Published online 8 September 2011 *Nucleic Acids Research, 2012, Vol. 40, No. 1* ***449–460***
doi:10.1093/nar/gkr707

Structure of the *Trichomonas vaginalis* Myb3 DNA-binding domain bound to a promoter sequence reveals a unique C-terminal β-hairpin conformation

Shu-Yi Wei[1], Yuan-Chao Lou[1], Jia-Yin Tsai[2], Meng-Ru Ho[1], Chun-Chi Chou[1], M. Rajasekaran[1,3,4], Hong-Ming Hsu[1,5], Jung-Hsiang Tai[1], Chwan-Deng Hsiao[2,*], and and Chinpan Chen[1,*]

[1]Institute of Biomedical Sciences, [2]Institute of Molecular Biology, Academia Sinica, Taipei 115, [3]Department of Life Science, National Tsing Hua University, Hsinchu 300, [4]Chemical Biology and Molecular Biophysics, Taiwan International Graduate Program, Academia Sinica, Taipei 115 and [5]Department of Parasitology, College of Medicine, National Taiwan University, Taipei 106, Taiwan, ROC

Received June 20, 2011; Revised August 8, 2011; Accepted August 15, 2011

ABSTRACT

Trichomonas vaginalis **Myb3 transcription factor (*tv*Myb3) recognizes the MRE-1 promoter sequence and regulates *ap65-1* gene, which encodes a hydrogenosomal malic enzyme that may play a role in the cytoadherence of the parasite. Here, we identified *tv*Myb3$_{53–180}$ as the essential fragment for DNA recognition and report the crystal structure of *tv*Myb3$_{53–180}$ bound to MRE-1 DNA. The N-terminal fragment adopts the classical conformation of an Myb DNA-binding domain, with the third helices of R2 and R3 motifs intercalating in the major groove of DNA. The C-terminal extension forms a β-hairpin followed by a flexible tail, which is stabilized by several interactions with the R3 motif and is not observed in other Myb proteins. Interestingly, this unique C-terminal fragment does not stably connect with DNA in the complex structure but is involved in DNA binding, as demonstrated by NMR chemical shift perturbation, ^{1}H-^{15}N heteronuclear-nuclear Overhauser effect and intermolecular paramagnetic relaxation enhancement. Site-directed mutagenesis also revealed that this C-terminal fragment is crucial for DNA binding, especially the residue Arg^{153} and the fragment K^{170}KRK^{173}. We provide a structural basis for MRE-1 DNA recognition and suggest a possible post-translational regulation of *tv*Myb3 protein.**

INTRODUCTION

Trichomonas vaginalis is one of the most widespread sexually transmitted human pathogens. Infection with *T. vaginalis* causes the disease trichomoniasis and is also associated with several adverse health consequences, including increased HIV transmission and infertility (1,2). With the increasing number of drug-resistant clinical *T. vaginalis* strains (3,4), the infection caused by *T. vaginalis* could become a major threat to public health. In *T. vaginalis*, the iron-responsive promoter region of the gene *ap65-1* has two discrete Myb protein-recognition sites, MRE-1/MRE-2r (*TAACGA*TA, MRE-1 in italic and MRE-2r underlined) and MRE-2f (TATCGT), interspersed among several closely spaced DNA elements (5,6). The *ap65-1* gene encodes a 65-kDa hydrogenosomal malic enzyme that is reputed to be a prominent trichomonad adhesin and plays a role in the cytoadherence of *T. vaginalis* (7,8). Three *T. vaginalis* Myb transcription factors, *tv*Myb1, *tv*Myb2 and *tv*Myb3 (Figure 1), were identified to regulate *ap65-1* (9–11). The *tv*Myb1 and *tv*Myb2 proteins, identified by Southwestern library screening with a probe of concatenated MRE-2f sequences, conferred antagonistic actions on basal and iron-inducible transcription of *ap65-1* through dual

*To whom correspondence should be addressed. Tel: +886 2 2788 2743; Fax: +886 2 2782 6085; Email: hsiao@gate.sinica.edu.tw
Correspondence may also be addressed to Chinpan Chen. Tel: +886 2 2652 3035; Fax: +886 2 2788 7641; Email: bmchinp@ibms.sinica.edu.tw

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

© The Author(s) 2011. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

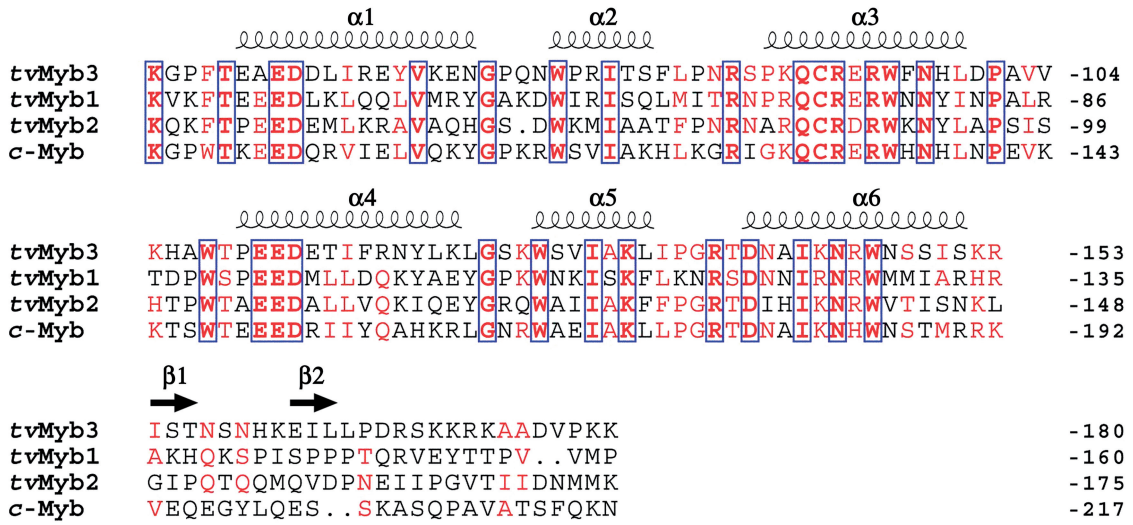

Figure 1. Sequence alignment of DBD of *tv*Myb3, *tv*Myb1, *tv*Myb2 and c-Myb. Secondary structure of *tv*Myb3$_{53–180}$ is shown on the top of the amino acid sequence. Amino acid numbering for each protein is indicated in the termini of sequence.

recognition and differential promoter selection toward both the MRE-1/MRE-2r and MRE-2f sites (10,11). Recently, *tv*Myb3 protein was identified with use of a probe with concatenated MRE-1/MRE-2r sequences. The *tv*Myb3 protein was found to recognize only the MRE-1 element to activate basal and iron-inducible *ap65-1* transcription. These studies demonstrate that *tv*Myb2 and *tv*Myb3 may coactivate basal and iron-inducible *ap65-1* transcription against *tv*Myb1 through conditional and competitive promoter entries (9).

Myb proteins are characterized by a highly conserved DNA-binding domain (DBD), the Myb domain (12). Myb DBD generally consists of 1–4 imperfect repeats (R), with each repeat typically containing approximately 50 amino acids divided by three regularly spaced Trp (or hydrophobic) residues that form a hydrophobic core in a 3D structure. The repeat forms a three-helix bundle structure, and the second and the third helices adopt a helix-turn-helix conformation, with the third helix (the recognition helix) directly contacting DNA in the major groove (12). In vertebrates, c-Myb protein was first identified to contain a three-repeat R1R2R3 DBD (13). Structural study showed that c-Myb R2 and R3 motifs are packed closely with each other so that the two recognition helices can cooperatively recognize specific DNA in the major groove. In *Arabidopsis thaliana*, the Myb protein family is large and contains more than 190 distinct members, most of which contain an R2R3 DBD (126 genes) with at least 40% sequence identity (12,14). Recently, Myb DBD from plant telomere-binding protein was found to contain a C-terminal extension required for binding plant telomeric DNA (15,16). Structural studies further revealed that this Myb DBD adopts a novel four-helix tetrahedron structure and the C-terminal extension forms the fourth helix, which, although not interacting with DNA directly, can stabilize the overall structure of DBD and is crucial for telomeric DNA recognition (17–19).

The *tv*Myb3 protein contains R1, R2 and R3 motifs and shares high sequence homology with *tv*Myb1, *tv*Myb2 and human c-Myb in the R2 and R3 motifs (Figure 1). However, *tv*Myb3 protein contains a C-terminal extension rich in basic amino acid residues (R^{168}SKKRK^{173}), which resembles the canonical nuclear localization signals (20,21). In the present study, we used electrophoretic mobility shift assay (EMSA) and isothermal titration calorimetry (ITC) and identified the DBD of *tv*Myb3 to be the fragment spanning residues 53–180, which encompasses R2, R3 motifs and a C-terminal extension (designated *tv*Myb3$_{53–180}$). This fragment shows about a 40 times increase in MRE-1 DNA binding ability than the shorter fragment (*tv*Myb3$_{53–161}$), which encompasses a similar R2R3 DBD of *tv*Myb1 that was shown in our previous study to interact with MRE-1 DNA strongly. The crystal structure of *tv*Myb3$_{53–180}$ in complex with MRE-1 DNA was then determined; the C-terminal extension forms a β-hairpin conformation followed by a flexible tail, with missing electron density for the last 10 residues. The roles of C-terminal extension, including the β-hairpin and the C-terminal tail, in DNA binding were further investigated by NMR and site-directed mutagenesis. Our study provides a structural basis for MRE-1 DNA recognition, identifies a novel β-hairpin conformation in the C-terminal extension and suggests a possible regulation of MRE-1 element by acetylation or phosphorylation at the C-terminal extension of the *tv*Myb3 DBD.

