figure 6. Structural comparison of NgTRF1$^{561-681}$ and the DNA-binding domain of hTRF1. Proteins are represented by ribbon diagrams. NgTRF1$^{561-681}$ is colored in green and the DNA-binding domain of hTRF1 is colored in yellow. The helices are numbered in sequence from N- to C-terminal. The X-ray crystal structures of the DNA-binding domain of hTRF1 in complex with telomeric DNA [PDB accession ID: 1W0T, (33)] was used for the structure of hTRF1.

Figure 7. Electrostatic surface potentials of NgTRF1$^{561-681}$. (**a**) Electrostatic surface potential of NgTRF1$^{561-681}$. DNA is shown in orange. Surface residues are color-coded according to their charge (blue for positively charged and red for negatively charged side chains). (**b**) Positively charged patches (in blue) formed by helix 3 (I), Arg-601 (II), the N-terminal arm (III), Arg-614 (IV) and Arg-638 (V) of NgTRF1$^{561-681}$ are indicated by yellow circles.
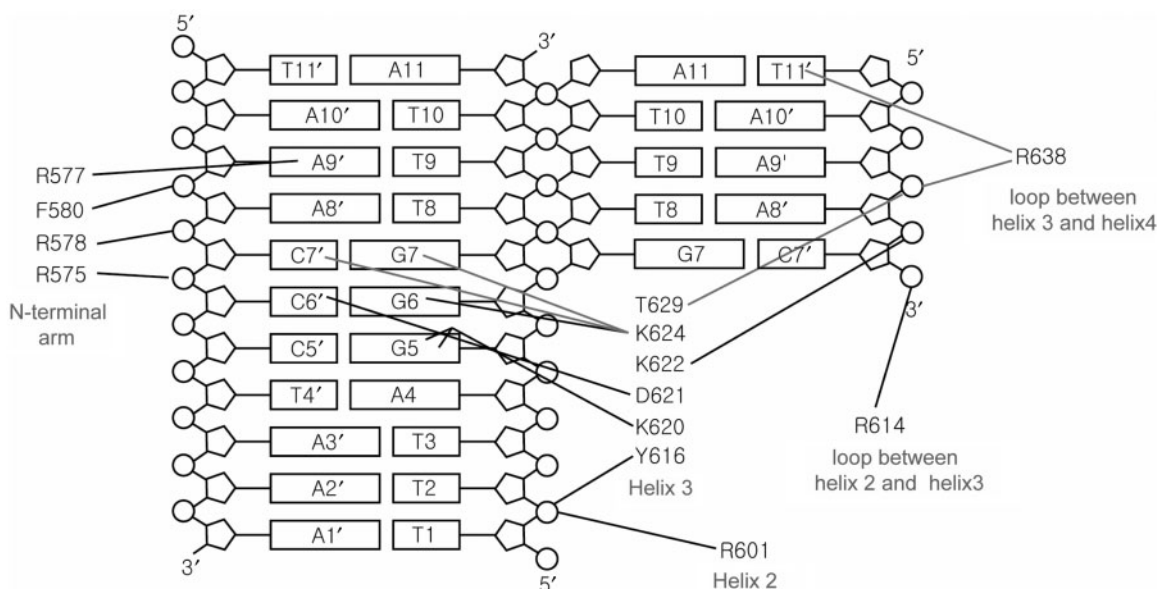
Figure 8. Schematic representation of the interactions between the indicated nucleotides of telomeric DNA and amino acids of NgTRF1$^{561-681}$. Black lines indicate common direct interactions that are also observed for the corresponding amino acid residues of hTRF1. Grey lines indicate interactions that are unique to NgTRF1$^{561-681}$.

patch at the interface with the minor groove (Figure 7b, V), and interacted with the phosphate group of the DNA backbone at nucleotide A9′ (Figure 8). This interaction is unique to the DNA-binding domains of plant telomere-binding proteins. These results demonstrated that the majority of the contacts with the DNA backbone appear to be conserved in NgTRF1$^{561-681}$ and the DNA-binding domain of hTRF1, and that two points of interaction in NgTRF1$^{561-681}$ are not in the DNA-binding domain of hTRF1. Given that the plant telomeric sequence is (TTTAGGG)$_n$, while that of human is (TTAGGG)$_n$ (54), the overall DNA-binding mechanism of NgTRF1 was well conserved compared to hTRF1 (Figure 8). Thus, although the DNA-binding domains of NgTRF1 contains one more α-helix compared to hTRF1, the orientation of the first three α-helices of the two proteins was very similar (Figure 6), and the mechanism of interaction with the DNA backbone of telomeric sequences appeared to be similar as well (Figure 8).

