8992–9008 *Nucleic Acids Research, 2011, Vol. 39, No. 20* *Published online 19 July 2011*
doi:10.1093/nar/gkr558

Molecular basis of the recognition of the ap65-1 gene transcription promoter elements by a Myb protein from the protozoan parasite *Trichomonas vaginalis*

Ingjye Jiang[1,2], Chen-Kun Tsai[1], Sheng-Chia Chen[1], Szu-huan Wang[1], Imamaddin Amiraslanov[3], Chi-Fon Chang[4], Wen-Jin Wu[1], Jung-Hsiang Tai[1], Yen-Chywan Liaw[3,*] and Tai-huang Huang[1,4,5,*]

[1]Institute of Biomedical Sciences, Academia Sinica, Taipei 115, [2]Institute of Bioinformatics and Structural Biology, National Tsing Hua University, Hsinchu 300, [3]Institute of Molecular Biology, [4]Genomics Research Center, Academia Sinica, Taipei 115 and [5]Department of Physics, National Taiwan Normal University, Taipei 116, Taiwan, ROC

Received September 20, 2010; Revised June 5, 2011; Accepted June 17, 2011

ABSTRACT

Iron-inducible transcription of the *ap65-1* gene in *Trichomonas vaginalis* involves at least three Myb-like transcriptional factors (*tv*Myb1, *tv*Myb2 and *tv*Myb3) that differentially bind to two closely spaced promoter sites, MRE-1/MRE-2r and MRE-2f. Here, we defined a fragment of *tv*Myb2 comprising residues 40–156 (*tv*Myb2$_{40–156}$) as the minimum structural unit that retains near full binding affinity with the promoter DNAs. Like c-Myb in vertebrates, the DNA-free *tv*Myb2$_{40–156}$ has a flexible and open conformation. Upon binding to the promoter DNA elements, *tv*Myb2$_{40–156}$ undergoes significant conformational re-arrangement and structure stabilization. Crystal structures of *tv*Myb2$_{40–156}$ in complex with promoter element-containing DNA oligomers showed that 5′-a/gACGAT-3′ is the specific base sequence recognized by *tv*Myb2$_{40–156}$, which does not fully conform to that of the Myb binding site sequence. Furthermore, Lys^{49}, which is upstream of the R2 motif (amino acids 52–102) also participates in specific DNA sequence recognition. Intriguingly, *tv*Myb2$_{40–156}$ binds to the promoter elements in an orientation opposite to that proposed in the HADDOCK model of the *tv*Myb1$_{35–141}$/MRE-1-MRE-2r complex. These results shed new light on understanding the molecular mechanism of Myb–DNA recognition and provide a framework to study the molecular basis of transcriptional regulation of myriad Mybs in *T. vaginalis*.

INTRODUCTION

The Myb proteins consist of a large family of proteins with diverse functions. These proteins have been demonstrated to play key roles in various important biological functions, including roles as regulators of stem and progenitor cells in the bone marrow, colonic crypts and a neurogenic region of the adult brain; they have also been identified as oncogenes involved in some human leukemias (1). In plants, Mybs were found to be involved in the control of plant-specific processes including metabolism, cell fate and identity, developmental processes and responses to biotic and abiotic stresses (2). Most Myb proteins function as transcription factors by binding to DNA via highly conserved Myb domains that recognize the t/cAACt/gG sequence, known as the Myb binding site (MBS) (3,4). The Myb domain generally consists of up to four imperfect amino acid sequence repeats (R) of about 50 amino acids, which form three α-helices. The second and third helices of each repeat form a helix turn helix (HTH) structure. The third helix of the second (R2) and third repeat (R3) are the recognition helices that bind to the major groove cooperatively and recognize the specific DNA sequence (3,5).

*To whom correspondence should be addressed. Tel: +886 2 2652 3036; Fax: +886 2 2788 7641; Email: bmthh@ibms.sinica.edu.tw
Correspondence may also be addressed to Yen-Chywan Liaw. Tel: +886 2 2789 9199; Fax: +886 2 782 6085; Email: mbycliaw@gate.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.

Trichomonas vaginalis is a protozoan parasite that causes trichomoniasis, one of the most common sexually transmitted diseases in humans (6).The *ap65-1* gene from *T. vaginalis* encodes a 65-kDa protein, which upon iron repletion has been reputed to be one of the surface adhesins that mediate cytoadherence of the parasite to vaginal epithelial cells (7–9). We have shown that transcription of the *ap65-1* gene is regulated in *cis* by two neighboring MBS, MRE-1/MRE-2r and MRE-2f, along with several other neighboring DNA regulatory elements in the *ap65-1* promoter. The MRE-1/MRE-2r site comprises three overlapping DNA elements, which are the binding sites for three Myb-like proteins: *tv*Myb1, *tv*Myb2 and *tv*Myb3 (Figure 1A) (10–12). *tv*Myb1 and *tv*Myb2 each comprises a R2R3 class, whereas *tv*Myb3 comprises a R1R2R3 class of the MBS domain and shorter but highly variable N- and C-termini (Figure 1B) (12,13). The constitutively expressed *tv*Myb1 may repress basal or iron-inducible, while enhance growth related, *ap65-1* transcription through differential bindings to the entire MRE-1/MRE-2r and/or MRE-2f (13). *tv*Myb2, which shares 33% identity with *tv*Myb1, may enhance *ap65-1* transcription by bindings to the MRE-2r moiety of MRE-1/MRE-2r and MRE-2f. It may preferentially target the *ap65-1* promoter in iron-replete cells to iron-depleted cells at an early growth stage when its expression is low, but the preference shifts to iron-depleted cells at a later growth stage when its expression is high (14). *tv*Myb3, which binds only to the MRE-1 moiety of MRE-1/ME-2r, may display a differential promoter selection antagonistic to that of *tv*Myb2, to activate transcription (12). It is intriguing why and how this simple protozoan without noticeable differentiation uses so many distinct Myb proteins for transcription.

