NMR structural analysis of DNA recognition by a novel Myb1 DNA-binding domain in the protozoan parasite *Trichomonas vaginalis*

Yuan-Chao Lou¹, Shu-Yi Wei¹, M. Rajasekaran¹,²,³, Chun-Chi Chou¹,⁴, Hong-Ming Hsu¹,⁵, Jung-Hsiang Tai¹ and Chinpan Chen¹,*

¹Institute of Biomedical Sciences, Academia Sinica, Taipei 115, ²Department of Life Science, National Tsing Hua University, Hsinchu 300, ³Chemical Biology and Molecular Biophysics, Taiwan International Graduate Program, Academia Sinica, Taipei 115, ⁴Graduate Institute of Life Sciences, National Defense Medical Center, Taipei 114, Taiwan, ROC and ⁵Department of Parasitology, College of Medicine, National Taiwan University, Taipei 106, Taiwan

Received November 4, 2008; Revised February 5, 2009; Accepted February 6, 2009

ABSTRACT

The transcription regulator, *tv*Myb1, is the first Myb family protein identified in *Trichomonas vaginalis*. Using an electrophoretic mobility shift assay, we defined the amino-acid sequence from Lys 35 to Ser 141 (*tv*Myb1₃₅₋₁₄₁) as the minimal DNA-binding domain, encompassing two Myb-like DNA-binding motifs (designated as R2 and R3 motifs) and an extension of 10 residues at the C-terminus. NMR solution structures of *tv*Myb1₃₅₋₁₄₁ show that both the R2 and R3 motifs adopt helix-turn-helix conformations while helix 6 in the R3 motif is longer than its counterpart in vertebrate Myb proteins. The extension of helix 6 was then shown to play an important role in protein stability as well as in DNA-binding activity. The structural basis for the *tv*Myb₁₃₅₋₁₄₁/DNA interaction was investigated using chemical shift perturbations, residual dipolar couplings, DNA specificity data and data-driven macromolecular docking by HADDOCK. Our data indicate that the orientation between R2 and R3 motifs dramatically changes upon binding to DNA so as to recognize the DNA major groove through a number of key contacts involving residues in helices 3 and 6. The *tv*Myb₁₃₅₋₁₄₁/DNA complex model furthers our understanding of DNA recognition by Myb proteins and this approach could be applied in determining the complex structures involving proteins with multiple domains.

INTRODUCTION

Transcription factors regulate the expression of genes at the level of transcription and control many critical biological processes. These factors typically recognize DNA sequences in the promoter regions of the target genes and regulate the frequency of transcription initiation of the genes. Transcription factors contain DNA-binding domains which bind to DNA with high sequence specificity and are classified based on the sequence similarity in the DNA-binding domain. Myb is one of the largest transcription factor families in plants (1), which contains DNA-binding domains composed of one, two or three repeated motifs of approximately 50 amino acids surrounded by three conserved tryptophan residues (2). These repeated motifs adopt a helix-turn-helix conformation to recognize the major groove of target DNA sequences. Vertebrate c-Myb protein contains three tandem repeated motifs, designated as R1, R2 and R3 motifs (3). Other Myb repeated motifs are categorized according to their similarity to the R1, R2 or R3 motifs.

Myb proteins regulate myriad gene-specific transcription in a wide range of eukaryotic systems. In vertebrates, there are three cellular Myb proteins, A-Myb, B-Myb and c-Myb (4), that are composed of 630–750 amino-acid residues and contain a highly conserved N-terminal DNA-binding domain (with ~90% identity) consisting of R1, R2 and R3 DNA-binding motifs. Vertebrate Myb proteins all recognize specific DNA stretches with a core pentanucleotide sequence, CNGTT, through three key base-contacting amino-acid residues: Lys₁²₈ in the R2 motif and Lys₁₈² and Asn₁₈³ in the R3 motif (3).
In plants such as *Arabidopsis thaliana*, the Myb protein family is expanded and contains more than 130 distinct members, most of which contain an R2R3 DNA-binding domain (125 genes) with at least 40% sequence identity (1). Although the key base-contacting amino-acid residues are also conserved, the sequence contexts of the Myb recognition elements in plants are not limited to those with a CNGTT core (5).

The transcription factor, *tv*Myb1 protein, is the first Myb family protein identified in the protozoan parasite *Trichomonas vaginalis* (6). *Trichomonas vaginalis* is responsible for the disease trichomoniasis which is one of the most common sexually transmitted infections in humans (7). The infection of *T. vaginalis* is also associated with several adverse health consequences including increased human immunodeficiency virus transmission, infertility, cervical intraepithelial neoplasia development in women, and nongonococcal urethritis and chronic prostatitis in men (8,9). With an increasing number of drug-resistant clinical *T. vaginalis* strains (10,11), the infection caused by *T. vaginalis* could become a major threat to public health. The ap65-1 gene, an iron-inducible virulence gene, encodes a 65 kDa protein that is reputed to be one of the surface adhesion proteins increasing the cytoadherence of *T. vaginalis* to the host cells which in turn induces the infection (12,13). Based on our previous studies, *tv*Myb1 protein, which contains a Myb-like R2R3 DNA-binding domain, can regulate multifarious *ap65-1* gene expression by recognizing multiple Myb recognition elements, MRE-1/MRE-2r (TAACCAGAT, MRE-1 in italic and MRE-2r underlined) and MRE-2f (TATCGGT) with a core hexanucleotide sequence, ACGATA (6,14,15).

Interestingly, the C-terminal fragment of *tv*Myb1 protein (residues 132–215) positively regulates binding of the R2R3 domain to MRE-1/MRE-2r but negatively regulates binding of the R2R3 domain to MRE-2f. However, the N-terminal fragment (residues 1–34) was shown to negatively regulate binding activity of the R2R3 domain to both MRE-1/MRE-2r and MRE-2f (J. H. Tai, unpublished data). Therefore, the availability of detailed structural information for the DNA-binding domain from *tv*Myb1 protein and its complexes with DNA target sites is essential for understanding its role and the mechanism of action in transcriptional control.

In the present study, the electrophoretic mobility shift assay (EMSA) was used to identify the DNA-binding domain of *tv*Myb1 spanning from residues 35 to 141 (referred as *tv*Myb135–141). This longer fragment showed a dramatic increase in protein stability as well as in DNA-binding ability when compared to the shorter fragment (*tv*Myb135–131), identified by Pfam database (16) and consisting of R2 and R3 motifs only. The NMR solution structure of *tv*Myb135–141 was then determined and it was found that the C-terminal 10-residue extension maintains the integrity of helix 6 at the R3 motif and thus increases both stability and DNA-binding activity of *tv*Myb135–141. The structural model of *tv*Myb135–141 in DNA-bound conformation was derived by refining the free structure with 1D NH residual dipolar couplings obtained from *tv*Myb135–141 in complex with DNA. The structures show that the relative orientation between R2 and R3 motifs is changed when bound to DNA. Finally, a data-driven structural model of the *tv*Myb135–141/DNA complex was calculated by HADDOCK (17–19) using chemical shift perturbation, residual dipolar couplings and DNA specificity data. The well-converged model reveals a number of specific contacts between the major groove of DNA and the residues at helices 3 and 6 of *tv*Myb135–141 as well as providing information about the DNA recognition by *tv*Myb1 DNA-binding domain. We also demonstrated that the inclusion of residual dipolar couplings in the data-driven macromolecular docking by HADDOCK is an efficient approach to determine the structure of complexes involving proteins with multiple domains.