MATERIALS AND METHODS

Sample preparation

The DNA fragments encoding *tv*Myb3$_{53–161}$ and *tv*Myb3$_{53–180}$ were cloned into the pET29b (Novagen) vector using NdeI and XhoI restriction sites to obtain a recombinant protein with a C-terminal His-tag. The various mutants of *tv*Myb3$_{53–180}$ were generated by the QuikChange site-directed mutagenesis protocol (Stratagene) and confirmed by DNA sequencing. The steps for expression and purification of *tv*Myb3

recombinant proteins are similar to those of *tv*Myb1 (22). All proteins were expressed in *Escherichia coli* BL21(DE3). For labeled (^{15}N, $^{15}N/^{13}C$ and $^{15}N/^{13}C/^{2}H$) protein samples, the cells were grown in H_2O- or D_2O-containing M9 minimal medium containing $^{15}NH_4Cl$ (1 g/l) and/or ^{13}C-glucose (2 g/l) at 37°C. The cells were lysed by microfluidizer, and the proteins were purified by nickel–nitrilotriacetic acid (Ni–NTA) affinity chromatography. Purity and authenticity of the recombinant proteins were verified by SDS–PAGE and mass analysis. Finally, the target protein was dialyzed and concentrated with buffer (20 mM NaH_2PO_4, 50 mM NaCl, 0.5 mM NaN_3 and 10 mM dithiothreitol) at pH 6.0 for study by NMR and circular dichroism (CD). The single-stranded DNAs of MRE-1, MRE-2f and biotinylated DNAs for surface plasmon resonance (SPR) experiments were purchased from MDBio Inc. (Taiwan). The double-stranded DNA was prepared by mixing equal amounts of two complementary deoxynucleotides, heating to 95°C for 10 min and cooling slowly to room temperature.

EMSA

Oligonucleotides were labeled at the 5′-end with [γ-^{32}P] ATP (Amersham Pharmacia Biotech) by T4 polynucleotide kinase as described by the supplier (Promega). All protein samples were incubated with γ-^{32}P-labeled 16-bp MRE-1 or MRE-2f. The mixtures were separated on a 10% acrylamide gel by electrophoresis. The DNA–protein complexes were detected and imaged by use of a Typhoon 9410 variable mode imager.

ITC

ITC experiments were performed at 25°C with a MicroCal iTC200 system (GE Healthcare). The DNA and protein samples were dialyzed in phosphate buffer containing 20 mM NaH_2PO_4, 150 mM NaCl, 0.5 mM NaN_3 and 1 mM dithiothreitol at pH 6.0. The proteins (30 μM) in the calorimetric cell were titrated with 300 μM DNA in the injection syringe. Titration was initiated by one 1-μl injection followed by 19 injections of 2 μl spaced by 180 s intervals. Control experiments involved titrating the DNA samples into buffer. The raw calorimetry data were collected and analyzed by use of the software Origin 7.0 (Microcal Inc., Northampton, MA, USA). The binding isotherms were fitted to the one-site binding model, giving values of the stoichiometry (n) of the interaction, the enthalpy of binding (ΔH) and the association constant (K_a).

Crystallization and data collection

Crystallization trials of the *tv*Myb3$_{53–180}$/16-bp MRE-1 DNA complex involved the hanging-drop vapor-diffusion method. The *tv*Myb3$_{53–180}$/DNA complex (1:1 molar ratio) was prepared in a buffer containing 20 mM HEPES, pH 7.0, 50 mM NaCl and 1 mM dithiothreitol. Initial crystals were obtained by use of crystal screen kits (Hampton Research) at 4 mg/ml. These crystals grew in a buffer of 0.1 M sodium cacodylate trihydrate, 0.1 M magnesium acetate tetrahydrate and 10% w/v PEG8000 at pH 6.4. The cryogenic data for the *tv*Myb3$_{53–180}$/DNA crystal were collected by the use of the BL13B1 beamline of the National Synchrotron Radiation Research Center (NSRRC) in Taiwan. Crystal was flash-frozen in a liquid-nitrogen stream at 100 K. The data were processed and scaled by use of the program HKL2000 (23). The crystal diffracted to 2.9 Å and belongs to the P1 space group with unit-cell parameters, $a = 45.81$, $b = 71.77$, $c = 87.82$, $\alpha = 94.68$, $\beta = 97.84$ and $\gamma = 99.28$, and R_{merge} 4.6%. The V_M was calculated as 2.71 Å^{3}Da^{-1}, corresponding to a solvent content of 45%, containing four molecules per asymmetric unit.

Structure determination and refinement

The crystal structure of the *tv*Myb3$_{53–180}$/16-bp MRE-1 DNA complex was determined by the molecular replacement method with MOLREP (24). The c-Myb/DNA complex (PDB code: 1H8A) with the highest protein sequence identity of 50% with *tv*Myb3$_{53–180}$/DNA was selected as the template model. Data between 8.0 and 4.0 Å and a Patterson radius of 25 Å were used for the rotation and translation function calculations. Significant rotation and translation solutions were obtained for the *tv*Myb3$_{53–180}$/DNA. Structural model building involved XtalView programs (25) and structural refinement was calculated by the use of CNS (26). A bulk solvent mask was calculated to improve the reflection data. The final model for the *tv*Myb3$_{53–180}$/DNA contains Lys^{53} to Lys^{170} in one molecule and Lys^{53} to Ser^{169} in the others. A total of 98 water molecules and four DNA molecules were located in the electron density map. The model was refined to a R_{factor} of 21.8% for all reflections above 2σ and a R_{free} of 27.3% using 5.0% randomly distributed reflections at 2.9 Å. Geometry of the structure checked by a Ramachandran plot with use of PROCHECK (27) showed 80.4% of protein residues in the most favored regions and no residues in disallowed regions. The statistics of data collection and structure refinement are in Table 2.

SPR

The real-time sensorgrams of *tv*Myb3$_{53–180}$ and several mutants interacting with a 21-bp MRE-1 DNA duplex were measured at 25°C on a BIAcore 3000 instrument (GE Healthcare). The 5′-biotinylated MRE-1 DNA duplex dissolved in PBS buffer with a concentration of 0.02 μM was immobilized to the streptavidin SA sensor chip at a flow rate of 10 μl/min. Protein samples were injected into the SA sensor chip immobilized with DNA at a constant flow rate of 30 μl/min for 3 min for association and then the running buffer containing 10 mM HEPES, 150 mM NaCl, 3.4 mM EDTA, pH 7.4, 0.05% (v/v) Tween-20 and 1 mM dithiothreitol was applied at the same rate for 3 min for dissociation. After each injection, the sensor chip was regenerated with 1 M NaCl. For analyzing the relative protein binding, the sensorgrams were evaluated by use of BIAevaluation v4.1 (BIAcore AB, Uppsala, Sweden). All values of the relative binding were the averages from three to five independent experiments.

NMR spectroscopy

NMR samples (unlabeled or ^{15}N, $^{15}N/^{13}C$ and $^{15}N/^{13}C/^{2}H$ isotopically labeled proteins) at pH 6.0 in phosphate buffer were prepared. Each sample contained ~0.30 ml of 1–2 mM protein in a Shigemi NMR tube. All NMR spectra were acquired with the use of Bruker AVANCE 600 or 800 MHz spectrometers equipped with a triple (^{1}H, ^{13}C and ^{15}N) resonance cryoprobe, which included a shielded *z*-gradient at 298 or 310 K. Traditional triple-resonance experiments [HNCO, HN(CA)CO, CBCA(CO)NH and HNCACB] and transverse relaxation-optimized spectroscopy-based triple-resonance experiments were used for backbone resonance assignments of free *tv*Myb3$_{53–180}$ and *tv*Myb3$_{53–180}$ in complex with DNA, respectively. Side-chain resonance assignment followed previously published procedures (28). Nuclear Overhauser effect (NOE) data were extracted from 3D ^{1}H-^{15}N NOE spectroscopy (NOESY)-heteronuclear single-quantum coherence (HSQC) spectra with 120 ms mixing time. The weighted chemical shift perturbations (CSPs) for backbone ^{15}N and $^{1}H_N$ resonances were calculated by the following equation: $\Delta\delta = [(\Delta\delta_{HN})^2 + (\Delta\delta_N/5)^2]^{0.5}$. ^{1}H-^{15}N steady-state heteronuclear NOE (HX-NOE) experiments involved a 5 ms separated train of 120° ^{1}H saturation pulses and a recycle delay of 5 s (29). To measure residual dipolar couplings (RDCs) of *tv*Myb3$_{53–180}$ in complex with DNA, the filamentous bacteriophage Pf1 (Asla Biotech Ltd, Latvia) was selected as the orienting medium. 2D ^{1}H-coupled (F1) IPAP ^{1}H-^{15}N HSQC spectra (30) were acquired with 256 complex t_1 (15 N) points and 64 scans per t_1 increment for both the isotropic and anisotropic samples. The quality factor of RDC values equals RMS($D^{calc} - D^{obs}$)/RMS(D^{obs}), where D^{calc} and D^{obs} are calculated and observed RDC values, respectively. Processing of all NMR spectra involved use of Bruker Topspin 2.0 (Karlsruhe, Germany) or NMRPipe (31), and data were analyzed by use of NMRView (32).