The solution structure of a truncated form of *tv*Myb1 (*tv*Myb1$_{35–141}$) was previously solved by nuclear magnetic resonance (NMR), and the structural basis for its interaction with the MRE-1/MRE-2r DNA duplex was modeled by a data-driven HADDOCK software (15). However, unambiguous intermolecular nuclear overhauser effects (NOE) between *tv*Myb1$_{35–141}$ and DNA could not be assigned; thus, the detailed interaction between *tv*Myb1 and the promoter DNA, as well as the molecular basis of sequence-specific recognition, remains elusive. Here we report the identification and dynamics characterization of a *tv*Myb2 fragment comprising amino acid residues 40–156 (*tv*Myb2$_{40–156}$) as the minimum structural unit that retains near full binding affinity with the promoter DNA. We also report the X-ray crystal structures of *tv*Myb2$_{40–156}$ in complex with DNA promoter elements. Structural comparison showed a remarkable similarity between c-Myb and *tv*Myb2$_{40–156}$, as well as features unique to *tv*Myb2.

MATERIALS AND METHODS

Cloning, protein expression and purification

A DNA fragment containing the coding region of the full-length *T. vaginalis Myb2* (designated *tvMyb2*) gene was cloned by polymerase chain reaction (PCR). The amplified DNA was digested with BamHI and SmaI and was inserted into the pGEX-6p3 vector (GE) containing a removable Glutathione-S-transferase (GST) tag at the N-terminus. The DNA encoding the truncated fragment comprising amino acid residues from 40 to 156 (designated *tv*Myb2$_{40–156}$) was amplified by PCR, and subcloned for ligation into the pET22b^{+} vector (Novagen), which contains a hexahistidine tag at the C-terminus. The forward primer was designed to contain an NdeI site and the reverse one a stop codon followed by an XhoI site.

The plasmids were each transformed into *Escherichia coli strain* BL21 star (DE3) competent cells (Invitrogen). The transformed cells were grown in 1 liter LB media supplemented with 50 μg/ml ampicillin in a 2.8 l Fernbach baffled flask with vigorous shaking (~200 rpm) at 37°C. The cells expressing *tv*Myb2 were induced at $OD_{600} \sim 0.8$ by adding IPTG to a final concentration of 0.1 mM, and the incubation was continued for another 4 h. *tv*Myb2$_{40–156}$ was similarly expressed except for a final concentration of IPTG to be 1 mM and post-induction incubation to be 6 h at 30°C.

The cells from 1 l culture were harvested by centrifugation, washed twice with 50 ml of phosphate buffered saline (PBS) containing 1 mM EDTA and 1 mM DTT, resuspended in 25 ml of a lysis buffer (50 mM Na_3PO_4, 500 mM NaCl, 0.1 mM EDTA) containing 1x Complete Protease inhibitor cocktail (Roche) and lysed with a microfluidizer (Microfluidics, MA, USA). To purify the GST–*tv*Myb2 fusion protein, the supernatant was incubated with gentle agitation at 4°C for 1 h in 10 ml of Glutathione Sepharose 4B resin (GE) in PBS and washed with 200 ml of a washing buffer (PBS containing 500 mM NaCl, 1 mM EDTA). The GST–*tv*Myb2 protein was cleaved from the resin by adding 20 μl of the Precision Protease to the resin in 25 ml of a cleavage buffer (50 mM Tris–HCl, pH 7, 150 mM NaCl, 1 mM EDTA, 1 mM DTT) at 4°C for 24 h.

To purify the His–*tv*Myb2$_{40–156}$, the supernatant was loaded to a column packed with 10 ml of Profinity Ni-charged IMAC resin (Bio-rad) equilibrated in a lysis buffer (50 mM Na_3PO_4, 500 mM NaCl, 0.1 mM EDTA, pH 8.0). After 1 h incubation at 4°C, the column was washed with 200 ml of a washing buffer (50 mM Na_3PO_4, 500 mM NaCl, 20 mM imidazole, pH 8.0) to remove unbound proteins. The His–*tv*Myb2$_{40–156}$ protein was eluted with 50 ml of an elution buffer (50 mM Na_3PO_4, 300 mM NaCl, 250 mM imidazole, pH 8.0).