**MATERIALS AND METHODS**

**Sample preparation**

The gene encoding *tv*Myb135–141 or *tv*Myb135–131 was PCR-amplified from the genomic DNA and subcloned to the pET29b (Novagen) vector. The mutants of *tv*Myb135–141 were constructed by using the kit QuikChange from Stratagene. Three mutants F38A, T67A and N126A were confirmed by DNA sequencing. The target protein with an N-terminal His-tag was expressed in *Escherichia coli* BL21(DE3) in LB broth. For labeled (15N and 15N/13C) samples, the cells were grown in M9 minimal medium containing 15NH4Cl (1 g l–1) and/or 13C-glucose (2 g l –1) at 37°C with 30 µg ml–1 kanamycin until OD 600 readings reach 0.8 and then were induced with 1 mM IPTG and grown for 4 additional hours. The cells were harvested and resuspended in buffer (20 mM Tris, pH 8.0, 500 mM NaCl) and the suspension was lysed by microfluidizer and centrifuged at 12000g for 30 min. The supernatant was applied to a nickel-nitrilotriacetic acid (Ni-NTA) affinity chromatographic column, washed with a washing buffer (20 mM Tris, pH 8.0, 500 mM NaCl and 60 mM imidazole) and eluted with an elution buffer (20 mM NaH2PO4, pH 6.0, 150 mM NaCl and 50 mM EDTA). Purity and authenticity of the recombinant proteins were verified by SDS–PAGE and mass analysis. Finally, the target protein was further dialyzed and concentrated with buffer at pH 6.0 (20 mM NaH2PO4, 50 mM NaCl, 5 mM NaN3 and 10 mM dithiothreitol) for NMR study. The single stranded DNA of MRE-1/MRE-2r (5′-AAGATAACGA TATTTA-3′) and biotinylated DNA for 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. The final concentrations of NMR samples are around 1.5 mM for free *tv*Myb135–141 and 0.7 mM for *tv*Myb135–141/DNA complex.

**EMSA**

All protein samples were incubated with γ-32P-labeled MRE-1/MRE-2r or MRE-2f. Probe labeling was performed as described previously (6). The mixture was separated in a 10% acrylamide gel by electrophoresis.
The DNA–protein complexes in reaction with probes were detected by autoradiograms. A BCA protein quantification kit was used to determine protein concentration by the supplier (Pierce).

Circular dichroism (CD)

All CD spectra were measured using an Aviv CD 202 spectrometer (Lakewood, NJ) calibrated with (+)-10-camphorsulfonic acid. All spectra were acquired at 298 K with ~20 μM protein samples in a 1 mm path-length cuvette. The signals from 195 nm to 260 nm were recorded three times with wavelength step of 0.5 nm and bandwidth of 1 nm. The average signals were converted from CD signal (millidegree) into mean residue ellipticity after subtraction of the background signals. Equilibrium thermal-denaturing experiments were obtained by measuring the change of CD signal at 222 nm from 4°C to 95°C with a 1°C interval and 3 min for equilibrium. The spectra were displayed and analyzed by SigmaPlot 8.02 (SPSS Inc.).

Surface plasmon resonance

The real-time association and dissociation of trvMyb135–141, tvMyb135–141 and three mutants with MRE-1/MRE-2r DNA duplex were measured at 25°C on a BIAcore 3000 instrument (BIAcore AB, Uppsala, Sweden). The 16-bp MRE-1/MRE-2r DNA duplex, biotinylated at the 5' end and dissolved in PBS buffer (137 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄ and 2 mM KH₂PO₄, pH 7.4) with a concentration of 0.02 μM, was applied to the streptavidin SA sensor chip (BIAcore AB, Uppsala, Sweden) at a flow rate of 10 μl min⁻¹, which resulted in the capture of 100–150 response units. Protein samples (25, 12.5, 6.3, 3.1, 1.7 or 0.8 nM) were injected to the SA sensor chip immobilized with DNA at a constant flow rate of 30 μl min⁻¹ for 2 min for association and then the running buffer (10 mM HEPES, 150 mM NaCl, 3.4 mM EDTA, pH 7.4, containing 0.005% Tween 20) was applied at the same rate for 3 min for dissociation. After each binding experiment, the sensor chip was regenerated with 1 M NaCl. For fitting the binding kinetics, the sensorgrams were evaluated by BIAevaluation version 4.1 (BIAcore AB, Uppsala, Sweden) using a 1:1 (Langmuir) binding model.

NMR experiments

All NMR spectra were carried out using Bruker AVANCE 600 or 800 MHz spectrometers equipped with a triple (¹H, ¹³C and ¹⁵N) resonance cryoprobe which included a shielded z-gradient. The triple-resonance experiments [HNCO, HN(CA)CO, CBCA(CO)NH and HNCA CB] were used for backbone resonance assignment of free tvMyb135–141 and tvMyb135–141 in complex with DNA. The weighted chemical shift perturbations for backbone ¹⁵N and ¹H residues were calculated by the equation: \( \Delta \delta = [(\Delta \delta_{1H})^2 + (\Delta \delta_{N})^2]^{0.5} \). Side-chain resonance assignment of tvMyb135–141 was achieved following similar procedures published previously (20). To measure residual dipolar couplings (RDC) of tvMyb135–141 in complex with DNA, the filamentous bacteriophage Pf1 (Asla Biotech Ltd, Latvia) was selected as the orienting medium. Pf1 (10 mg ml⁻¹) was added to the ¹⁵N-labeled protein/DNA complex sample at pH 7.0, to produce weak alignment of the complex. No significant perturbations in amide chemical shifts were observed in the presence of Pf1 phage, suggesting that the phage caused little structural change. 2D ¹H-coupled (F1) IPAP ¹H,¹⁵N HSQC spectra (21) were acquired with 256 complex ²f₁(¹⁵N) points and 64 scans per ²f₁ increment for both the isotropic and anisotropic samples. All of the NMR spectra were processed using Bruker XWINNMR or NMRPipe package (22), and analyzed using NMRView 5.0 (23) or Sparky (24).

Structure determination of trvMyb135–141

The backbone dihedral angle restraints were predicted using the program TALOS (25). The hydrogen bond restraints were introduced for residues that exhibit slow amide proton exchange rates. The nuclear Overhauser effect restraints from the ¹H-¹⁵N NOESY-HSQC and ¹H-¹³C NOESY-HSQC spectra were automatically assigned by the CANDID module of CYANA (26) and checked manually. Secondary structure was identified based on chemical shift index, amide proton exchange rate and NOEs connectivities. NMR structures were calculated from all experimental restraints by dynamical simulated annealing procedure using a modified protocol of Xplor-NIH program (27). In this protocol, the final van der Waals radii in the cooling step was increased (Sfin_rad = 0.80; the original value is 0.75) to reduce the numbers of close-contacts between heavy atoms. The final 20 structures with no distance restraint violation greater than 0.4 Å, and no dihedral angle restraint violations larger than 3° were chosen on the basis of total energy. The program PROCHECK-NMR (28) was applied for analyzing the generated structures.