Paramagnetic relaxation enhancement measurements

To observe the intermolecular paramagnetic relaxation enhancement (PRE) effects on *tv*Myb3$_{53–180}$, an EDTA was conjugated with deoxythymidine (dT-EDTA) and introduced into DNA as described (33). The DNA fragments with dT-EDTA were purchased from Yao-Hong Biotechnology Inc. (Taiwan). Purity and authenticity of individual strands of DNA were verified by chromatography and mass analysis. The double-stranded DNA was prepared by mixing equal amounts of two complementary deoxynucleotides, heating to 95°C for 10 min and cooling slowly to room temperature. The EDTA-conjugated DNA duplex was mixed with equal amounts of metal ions (Mn^{2+} or Ca^{2+}), and *tv*Myb3$_{53–180}$ (0.2 mM) was added into the mixture at a protein–DNA ratio of 1:1.5. PRE $^{1}H_N$-Γ_2 measurement involved 2D ^{1}H-^{15}N correlated spectra with a two time-point delay T of 10 ms as described (34). Values of $^{1}H_N$-Γ_2 were calculated from peak intensities for the diamagnetic (chelated with Ca^{2+}) and paramagnetic (chelated with Mn^{2+}) state as described (34).

RESULTS

Identification of the DBD of *tv*Myb3

Previously, we mapped the DBD of *tv*Myb1 as the fragment from residue 35 to 141 and determined its solution structure as well as its complex model with MRE-1 DNA (22). To identify the *tv*Myb3 DBD, we generated two protein fragments and examined their binding interactions with the 16-bp DNA containing the MRE-1 of 5′-ATAAC GAT-3′ or MRE-2f of 5′-GTATCGTC-3′ by EMSA. The shorter fragment, *tv*Myb3$_{53–161}$, which includes similar R2R3 motifs as *tv*Myb1$_{35–141}$ (Figure 1), is not sufficient for binding to both DNAs (Figure 2A). Interestingly, *tv*Myb3$_{53–180}$ that contains 19 additional C-terminal amino acids and displays similar helical conformation and slightly higher thermal stability than the shorter *tv*Myb3 (Supplementary Figure S1), exhibits the DNA-binding ability.

To confirm the EMSA results and to quantitate the DNA-binding affinity of the two *tv*Myb3 fragments, we monitored the protein–DNA interactions at 150 mM NaCl by ITC (Figure 2B). Fitting of ITC thermograms showed that *tv*Myb3 binds with 1:1 stoichiometry to the specific MRE-1 or MRE-2f DNA. The formation of the *tv*Myb3/MRE-1 complex is enthalpically driven and that of the *tv*Myb3/MRE-2f complex entropically driven. The association constant (K_a) value of *tv*Myb3$_{53–180}$ bound to 16-bp MRE-1 was $1.51 \pm 0.20 \times 10^7\,M^{-1}$ and was ~10 times weaker for MRE-2f DNA (Table 1). Also, the shorter fragment bound to MRE-1 DNA ~40 times weaker than the longer *tv*Myb3, and the interaction with MRE-2f DNA was too weak to be determined by ITC, which agrees with the EMSA results. Interestingly, with thymine replaced at position 8 by adenine (MRE-2f_T8A), which changes the sequence to 5′-G***TA<u>A</u>C G***TC-3′ (the mutated base is underlined and the conserved bases with MRE-1 are in bold italics), the K_a value of *tv*Myb3$_{53–180}$ bound to MRE-2f_T8A was similar to that of the *tv*Myb3$_{53–180}$/MRE-1 complex formation, which suggests that *tv*Myb3$_{53–180}$ recognizes the sequence 5′-TA ACG-3′ specifically. The longer fragment, *tv*Myb3$_{53–180}$, and the MRE-1 DNA were hence selected for complex structure study.

Crystal structure of *tv*Myb3$_{53–180}$ bound to MRE-1 DNA

The overall structures of *tv*Myb3$_{53–180}$ bound to a 16-bp MRE-1 DNA were determined at 2.9 Å (Table 2) by X-ray crystallography. The asymmetric unit of the crystal contains four protein molecules in complex with four DNA duplexes (Figure 3A). The N-terminal parts, from residues Lys^{53} to Ser^{151}, of all four protein structures revealed the classical R2R3 Myb DBD, with the third helices of R2 and R3 motifs intercalating in the major groove of DNA. Interestingly, residues from Ile^{154} to Leu^{164} form a β-hairpin conformation, which was not previously observed in Myb proteins. Additionally, the electron density was not observed for the C-terminal 10 or 11 residues, which suggests that they are flexible in the crystals. The superimposition of four protein structures for all backbone atoms shows a high degree of structural

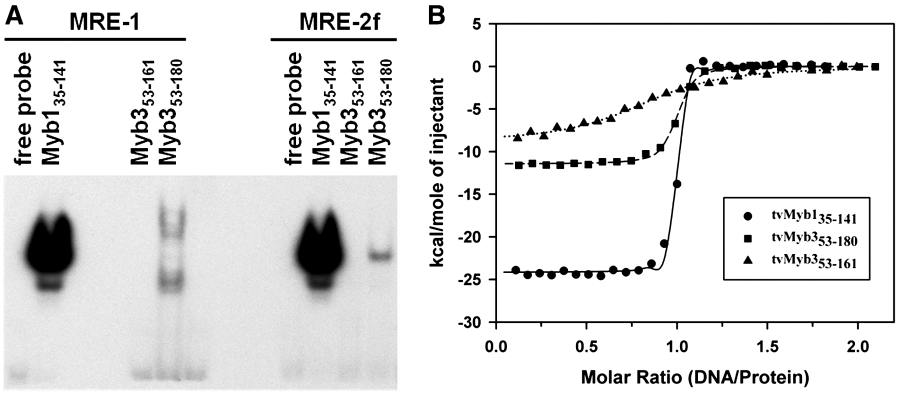

Figure 2. DNA-binding activity of *tv*Myb1$_{35–141}$, *tv*Myb3$_{53–161}$ and *tv*Myb3$_{53–180}$ probed by EMSA and ITC. (**A**) EMSA experiments were performed using 16-bp oligonucleotides. Three pico mole Myb proteins were incubated with a γ-^{32}P-labeled MRE-1 DNA. Three pico mole *tv*Myb1$_{35-141}$ and 30 pmol *tv*Myb3 proteins were incubated with a γ-^{32}P-labeled MRE-2f DNA. All protein–DNA mixtures were separated by 10% PAGE. (**B**) ITC thermograms for sequential 2-μl additions of 16 bp MRE-1 DNA (300 μM) to *tv*Myb proteins (30 μM). The lines were fitted to the integrated heats using a single-site binding model.

Table 1. Thermodynamic parameters of Myb DBDs binding to DNAs

Sample	*n*	K_a (M^{-1})	ΔG (kcal mol^{-1})	ΔH (kcal mol^{-1})	ΔS (cal mol^{-1} K^{-1})
MRE-1 (16 bp)					
*tv*Myb1$_{35–141}$	1.01 ± 0.08	1.48 (±0.45) × 10^8	−11.14	−24.18 ± 0.09	−43.77 ± 0.55
*tv*Myb3$_{53–161}$	0.98 ± 0.05	3.40 (±0.43) × 10^5	−7.54	−8.75 ± 0.07	−4.07 ± 0.02
*tv*Myb3$_{53–180}$	0.99 ± 0.05	1.51 (±0.20) × 10^7	−9.79	−11.41 ± 0.04	−5.44 ± 0.23
AAAA	1.10 ± 0.01	2.86 (±0.46) × 10^6	−8.80	−10.85 ± 0.14	−6.87
MRE-1 (21 bp)					
*tv*Myb3$_{53–180}$	1.09 ± 0.00	6.02 (±0.96) × 10^7	−10.62	−13.00 ± 0.05	−7.99
S169D	1.01 ± 0.00	2.67 (±0.41) × 10^7	−10.14	−12.93 ± 0.08	−9.38
AAAA	1.01 ± 0.00	5.12 (±0.41) × 10^6	−9.16	−13.21 ± 0.07	−13.60
MRE-2f_T8A (16 bp)					
*tv*Myb3$_{53–180}$	1.05 ± 0.03	1.17 (±0.59) × 10^7	−9.60	−6.69 ± 0.08	9.77 ± 1.32
MRE-2f (16 bp)					
*tv*Myb1$_{35–141}$	1.10 ± 0.09	1.17 (±0.14) × 10^8	−11.01	−23.24 ± 0.24	−41.03 ± 0.96
*tv*Myb3$_{53–161}$	ND[b]	ND[b]	ND[b]	ND[b]	ND[b]
*tv*Myb3$_{53-180}$	1.03 ± 0.13	1.39 (±0.32) × 10^6	−8.387	−3.49 ± 0.22	16.4 ± 0.95

[a]$\Delta G = -RT \cdot \ln K_a$, R = 1.985 cal·$K^{-1}mol^{-1}$, gas constant.
[b]ND, not determined due to the small ΔH.

similarity, with RMSD of 0.47 ± 0.05 Å for backbone atoms. However, fitting only the backbone atoms of the R2 and R3 motifs (Lys^{53} to Ser^{151}) showed that the RMSD for this region can be lowered to 0.35 ± 0.03 Å, but the deviation of the C-terminal residues (Lys^{152} to Ser^{169}) was increased to 0.95 ± 0.19 Å (Supplementary Figure S2), which implies that the structure of the C-terminal fragment in the complex state is relatively more flexible than that of R2 and R3 motifs.