NgTRF1$^{561-681}$ specifically recognized telomeric sequences through interaction with bases at three parts: Part 1: helix3, Part 2: N-terminal arm and Part 3: loop between helix 3 and helix 4. The helix 3 of NgTRF1$^{561-681}$ (part 1) interacted with bases of the major groove through hydrogen bonds. N-terminal arm of NgTRF1$^{561-681}$ (part 2) and the loop between helix 3 and helix 4 (part 3) recognized the base of the minor groove (Table 3, Figure 8). Especially, the recognition of T11′ using part 3 is unique in NgTRF1$^{561-681}$ and absent in the DNA-binding domain of hTRF1. Thus, a specific recognition of the telomeric sequences by Myb-like domain is well conserved in NgTRF1$^{561-681}$ and the DNA-binding domain of hTRF1. But one additional interaction in the C-terminal Myb-extension between Arg-638 and T11′ is unique in NgTRF1$^{561-681}$.

Mutational analysis of the interaction with telomeric DNA

Based on the structure of NgTRF1$^{561-681}$ bound to telomeric DNA, 13 amino acids residues appeared to interact directly with the DNA (Figure 8). We carried out site-directed mutagenesis to change three residues in the N-terminal arm (Arg-575, Arg-577 and Arg-578), five residues in the helix3 (Lys-620, Asp-621, Lys-622, Lys-624 and Thr-629) and one residue in the loop between helix 3 and helix 4 (Arg-638) to alanine (Figure 9), and then assayed the telomeric DNA-binding activity of the mutant proteins. In addition, Arg-574, which is located in the most N-terminal in the structure of NgTRF1$^{561-681}$ bound to telomeric DNA, was also mutated to alanine to examine its binding to DNA. Binding activity was almost completely abolished by substitution of alanine at each of these ten positions (Figure 9b), which demonstrated that these residues are important for binding to telomeric DNA. The interaction of the DNA-binding domains of hTRF1 and hTRF2 with human telomeric DNA has been examined by mutational analysis of the telomeric sequence (38,55). The DNA-binding domains of both hTRF1 and hTRF2 showed a very low tolerance for single-base changes in the telomeric DNA sequences. Taken together, these results indicate that the DNA-binding activity of the telomere-binding proteins and telomeric DNA sequences have been well conserved throughout evolution.

In the current study, we could see that 12 additional N-terminal amino acids of the Myb-like domain (N-terminal extension) are necessary in the binding of NgTRF1$^{561-681}$ to telomeric DNA (Figure 1). To further characterize their role in telomeric DNA binding, we carried out site-directed mutagenesis to change two positively charged residues, Lys-567 and Arg-568, to methionine (K567M) and isoleucine (R568I), respectively (Supplementary Figure 1a). These mutations completely abolished the DNA-binding activity of NgTRF1$^{561-681}$ (Supplementary Figure 1b), which indicated that these N-terminal residues are crucial for DNA binding. Arg-575 in the N-terminal arm and Thr-629 in helix 3 interact directly with telomeric DNA (Figure 8) and their mutation to isoleucine (R575I) and valine (T629V) also abolished DNA-binding activity of NgTRF1$^{561-681}$. In contrast, mutations of the residues in the C-terminal loop (H665G, K670M and Q677E) of NgTRF1$^{561-681}$ did not show

Table 3. DNA interaction comparison of NgTRF$^{561-681}$ and the DNA-binding domain of hTRF1

DNA	Protein	NgTRF$^{561-681}$	hTRF1 (The DNA binding domain)	DNA Interaction site	Structure conservation
Phosphate group interaction	Patch 1	Tyr616-T2	Ser417-T2	Major groove	O^a
		Lys622-A8′	Arg423-A7′	Major groove	O
		Thr629-A9′	–	Major groove	X^b
	Patch 2	Ala601-T2	Ser404-T2	Major groove	O
	Patch 3	Arg575-C6′	–	Minor groove	X
		Arg578-C7′	Gln381-C6′	Minor groove	O
		Phe580-A8′	Trp383-A7′	Minor groove	O
	Patch 4	Arg614-C7′	Arg415-C6′	Minor groove	O
	Patch 5	Arg638-A9′	–	Minor groove	X
Base interaction	Part 1	Lys620-G5(N$_7$, C$=$O)	Lys421-A4, G5	Major groove	O
		Asp621-C6′ (N$_4$)	Asp422-C7′ (N$_4$), and C8′ (N$_4$)	Major groove	O
		Lys624-G6(C$=$O), G7(C$=$O) and C7′ (N$_4$)	Arg425-G6(N7, C$=$O)	Major groove	O
	Part 2	Arg577-A9′ (N$_3$)	Arg380-A6′ (N$_3$) and T9(O$_2$)	Major groove	O
	Part 3	Arg638-T11′ (C$=$O)	–	Major groove	X

^{a}O indicates the structural conservation in the two structures.
^{b}X indicates the structural difference in the two structures.