The eluted proteins were further purified through a Superdex75 size exclusion column (SEC) connected to an AKTA Explorer system (GE), and the protein authenticities were verified by peptide mass fingerprinting (PMF). To prepare the uniformly ^{15}N,^{13}C-labeled protein for NMR experiments, cells were grown in M9 minimal medium supplemented with $^{15}NH_4Cl$ (1 g/l), ^{13}C-glucose (2 g/l) and ^{15}N,^{13}C-isogro (Sigma-Aldrich). DNA oligonucleotides containing MRE-1/MRE-2r or MRE-2f promoters were purchased from Genomics Inc. (Taiwan) (Figure 1C). The concentrations of proteins and double-stranded DNAs (dsDNA) were determined by absorbance at 280 and 260 nm, respectively, with an

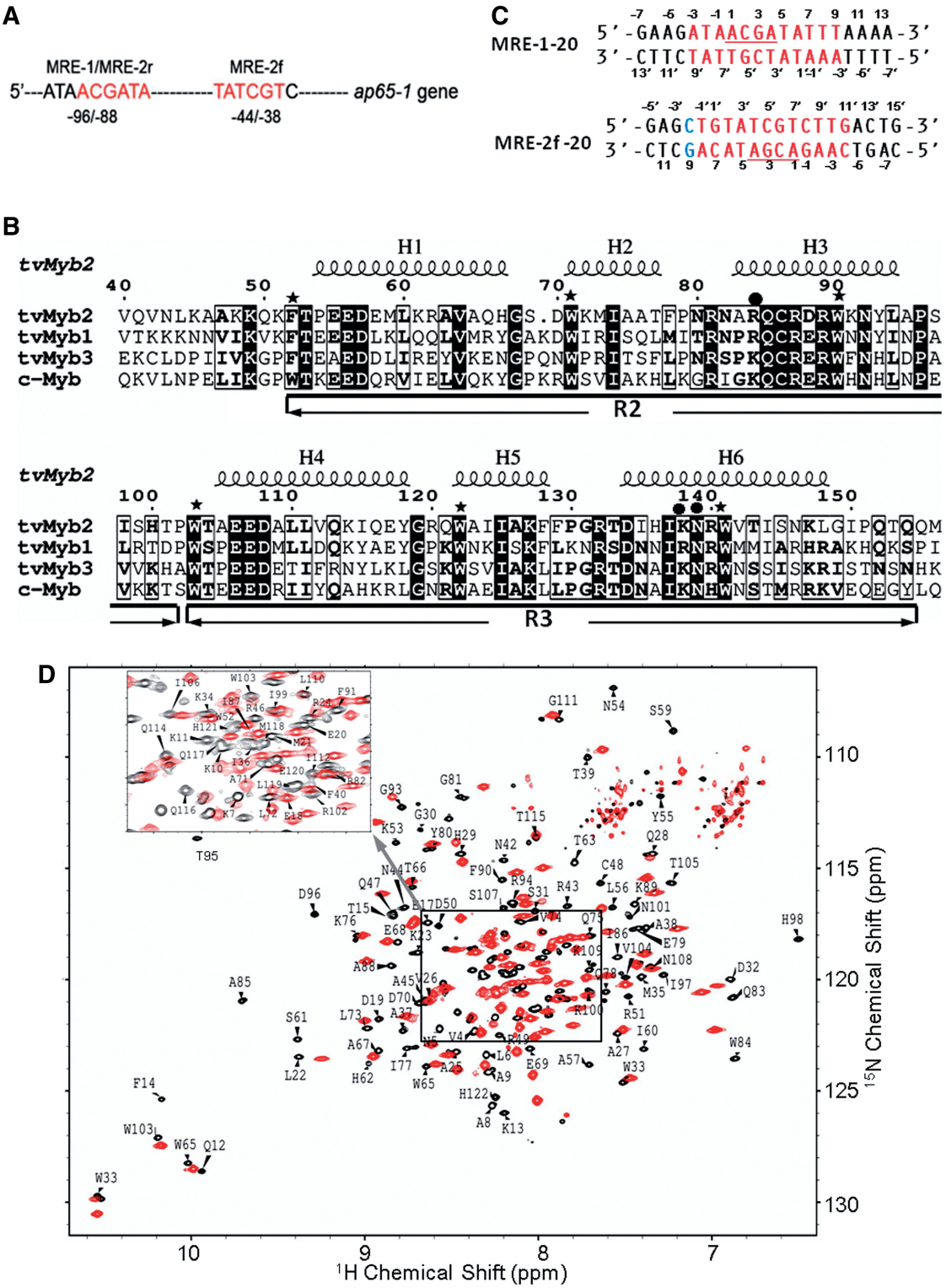

Figure 1. (**A**) The promoter organization of the *ap65-1* gene. (**B**) Amino acid sequence alignment of the R2R3 motif regions of *tv*Myb1, *tv*Myb2, *tv*Myb3 and mouse c-myb (SwissProt ID number P06876). The sequence is numbered according to that of *tv*Myb2. The R2 (amino acids 52–102) and R3 (amino acids 103–154) motif regions are underlined. The stars identify the conserved aromatic residues and the dotted residues are the conserved residues participating in specific DNA sequence recognition. The helical regions of *tv*Myb2$_{40–156}$ are also shown. (**C**) DNA 20-mer duplexes used for ITC studies. The sequences span the −81 or −100 and −31 or −50 regions of the promoter containing the MRE-1/MRE-2r and MRE-2f recognition sites, respectively. The red-colored bases are the 12-mer DNA sequences (MRE-1-12 and MRE-2-12) used for NMR studies. The red-colored bases plus the blue-colored bases in MRE-2-20 are the DNA sequences of the 13-mer DNA (MRE-2-13) used for crystallography studies. For the convenience of structural comparison, the bases in both MRE-1-20 and MRE-2-20 are numbered such that the specific sequence recognized by *tv*Myb2 (i.e. $A_1C_2G_3A_4T_5A_6$) is numbered the same in the two oligomers. (**D**) Superposition of the ^{15}N-HSQC spectra of *tv*Myb2$_{40–156}$ (red) and *tv*Myb2$_{40–156}$/MRE-1-12 complex (black).

ND-1000 UV-Vis spectrophotometer (NanoDrop Technologies, Inc.).

Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) was performed with an ITC_{200} calorimeter (MicroCal Inc.) at 25°C. Protein and DNA solution were dialyzed overnight against the same reaction buffer (20 mM Na_3PO_4, 150 mM NaCl, 1 mM EDTA, pH 7.4). The titration was carried out by injecting 1 µl (first injection) or 2.5 µl (2nd to 15th injection) of *tv*Myb2$_{40–156}$ solution at 100 µM concentration into the sample cell filled with 10 µM DNA solution. The initial delay was 300 s, with a 120 s interval between two successive injections. Since the T_m values of MRE-1-12 and MRE-2-12 are near room temperature, the longer DNA duplexes of MRE-1-20 and MRE-2-20 were used for ITC studies. ITC binding curves were fitted to the single-site binding equation in the Microcal Origin software package, with the last two points taken as dilution heat.