To select a representative complex structure, we determined the one-bond RDC constant, $^1D_{NH}$, of *tv*Myb3$_{53–180}$ in complex with DNA by NMR spectroscopy in solution. With the four crystal structures used as templates, the theoretical $^1D_{NH}$ RDC values can be calculated. For the residues in secondary structural regions, the experimental RDC values fit well to the theoretical values from four protein structures with quality factors between 0.37 and 0.43 (Supplementary Figure S3). However, the experimental RDC values of all residues fit best to the values from the fourth complex structure with a quality factor of 0.65, which suggests that the fourth complex structure in crystalline state is similar to that in solution state.

In the fourth complex structure, *tv*Myb3$_{53–180}$ consists of six α-helices, a β-hairpin and a C-terminal tail (Figure 3B), with α1 (Glu^{58} to Asn^{71}), α2 (Trp^{76} to Ser^{81}), α3 (Pro^{88} to Leu^{99}), α4 (Pro^{110} to Lys^{122}), α5 (Trp^{127} to Leu^{133}), α6 (Asp^{139} to Ser^{151}), β1 (Ile^{154} to Thr^{156}) and β2 (Glu^{162} to Leu^{164}). The three-helix bundles of R2 and R3 motifs are stabilized by hydrophobic and electrostatic interactions. Two hydrogen bonds (Val^{104} and Gly^{136}, Asn^{97} and Thr^{138}) and 1 van der Waals interaction between the side chains of Val^{104} and Arg^{137} span the R2 and R3 motifs. Notably, the C-terminal fragment makes several contacts with the R3 motif (Figure 3C). Leu^{164} on β2 takes part in the formation of the R3 hydrophobic core. Arg^{153} and Asp^{167} form a salt bridge network with Asp^{113} on α4, Arg^{145} on α6 and O2P of THY5. In addition, O′ of Glu^{162} on β2 constitutes a hydrogen bond with H$^{\eta 12}$ of

Table 2. Data collection and refinement statistics

Crystals	$Myb3_{53–180}$/DNA
Data collection	
Space group	*P*1
Cell dimensions	
a (Å)	45.81
b (Å)	71.77
c (Å)	87.82
α (°)	94.68
β (°)	97.84
r (°)	99.28
Wavelength (Å)	1
Resolution (Å)	30–2.9 (3.0–2.9)[a]
Number of total reflections	88 446
Number of unique reflections	23 110
Completeness (%)	95.5 (87.5)
Redundancy	3.8 (3.5)
R_{merge} (%)[b]	4.6 (22.7)
I/σ (I)	29.46 (5.02)
Refinement	
Number of reflections	19 482
R_{work} (%)	21.8
R_{free} (%)[c]	27.3
Geometry deviations	
Bond length (Å)	0.009
Bond angles (°)	1.6
Mean B-values (Å) / No.	
Protein atoms	63.01/3929
DNA	59.82/2600
Water molecules	36.79/98
Ramachandran plot (%)	
Most favored	80.4
Additionally allowed	19.6

[a]Values for the highest resolution shell are shown in parentheses.
[b]$R_{merge} = \Sigma|I - <I>| / \Sigma I$, where I = observed intensity and $<I>$ = average intensity from multiple observations f symmetry related reflections.
[c]R_{free} was calculated on the basis of 5% of the total number of reflections randomly omitted from the refinement.

Arg^{118} on α4. The C-terminal β-hairpin conformation seems to be stabilized by these interactions.

Structural basis for MRE-1/MRE-2 r recognition by *tv*Myb3$_{53–180}$

In the representative *tv*$Myb3_{53–180}$/DNA complex structure, the recognition helices (helices 3 and 6) insert into the major groove of the MRE-1 DNA duplex. All interactions between protein and DNA are illustrated in Figure 3D. The side chains of Lys^{89}, Glu^{93} in helix 3 and Asn^{140}, Lys^{143} and Asn^{144} in helix 6 form specific hydrogen bonds with the bases of GUA9, CYT8, ADE7, GUA8′ and ADE6, respectively (Supplementary Figure S4). Moreover, several residues, including Gln^{74}, Gln^{90}, Arg^{92}, Arg^{94} and His^{98} in the R2 motif and Trp^{108}, Ser^{125}, Trp^{127} and Ser^{128} in the R3 motif contact with the phosphate groups of the DNA backbone in the major groove. Additionally, the backbone atoms of Gly^{54}, Pro^{55} and Phe^{56} interact with the phosphate groups of ADE7 and CYT8 in the minor groove. However, there are no direct contacts between DNA and the β-hairpin or the C-terminal tail. The nearest distance is 4.42 Å between $N^{\eta 1}$ of Arg^{168} and O1P of GUA3.

In the preceding result of identification of the *tv*Myb3 DBD, the DNA-binding affinity of *tv*$Myb3_{53–180}$ is 40 times stronger than that of *tv*$Myb3_{53–161}$. Therefore, the observation of no interaction between DNA and the C-terminal residues of *tv*$Myb3_{53–180}$ in the X-ray structure is a surprise. The roles of the C-terminal fragment, including the β-hairpin, the C-terminal tail and the missing residues, in DNA binding need further investigation as described below.

NMR investigations of the C-terminal fragment

We observed no electron density for the C-terminal 10 residues 'K^{171}RKAADVPKK^{180}' in the crystal structure of the *tv*$Myb3_{53–180}$/DNA complex, which may due to its high flexibility. However, the DNA length in the crystal may be too short for contacting with the last 10 residues. Hence, in addition to the 16-bp DNA, 21-, 25- and 30-bp DNAs were titrated into ^{15}N-labeled *tv*$Myb3_{53–180}$ to characterize their interactions by NMR spectroscopy. 2D ^{1}H-^{15}N HSQC spectra showed backbone amide resonances of several residues in complex with 21-bp DNA significantly different from those in complex with 16-bp DNA (Figure 4A, left panel), and longer DNA did not cause substantial further perturbations in these residues (Figure 4A, right panel), which suggests that the 21-bp MRE-1 sequence is suitable for *tv*$Myb3_{53–180}$/DNA complex formation.

The backbone resonance assignments of *tv*$Myb3_{53–180}$ in the free state and in complex with 16- and 21-bp DNAs were achieved by several 3D triple-resonance experiments (Supplementary Figure S5). Backbone resonances of residues in the R2 and R3 motifs (Lys^{53} to Ser^{151}) undergo similar CSPs in complex with 16- and 21-bp DNAs (Figure 4B). Most of the largely perturbed residues ($\Delta\delta > \Delta\delta_{average} + SD = \sim 0.5$ ppm) in the R2R3 fragment are located at the N terminus, the loop connecting helixes 1–3, the loop connecting R2 and R3 motifs, and helix 6 (Figure 4C). One additional strong CSP was observed for Ser^{128} located in helix 5. For the residues located at the C-terminal fragment (Lys^{152} to Lys^{180}), the complex sample with 21-bp DNA shows significantly larger CSP values than those with 16-bp DNA, which suggests that these residues have stronger interactions with the longer DNA than the shorter one.

We then checked the backbone rigidity of *tv*$Myb3_{53–180}$ in the absence and presence of DNA by ^{1}H-^{15}N HX-NOE measurements (Figure 4D). HX-NOE values of almost all *tv*$Myb3_{53–180}$ residues were significantly increased in the presence of DNA; especially the N- (Lys^{53} to Thr^{57}) and C-terminal residues (Ile^{154} to Lys^{180}), which are highly flexible in the absence of DNA, adopt a more rigid conformation on binding to DNA. The average HX-NOE values of the R2 and R3 motifs (Lys^{53} to Ser^{151}) in the absence and presence of 16- and 21-bp DNAs were 0.78 ± 0.08, 0.84 ± 0.07 and 0.85 ± 0.08, respectively. However, the average values of the C-terminal fragment (Lys^{152} to Lys^{180}) were increased from 0.40 ± 0.26 for the free state to 0.52 ± 0.18 for the complex with 16-bp DNA and 0.57 ± 0.18 for binding to 21-bp DNA. Although the C-terminal fragment of *tv*$Myb3_{53–180}$ was still flexible, the