(a)

(b)
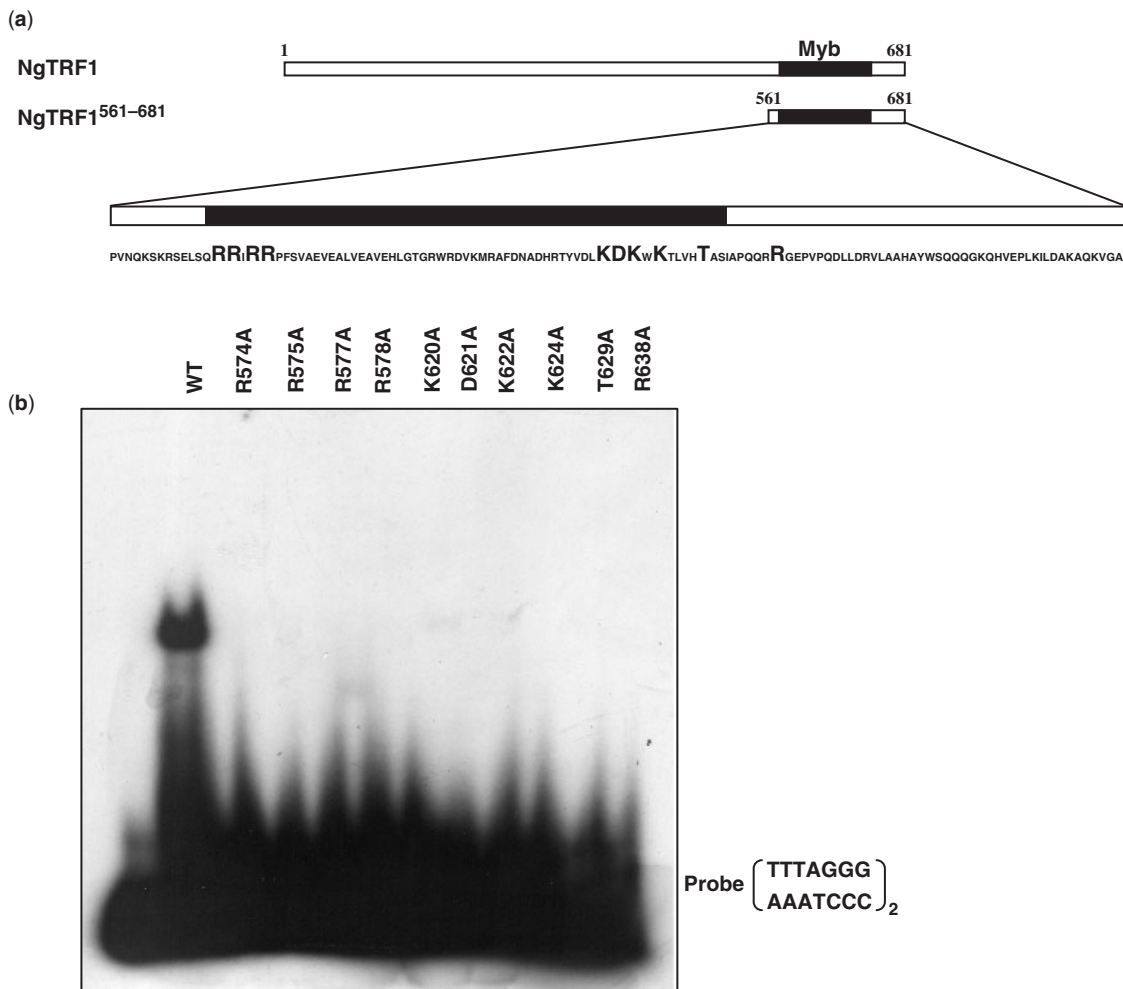
Figure 9. Mutational analysis of telomeric DNA-binding of NgTRF1[561–681]. (**a**) Schematic representation of full-length NgTRF1 and NgTRF1[561–681], and the sequence of amino acids residues 561–681, with residues that were mutated indicated in bold. (**b**) Gel retardation assay of the indicated substitution mutants of NgTRF1[561–681]. Wild-type and mutant proteins were incubated with radiolabeled double-stranded telomeric DNA (TTTAGGG)$_2$, and then subjected to electrophoresis on a nondenaturing polyacrylamide gel. Free and protein-complexed probes were separated on nondenaturing gels and visualized by autoradiography.