Structure determination of tvMyb135–141 in the DNA-bound conformation

The structure of tvMyb135–141 in the DNA-bound conformation was calculated by refining the free structure against 74 ¹⁵N RDC constraints obtained from ¹⁵N-labeled protein in complex with DNA in the presence of Pf1 (10 mg ml⁻¹) at pH 7.0. The NOEs, dihedral angles and hydrogen bond restraints that define the helical structure of free tvMyb135–141 were used in the refinement protocol; the NOE restraints that define the distances of atoms between the R2 and R3 motifs and the restraints of loop regions were excluded. The measured ¹D NH RDC values ranged from ~25 to 29 Hz. The axial and rhombic components of the alignment tensor were determined by the grid search method proposed by Clore and coworkers (29) to be 15.4 Hz and 7 Hz, respectively. A set of 20 structures with no distance restraint violations greater than 0.3 Å, and no dihedral angle restraint violations larger than 5° were selected based on the energy of RDC constraints.
HADDOCK docking

The model of the \( vtvMyb1_{135-141}/DNA \) complex was calculated using the information-driven method called HADDOCK v2.0 (17–19). Chemical shift perturbations (CSP) and DNA specificity data were translated into ambiguous interaction restraints (AIRs) to drive the docking process. The AIRs can also be combined with RDC data to allow a better definition of the relative orientations of the components (18,30). The starting structures for the docking were a B form model of the 16-bp MRE-1/MRE-2r DNA duplex built by the Discovery studio 2.0 (Accelrys) and the 20 structures of \( vtvMyb1_{135-141} \) in the DNA-bound conformation. For \( vtvMyb1_{135-141} \) protein, residues having a weighted chemical shift perturbation upon complex formation greater than 0.5 ppm and displaying high solvent accessibility (>50%) were selected as active residues. Solvent accessibility for the active residues was calculated using the program NACCESS (31). The selected active residues are Val100, Phe108, Thr109, Asn110, Gln112, Glu113, Asn116 and Asn117 and the semi-flexible regions were defined for residues 34–41, 67–77, 108–112 and 120–128. According to our previous results on MRE-1/MRE-2r DNA specificity, several base replacements disrupt the interactions between \( vtvMyb1 \) protein and DNA, which include Ade6 to CyT, CyT8 to ThY, GUA9 to Ade, Ade10 to CyT and ThY11 to Gua (6). These nucleotides and their complementary bases were selected as active bases and the semi-flexible regions were defined for bases 6–11 and 22–27. The specific AIR restraints were defined between suitable atoms of active residues to unique base atoms of active bases. For example, the base replacement, CyT8 to ThY, loses H41 of CyT8 and O6 of GUA25 for the specific hydrogen bond interaction. The AIR restraints were thus defined between H41 of CyT8 to carbonyl O of backbone, O4 of Thr, or to carbonyl O of side-chains of Asn, Gln and Glu and between O6 of GUA25 to amide protons of backbone or to amide protons of side-chains of Asn and Gln. This rule was applied to define all the specific AIR restraints and a total of 31 AIRs with 2 A\(^{\circ} \) were introduced as intervector projection angle restraints, VEAN energy terms (32), during stages (i) rigid body docking and (ii) semi-flexible simulated annealing. The floating alignment tensor, SANI energy term (33), was used in the stage of final water refinement. During the rigid body energy minimization, 10,000 structures were calculated and the 200 best solutions based on the intermolecular energy were used for the semi-flexible, simulated annealing followed by an explicit water refinement. Docked structures corresponding to the 200 best solutions with lowest intermolecular energies were generated. The 200 solutions were clustered using a 1.0 A\(^{\circ} \) RMSD cut-off criterion. The clusters were ranked based on the averaged HADDOCK score of their top 10 structures.

Data bank accession number

The chemical shifts of \( vtvMyb1_{135-141} \) at pH 6.0 and 298K have been deposited in the BioMagResBank under accession number BMRB-15989. The best 20 structures of \( vtvMyb1_{135-141} \) and the best 10 structures of \( vtvMyb1_{135-141}/DNA \) complex have been deposited in the RCSB Protein Data Bank under accession number 2k9n and 2kdz, respectively.

RESULTS

Identification of the DNA-binding domain of \( vtvMyb1 \) for structural investigation

The transcription factor, \( vtvMyb1 \), has been found to regulate the multifarious transcription of the \( ap65-1 \) gene by binding to two Myb recognition elements, MRE-1/MRE-2r and MRE-2f, with a core hexanucleotide sequence, ACAGATA (6). Domain analysis in the Pfam database (16) indicated that the segments from Lys164 to Ile167 and Thr187 to Ile191 of \( vtvMyb1 \) are classified as Myb-like DNA-binding motifs. The fragment, \( vtvMyb1_{135-141} \) was hence cloned but it was soon found to be highly susceptible to degradation and precipitation. Previous reports showed that several plant Myb proteins contain an additional C-terminal extension of the DNA-binding domain (34–36). Therefore, another fragment, \( vtvMyb1_{135-141} \) was constructed to test for the necessary DNA-binding ability and structural stability. The binding ability of these two fragments to the Myb recognition elements, MRE-1/MRE-2r and MRE-2f, were examined by EMSA. The shorter fragment, \( vtvMyb1_{135-131} \), showed little interaction with either element. In contrast, the longer fragment, that contained 10 additional C-terminal residues, exhibited the desired DNA-binding activity (Figure 1A).

To quantitate the DNA-binding affinity of the two fragments, we performed surface plasmon resonance (SPR) experiments on a BIACore 3000 biosensor system. The 16-bp MRE-1/MRE-2r DNA duplex (nucleotide sequence: AAGATAACGATATTTA) was immobilized onto the streptavidin SA sensor chip and probed with different concentrations of proteins (from 0.78 nM to 25 nM). The sensorgrams of \( vtvMyb1_{135-141} \) showed a concentration-dependent binding (Figure 1B) and the traces were analyzed with a 1:1 Langmuir binding (with mass transfer) model by BIATool software, giving an equilibrium dissociation constant (\( K_D \)) of 1.24 × 10\(^{-9} \) M for \( vtvMyb1_{135-141} \) interacting with immobilized MRE-1/MRE-2r, indicating a high-affinity interaction that typically observed for a sequence-specific DNA-binding protein. Similar SPR experiments for the shorter fragment \( vtvMyb1_{135-131} \) did not give reliable results possibly due to its high propensity to aggregate. Hence the relative DNA-binding ability of \( vtvMyb1_{135-131} \) was calculated from SPR responses, which indicated that the DNA-binding ability of \( vtvMyb1_{135-131} \) is lower than 10% of that of \( vtvMyb1_{135-141} \) (Figure 1C), in good agreement with the results from EMSA assays. The DNA bindings of three mutants of \( vtvMyb1_{135-141} \) were also checked by EMSA.
and SPR experiments and are described later in this section.

The secondary structures and structural stabilities of tvMyb135–131 and tvMyb135–141 were monitored by circular dichroism spectra. The wavelength scans of these two fragments showed similar absorptions at 222 and 208 nm, indicating that they exhibit similar helical structures (Figure 2A). However, temperature denaturation experiments indicated that tvMyb135–141 exhibits higher thermal stability than tvMyb135–131 (Figure 2B). The structures of the two fragments in solution were further checked by 2D 1H-15N heteronuclear single quantum coherence (HSQC) spectra. Both spectra showed well-dispersed cross peaks that overlap significantly with each other, indicating that each fragment is well structured in solution (Supplementary Figure S1). In addition, the cross peaks of the additional C-terminal 10 residues of tvMyb135–141 are also well dispersed, suggesting a stable conformation rather than an unstructured state. Based on all these results, the longer fragment (tvMyb135–141) was identified as the DNA-binding domain of tvMyb1 protein and was selected for further NMR structural investigation.