NMR experiments

The NMR spectra were acquired on Bruker AV500, AV600 or AV800 spectrometers equipped with 5-mm triple resonance cryoprobes and a single-axis pulsed field gradient at 310 K. NMR data were acquired in Shigemi tubes on 0.7 mM protein samples in 20 mM Na_3PO_4, 50 mM NaCl, 1 mM EDTA, 1 mM DTT, 10% (v/v) D_2O at pH 6. The ^{15}N,^{13}C-labeled *tv*Myb2$_{40–156}$-DNA complex samples were prepared at a protein–DNA molar ratio of 1:1.2. As 12-mer DNA duplexes gave the best NMR spectra, they were used for all NMR studies. Protein backbone resonance assignments were achieved by standard triple resonance experiments, including HNCA, HN(CO)CA, HNCACB, CBCA(CO)NH, HNCO and HN(CA)CO (16–20). Aliphatic side chain assignments primarily involved HCCH-COSY and HCCH-TOCSY, with the help of complementary HCC(CO)NH, (H)CC(CO)NH, HBHA(CO)NH and (H)CCH-TOCSY experiments (21–23). ^{15}N-edited TOCSY-HSQC acquired with 60-ms mixing time was used to resolve ambiguity due to spectral overlap. Aromatic side chains were assigned with the use of proton homo-nuclear TOCSY and NOESY experiments, with additional verification from ^{13}C-HMQC, (HB)CB(CGCD)HD and (HB)CB(CGCDCE)HE experiments (24,25). ^{1}H chemical shifts were externally referenced to 0 ppm methyl resonance of 2,2-dimethyl-2-silapentane-5-sulfonate (DSS), whereas ^{13}C and ^{15}N chemical shifts were indirectly referenced according to the recommendations of the International Union of Pure and Applied Chemistry (26). The NMR spectra were processed using Bruker TOPSPIN 2.0 and analyzed by Sparky (27) and CARA (Available from http://www.nmr.ch).

^{15}N-T_1, ^{15}N-T_2 and [^{1}H-^{15}N] NOE were determined by using standard pulse sequences (28). Ten ^{15}N spin–lattice/longitudinal relaxation rate constant (T_1) experiments were performed in random order, with relaxation delays of 0, 195.6, 301.0, 496.6 (duplicate), 707.3, 1098.5 (in duplicate), 1610.2 and 2197.1 ms. Similarly, spin–spin/transverse relaxation rate constant (T_2) experiments were performed randomly with relaxation delays of 0 (in duplicate), 15.8, 31.5, 47.3 (duplicate), 63.0, 78.8, 94.5 and 126.0 ms. Both rate constants were determined using the program Curvefit, assuming mono-exponential decay of the peak intensities. The errors in peak intensities were calculated from two duplicate experiments. The steady-state heteronuclear [^{1}H-^{15}N] NOE experiment was carried out in duplicate in an interleaved manner, with and without proton saturation. The NOE was calculated as the error-weighted average ratio of peak intensities, with error estimate by propagating the base-plane noise. The reduced spectral density analysis was performed as previously described (29–35).