severe effect in the DNA-binding activity. One negatively charged amino acid, Glu-570, was mutated to glutamine (E570Q) and this mutation increased the DNA-binding affinity of NgTRF1[561–681]. Ser-572 was not resolved in the crystal structure of NgTRF1[561–681] and it exhibited a high chemical-shift perturbation upon DNA binding (Figure 4c). Mutation of Ser-572 to alanine (S572A) also increased the DNA-binding activity (Supplementary Figure 1). Taken together these results, the residues in the N-terminal extension of the Myb-like domain including charged amino acids also play an important role in the interaction of NgTRF1[561–681] with telomeric DNA.

Structural comparison of NgTRF1[561–681] with the NMR structure of AtTRP1[464–560]

NgTRF1 shows the highest level of sequence similarity with AtTRP1 compared to other telomere-binding proteins in the database (27). The NMR structure of the DNA-binding domain of AtTRP1, AtTRP1[464–560], has

been reported (39). There is ~86% sequence similarity between NgTRF1[561–681] and AtTRP1[464–560], which suggests that the architecture of the peptide backbone of the two molecules would be similar as well. However, superimposition of the structures of the two DNA-binding domains shows that there is significant deviation between them (RMSD of 7.437) (Supplementary Figure 2). In the solution structure of AtTRP1[464–560], the N-terminal arm, helix 3 and loop between helix 3 and helix 4 were implicated in the interaction with DNA, based on chemical shift perturbation data and analysis of the surface charge distribution (39). Specifically, four arginine residues in the N-terminal arm, Arg-465, Arg-466, Arg-468 and Arg-469 of AtTRP1[464–560], were suggested to be involved in interacting with DNA. In the current study, we show that the equivalent positions in the N-terminal arm of NgTRF1[561–681], Arg-574, Arg-575, Arg-577 and Arg-578, also interacted directly with DNA. In helix 3 of AtTRP1[464–560], three positively charged residues, Lys-511, Lys-513 and Lys-515, were suggested to

interact with DNA. The equivalent residues of NgTRF1$^{561–681}$, Lys-620, Lys-621 and Lys-623, also interacted with DNA, as well as three additional residues of helix 3, Tyr-616, Asp-621 and Thr-629. In the loop between helix 3 and helix 4, three positively charged amino acids of AtTRP1$^{464–560}$, Lys-522, Arg-528 and Arg-529, were suggested to interact with DNA, while in NgTRF1$^{561–681}$, only Arg-638, which is equivalent to Arg-529 of AtTRP1$^{464–560}$ made contact with the DNA. These differences could be explainable by experimental conditions or data analysis between two systems. Crystallographic study showed more clearly the interaction between NgTRF1$^{561–681}$ and DNA in the structure of NgTRF1$^{561–681}$ in complex with telomeric DNA, whereas NMR approach showed rather indirectly the interaction between DNA and AtTRP1$^{464–560}$. Thus, the residues that do not interact directly with DNA, but whose conformation changes upon binding, might show a high chemical shift perturbation in NMR study. However, we could not clarify why two structures are different in the absence of telomere DNA in solution.

DISCUSSION

Unique features of the DNA-binding domain of plant double-stranded telomere-binding proteins

The Myb family of proteins, of which c-Myb was the first identified, are found in a wide spectrum of eukaryotes including yeasts, vertebrates and higher plants (56). c-Myb contains three repeats of a sequence called the Myb domain (termed R1, R2 and R3). In plants, the largest Myb sub-family contains two Myb domain repeats (termed R2R3 proteins). Most double-stranded telomere-binding proteins contain a single Myb-like domain. These proteins form dimers in which each Myb-like domain independently recognizes the target DNA sequence (37). In c-Myb, both R2 and R3 are necessary for DNA binding, and act cooperatively to recognize and bind to specific DNA sequences (57). Based on structural studies, the Myb-like domains of double-stranded telomere-binding proteins adopt different conformations when they bind to DNA compared to the Myb domains of c-Myb. The helix–turn–helix motif of the Myb domains of c-Myb is involved in DNA recognition in the major groove of the DNA. In double-stranded telomere-binding proteins, on the other hand, the short N-terminal arm of the Myb-like domain also interacts directly with the minor groove of the DNA. This mechanism of DNA binding is characteristic of homeodomains, which are another class of three-helix bundle-containing DNA-binding domains (14). Plant double-stranded telomere-binding proteins, including NgTRF1, are unique in that the Myb-like domain alone is not sufficient for telomere binding (27,29,52). In the current study, we determined the structure of the DNA-binding domain of NgTRF1, NgTRF1$^{561–681}$ and the complex bound to telomeric DNA in order to gain a better understanding of the mechanism of interaction of plant double-stranded telomere-binding proteins with DNA.