**NMR resonance assignment and structural determination of tvMyb135–141**

NMR resonance assignment of tvMyb135–141 was achieved following the procedures published previously (20) and is described briefly in ‘Materials and Methods’ section. All backbone resonances were clearly identified except those of Ile81 and Met129 (Supplementary Figure S2) and around 90% of the side-chain resonances were unambiguously assigned. From the consensus chemical shift indices and nuclear Overhauser effect (NOE) patterns, the secondary structure topology of tvMyb135–141 (Figure 2B). The DNA-binding domain tvMyb135–141 contains six helices, designated as: helix 1 (H1), Glu40 to Tyr53; helix 2 (H2), Trp58 to Leu64; helix 3 (H3), Pro92 to Glu104; helix 5 (H5), Trp109 to Leu116 and helix 6 (H6), Asp121 to Arg135. The first three helices of tvMyb135–141 resemble the R2 motif of c-Myb and the last three helices correspond to the R3 motif of c-Myb. The R2 motif is connected to the R3 motif by a 11-residue loop (L1). However, it is worth noting that H6 of tvMyb135–141 is longer than the corresponding helix in c-Myb. In the preceding results,
we found that the C-terminal 10-residue extension (Ala132-Ser141) increases the DNA-binding ability as well as the protein stability significantly. Accordingly, it seems that the C-terminal extension maintains the integrity of H6 and thus increases both stability and DNA-binding activity.

The tertiary structure of tvMyb135–141 was subsequently determined based on a set of 1394 NOE distance restraints, 28 hydrogen bond restraints and 144 backbone φ and ψ dihedral angle restraints. An ensemble of 20 structures with no distance restraint violations greater than 0.4 Å and no dihedral angle restraint violations greater than 3° was selected based on the total energy (Figure 3A and B). The structural statistics are listed in Table 1. The tertiary structures of the R2 and R3 motifs are in particular well defined, with root-mean-square deviations (RMSD) of 0.32 (±0.11) Å and 0.52 (±0.19) Å for the backbone atoms of helical residues in the R2 and R3 motifs, respectively. The helix-turn-helix structure of the R2 and R3 motifs are highly stabilized by the hydrophobic cores composed of Phe38, Leu46, Leu49, Val50, Tyr53, Trp58, Ile61 and Trp77 for R2 and Trp90, Leu98, Tyr102, Trp109, Ile112, Leu116, Ile124, Trp128, Ile131 and Ala132 for R3. It was also found that H6 of the R3 motif is amphipathic and its hydrophobic residues Ile124, Trp128, Ile131 and Ala132 constitute the major part of the hydrophobic core of the R3 motif, suggesting that H6 is important for the stability of the R3 motif.

Although the R2 and R3 motifs are well defined, their relative orientation was not fixed in the beginning of
structural calculation because the loop (L1) between H3 and H4 was not well defined due to fewer NOE restraints observed in this region. After the NOE assignment was finished, we found several NOE cross peaks between the R2 and R3 motifs as listed in Supplementary Table S1. The RMSD values of the 20 final NMR structures were 0.70 (±0.15) Å and 1.56 (±0.14) Å for the backbone atoms and the heavy atoms of all helical residues, respectively. Even without the addition of backbone atoms and the heavy atoms of all helical residues and negatively charged residues, four possible salt bridges of Glu\(^{40}\)(R2)-Arg\(^{125}\)(R3), Glu\(^{75}\)(R2)-Lys\(^{114}\)(R3), Arg\(^{76}\)(R2)-Asp\(^{121}\)(R3) and Asp\(^{88}\)(L1)-Arg\(^{117}\)(R3) were observed to stabilize the relative orientation between the R2 and R3 motifs as shown in Figure 3C. From our NMR solution structure, it was observed that tvMyb\(^{135-141}\) is highly stabilized by hydrophobic and electrostatic interactions, which contribute to its high stability toward thermal denaturation.

Mapping the tvMyb\(^{135-141}/DNA interaction by NMR spectroscopy

To map the DNA-binding sites of tvMyb\(^{135-141}\) with the 16-mer MRE-1/MRE-2r DNA duplex, the \(^{15}\)N-labeled protein sample was titrated with the DNA duplex in a 1:1 ratio and 2D \(^1\)H-\(^{15}\)N HSQC spectra were acquired to monitor the chemical shift changes of backbone amide resonances. As shown in Figure 4A, the cross peaks in the 2D \(^1\)H-\(^{15}\)N HSQC of tvMyb\(^{135-141}/DNA complex in general are broader than the peaks of free tvMyb\(^{135-141}\) and the backbone amide resonances of Asp\(^99\), Gln\(^{100}\) and Trp\(^{128}\) to Ala\(^{132}\) could not be assigned due to peak broadening or peak absence, suggesting that these residues may experience intermediate exchange. A variety of experimental conditions have been tested, including different pH values, temperatures and DNA lengths, but the resulting NMR spectra did not show substantial improvements. Based on the best assignment of tvMyb\(^{135-141}\) in complex with DNA we can obtain, it was found that the backbone resonances undergo significant perturbations in complex with DNA (Figure 4B). The majority of residues that exhibit large changes in chemical shift (∆δ > ∆δ\(_{\text{average}}\) + SD ~0.5 ppm) are located at the N-terminus, H3, L1 loop and H6. We mapped these residues onto the free tvMyb\(^{135-141}\) structure (Figure 5C) and found that these residues lie mainly in the middle part of the structure and constitute several discontinuous faces, which are unlikely to be the DNA-binding surfaces.

### Table 1. Structural statistics of tvMyb\(^{135-141}\) and tvMyb\(^{135-141}/DNA complex

| Parameter                      | tvMyb\(^{135-141}\) | DNA-bound conformation | tvMyb\(^{135-141}/DNA complex |
|--------------------------------|----------------------|------------------------|--------------------------------|
| NMR restraints used            |                      |                        |                                |
| Intraresidue \((i - j = 0)\)   | 289                  | 196                    |                                |
| Sequential \((i - j = 1)\)     | 551                  | 403                    |                                |
| Medium range \((1 < i - j < 5)\)| 4500                 | 405                    |                                |
| Long range \((i - j ≥ 5)\)    | 104                  | 83                     |                                |
| Total NOE restraints           | 1394                 | 1087                   |                                |
| Total AIRs                     |                      |                        | 20                             |
| Hydrogen bonds                 | 28 x 2               | 28 x 2                 | 120                            |
| Dihedral angles                | 144                  |                        |                                |
| \(^{1}\)DHN RDCs               |                      |                        |                                |
| Energy statistics              |                      |                        |                                |
| \(E_{\text{total}}\)          | 335.1 ± 11.5         | 221.1 ± 9.5            | −4462.1 ± 116.5                |
| \(E_{\text{bond}}\)           | 29.3 ± 1.6           | 10.9 ± 0.9             | 30.1 ± 3.2                     |
| \(E_{\text{angle}}\)          | 114.6 ± 3.4          | 74.6 ± 2.2             | 81.3 ± 39.1                    |
| \(E_{\text{improper}}\)       | 16.4 ± 0.5           | 20.8 ± 1.2             | 67.4 ± 21.9                    |
| \(E_{\text{vdW}}\)            | 100.9 ± 2.8          | 73.2 ± 3.1             | −629.9 ± 38.3                  |
| \(E_{\text{NOE}}\)            | 70.6 ± 2.1           | 38.2 ± 1.8             | 20.8 ± 1.2                     |
| \(E_{\text{disub}}\)          | 3.4 ± 0.3            | 3.9 ± 0.4              |                                |
| \(E_{\text{disub}}\)          | 10.1 ± 1.4           |                        |                                |
| RDC Q-factor\(^a\)            | 0.67 ± 0.04          | 0.029 ± 0.003          | −4862.4 ± 63.2                 |
| Deviation from ideal geometry (RMSD) |                |                        | 0.102 ± 0.007                 |
| Bond (Å)                       | 0.0041 ± 0.0001      | 0.0025 ± 0.0001        | 0.0032 ± 0.0002                |
| Angle (°)                      | 0.48 ± 0.01          | 0.38 ± 0.007           | 0.67 ± 0.03                    |
| Mean global RMSD of helical residues (Å) | |                        |                                |
| Backbone atoms                 | 0.70 ± 0.15          | 0.96 ± 0.25            | 0.43 ± 0.09                    |
| Heavy atoms                    | 1.56 ± 0.14          | 1.89 ± 0.22            | 0.88 ± 0.16                    |
| Ramachandran plot              |                      |                        |                                |
| Most favored (%)               | 78.9                 | 80.9                   | 77.0                           |
| Allowed (%)                    | 18.4                 | 17.6                   | 17.5                           |
| Generously allowed (%)         | 2.5                  | 1.5                    | 4.4                            |
| Disallowed (%)                 | 0.2                  | 0.0                    | 1.1                            |

\(^a\)Q-factor = RMS(D\(^{\text{calc}}\) - D\(^{\text{obs}}\))/RMS(D\(^{\text{obs}}\)), where D\(^{\text{calc}}\) and D\(^{\text{obs}}\) are calculated and observed RDC values, respectively.