To measure one-bond ^{1}H-^{15}N residual dipolar couplings (RDC), *tv*Myb2$_{40–156}$ was partially aligned in diluted liquid crystalline phase of 5% C12E5 polyethyelene–glycol/*n*-hexanol ($r = 0.96$) (36). The *tv*Myb2$_{40–156}$/MRE-1-12 complex was partially aligned in liquid crystalline phase containing 15 mg/ml Pf1 bacteriophage (Asla Biotech Ltd, Latvia). Changes in splitting relative to the isotropic $^{1}J_{NH}$ values were measured using DSSE-HSQC experiments to obtain one bond ^{1}H-^{15}N RDC (37). The measured RDCs were analyzed using the program PALES (38).

Crystallization of protein–DNA complex

Samples of protein–DNA complex were prepared by mixing *tv*Myb2$_{40–156}$ or Se-Met-*tv*Myb$_{40–156}$ with DNAs at 1:1 molar ratio at a final concentration of 15 mg/ml in a buffer containing 20 mM Tris–HCl (pH 7.5) and 100 mM NaCl. The mixtures were further purified through Superdex75 size exclusion column. The 12-mer duplex DNA with a sequence of 5′-ATAACGATA TTT-3′ (MRE-1-12) and the 13-mer duplex DNA with a sequence of 5′-CTGTATCGTCTTG-3′ (MRE-2-13) (Figure 1C) were purchased from Genomics, Inc. (Taiwan). The *tv*Myb2$_{40–156}$/MRE-1-12 and *tv*Myb2$_{40–156}$/MRE-2-13 complexes produced best quality crystals for structural analysis. The crystallization experiments were carried out at 22°C by the vapor-diffusion method. The crystals of the *tv*Myb2$_{40–156}$/MRE-2-13 complex were obtained by mixing 1 µl of the protein–DNA complex solution and 1 µl of a crystallization solution composed of 0.1 M sodium cacodylate (pH 6.0), 15% isopropanol, 25% PEK4000 and 10% glycerol. The crystals of the Se-Met *tv*Myb2$_{40–156}$/MRE-2-13 complex were obtained by mixing 1 µl of the protein solution with 1 µl of a solution contain 0.2 M ammonium nitrate and 2.2 M ammonium sulfate and the crystals of the *tv*Myb2$_{40–156}$/MRE-1-12 complex were obtained by mixing 1 µl protein solution with 1 µl of a solution contain 0.1 M HEPES (pH 7.0) and 1.5–1.7 M ammonium sulfate.

Data collection and structure determination

The crystals of the Se-Met *tv*Myb2$_{40–156}$/MRE-2-13 complex were flash-cooled in liquid nitrogen without a cryo-protectant, whereas the crystals of the *tv*Myb2$_{40–156}$/MRE-1-12 complex were flash-cooled in the liquid nitrogen using 20% glycerol as a cryo-protectant. Data

were collected on a quantum 315 CCD detector at beamlines 13B and 13C of the National Synchrotron Radiation Research Center (NSRRC) in Taiwan, the Republic of China and the Mar225HE detector at beamline 44XU of the Super Photon Ring-8 Synchrotron Radiation Center in Japan. Data were processed and scaled with HKL2000 (39). The structure of the Se-Met-*tv*Myb2$_{40–156}$/MRE-2-13 complex was determined by Se-MAD phasing. Four of the expected selenium sites were found, and the initial phase was calculated by the use of SOLVE (40). Density modification and auto-model building were performed with RESOLVE (40). The structure of the *tv*Myb2$_{40–156}$/MRE-1-12 complex was determined by molecular replacement with the program phaser in PHENIX (41,42), with the solved structure of the *tv*Myb2$_{40–156}$/MRE-2-13 complex as a model. Iterative cycles of manual model building in COOT (43) and refinement in PHENIX yielded the final model. Table 2 gives the refinement statistics. Overall geometry of the model was assessed by Molprobity (44,45). The structures were analyzed by the use of PyMol (PyMOL Molecular Graphics System, v1.3, Schrödinger, LLC.) The structure factors and coordinates were deposited in the Protein Data Bank under (PDB) under PDB ID codes 3OSF (*tv*Myb2$_{40–156}$/MRE-2-13) and 3OSG (*tv*Myb2$_{40–156}$/MRE-1-12).

RESULTS

Mapping the minimum structured fragment of tvMyb2

Full-length *tv*Myb2 (*tv*Myb2) is most likely to exist as monomer, as predicted from analytical ultracentrifugation (Supplementary Figure S1A). However, the molecular weight of 36.1 kDa deduced from size exclusion chromatogram is much larger than the 21 kDa calculated from the sequence (Supplementary Figure S1B), suggesting that the protein may not be folded as a compact globular shape. To map the structured region essential for DNA binding, we performed a Predictor of Natural Disordered Regions (PONDR) prediction of the order–disorder profile (46) and showed that the N-terminal 50 amino acid and the C-terminal 20 amino acid are disordered (Supplementary Figure S1C). Gene fragments encoding various lengths of the *tv*Myb2 were constructed and expressed. The foldings of these fragments were assessed by ^{15}N-HSQC NMR. The spectrum of *tv*Myb2 contains many well-dispersed peaks, (red peaks in Supplementary Figure S2), the characteristic of a folded protein, but many resonances were crowded in the 7.6–8.6 ppm region in the proton dimension (red peaks in the boxed region in Supplementary Figure S2), characteristics of a disordered protein. In contrast, the peaks in the spectrum of *tv*Myb2$_{40–156}$ are well dispersed (black peaks in Figure 1D). The addition of MRE-1-12 (red peaks in Figure 1D) or MRE-2-12 (data not shown) to *tv*Myb2$_{40–156}$ resulted in the disappearance of a set of resonances with the concurrent appearance of another set of resonances, the characteristics of a high-affinity DNA binding. Extending towards N- or C-terminus did not give more resonances outside of the 7.6–8.6 ppm region (data not shown), suggesting that *tv*Myb2$_{40–156}$ is the minimal fragment with an ordered structure. Circular dichroism studies showed that *tv*Myb2$_{40–156}$ is a helical protein (Supplementary Figure S1D).