The presence of a highly conserved region C-terminal to the Myb-like domain called the C-terminal Myb-extension is characteristic of plant double-stranded telomere-binding proteins (27,29,52). In the current study, we could see two distinctive roles of C-terminal Myb-extension of NgTRF1$^{561–681}$ from the structural analysis. First, residues in helix 4 of the C-terminal Myb-extension of NgTRF1$^{561–681}$ made extensive contacts with residues in other parts of the molecule, which suggests that the role of the C-terminal Myb-extension is to stabilize the overall structure of the DNA-binding domain. According to this view, the DNA-binding domain of NgTRF1 shows structural similarity to the DNA-binding domain of the budding yeast double-stranded telomere-binding protein scRAP1, although NgTRF1 and scRAP1 are not homologous. The DNA-binding domain of scRAP1 consists of a tandem array of two structurally similar domains, domain 1 and domain 2, which recognize two tandem sequence repeats of telomeric DNA (36). Each domain contains a three-helix bundle and an N-terminal arm, which make specific contacts with nucleotides in the major and minor groove of the DNA, respectively, similar to the DNA-binding domain of other double-stranded telomere-binding proteins. The C-terminal part of domain 1, which links domain 1 with the N-terminal arm of domain 2, is closely associated with the three-helix bundle of domain 1, and stabilizes the α-helical core of domain 1 through various hydrophobic contacts and hydrogen bonds, although it does not adopt a well-defined secondary structure (36). Domain 2 contains a fourth helix which stabilizes the overall structure of the domain, similar to the DNA-binding domain of NgTRF1. The relative orientation of the helices of NgTRF1$^{561–681}$ and domain 2 of scRAP1 were comparable when the two molecules were superimposed (data not shown). Second, in the structure of NgTRF1$^{561–681}$ bound to telomeric DNA, Arg-638 in the loop between helix 3 and helix 4 interacted directly with the minor groove of the DNA, indicating that the C-terminal Myb-extension is also involved in the interaction of NgTRF1 with telomeric DNA. This was confirmed by analysis of an alanine substitution mutant of Arg-638 (Figure 9). The notion that the C-terminal Myb-extension plays a specific role in the binding of plant double-stranded telomere-binding proteins to telomeric DNA was suggested previously from the DNA-binding study of TRF-like proteins of *Arabidopsis* (29). The 12 *Arabidopsis* TRFL genes could be grouped into two distinct gene families based on the presence or absence of the C-terminal Myb-extension. TRFL family 1 proteins contain the C-terminal Myb-extension, and a recombinant protein consisting of the Myb-like domain and the C-terminal extension bound to telomeric DNA *in vitro*. Deletion of the C-terminal Myb-extension from TRFL1, a TRFL family 1 protein, abolished its DNA-binding activity, and introduction of the C-terminal Myb-extension of TRFL1 into the TRFL3, a TRFL family 2 protein, conferred DNA-binding activity (29). In the structure of NgTRF1$^{561–681}$ in complex with telomeric DNA, Arg-638 interacted with T11′ and the phosphate group of the DNA backbone at nucleotide A9′ (Figure 8). This interaction may provide the

structural basis of the reason why the C-terminal Myb-extension is essential for plant double-stranded telomere-binding proteins to bind plant telomeric DNA, $(TTTAGGG)_n$.

The relationship between the C-terminal Myb-extension and plant telomere sequence