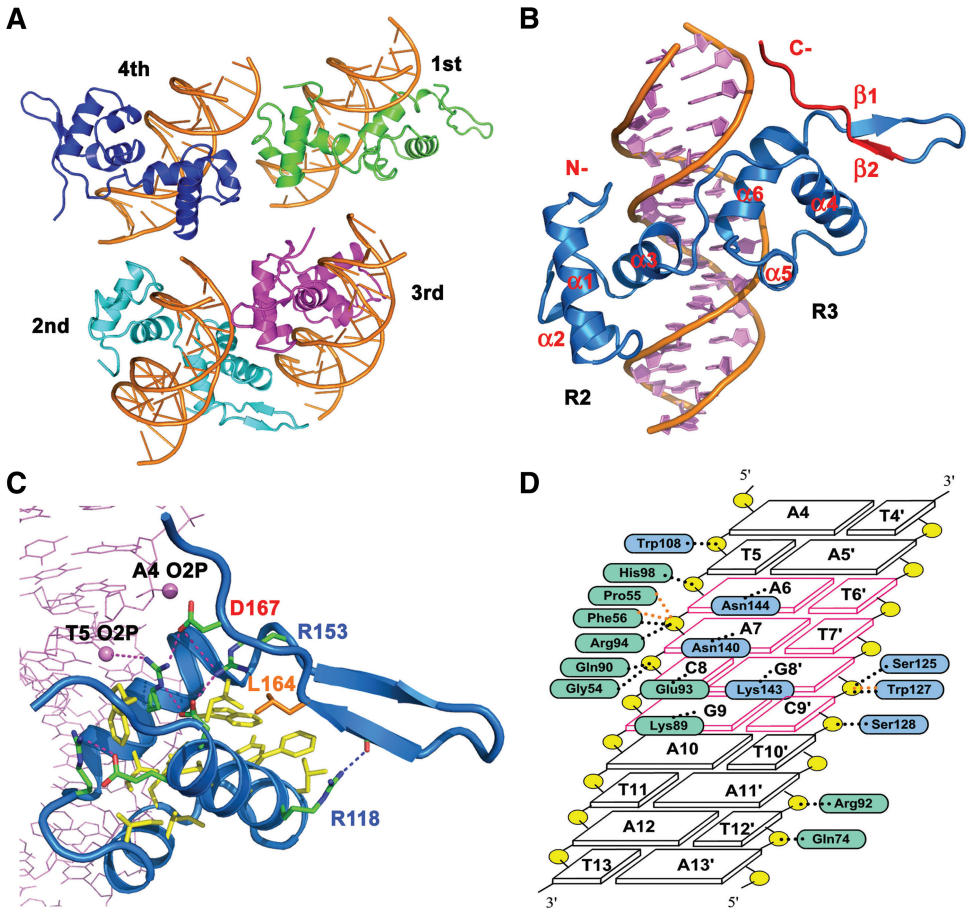

Figure 3. Crystal structures and interactions of *tv*Myb3$_{53–180}$ in complex with 16-bp MRE-1 DNA. (**A**) There are four proteins in complex with four DNA duplexes in the asymmetric unit of the crystal. (**B**) The representative complex structure, in which, *tv*Myb3$_{53–180}$ consists of six α-helices, a β-hairpin and a C-terminal tail. The electron density for the C-terminal 10 residues was not observed, and the residues from Glu^{162} to Lys^{170} are colored in red. (**C**) The hydrophobic, electrostatic and hydrogen-bonding interactions between the R3 motif and C-terminal β-hairpin. The involved residues in the C-terminal β-hairpin are labeled. Two O2P atoms of ADE4 and THY5 are shown. The salt bridges are linked with dashed lines in magenta and hydrogen bond in blue. (**D**) Schematic representation illustrating the contacts between *tv*Myb3$_{53–180}$ and MRE-1 DNA. The amino acid residues in R2 and R3 motifs are shown with green and blue background colors, respectively. The van der Waals contacts and hydrogen-bonding interactions are indicated by orange and black dotted lines, respectively. The specific AACG base pairs are colored in pink. The phosphates are indicated in yellow circles.

amide pico-second to nano-second time scale motions indeed are stabilized by the extension of DNA length.

The results from NMR CSPs and backbone HX-NOE values reveal that the interactions between DNA and the C-terminal fragment, including the last 10 residues missing in the X-ray structure, were increased when the DNA length was extended from 16 to 21 bp. So we tried to crystallize the *tv*Myb3$_{53–180}$/21-bp MRE-1 DNA complex sample, but the resolution of resulting crystals was too low to determine the structure. Instead, we used intermolecular PRE to investigate the interactions between DNA and the C-terminal region of *tv*Myb3$_{53–180}$. We purchased the 21-bp DNA duplex containing the EDTA-derivatized deoxythymidine (dT-EDTA) at the desired site. Values of PRE of amide protons (^1H_N-Γ_2) for *tv*Myb3$_{53–180}$ were determined from a two time-point measurement in both the paramagnetic (chelated with Mn^{2+}) and diamagnetic (chelated with Ca^{2+}) states (33,34). The residues exhibiting ^1H_N-$\Gamma_2 > 5\,s^{-1}$ are located at the N terminus, the loop connecting the R2 and R3 motifs, helix 6 and the C-terminal tail (Figure 5A). A structural model of *tv*Myb3$_{53–180}$ in complex with a 21-bp DNA duplex was generated by extending the 16-bp DNA in the crystal structure with a standard B-form conformation. We mapped the intermolecular PRE ^1H_N-Γ_2 values on the structural model (Figure 5B), and found that most of the PRE ^1H_N-Γ_2 values correlated well with the distances measured in the model structure, with ^1H_N-Γ_2 between 14 and $10\,s^{-1}$ corresponding to the distance from 18 to 25 Å and ^1H_N-Γ_2 between 10 and $5\,s^{-1}$ corresponding to the distance from 24 to 29 Å (35). The C-terminal residue Arg^{168} exhibits the largest ^1H_N-Γ_2 value of $13.3\,s^{-1}$, which suggests that it is close to the DNA although it does not form any direct contact with DNA in the crystal structure. Also, the PRE values of the C-terminal 12 residues, like Ser^{169} and Lys^{170}, which should be in regions closest to Mn^{2+} if they formed stable contacts with the extended DNA, are all $>5\,s^{-1}$ but smaller than

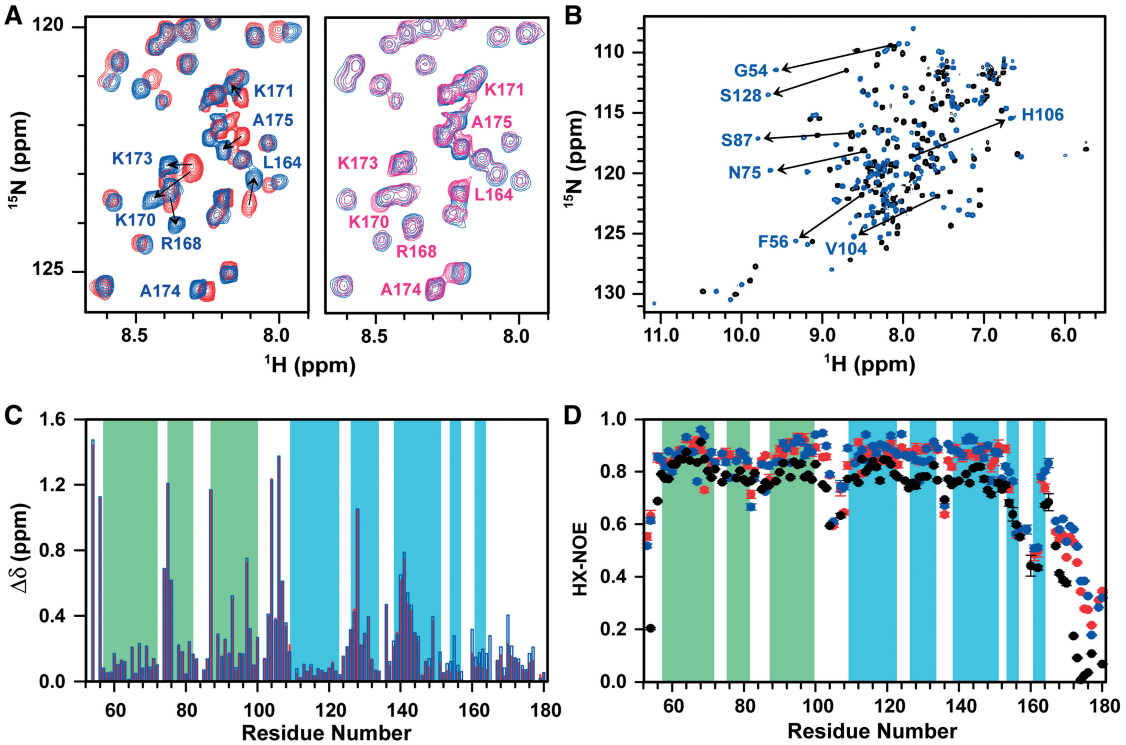

Figure 4. NMR investigations on tvMyb3$_{53–180}$/DNA complexes. (**A**) The overlaid 2D ^{1}H-^{15}N HSQC spectra of tvMyb3$_{53–180}$ in complex with 16-bp (red, left panel), 21-bp (blue, left and right panels) and 25-bp (purple, right panel) MRE-1 DNAs. (**B**) The overlay of the ^{1}H-^{15}N HSQC spectra of free *tv*Myb3$_{53–180}$ (black) and in complex with 21-bp MRE-1 DNA (blue) at 310 K. The residues with large chemical shifts are indicated. (**C**) Weighted CSPs for backbone ^{15}N and ^{1}H$_{N}$ resonances of tvMyb3$_{53–180}$ in complex with 16-bp (red) or 21-bp (blue) DNA as calculated by the following equation: $\Delta\delta = [(\Delta\delta_{HN})^2 + (\Delta\delta_N/5)^2]^{0.5}$. (**D**) ^{1}H-^{15}N heteronuclear NOEs of tvMyb353-180 in free form (black circles), and in complex with 16-bp (red circles) or 21-bp (blue circles) MRE-1 DNAs. The secondary structures are highlighted in (C and D).

that of Arg^{168}, which suggests that these residues are close to DNA but their mobility reduces the observed PRE values.