NMR characterization of tvMyb2$_{40–156}$-promoter interaction

To gain insight into the structure and interaction with promoter DNAs, heteronuclear 3D NMR spectra of DNA-free *tv*Myb2$_{40–156}$ and *tv*Myb2$_{40–156}$ bound to MRE-1-12 or MRE-2-12 were obtained (Figure 1D), and the resonances were assigned and deposited in the Biological Magnetic Resonance Data Bank (accession number 17170). The amide resonances of Lys^{91}–Ser^{97} in the ^{15}N-HSQC spectrum were observed in the DNA–protein complexes, but not in the DNA-free protein form (Figure 1D). Further analysis of the relaxation data revealed that these residues reside in a flexible linker region with the structure of the DNA-free form much more flexible than that of the DNA-bound form (to be discussed later). The secondary structure of *tv*Myb2$_{40–156}$ was deduced from the consensus chemical shift index derived from the spectra of the promoter DNA-bound forms (47). The results showed that *tv*Myb2$_{40–156}$ consists of six helices, designated to be H1 (Glu^{55}–Gly^{68}), H2 (Trp^{71}–Ala^{76}), H3 (Ala^{83}–Leu^{94}), H4 (Ala^{105}–Tyr^{118}), H5 (Trp^{122}–Phe^{127}) and H6 (Ile^{137}–Gly^{149}) (Supplementary Figure S3).

To probe the Myb binding site, we further determined the changes of DNA binding-induced chemical shift, $\Delta\delta$, as calculated from the equation $[\delta_{NH}{}^2 + (0.154\Delta\delta_N)^2]^{1/2}$, where δ_{NH} is the chemical shift change of the amide proton and δ_N is the chemical shift change of the amide ^{15}N (Figure 2) (48). The chemical shift perturbation patterns induced by binding of MRE-1-12 or MRE-2-12 are nearly identical. The residues with $\Delta\delta > 0.2$ ppm clustered around the N-terminus; H2–H3 linker, H3; H3–H4 linker and the H5 and H6. To ensure that the 12-mer DNAs were sufficient for protein recognition, we also obtained ^{1}H-^{15}N-HSQC spectra of (u-^{15}N)-*tv*Myb2$_{40–156}$ in complex MRE-1-20 or MRE-2-20 (data not shown). Although the spectral quality of these two complexes were not as good, nevertheless, their chemical shift perturbation patterns were practically identical to those observed for the corresponding 12-mers, suggesting that the bases beyond the 12-mer region are not essential for interaction with *tv*Myb2$_{40–156}$.

Promoter DNA-binding affinity and thermodynamics of tvMyb2 variants

MRE-1-12 and MRE-2-12 are not suitable for *tv*Myb2 binding study by ITC since their thermal melting temperatures are close to room temperature. Thus, MRE-1-20 and MRE-2-20 were used for ITC studies. As shown in Figure 3 and Table 1, *tv*Myb2 and *tv*Myb2$_{40–156}$ bind to MRE-1-20 of similar affinities with $K_D = 100 \pm 35$ nM and 110 ± 23 nM, corresponding to binding free energy of $\Delta G = -9.7$ and -9.6 kcal/mol, respectively. In contrast, the binding affinity towards MRE-2-20 was higher for *tv*Myb2 ($K_D = 12.4 \pm 2.6$ nM) than *tv*Myb2$_{40–156}$ ($K_D = 32.3 \pm 6.1$ nM), suggesting that amino acid