C-terminal Myb-extensions are highly conserved and unique in the plant double-stranded telomere-binding proteins (29). Telomere sequences of most plants are $(TTTAGGG)_n$. Therefore, it is likely that the C-terminal Myb-extension and plant telomere sequence are related to each other evolutionarily. In the previous report by Sue *et al.* (39) C-terminal Myb-extension was known to be involved in the telomere sequence binding but the exact role of it in the recognition of plant telomere sequence was not identified. The structural comparison of NgTRF1$^{561-681}$ and the DNA-binding domain of hTRF1 in complex with telomere sequence showed that a specific recognition of the telomeric sequences by Myb-like domain is well conserved in them. Therefore, if the C-terminal Myb-extension and plant telomere sequence are correlated during evolution, the interaction between the C-terminal Myb-extension and telomere sequences would be essential in recognition of plant telomere sequences by the plant double-stranded telomere-binding proteins. From the structure of NgTRF$^{561-681}$ in complex with plant telomeric sequence, we could see that the Arg-683 in the C-terminal Myb-extension interacts with the base of T11. The recognition of T11 by Arg-683, which is located in the plant-specific C-terminal Myb-extension, is expected to contribute to the specific interaction of NgTRF1 with the plant telomeric sequence. Indeed, we found that NgTRF$^{561-681}$ did not bind to the vertebrate telomeric sequence, (data not shown), suggesting that the plant telomeric-binding proteins and the telomeric sequence have indeed co-evolved to maintain a highly specific protein–DNA interaction.

Accession number

The coordinates and structure factors for the crystal structures of NgTRF1$^{561-681}$ and the complex with telomeric DNA have been deposited to the Protein Data Bank with codes 2CKX and 2QHB, respectively. The coordinate for the solution structure of NgTRF1$^{561-681}$ is available from the Protein Data Bank with code 2JUH.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

REFERENCES

1. Zakian,V.A. (1995) Telomeres: beginning to understand the end. *Science*, **270**, 1601–1607.
2. Hastie,N.D., Dempster,M., Dunlop,M.G., Thompson,A.M., Green,D.K. and Allshire,R.C. (1990) Telomere reduction in human colorectal carcinoma and with ageing. *Nature*, **346**, 866–868.
3. Bodnar,A.G., Ouellette,M., Frolkis,M., Holt,S.E., Chiu,C.P., Morin,G.B., Harley,C.B., Shay,J.W., Lichtsteiner,S. and Wright,W.E (1998) Extension of life-span by introduction of telomerase into normal human cells. *Science*, **279**, 349–352.
4. Blackburn,E.H. (1991) Structure and function of telomeres. *Nature*, **350**, 569–573.
5. Greider,C.W. and Blackburn,E.H. (1985) Identification of a specific telomere terminal transferase activity in Tetrahymena extracts. *Cell*, **43**, 405–413.
6. Greider,C.W. and Blackburn,E.H. (1987) The telomere terminal transferase of Tetrahymena is a ribonucleoprotein enzyme with two kinds of primer specificity. *Cell*, **51**, 887–898.
7. Smogorzewska,A., van Steensel,B., Bianchi,A., Oelmann,S., Schaefer,M.R., Schnapp,G. and de Lange,T. (2000) Control of human telomere length by TRF1 and TRF2. *Mol. Cell. Biol.*, **20**, 1659–1668.
8. Celli,G.B. and de Lange,T. (2005) DNA processing is not required for ATM-mediated telomere damage response after TRF2 deletion. *Nat. Cell Biol.*, **7**, 712–718.
9. Herbig,U., Jobling,W.A., Chen,B.P., Chen,D.J. and Sedivy,J.M. (2004) Telomere shortening triggers senescence of human cells through a pathway involving ATM, p53, and p21(CIP1), but not p16(INK4a). *Mol. Cell*, **14**, 501–513.
10. Karlseder,J., Broccoli,D., Dai,Y., Hardy,S. and de Lange,T. (1999) p53- and ATM-dependent apoptosis induced by telomeres lacking TRF2. *Science*, **283**, 1321–1325.
11. van Steensel,B., Smogorzewska,A. and de Lange,T. (1998) TRF2 protects human telomeres from end-to-end fusions. *Cell*, **92**, 401–413.
12. de Lange,T. (2005) Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes Dev.*, **19**, 2100–2110.
13. Chong,L., van Steensel,B., Broccoli,D., Erdjument-Bromage,H., Hanish,J., Tempst,P. and de Lange,T. (1995) A human telomeric protein. *Science*, **270**, 1663–1667.
14. Konig,P., Fairall,L. and Rhodes,D. (1998) Sequence-specific DNA recognition by the myb-like domain of the human telomere binding protein TRF1: a model for the protein-DNA complex. *Nucleic Acids Res.*, **26**, 1731–1740.

15. van Steensel,B. and de Lange,T. (1997) Control of telomere length by the human telomeric protein TRF1. *Nature*, **385**, 740–743.

16. Bilaud,T., Brun,C., Ancelin,K., Koering,C.E., Laroche,T. and Gilson,E. (1997) Telomeric localization of TRF2, a novel human telobox protein. *Nat. Genet.*, **17**, 236–239.