NMR studies shows that the C-terminal extension is flexible but sensitive to the length of interacted DNA, suggesting that it is involved in DNA binding. The increased intermolecular PRE values of C-terminal missing residues further imply that these residues have transient interactions with DNA.

Mutations on the C-terminal fragment

From the CSPs, backbone HX-NOE and intermolecular PRE, the C-terminal fragment (Lys^{152} to Lys^{180}) was found to be highly involved in DNA binding, although it does not form stable contacts with DNA in the crystal structure. To further validate the roles of the C-terminal residues in DNA binding, we generated various mutants of *tv*Myb3$_{53–180}$ and monitored their interactions with DNA by SPR at 150 mM NaCl (Figure 5C and Supplementary Figure S6). The mutants R153A and L164A, which disrupt the electrostatic and hydrophobic interactions between the β-hairpin and R3 motif, show a significant decrease in DNA binding ability. Interestingly, the mutants that change one positively charged residue at the C terminus to alanine bind to DNA as strongly as does the wild-type protein. However, the mutants AA1 (K170A, K171A) and AA2 (R172A, K173A) show notable decreases, and AAAA (K170A, K171A, R172A, K173A) exhibits severely reduced DNA binding activity.

The association constants (K_a) of the AAAA mutant bound to 16- and 21-bp MRE-1 DNAs at 150 mM NaCl were measured by ITC (Table 1). For 21-bp DNA, the affinity between DNA and the AAAA mutant ($5.12 \pm 0.41 \times 10^6\,M^{-1}$) was ~12 times weaker than that of the wild-type protein ($60.2 \pm 9.6 \times 10^6\,M^{-1}$). For 16-bp DNA, the reduction in association constant was ~5-fold for the mutant protein. In addition, the K_a value of the S169D mutant bound to 21-bp DNA was about half of that of the wild-type protein (Table 1). The interactions between DNA and S169D, AA1, AA2 or AAAA mutants were all weaker than that between DNA and wild-type protein, so acetylation at the Lys residue or phosphorylation at the Ser residue in the C-terminal fragment will lower the binding affinity between *tv*Myb3$_{53–180}$ and DNA.

Thus, the structural integrity of the R3 motif and the β-hairpin and the positively charged residues at the C terminus are critical for DNA binding.

The β-hairpin conformation is preserved in free *tv*Myb3$_{53–180}$

According to 3D structure comparison by the Dali server, the C-terminal β-hairpin conformation of *tv*Myb3$_{53–180}$ is novel as compared with other Myb-like DBD. Hence,

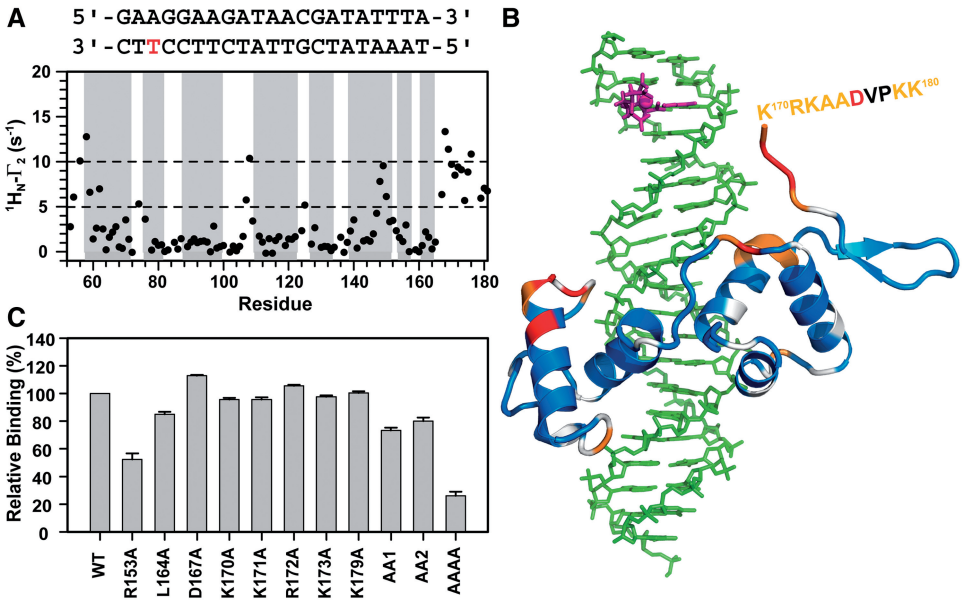

Figure 5. Intermolecular PRE for *tv*Myb3$_{53–180}$/DNA complex and DNA binding activities of *tv*Myb3$_{53–180}$ mutants. (**A**) EDTA labeled site on 21-bp MRE-1 DNA (in red) and intermolecular PRE ^1H_N-Γ_2 values obtained with EDTA chelating with Mn^{2+}. (**B**) The intermolecular PRE values are mapped on a structural model of *tv*Myb3$_{53–180}$ in complex with a 21-bp DNA duplex, which is generated by extending the 16-bp DNA in the crystal structure with standard B-form conformation. EDTA and Mn^{2+} are shown in magenta. The residues exhibiting ^1H_N-Γ_2 values $>10\,s^{-1}$ are in red, 5–10 s^{-1} in orange and $<5\,s^{-1}$ in blue. The missing C-terminal 10 residues are indicated with one-letter abbreviations. (**C**) DNA binding activities of *tv*Myb3$_{53–180}$ mutants probed with SPR. The relative DNA binding activities of mutants are calculated from SPR responses during 15 s before the end of protein injection.

determining whether this unique structure is formed only in the DNA-bound form or is preserved in the free form is of interest. From backbone HX-NOE experiments of free *tv*Myb3$_{53–180}$ (Figure 4D), the residues in the β-strands, especially Leu^{164} and Leu^{165}, exhibit backbone HX-NOE values larger than those of N- and C-terminal residues, which suggests that this region adopts a conformation more stable than a random coil. As well, ^{15}N-edited NOESY-HSQC spectra revealed several long-range NOEs between these C-terminal residues (Arg^{153} to Leu^{165}), like the NOEs between H_α of Thr^{156} and H_N of Lys^{161}–Ile^{163} (Figure 6A). The backbone NOE connectivity among these residues suggests that this fragment adopts a hairpin conformation in free *tv*Myb3$_{53–180}$ (Figure 6B).

DISCUSSION

The C-terminal β-hairpin extension is unique and crucial for MRE-1 DNA recognition

Myb protein, a member of the transcription factor family, represents a functional DBD that is conserved in a wide range of eukaryotic systems. c-Myb protein was first identified in vertebrates and found to contain a three-repeat R1R2R3 DBD. Each repeat of c-Myb forms a three-helix bundle conformation, but only the third helices of R2 and R3 motifs recognize specific DNA in the major groove (13). In plants, most of the double-stranded telomere-binding proteins contain only one Myb-like domain. However, this Myb-like domain alone is not sufficient for telomere binding. A C-terminal extension was found to be highly involved in DNA binding (15,16). Structural studies revealed that instead of adopting the typical fold of Myb DBD, the Myb-like domain of plant double-stranded telomere-binding protein folds into a four-helix tetrahedron structure, and the fourth helix, which is formed by the C-terminal extension, participates in the formation of a hydrophobic core and stabilizes the whole DBD structure for telomere binding (17–19). We determined the structure of *tv*Myb3$_{53–180}$ in complex with MRE-1 DNA and showed that the C-terminal extension forms a β-hairpin conformation, which was demonstrated by NMR to participate in DNA binding and by site-directed mutagenesis to be crucial for DNA binding. Also, NMR studies indicate that the C-terminal β-hairpin conformation is formed in the DNA-bound state as well as in free protein. The unique features of this C-terminal β-hairpin conformation are described below.