\(^b\)For backbone atoms of all helical residues and all phosphate backbone atoms of DNA.
Besides, several residues that exhibited large chemical shift changes are located in the interface between the R2 and R3 motifs (such as Asn110 and Asn122). These data suggest that the DNA-bound conformation is different from the free structure. We then checked the backbone rigidity of tvMyb135–141 in the presence and absence of DNA by 1H-15N hetero-nuclear NOE (HX-NOE) measurements (Figure 4C). The data showed that the N-terminal residues, Val10–Phe38, are highly flexible in the absence of DNA and adopt a much more rigid conformation upon binding to DNA. The rigidity of residues at H3, L1 loop and the C-terminus are also increased, although not as dramatically as are the N-terminal residues. In addition, almost all the residues in the DNA-bound conformation demonstrate stronger HX-NOE values and the mean values of HX-NOE were increased from 0.67 (±0.14) to 0.77 (±0.12) when the protein binds to DNA, indicating that DNA binding highly stabilized the tvMyb135–141 protein and suggesting the occurrence of certain conformational changes.

Structure of tvMyb135–141 in DNA-bound conformation

To gain more structural information of the tvMyb135–141 in complex with MRE-1/MRE-2r, we measured the 1DNH residual dipolar coupling (RDC) from the 15N-labeled tvMyb135–141/DNA complex sample partially aligned in medium containing 10 mg ml–1 pf1 phage. 74 1DNH RDC values ranging from –25 to 29 Hz were measured unambiguously. To see if the structure of tvMyb135–141 was changed when bound to DNA, the correlation between the measured 1DNH RDC values and the back-calculated values derived from the ensemble of 20 structures of tvMyb135–141 were analyzed by the program Pales and the goodness of fit was assessed with the RDC Q-factor (37). We first analyzed the residues located in the helices of the R2 motif (23 1DNH RDC values) and found that most of the back-calculated RDC values derived from the 20 structures of R2 motif fitted well to the experimental RDC values with an average Q-factor = 0.30 ± 0.03 (0.259–0.348), suggesting that the backbone conformation of the R2 motif is similar in the presence and absence of DNA. Similarly, the fitting focused on the helical residues of the R3 motif (18 1DNH RDC values) also gave a low average Q-factor (0.31 ± 0.04; 0.252–0.399), indicating that the fold of the R3 motif is also maintained upon DNA binding. However, unlike the fitting to individual R2 or R3 motifs, the fitting of the measured RDC values of all helical residues to the calculated RDC values derived from 20 structures gave a high average Q-factor (0.67 ± 0.04; 0.611–0.752). These results suggest that when tvMyb135–141 binds to DNA the backbone conformations of individual R2 and R3 motifs are mostly preserved but the relative orientation between the two motifs is dramatically changed.

The structure of DNA-bound tvMyb135–141 was generated by refining the free structure with all 1DNH RDC constraints obtained from protein sample in complex with DNA. The NOE, dihedral angle and hydrogen bond restraints of helical residues of free tvMyb135–141 were also used in the structural refinement protocol since these restraints define the individual folds of the R2 and R3 motifs, which agree with the RDC values from the DNA-bound conformation. A set of 20 structures was selected based on the energy of RDC restraints (Figure 5A) with RMSD values of 0.96 ± 0.25 Å and
1.89 ± 0.22 Å for the backbone atoms and heavy atoms of all helical residues, respectively (Table 1). The RDC values calculated from the final 20 structures correlated with observed RDC values very well with an average RMSD value = 0.37 ± 0.02 Hz. Figure 5B shows the structure with the lowest RDC energy. The angles between H1, H2 and H3 of the R2 motif in DNA-bound conformation are similar to those in the free structure as are the angles between H4, H5 and H6 in the R3 motif (Supplementary Table S2). However, the angles between the R2 helices and the R3 helices are dramatically different between the free structure and the DNA-bound conformation. Figure 5D shows that the difference between the two R3 motifs is about 50 degrees of rotation if the backbone atoms of the R2 motifs are superimposed. After the rotation, H3 and H6 form a V-shaped surface and the residues that exhibited significant chemical shift changes upon DNA binding are shown in stick and are colored in red for those selected as active residues in HADDOCK docking. The residues that exhibit significant chemical shift changes are also shown. (D) Structural superimposition of tvMyb135-141 R2 motifs in free form (light green) and DNA-bound form (light blue) revealed that the orientation of R3 motif in DNA-bound conformation rotated 50° relative to that in free form.

**Figure 5.** The tvMyb135-141 solution structure in free form and in DNA-bound conformation. (A) The ensemble of 20 final structures of tvMyb135-141 in DNA-bound conformation. The structures were derived from refining the free structure with 1DNH RDC constraints obtained from the tvMyb135-141/DNA complex sample. The structures with lowest RDC energies were selected. (B) Cartoon representation of tvMyb135-141 in DNA-bound conformation. The residues that exhibit significant chemical shift changes upon DNA binding are shown in stick and are colored in red for those selected as active residues in HADDOCK docking. (C) Cartoon representation of tvMyb135-141 in free form. The residues that exhibit significant chemical shift changes are also shown. (D) Structural superimposition of tvMyb135-141 R2 motifs in free form (light green) and DNA-bound form (light blue) revealed that the orientation of R3 motif in DNA-bound conformation rotated 50° relative to that in free form.

The structural model of tvMyb135-141/DNA complex

Although most of the backbone resonance assignment of tvMyb135-141 in complex with DNA was completed, no intermolecular NOEs could be unambiguously assigned from either a 3D [F1] 13C, 15N-filtered, [F2, F3] 15N-edited NOESY-HSQC spectrum or a 3D 15N-edited NOESY-HSQC spectrum obtained from a 95% 2H, 13C, 15N-labeled tvMyb135-141/DNA complex sample. Therefore, to gain insight into the likely DNA-bound conformation of tvMyb135-141, we used the chemical shift perturbations, 1DNH RDC values and DNA specificity data to calculate a structural model of the protein/DNA complex using the program HADDOCK (17–19). The active residues of tvMyb135-141 were defined for those that have weighted 1H and 15N chemical shift perturbations upon complex formation greater than 0.5 ppm (\(\Delta \delta > \Delta \delta_{\text{average}} + \text{SD} \approx 0.5\) ppm) and display high solvent accessibility (>50%). According to our previous results on the MRE-1/MRE-2r DNA specificity (6), the base replacements that disrupt the interactions between tvMyb1 protein and the DNA were selected as active bases, which include ADE6 and CYT8 to THY11 and their complementary bases (the strand, AAGATAACG ATATTTA, is numbered 1–16 and the complementary strand, TAAATATCGTTATCTT is numbered 17–32). The specific AIR restraints (provided as Supplementary Data) were defined between suitable atoms of active residues to unique base atoms of active bases as described in ‘Materials and Methods’ section.