17. Broccoli,D., Smogorzewska,A., Chong,L. and de Lange,T. (1997) Human telomeres contain two distinct Myb-related proteins, TRF1 and TRF2. *Nat. Genet.*, **17**, 231–235.

18. Loayza,D. and De Lange,T. (2003) POT1 as a terminal transducer of TRF1 telomere length control. *Nature*, **423**, 1013–1018.

19. Conrad,M.N., Wright,J.H., Wolf,A.J. and Zakian,V.A. (1990) RAP1 protein interacts with yeast telomeres in vivo: overproduction alters telomere structure and decreases chromosome stability. *Cell*, **63**, 739–750.

20. Gilson,E., Roberge,M., Giraldo,R., Rhodes,D. and Gasser,S.M. (1993) Distortion of the DNA double helix by RAP1 at silencers and multiple telomeric binding sites. *J. Mol. Biol.*, **231**, 293–310.

21. Longtine,M.S., Wilson,N.M., Petracek,M.E. and Berman,J. (1989) A yeast telomere binding activity binds to two related telomere sequence motifs and is indistinguishable from RAP1. *Curr. Genet.*, **16**, 225–239.

22. Krauskopf,A. and Blackburn,E.H. (1998) Rap1 protein regulates telomere turnover in yeast. *Proc. Natl Acad. Sci. USA*, **95**, 12486–12491.

23. Marcand,S., Gilson,E. and Shore,D. (1997) A protein-counting mechanism for telomere length regulation in yeast. *Science (New York, NY)*, **275**, 986–990.

24. Cooper,J.P., Nimmo,E.R., Allshire,R.C. and Cech,T.R. (1997) Regulation of telomere length and function by a Myb-domain protein in fission yeast. *Nature*, **385**, 744–747.

25. Kim,J.H., Kim,W.T. and Chung,I.K. (1998) Rice proteins that bind single-stranded G-rich telomere DNA. *Plant Mol. Biol.*, **36**, 661–672.

26. Lee,J.H., Kim,J.H., Kim,W.T., Kang,B.G. and Chung,I.K. (2000) Characterization and developmental expression of single-stranded telomeric DNA-binding proteins from mung bean (Vigna radiata). *Plant Mol. Biol.*, **42**, 547–557.

27. Yang,S.W., Kim,D.H., Lee,J.J., Chun,Y.J., Lee,J.H., Kim,Y.J., Chung,I.K. and Kim,W.T. (2003) Expression of the telomeric repeat binding factor gene NgTRF1 is closely coordinated with the cell division program in tobacco BY-2 suspension culture cells. *J. Biol. Chem.*, **278**, 21395–21407.

28. Zentgraf,U. (1995) Telomere-binding proteins of Arabidopsis thaliana. *Plant Mol. Biol.*, **27**, 467–475.

29. Karamysheva,Z.N., Surovtseva,Y.V., Vespa,L., Shakirov,E.V. and Shippen,D.E. (2004) A C-terminal Myb extension domain defines a novel family of double-strand telomeric DNA-binding proteins in Arabidopsis. *J. Biol. Chem.*, **279**, 47799–47807.

30. Marian,C.O., Bordoli,S.J., Goltz,M., Santarella,R.A., Jackson,L.P., Danilevskaya,O., Beckstette,M., Meeley,R. and Bass,H.W. (2003) The maize Single myb histone 1 gene, Smh1, belongs to a novel gene family and encodes a protein that binds telomere DNA repeats in vitro. *Plant Physiol.*, **133**, 1336–1350.

31. Schrumpfova,P., Kuchar,M., Mikova,G., Skrisovska,L., Kubicarova,T. and Fajkus,J. (2004) Characterization of two Arabidopsis thaliana myb-like proteins showing affinity to telomeric DNA sequence. *Genome Natl Res. Council Can.*, **47**, 316–324.

32. Kuchar,M. and Fajkus,J. (2004) Interactions of putative telomere-binding proteins in Arabidopsis thaliana: identification of functional TRF2 homolog in plants. *FEBS Lett.*, **578**, 311–315.

33. Yang,S.W., Kim,S.K. and Kim,W.T. (2004) Perturbation of NgTRF1 expression induces apoptosis-like cell death in tobacco BY-2 cells and implicates NgTRF1 in the control of telomere length and stability. *Plant Cell*, **16**, 3370–3385.

34. Hong,J.P., Byun,M.Y., Koo,D.H., An,K., Bang,J.W., Chung,I.K., An,G. and Kim,W.T. (2007) Suppression of RICE TELOMERE BINDING PROTEIN 1 results in severe and gradual developmental defects accompanied by genome instability in rice. *Plant Cell*, **19**, 1770–1781.