First, like the fourth helix in the Myb-like domain of plant double-stranded telomere-binding protein, the C-terminal β-hairpin conformation of *tv*Myb3$_{53–180}$, although not making any direct contact with DNA, plays a role in stabilizing the overall structure and is critical for DNA binding. This observation was confirmed by determining the thermal stability, whereby *tv*Myb3$_{53–180}$ is 3°C more stable than the shorter fragment *tv*Myb3$_{53–161}$, and by the analysis of the mutants R153A and L164A, which disrupt the electrostatic and hydrophobic interactions between the β-hairpin and R3 motif, showing a significant decrease in DNA binding ability. The one exception is the

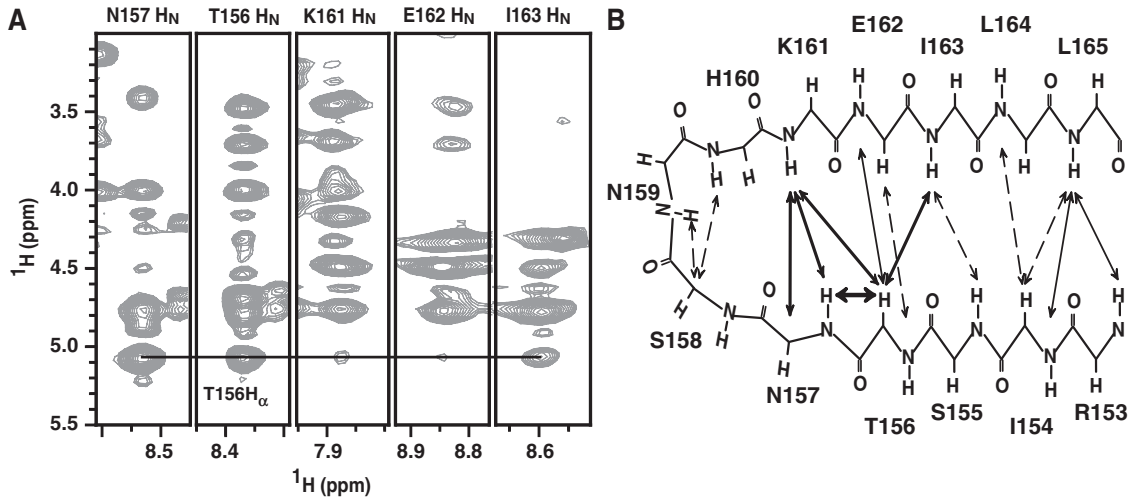

Figure 6. The C-terminal fragment of *tv*Myb3$_{53–180}$ forms a β-hairpin conformation without DNA. (**A**) Strip plots from ^{15}N-edited NOESY-HSQC spectrum of free *tv*Myb3$_{53–180}$, in which NOE cross-peaks between H_α of Thr^{156} and H_N of Asn^{157}–Ile^{163} are connected by a black line. (**B**) NOE connectivity of residues from Arg^{153} to Leu^{165}. Strong, medium and weak NOEs are indicated by solid-thick, solid and solid-thin arrows, respectively. The dashed arrows indicate the ambiguous NOEs, which cannot be clearly identified because of overlapping.

mutant D167A, which, although breaking the salt bridge networks among Asp^{167}, Arg^{153} and Arg^{145} (Figure 3C), showed an increase in DNA binding activity. The mutant D167A may reduce the repulsion forces between $O^{\delta1}$ of Asp^{167} and O2P of ADE4, which is 4.4-Å apart in the crystal structure, and the salt bridge networks can be supplemented by the interactions between $N^{\eta2}$ of Arg^{145} and O2P of THY5, which is 4.1-Å apart in the crystal structure.

Second, the C-terminal tail, which is mostly not observed in the complex structure, is also important for DNA binding, which suggests that the DNA recognition by *tv*Myb3 may be regulated by post-translational modification at these residues. Our NMR studies demonstrated that the C-terminal tail, although flexible, has some interactions with the extended DNA. As well, the AA1, AA2 and AAAA mutants, which replaced the positively charged residues in the C-terminal tail to alanine and S169D, which introduces a negative charge to mimic phosphorylation on Ser residue, all show a significant decrease in DNA binding affinity. In the complex structure, the third helix of the R3 motif contacts with DNA, and the C-terminal basic residues are restricted by the β-hairpin to be in a position next to the third helix of the R3 motif, where these basic side chains may have transient interactions with the DNA phosphate backbone or the minor groove. According to this view, the interaction mode shows similarity with the C-terminal fragment of another DBD, the winged helix or the forkhead box (termed FOX). The FOX motif consists of two wings (W1 and W2), three α-helices (α1, α2 and α3) and three β-strands (β1, β2 and β3), arranged in order of α1–β1–α2–α3–β2–W1–β3–W2. Like the C-terminal half of *tv*Myb3$_{53–180}$, the second and third helices of FOX adopt a helix-turn-helix conformation, β2 and β3 forms an anti-parallel β-sheet and the loop or wing W2 extends from β3 to the C terminus of the DBD. In DNA recognition by canonical FOX proteins, the third helix makes direct contact with DNA in the major groove and the wing W1 interacts with DNA in the minor groove (36). In the FOXO3a/DNA complex structure, the C-terminal W2 basic residues forms hydrogen bonds with the DNA phosphate backbone (37). For FOXO1 protein, the flexible C-terminal fragment is not observed in the complex structure (38). However, the C-terminal coil is necessary for DNA binding in both FOXO3a and FOXO1 proteins, and their DNA binding activities are attenuated by phosphorylation of Ser or acetylation of Lys residues in the C-terminal coil (37,38). According to our study of *tv*Myb3 and its similarity with FOX protein, MRE-1 element recognition by *tv*Myb3 protein may be regulated by phosphorylation or acetylation of the C-terminal residues.

Structural comparison of the R2R3 DBD with other Myb proteins

The *tv*Myb3 protein exhibits high amino acid sequence identity with human c-Myb and *tv*Myb1 in the R2R3 DBD (Figure 1). The structure of c-Myb DBD in complex with DNA has been reported (13,39). In our previous study, we determined the solution structure of free *tv*Myb1$_{35–141}$ and proposed the *tv*Myb1$_{35–141}$/MRE-1 DNA complex model using CSPs, residual dipolar couplings, DNA specificity data and data-driven macromolecular docking by HADDOCK (22). We found that the orientation between R2 and R3 motifs of *tv*Myb1 greatly changes on binding to DNA to recognize the DNA major groove through a number of contacts involving residues in two recognition helices. A superimposition of the six helices from *tv*Myb3 R2R3 DBD with those from *tv*Myb1 and c-Myb showed high similarity in backbone structure, with RMSD of 0.91 and 1.79 Å for C^α atoms, respectively (Supplementary Figure S7), so the three R2R3 DBDs may adopt a similar architecture of protein backbone to recognize the DNA major groove.

For the residues in two recognition helices, *tv*Myb3 protein shows 83% amino acid sequence identity with human c-Myb. The R2R3 DBD of c-Myb recognizes the base sequence 5′-TAACNG-3′ specifically and that of *tv*Myb3 binds to the MRE-1 element containing 5′-TAA CGA-3′ with high specificity. High similarities in both protein and DNA sequences suggest that the protein–DNA interactions of *tv*Myb3 and c-Myb are similar. Indeed, close examination of the two protein–DNA complex structures reveals that most of the residues identified for DNA recognition in *tv*Myb3 can also be found in the c-Myb/DNA complex (Figure 3D and Supplementary Figure S8A). One difference is that Lys^{89} in helix 3 forms specific hydrogen bonds with the bases of GUA9 in the *tv*$Myb3_{53–180}$/DNA complex structure, but the corresponding Lys residue (Lys^{128}) in c-Myb does not contact with the corresponding GUA (GUA21) but rather with GUA22. Also, the residues Trp^{108}, Gly^{54} and Gln^{74} in *tv*Myb3 contact with the DNA phosphate backbone, and these interactions are not conserved in the c-Myb/DNA complex.

The amino acid sequence identity for the residues in two recognition helices between *tv*Myb3 and *tv*Myb1 is only 57%. *tv*Myb1 recognizes the whole MRE-1 element containing 5′-TAACGAT-3′, whereas *tv*Myb3 binds only to 5′-AACG-3′. The examination of two protein–DNA interactions reveals that three specific hydrogen bonds with bases and four hydrogen bonds with a phosphate backbone in the *tv*$Myb3_{53–180}$/DNA complex structure can be found in the *tv*$Myb1_{35–141}$/DNA complex model (Figure 3D and Supplementary Figure S8B). However, two Lys residues in *tv*Myb3, Lys^{89} and Lys^{143}, which recognize bases specifically, are Arg residues in *tv*Myb1 (Arg^{71} and Arg^{125}) and form hydrogen bonds with bases of ADE10 and ADE11′ and the phosphate backbone of CYT9′, respectively, in the complex model. In addition, the specific interactions of Asn^{69} and Gln^{72} identified in the *tv*$Myb1_{35–141}$/DNA complex are not found in the *tv*Myb3 complex structure. The corresponding residue of Asn^{69} in *tv*Myb3 is Ser^{87}, which does not recognize the base of ADE10, perhaps because of its shorter side chain. However, without the crystal structure of the *tv*$Myb1_{35–141}$/DNA complex, confirming the interactions observed in the model is not easy. The comparison of *tv*Myb1 and *tv*Myb3 complexes suggests that the strategy that combined the free form structure with CSPs, residual dipolar couplings, DNA specificity data and data-driven macromolecular docking can generate a model with reliable protein backbone architecture and some conserved protein–DNA interactions. This method is an alternative way to obtain the structural basis of protein–DNA interactions if the crystal structure is not available.

ACCESSION NUMBER

The atomic coordinates of *tv*$Myb3_{53–180}$/16-bp MRE-1 DNA complex were deposited in the Protein Data Bank under accession code 3zqc.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We are grateful for the access to the synchrotron radiation beamline 13B1 at the National Synchrotron Radiation Research Center (NSRRC) in Taiwan. The NMR spectra were obtained at the High-Field Nuclear Magnetic Resonance Center (HFNMRC) supported by the National Research Program for Genomic Medicine. We thank Dr Shu-Chuan Jao of the Biophysics Core Facility, Scientific Instrument Center at Academia Sinica for providing technical assistance of ITC experiments. We also thank Dr Chi-Fon Chang for setting up the experiments for PRE measurements and Laura Smales for copyediting the manuscript.