After the semi-flexible simulated annealing and explicit water refinement protocol, the final 200 structures were clustered based on the pair-wise RMSD matrix using a 1.0 Å cutoff and resulted in 11 different clusters. Table 2 shows the statistics of the top seven clusters based on the averaged HADDOCK score of their top 10 structures. Cluster 4 has the best average HADDOCK score as well as the most favorable intermolecular energies of the van der Waals’ and the electrostatic interactions.
The 10 lowest energy structures from this cluster were selected to represent the model of the tvMyb135–141/DNA complex (Figure 6A). The structural ensemble has an RMSD value of 0.43 ± 0.09 Å over all backbone atoms of the secondary structural residues of the protein and an RMSD value of 0.68 ± 0.25 Å if all DNA phosphate backbone atoms are also included. The DNA-recognition surface of tvMyb135–141 comprises mainly residues of H3.

Table 2. Statistics of the top seven clusters obtained with HADDOCK

| Cluster | Haddock score | RMSD-\(E_{\text{vdw}}^{\text{c}}\) | \(N^{d}\) | \(E_{\text{vdw}}^{\text{c}}\) | \(E_{\text{elec}}^{\text{c}}\) | \(E_{\text{AIR}}^{f}\) | \(E_{\text{sani}}^{g}\) | BSA\(^{h}\) | \(E_{\text{desolv}}^{i}\) |
|---------|---------------|-------------------------------|------|----------------|----------------|----------------|----------------|----------|----------------|
| 4       | 700 ± 33      | 0.55 ± 0.2                    | 19   | −64 ± 4        | −612 ± 31      | 1313 ± 21      | 23 ± 3         | 2120 ± 74 | 40 ± 3        |
| 3       | 853 ± 35      | 3.97 ± 0.1                    | 23   | −50 ± 4        | −406 ± 24      | 1259 ± 22      | 27 ± 12        | 1762 ± 57 | 25 ± 7        |
| 7       | 919 ± 92      | 2.33 ± 0.1                    | 8    | −67 ± 5        | −491 ± 44      | 1398 ± 39      | 39 ± 40        | 2190 ± 64 | 40 ± 4        |
| 10      | 967 ± 102     | 4.26 ± 0.1                    | 4    | −52 ± 7        | −429 ± 28      | 1391 ± 71      | 29 ± 11        | 1839 ± 43 | 29 ± 4        |
| 2       | 993 ± 19      | 5.26 ± 0.1                    | 36   | −73 ± 4        | −392 ± 18      | 1411 ± 17      | 24 ± 2         | 2151 ± 40 | 24 ± 4        |
| 5       | 1006 ± 31     | 17.42 ± 0.1                   | 15   | −71 ± 4        | −402 ± 19      | 1430 ± 38      | 22 ± 6         | 2237 ± 72 | 26 ± 6        |
| 8       | 1015 ± 90     | 17.35 ± 0.1                   | 10   | −58 ± 5        | −415 ± 40      | 1436 ± 71      | 28 ± 26        | 2195 ± 57 | 23 ± 3        |

\(^a\)The final 200 structures were clustered based on the pair-wise RMSD matrix using a 1.0 Å cutoff. The statistics are for the 10 lowest energy structures.

\(^b\)The HADDOCK score was calculated as the sum of: \(E_{\text{vdw}} + E_{\text{elec}} + E_{\text{AIR}} + E_{\text{sani}} + E_{\text{desolv}}\).

\(^c\)Overall backbone RMSD from the lowest energy structure.

\(^d\)Number of structures in a given cluster.

\(^e\)Intermolecular energies (kcal mol\(^{-1}\)) were calculated with the OPLS parameters using a 8.5 Å cut-off.

\(^f\)HADDOCK ambiguous interaction restraint energy (kcal mol\(^{-1}\)).

\(^g\)Energy for the direct RDC constraints.

\(^h\)Buried surface area (Å\(^2\)).

\(^i\)The desolvation energy (kcal mol\(^{-1}\)).

Figure 6. Structural model of the tvMyb135–141/DNA complex. (A) Backbone illustration of the 10 lowest energy structures from HADDOCK. (B) Surface representation of the DNA and ribbon display of tvMyb135–141 clearly showed that H3 and H6 insert into the major groove of DNA and the N-terminus contacts with the DNA minor groove. (C) Stereo view of the specific hydrogen bonds in tvMyb135–141/DNA complex. The residues involved in specific hydrogen bond interactions are shown in stick and the atoms of DNA are shown as sphere (with C in green, N blue, O red and H white). The hydrogen bonds are indicated as black dash lines.
and H6 that insert into the major groove of the hexanucleotide AACGAT and some residues at the N-terminus that contact the DNA minor groove (Figure 6B). A lot of hydrogen bond and hydrophobic interactions between protein and DNA molecules are observed in >40% of the final docking structures (Table 3). The side-chains of Asn⁶⁹, Arg⁷¹, Gln⁷², Glu⁷⁵, Asn¹²² and Asn¹²⁶ play important roles in DNA specific recognition, in which the side-chains of these residues form hydrogen bonds with bases of the DNA (Figure 6C). A number of hydrogen bonds are also found mainly between the positively charged side-chains and the backbone phosphates of the DNA. In addition, two hydrogen bonds are observed between the protein backbone amide protons and the DNA backbone phosphates (Asn⁶⁹-GUA⁹ and Asn¹¹⁰-THY²³). The observation of plenty of hydrogen bonds and hydrophobic contacts between the protein and the DNA molecules agrees well with the high affinity found between tvMyb1₃₅-₁₄₁ and the MRE-1/MRE-2r DNA duplex (K D = 1.24 x 10⁻⁹ M).

The conformations of the MRE-1/MRE-2r DNA duplex in the final complex structures were analyzed by the program 3DNA (38). The average helical parameters for DNA bound to tvMyb1₃₅-₁₄₁ are shown in Supplementary Figure S3. Basically, the overall DNA structure remains canonical double-stranded base-pairing geometry. The major groove width between ADE₆-ADE₇ is at maximun, the roll angles between ADE₇ to ADE₁₂ and the twist angles between CYT₈ to ADE₁₀ are deviated from standard B-form DNA, indicating the interaction between tvMyb1₃₅-₁₄₁ and the DNA.

To validate our model, three point mutations of tvMyb1₃₅-₁₄₁ were generated (F₃₈A, T₆₇A and N₁₂₆A). The CD spectra showed that they are all as well-folded as tvMyb1₃₅-₁₄₁ (data not shown). Their binding affinities toward the MRE-1/MRE-2r DNA duplex were checked by EMSA and SPR experiments (Figure 1A and C). The mutant F₃₈A exhibits the lowest DNA-binding ability and N₁₂₆A shows a significant decrease in DNA binding, agreeing with our complex structure in which Phe³₈ and Asn₁₂₆ forms contacts with the DNA. The mutant T₆₇A binds DNA as strongly as tvMyb1₃₅-₁₄₁ does and Thr⁶₇ shows no contact with DNA in the complex structure. These data highly support the accuracy of the tvMyb1₃₅-₁₄₁/DNA complex structure.