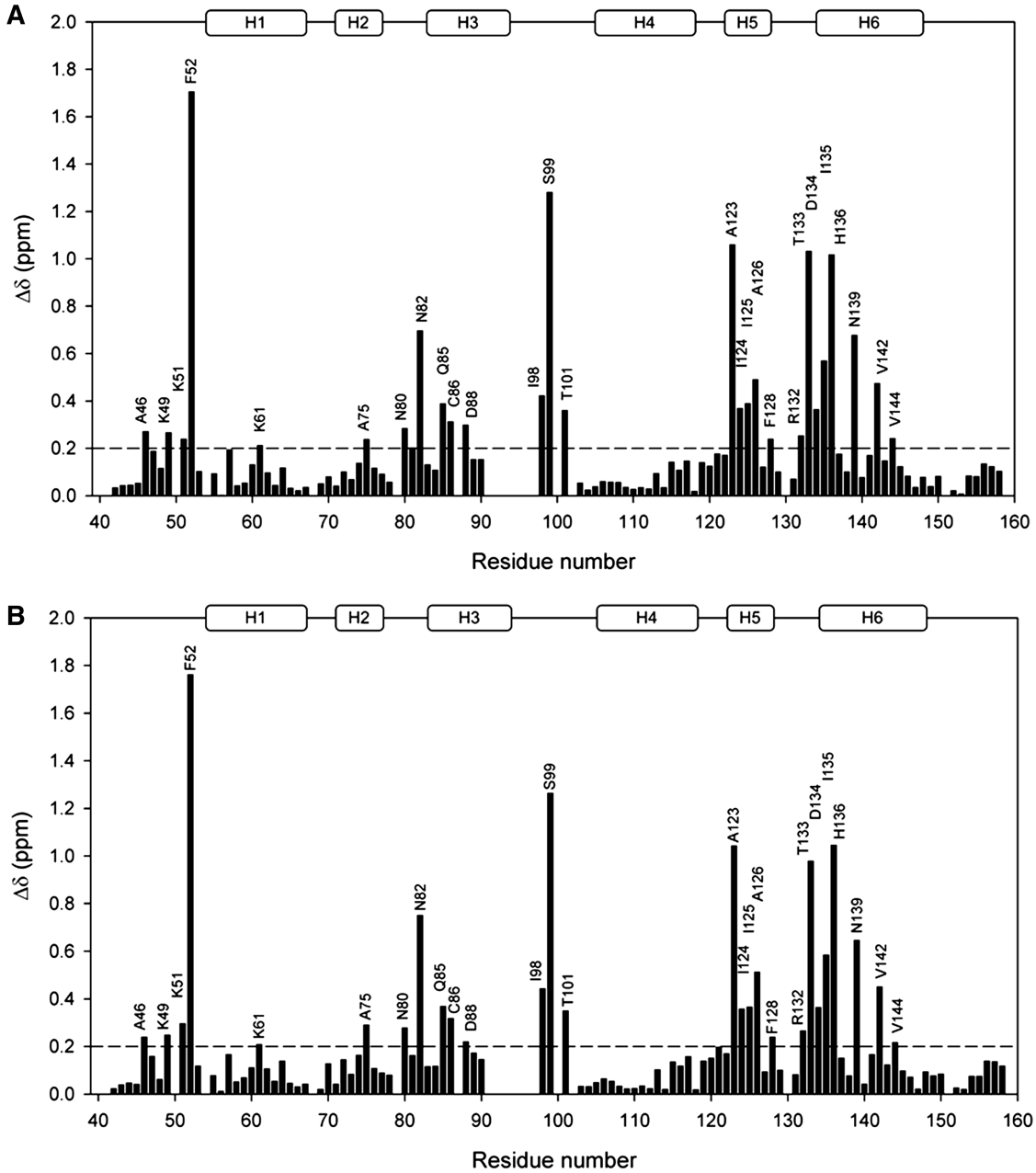

Figure 2. The normalized chemical shift changes induced by binding of MRE-1-12 (**A**) or MRE-2-12 to *tv*Myb2$_{40–156}$ (**B**). The assigned secondary structure regions are shown on top and the residues with chemical shift changes larger than 0.2 ppm are labeled.

residues flanking *tv*Myb2$_{40–156}$ also contribute to the binding of *tv*Myb2 to MRE-2-20. Nonetheless, the difference in their binding energies is small compared with the overall binding energy (−10.8 and −10.3 kcal/mol for *tv*My2 and *tv*Myb2$_{40–156}$, respectively). Therefore, *tv*Myb2$_{40–156}$ was used for structural studies by NMR and X-ray crystallography. This fragment encompasses the R2 and R3 motifs corresponding to vertebrate c-Myb. Interestingly, both entropy and enthalpy changes at room temperature contributed to a similar free energy of ∼4.8 kcal/mol toward the binding of *tv*Myb2 and *tv*Myb2$_{40–156}$ to MRE-1-20, whereas the binding to MRE-2-20 was primarily enthalpy driven, with a small unfavorable entropy change.

Crystal structure of tvMyb2$_{40–156}$ in complex with MRE-1-12

The crystal structure of *tv*Myb2$_{40–156}$ in complex with MRE-1-12 was determined by molecular replacement, using the structure of the *tv*Myb2$_{40–156}$/MRE-2-13 complex as a template and refining it to 2.0 Å resolution (Figure 4A and B). Table 2 shows the diffraction parameters and refinement statistics. The protein was crystallized in $P2_1$ symmetry. Each asymmetric unit consisted of two *tv*Myb2$_{40–156}$–DNA complexes of practically identical structures, with an rmsd of 0.464 Å for the two *tv*Myb2$_{40–156}$ proteins. The structure of *tv*Myb2$_{40–156}$ consists of six α-helices, denoted H1 (Pro^{54}–His^{67}), H2