35. Court,R., Chapman,L., Fairall,L. and Rhodes,D. (2005) How the human telomeric proteins TRF1 and TRF2 recognize telomeric DNA: a view from high-resolution crystal structures. *EMBO Rep.*, **6**, 39–45.

36. Konig,P., Giraldo,R., Chapman,L. and Rhodes,D. (1996) The crystal structure of the DNA-binding domain of yeast RAP1 in complex with telomeric DNA. *Cell*, **85**, 125–136.

37. Nishikawa,T., Nagadoi,A., Yoshimura,S., Aimoto,S. and Nishimura,Y. (1998) Solution structure of the DNA-binding domain of human telomeric protein, hTRF1. *Structure*, **6**, 1057–1065.

38. Nishikawa,T., Okamura,H., Nagadoi,A., Konig,P., Rhodes,D. and Nishimura,Y. (2001) Solution structure of a telomeric DNA complex of human TRF1. *Structure*, **9**, 1237–1251.

39. Sue,S.C., Hsiao,H.H., Chung,B.C., Cheng,Y.H., Hsueh,K.L., Chen,C.M., Ho,C.H. and Huang,T.H. (2006) Solution structure of the Arabidopsis thaliana telomeric repeat-binding protein DNA binding domain: a new fold with an additional C-terminal helix. *J. Mol. Biol.*, **356**, 72–85.

40. Richards,E.J. and Ausubel,F.M. (1988) Isolation of a higher eukaryotic telomere from Arabidopsis thaliana. *Cell*, **53**, 127–136.

41. Ganal,M.W., Lapitan,N.L. and Tanksley,S.D. (1991) Macrostructure of the tomato telomeres. *Plant Cell*, **3**, 87–94.

42. Wu,K.S. and Tanksley,S.D. (1993) Genetic and physical mapping of telomeres and macrosatellites of rice. *Plant Mol. Biol.*, **22**, 861–872.

43. Delaglio,F., Grzesiek,S., Vuister,G.W., Zhu,G., Pfeifer,J. and Bax,A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, **6**, 277–293.

44. Cornilescu,G., Delaglio,F. and Bax,A. (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR*, **13**, 289–302.

45. Laskowski,R.A., Rullmannn,J.A., MacArthur,M.W., Kaptein,R. and 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.

46. Koradi,R., Billeter,M. and Wuthrich,K. (1996) MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graphics*, **14**, 29–32, 51–55.

47. Otwinowski,Z. and Minor,W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, **276**, 307–326.

48. Schneider,T.R. and Sheldrick,G.M. (2002) Substructure solution with SHELXD. *Acta Crystallogr.*, **58**, 1772–1779.

49. Jones,E.Y., Walker,N.P. and Stuart,D.I. (1991) Methodology employed for the structure determination of tumour necrosis factor, a case of high non-crystallographic symmetry. *Acta Crystallogr.*, **47(Pt 6)**, 753–770.

50. Brunger,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 & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr.*, **54**, 905–921.

51. Navaza,J. (1994) AMoRe: an Automated Package for molecular replacement. *Acta Crystallogr.*, **D50**, 157–163.

52. Chen,C.M., Wang,C.T. and Ho,C.H. (2001) A plant gene encoding a Myb-like protein that binds telomeric GGTTAG repeats in vitro. *J. Biol. Chem.*, **276**, 16511–16519.

53. Harrison,S.C. and Aggarwal,A.K. (1990) DNA recognition by proteins with the helix-turn-helix motif. *Annu. Rev. Biochem.*, **59**, 933–969.

54. Smogorzewska,A. and de Lange,T. (2004) Regulation of telomerase by telomeric proteins. *Ann. Rev. Biochem.*, **73**, 177–208.

55. Bianchi,A., Smith,S., Chong,L., Elias,P. and de Lange,T. (1997) TRF1 is a dimer and bends telomeric DNA. *EMBO J.*, **16**, 1785–1794.

56. Graf,T. (1992) Myb: a transcriptional activator linking proliferation and differentiation in hematopoietic cells. *Curr. Opin. Genet. Dev.*, **2**, 249–255.

57. Ogata,K., Morikawa,S., Nakamura,H., Hojo,H., Yoshimura,S., Zhang,R., Aimoto,S., Ametani,Y., Hirata,Z., Sarai,A. et al. (1995) Comparison of the free and DNA-complexed forms of the DNA-binding domain from c-Myb. *Nat. Struct. Biol.*, **2**, 309–320.