### DISCUSSION

In the present study, we have defined the DNA-binding domain of the first Myb family protein identified in Table 3. Intermolecular hydrogen bonds between tvMyb1₃₅-₁₄₁ and DNA

| Protein | Atom | Base | Atom | Distance (Å) | No. (of 10) |
|---------|------|------|------|-------------|-------------|
| Hydrogen bonds | | | | | |
| Lys³⁵ | HZ⁺ | ADE₂⁸ | O₁P | 1.92 | 4 |
| Lys³⁷ | HZ⁺ | GUA⁹ | O₁P | 1.88 | 7 |
| Arg⁶⁸ | HE | CYT₈ | O₂P | 1.96 | 10 |
| Arg⁶⁸ | HH⁺ | CYT₈ | O₂P | 1.89 | 5 |
| Asn⁶⁹ | HD² | ADE₁₀ | N₇ | 2.13 | 10 |
| Asn⁶⁹ | HN | GUA⁹ | O₂P | 1.92 | 10 |
| Asn⁷¹ | HH⁺ | ADE₁₀ | N₆ | 2.18 | 5 |
| Asn⁷¹ | NH⁺ | ADE₁₀ | H₆¹ | 2.13 | 4 |
| Gln⁷² | HE²⁺ | GUΑ⁹ | O₆ | 2.21 | 10 |
| Gln⁷² | HE²⁺ | THY₂³ | O₄ | 1.90 | 5 |
| Arg⁷₄ | HH⁺ | ADE₂₂ | O₁P | 2.08 | 10 |
| Glu⁷⁵ | OE⁻ | CYT₂⁴ | H₄⁺ | 2.26 | 8 |
| Glu⁷⁵ | OE⁻ | CYT₂⁴ | H₄¹ | 2.0¹ | 8 |
| Arg⁷₆ | HH⁺ | ADE₇ | O₁P | 2.02 | 10 |
| Asn¹¹⁰ | HD₂¹ | ADE₂₂ | O₃⁻ | 1.91 | 9 |
| Asn¹¹⁰ | NH | THY₂³ | O₂P | 2.21 | 4 |
| Asn¹²² | OD₁ | CYT₂⁴ | H₄¹ | 2.11 | 9 |
| Arg¹₂⁵ | HH⁺ | CYT₂⁴ | O₁P | 2.05 | 7 |
| Asn¹₂⁶ | HD₂² | ADE₆ | N₇ | 2.0³ | 10 |
| Arg¹₂⁷ | HH⁺ | THY₅ | O₂P | 1.9⁴ | 5 |
| Arg¹₂⁷ | HH⁺ | ADE₆ | O₁P | 1.8⁶ | 8 |
| Arg¹₃₃ | HH⁺ | ADE₄ | O₂P | 1.8⁵ | 9 |

| Hydrophobic contacts | | | | |
| Side chains | | | | |
| Val₃⁶ | ADE₂⁸ | CYT₈ | 3.6⁰ | 10 |
| Phe₃⁸ | ADE₇, CYT₈ | 3.₅⁵ | 10 |
| Arg⁷¹ | THY₂¹, ADE₂² | 3.₄⁴ | 10 |
| Pro¹₀⁷ | CYT₂⁴ | 3.₄⁹ | 10 |
| Trp¹⁰⁹ | CYT₂⁴ | 3.₅² | 10 |
| Asp¹₂¹ | THY₂³ | 3.₆² | 10 |
| Asn¹₂⁶ | ADE₆ | 3.₁⁴ | 10 |
| Met¹₃₀ | THY₅ | 3.₅₈ | 10 |

The specific interactions are in bold.

*Indicates all possible atoms at the position.
the protozoan parasite *T. vaginalis* and determined its NMR solution structure with high resolution. In order to understand the rearrangement of the tertiary fold of the DNA-binding domain for DNA recognition, the $^1$D$_{NH}$ residual dipolar couplings of *tvMyb* in complex with DNA were acquired and used for structural refinement. Finally, based on the chemical shift perturbation, residual dipolar couplings and DNA specificity data, the molecular basis for the DNA recognition of this domain was suggested.

To define the DNA-binding domain of *tvMyb* protein, two clones were constructed and tested for DNA recognition. The shorter fragment, *tvMyb*$_{135-141}$, which contains two classical Myb-like DNA-binding domain (one from Lys$^3$ to Ile$^{81}$ and the other Thr$^{87}$ to Ile$^{131}$) as suggested by the Pfam database, was found to be prone to degradation and aggregation. Its thermal stability and DNA-binding ability was also much lower than those of the longer fragment, *tvMyb*$_{135-141}$. It then becomes important to discuss the roles of the C-terminal 10 residues A$^{132}$RHRAKHQS$^{141}$ of the protein. The last helix of *tvMyb*$_{135-141}$, H6, spanning from Asp$^{121}$ to Arg$^{135}$, was found to be amphipathic. Its hydrophobic residues Ile$^{124}$, Trp$^{128}$, Ile$^{131}$ and Ala$^{132}$ constitute the major part of the hydrophobic core of the R3 motif and thus the integrity of H6 is important for the hydrophobic packing of the R3 motif. It is well known that an α-helix has an overall dipole moment from C-terminal to N-terminus caused by the cumulative effect of each residue backbone unit dipole contribution. As a result, α-helices are often capped at the N-terminal end by a negatively charged amino acid or at the C-terminus with a positively charged amino acid in order to neutralize this helix dipole (39,40). In *tvMyb*$_{135-141}$, H6 is capped and stabilized by Asp$^{121}$ and Arg$^{135}$ at the two termini. However, in the shorter fragment, H6 is terminated at Ile$^{131}$; the C-terminal capping by Arg$^{135}$ is lost, which may disrupt the integrity of H6 and destabilizes the hydrophobic core of the R3 motif. This explains why *tvMyb*$_{135-141}$ tends to degrade and to form amorphous aggregates. For DNA-binding ability, the C-terminal 10-residues extension introduces four additional positively charged residues to the protein molecule, which may increase the electrostatic interaction between the highly positively charged protein molecule and the highly negatively charged DNA backbone. In the $^1$H,15$N$ HX-NOE experiments, the C-terminal residues in the complex state showed higher HX-NOE values than those in the free form, which is likely due to electrostatic interactions between C-terminal residues and DNA. In our complex structure, Arg$^{133}$ forms a hydrogen bond with the backbone phosphate of ADE4. In addition, SPR experiments also indicated that a higher salt concentration decreases the binding response between *tvMyb*$_{135-141}$ and DNA (data not shown). Taking all these together, the C-terminal 10-residues extension can stabilize the R3 motif by hydrophobic residues and enhance the DNA binding through positively charged residues.

The structure of *tvMyb*$_{135-141}$ in the DNA-bound form was acquired by refining the free structure with the complex $^1$D$_{NH}$ RDC constraints and the *tvMyb*$_{135-141}$/DNA complex structural model was built based on the chemical shift perturbation, residual dipolar couplings and DNA specificity data by HADDOCK (17–19). The results showed that the relative orientation between the R2 and R3 motifs are changed upon binding to the DNA major groove. The large changes in amide chemical shifts of the residues in L1 loop suggest significant conformational rearrangements of this loop in the DNA-bound form. In the free structure, four possible salt bridges were observed to restrict the architecture of R2R3 domain. But in DNA-bound conformation, all of them were lost. Instead, four hydrogen bonds were observed between residues Arg$^{125}$, Glu$^{135}$ and Arg$^{76}$, and DNA molecule. And after the rotation between R2 and R3 motifs, the complex conformation becomes complementary to the DNA major groove. It is then possible to calculate the complex structure by the docking program, HADDOCK.