FUNDING

Academia Sinica and the National Science Council, Taiwan, ROC (NSC 95-2320-B-001-040-MY2 to C.C. and NSC98-2311-B-001-009-MY3 to C.-D.H.). Funding for open access charge: Academia Sinica.

Conflict of interest statement. None declared.

REFERENCES

1. Sorvillo,F., Smith,L., Kerndt,P. and Ash,L. (2001) Trichomonas vaginalis, HIV, and African-Americans. *Emerg. Infect. Dis.*, **7**, 927–932.
2. Soper,D. (2004) Trichomoniasis: under control or undercontrolled? *Am. J. Obstet. Gynecol.*, **190**, 281–290.
3. Cudmore,S.L., Delgaty,K.L., Hayward-McClelland,S.F., Petrin,D.P. and Garber,G.E. (2004) Treatment of infections caused by metronidazole-resistant *Trichomonas vaginalis*. *Clin. Microbiol. Rev.*, **17**, 783–793.
4. Dunne,R.L., Dunn,L.A., Upcroft,P., O'Donoghue,P.J. and Upcroft,J.A. (2003) Drug resistance in the sexually transmitted protozoan *Trichomonas vaginalis*. *Cell Res.*, **13**, 239–249.
5. Ong,S.J., Huang,S.C., Liu,H.W. and Tai,J.H. (2004) Involvement of multiple DNA elements in iron-inducible transcription of the *ap65-1* gene in the protozoan parasite *Trichomonas vaginalis*. *Mol. Microbiol.*, **52**, 1721–1730.
6. Tsai,C.D., Liu,H.W. and Tai,J.H. (2002) Characterization of an iron-responsive promoter in the protozoan pathogen *Trichomonas vaginalis*. *J. Biol. Chem.*, **277**, 5153–5162.
7. Alderete,J.F., O'Brien,J.L., Arroyo,R., Engbring,J.A., Musatovova,O., Lopez,O., Lauriano,C. and Nguyen,J. (1995) Cloning and molecular characterization of two genes encoding adhesion proteins involved in *Trichomonas vaginalis* cytoadherence. *Mol. Microbiol.*, **17**, 69–83.
8. Garcia,A.F., Chang,T.H., Benchimol,M., Klumpp,D.J., Lehker,M.W. and Alderete,J.F. (2003) Iron and contact with host cells induce expression of adhesins on surface of *Trichomonas vaginalis*. *Mol. Microbiol.*, **47**, 1207–1224.
9. Hsu,H.M., Ong,S.J., Lee,M.C. and Tai,J.H. (2009) Transcriptional regulation of an iron-inducible gene by differential and alternate promoter entries of multiple Myb proteins in the protozoan parasite *Trichomonas vaginalis*. *Eukaryot. Cell*, **8**, 362–372.
10. Ong,S.J., Hsu,H.M., Liu,H.W., Chu,C.H. and Tai,J.H. (2006) Multifarious transcriptional regulation of adhesion protein gene *ap65-1* by a novel Myb1 protein in the protozoan parasite *Trichomonas vaginalis*. *Eukaryot. Cell*, **5**, 391–399.

11. Ong,S.J., Hsu,H.M., Liu,H.W., Chu,C.H. and Tai,J.H. (2007) Activation of multifarious transcription of an adhesion protein *ap65-1* gene by a novel Myb2 protein in the protozoan parasite *Trichomonas vaginalis*. *J. Biol. Chem.*, **282**, 6716–6725.
12. Dubos,C., Stracke,R., Grotewold,E., Weisshaar,B., Martin,C. and Lepiniec,L. (2010) MYB transcription factors in Arabidopsis. *Trends Plant Sci.*, **15**, 573–581.
13. Ogata,K., Morikawa,S., Nakamura,H., Sekikawa,A., Inoue,T., Kanai,H., Sarai,A., Ishii,S. and Nishimura,Y. (1994) Solution structure of a specific DNA complex of the Myb DNA-binding domain with cooperative recognition helices. *Cell*, **79**, 639–648.
14. Stracke,R., Werber,M. and Weisshaar,B. (2001) The R2R3-MYB gene family in Arabidopsis thaliana. *Curr. Opin. Plant Biol.*, **4**, 447–456.
15. Chen,C.M., Wang,C.T. and Ho,C.H. (2001) A plant gene encoding a Myb-like protein that binds telomeric GGTTTAG repeats in vitro. *J. Biol. Chem.*, **276**, 16511–16519.
16. 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.
17. Ko,S., Yu,E.Y., Shin,J., Yoo,H.H., Tanaka,T., Kim,W.T., Cho,H.S., Lee,W. and Chung,I.K. (2009) Solution structure of the DNA binding domain of rice telomere binding protein RTBP1. *Biochemistry*, **48**, 827–838.
18. Ko,S., Jun,S.H., Bae,H., Byun,J.S., Han,W., Park,H., Yang,S.W., Park,S.Y., Jeon,Y.H., Cheong,C. *et al.* (2008) Structure of the DNA-binding domain of NgTRF1 reveals unique features of plant telomere-binding proteins. *Nucleic Acids Res.*, **36**, 2739–2755.
19. Sue,S.C., Hsiao,H.H., Chung,B.C.P., 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.
20. Boulikas,T. (1993) Nuclear localization signals (NLS). *Crit. Rev. Eukaryot. Gene Expr.*, **3**, 193–227.
21. Lange,A., Mills,R.E., Lange,C.J., Stewart,M., Devine,S.E. and Corbett,A.H. (2007) Classical nuclear localization signals: definition, function, and interaction with importin alpha. *J. Biol. Chem.*, **282**, 5101–5105.
22. Lou,Y.C., Wei,S.Y., Rajasekaran,M., Chou,C.C., Hsu,H.M., Tai,J.H. and Chen,C. (2009) NMR structural analysis of DNA recognition by a novel Myb1 DNA-binding domain in the protozoan parasite *Trichomonas vaginalis*. *Nucleic Acids Res.*, **37**, 2381–2394.
23. Otwinowski,Z. and Minor,W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, **276**, 307–326.
24. Vagin,A. and Teplyakov,A. (1997) MOLREP: an Automated Program for Molecular Replacement. *J. Appl. Crystallogr.*, **30**, 1022–1025.
25. McRee,D.E. (1999) XtalView/Xfit–A versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.*, **125**, 156–165.
26. 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. D Biol. Crystallogr.*, **54**, 905–921.
27. Laskowski,R.A., MacArthur,M.W., Moss,D.S. and Thornton,J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*, **26**, 283–291.
28. Lin,T.H., Chen,C., Huang,R.F., Lee,Y.L., Shaw,J.F. and Huang,T.H. (1998) Multinuclear NMR resonance assignments and the secondary structure of *Escherichia coli* thioesterase/protease I: a member of a new subclass of lipolytic enzymes. *J. Biomol. NMR*, **11**, 363–380.
29. Barbato,G., Ikura,M., Kay,L.E., Pastor,R.W. and Bax,A. (1992) Backbone dynamics of calmodulin studied by 15N relaxation using inverse detected two-dimensional NMR spectroscopy: the central helix is flexible. *Biochemistry*, **31**, 5269–5278.
30. Ottiger,M., Delaglio,F. and Bax,A. (1998) Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. *J. Magn. Reson.*, **131**, 373–378.
31. 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.
32. Johnson,B.A. and Blevins,R.A. (1994) NMR View: A computer program for the visualization and analysis of NMR data. *J. Biomol. NMR*, **4**, 603–614.
33. Iwahara,J., Anderson,D.E., Murphy,E.C. and Clore,G.M. (2003) EDTA-derivatized deoxythymidine as a tool for rapid determination of protein binding polarity to DNA by intermolecular paramagnetic relaxation enhancement. *J. Am. Chem. Soc.*, **125**, 6634–6635.
34. Iwahara,J., Tang,C. and Clore,G.M. (2007) Practical aspects of H-1 transverse paramagnetic relaxation enhancement measurements on macromolecules. *J. Magn. Reson.*, **184**, 185–195.
35. Donaldson,L.W., Skrynnikov,N.R., Choy,W.Y., Muhandiram,D.R., Sarkar,B., Forman-Kay,J.D. and Kay,L.E. (2001) Structural characterization of proteins with an attached ATCUN motif by paramagnetic relaxation enhancement NMR spectroscopy. *J. Am. Chem. Soc.*, **123**, 9843–9847.
36. Gajiwala,K.S. and Burley,S.K. (2000) Winged helix proteins. *Curr. Opin. Struct. Biol.*, **10**, 110–116.
37. Tsai,K.L., Sun,Y.J., Huang,C.Y., Yang,J.Y., Hung,M.C. and Hsiao,C.D. (2007) Crystal structure of the human FOXO3a-DBD/DNA complex suggests the effects of post-translational modification. *Nucleic Acids Res.*, **35**, 6984–6994.
38. Brent,M.M., Anand,R. and Marmorstein,R. (2008) Structural basis for DNA recognition by FoxO1 and its regulation by posttranslational modification. *Structure*, **16**, 1407–1416.
39. Tahirov,T.H., Sato,K., Ichikawa-Iwata,E., Sasaki,M., Inoue-Bungo,T., Shiina,M., Kimura,K., Takata,S., Fujikawa,A., Morii,H. *et al.* (2002) Mechanism of c-Myb-C/EBP beta cooperation from separated sites on a promoter. *Cell*, **108**, 57–70.