In our docking structure, many hydrogen bonds and hydrophobic interactions are observed between *tvMyb*$_{135-141}$ and the DNA duplex which correlate well with our experimental results. From the analysis of the $^1$H-15$N$ HSQC titration and $^1$H-15$N$ HX-NOE experiments, we found that the residues located in the N-terminus, H3, H5 and H6 exhibit large chemical shift perturbations and increased HX-NOE values in the presence of DNA. These residues make hydrogen bonds and hydrophobic contacts with the DNA in our structure. Especially for the residues in the N-terminus, they are flexible in the free structure but showed the highest chemical shift perturbation and largest increase in HX-NOE value in complex with DNA, implying that this region adopts a much more rigid conformation compared to the free structure. In the complex model, the N-terminal residues do display several hydrogen bonds and hydrophobic interactions with DNA. It is also notable that the chemical shifts of the amide resonances of Asn$^{106}$ and Asn$^{110}$ are greatly perturbed with downfield shifting, which agrees well with the observation that the amide protons of Asn$^{106}$ and Asn$^{110}$ form direct hydrogen bonds with the phosphate backbone of DNA. In addition, based on our previous studies on MRE-1/MRE-2f specificity (6), the *tvMyb* protein binds specifically to ADE6 and CYT8 to THY11. These nucleotides display nine hydrogen bonds between the bases and the protein in the complex structure. Although we failed to assign the inter-molecular NOEs between the protein and the DNA duplex, the complex model agrees well with all our experimental results and may reflect the specific DNA recognition by *tvMyb*$_{135-141}$.

From the searching of RCSB protein data bank, there is only one R2R3 domain/DNA complex structure determined by NMR, the c-Myb R2R3 domain/DNA complex structure (3). The main difference in the DNA recognition domain between *tvMyb* and c-Myb is the length of H6. The H6 in *tvMyb*$_{135-141}$ is much longer than that in c-Myb (Supplementary Figure S4) and we have demonstrated that this longer helix is important in protein stability and DNA-binding activity for *tvMyb*$_{135-141}$. In our complex structure, some of the intermolecular contacts between *tvMyb*$_{135-141}$ and DNA are very similar to
those observed in the homologous c-Myb R2R3 domain/DNA complex structure, but some are unique which correlates with DNA specificity of tvMyb135–141. For clarity, our DNA sequence (ATAACGAT) is numbered from 4 to 11 (the complementary strand, ATCGTAT, is numbered from 22 to 29), and the DNA sequence in c-Myb/DNA complex (CTAACCTGA) is numbered from 2 to 9 (the complementary strand, TCAGTTAG, is numbered from 14 to 21). The underlined sequence is the binding sequence of the two proteins. In c-Myb/DNA complex structure, Glu\textsuperscript{152} forms hydrogen bonds with CYT6 and CYT15; Asn\textsuperscript{183} forms a hydrogen bond with ADE4; Arg\textsuperscript{153} forms a hydrogen bond with phosphate of ADE5; Arg\textsuperscript{131} forms a hydrogen bond with phosphate of THY14; and Ala\textsuperscript{167} forms a hydrogen bond with phosphate of ADE16. Similar hydrogen bond contacts are also observed in our complex model; the aforementioned interaction between Glu\textsuperscript{75}, Asn\textsuperscript{126}, Arg\textsuperscript{76}, Arg\textsuperscript{74} and Asn\textsuperscript{10} of tvMyb135–141 with DNA. However, three hydrogen bonds in c-Myb complex, Asn\textsuperscript{186} with THY19 and THY18 and Ser\textsuperscript{187} with THY3, which are important for specific binding of c-Myb, are not observed in our model because the corresponding residues in tvMyb135–141 are Met\textsuperscript{129} and Met\textsuperscript{130}. In our complex structure, three additional hydrogen bonds are observed; Asn\textsuperscript{69} with ADE10 and Gln\textsuperscript{72} with GU9 and THY23. The similarities and differences of hydrogen bond contacts observed in c-Myb and tvMyb135–141 complexes explain the difference in DNA-binding specificity.

The structure determination of multi-domain proteins in complex with DNA is biologically relevant. However, this kind of complex normally does not provide NMR spectra of enough quality for complex structure determination. The residues at binding interfaces may experience intermediate exchange and give broadened peaks or no peaks for assignment. Different buffer conditions, protein lengths or even mutations must be tested in order to improve the quality of the NMR spectra. In addition, the identification of inter-molecular NOEs is tedious and time consuming. In this study, we obtained several intermolecular NOEs especially in the 3D \textsuperscript{15}N-edited NOESY-HSQC spectrum obtained from a 95% \textsuperscript{2}H, \textsuperscript{13}C, \textsuperscript{15}N-labeled tvMyb135–141/DNA complex sample. However, the assignment of proton resonances for the protein/DNA complex could not be finished due to the limited stability of the complex sample, the lack of stable isotope labeling on the DNA duplex, peak overlapping, and the short $T_2$ relaxation time. The average transverse relaxation time, $T_2$, of the amide protons of tvMyb135–141/DNA complex sample at 298K, which was measured by the 1-1 echo sequence (41), is around 12.9 ms, similar to what is observed for a protein with molecular weight of 30 kDa. The sensitivity and resolution of NMR spectra of the complex sample were therefore largely reduced. Hence, the assignments for the proton resonances of the complex sample and for the intermolecular NOEs could not be well established. Here, we used an alternative approach, combining the free form structure, $^1$D$_{NH}$ RDC values from the complex sample, chemical shift perturbations upon binding, DNA specificity data from EMSA and data-driven macro-molecular docking, to understand the molecular basis for DNA recognition by tvMyb135–141. This method could be especially useful for revealing the structural information of a multi-domain protein in complex with macro-molecules. First, it is much easier to assign the backbone resonances in HSQC or TROSY spectra of the complex than to identify the inter-molecular NOEs. Second, residual dipolar couplings have been shown to be powerful sources of long-range structural information, especially domain orientations (42,43). The RDCs have been successfully applied in determining the first solution structure of Lys48-linked di-ubiquitin by HADDOCK (18). In our case, although we did not provide any direct inter-molecular NOE restraints, the AIRs from chemical shift perturbations and the DNA specificity data are sufficient to define many reasonable interactions between tvMyb135–141 and the DNA that correlate well with all the experimental results. These data further our understanding of DNA recognition by Myb protein. This approach should be very useful when applied to determine the complex structures involving proteins with multiple domains.

**SUPPLEMENTARY DATA**

Supplementary Data are available at NAR Online.

**ACKNOWLEDGEMENTS**

We would like to thank the High-field Biomacromolecular NMR Core Facility at the Academia Sinica supported by the National Research Program for Genomic Medicine for obtaining all the NMR spectra. We also thank D. Victoria Williams from Department of Chemistry, University of Washington, Seattle, USA for proofreading the manuscript.

**FUNDING**

Academia Sinica; National Science Council, Taiwan, ROC (NSC 95-2320-B-001-040-MY2). Funding for open access charge: Academia Sinica, Taiwan.

**Conflict of interest statement.** None declared.

**REFERENCES**

1. Stracke,R., Werber,M. and Weisshaar,B. (2001) The R2R3-MYB gene family in Arabidopsis thaliana. *Curr. Opin. Plant Biol.*, 4, 447–456.
2. Lipsick,J.S. (1996) One billion years of Myb. *Oncogene*, 13, 223–235.
3. 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.
4. Nomura,N., Takahashi,M., Matsu,M., Ishii,S., Date,T., Sasamoto,S. and Ishizaki,R. (1988) Isolation of human cDNA clones of myb-related genes, A-myb and B-myb. *Nucleic Acids Res.*, 16, 11075–11089.
5. Solano,R., Fuertes,A., Sanchez-Pulido,L., Valencia,A. and Paz-Ares,J. (1997) A single residue substitution causes a switch from the dual DNA binding specificity of plant transcription factor