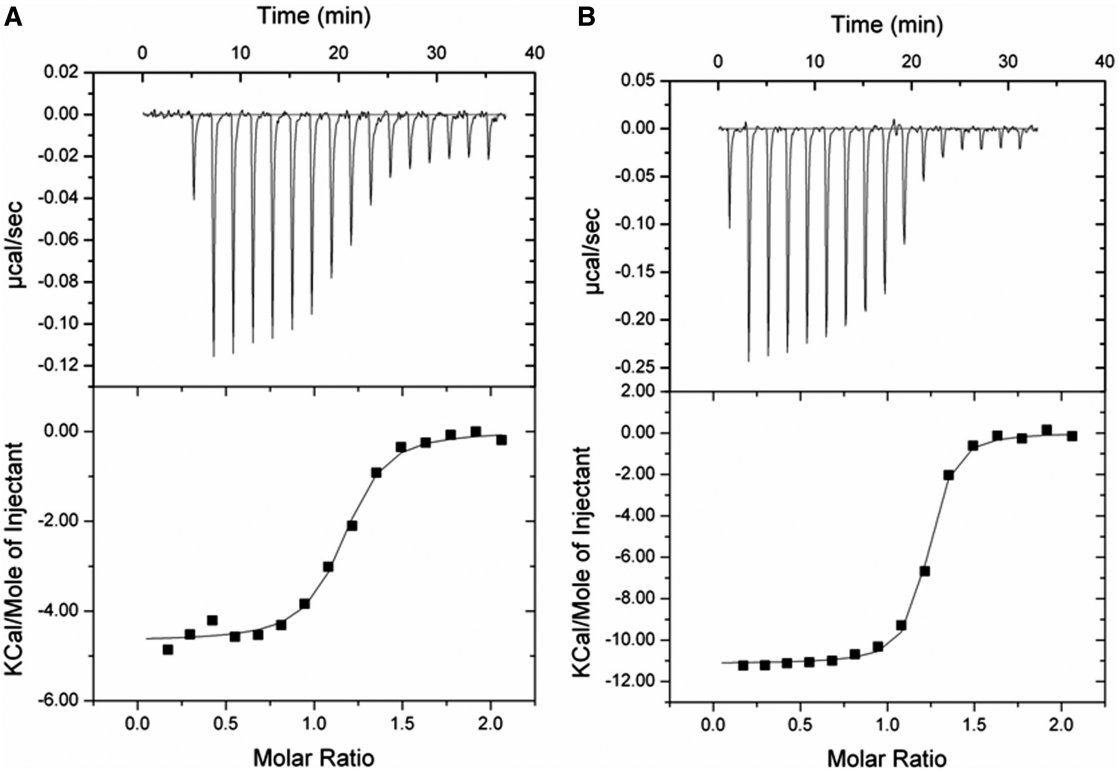

Figure 3. ITC studies of the binding of *tv*Myb2$_{40–156}$ with MRE-1-20 (**A**) or MRE-2-20 (**B**). ITC traces are shown on the upper panels, and the binding isotherms are fit with a single exponential shown on the corresponding lower panels.

Table 1. Thermodynamic data of *tv*Myb2$_{40–156}$- and mutant *tv*Myb2$_{40–156}$-promoter DNA binding deduced from ITC

Sample	ΔH (cal/mol)	$-T\Delta S$ (cal/mol)	K_D (or K_D^m) (nM)[a]	K_D^m/K_D[b]	ΔG (kcal/mol)
*tv*Myb2/MRE-1-20	−4819 ± 63	−4792 ± 213	100 ± 35	–	−9.61
*tv*Myb2/MRE-2-20	−11 730 ± 28	939 ± 102	12.4 ± 2.6	–	−10.79
*tv*Myb2$_{40–156}$/MRE-1-20	−4485 ± 250	−5040 ± 360	110 ± 23	–	−9.53
*tv*Myb2$_{40–156}$/MRE-20	−11 240 ± 160	1017 ± 45	32.3 ± 6.1	–	−10.22
K49A/MRE-1-20	−3571	−2795	21 459	195	−6.37
K49A/MRE-2-20	−9495	1743	2079	64	−7.75
K51A/MRE-1-20	−4014	−3129	5680	52	−7.14
K51A/MRE-2-20	−11 680	3457	962	30	−8.22
R84A/MRE-1-20[c]	n/a	n/a	n/a	n/a	n/a
R84A/MRE-2-20	−1628	−4669	21 142	654	−6.3
R87A/MRE-1-20	−4740	−1833	15200	138	−6.57
R87A/MRE-2-20	−9694	1842	1750	54	−7.85
K138A/MRE-1-20	−1464	−6020	3247	33	−7.48
K138A/MRE-2-20	−5588	−2175	2033	63	−7.76
N139A/MRE-1-20	−5004	−1910	8547	78	−6.91
N139A/MRE-2-20	−10 250	2314	1550	48	−7.94
F52A/MRE-1-20	−17 290	11 115	30 700	279	−6.18
F52A/MRE-2-20	−21 620	14 215	3970	123	−7.41

[a]K_D, the dissociation constants of wild-type (K_D) *tv*Myb2$_{40–156}$; K_D^m, the dissociation constant pf, the mutant *tv*Myb2$_{40–156}$.
[b]K_D^m/K_D, the ratio of the dissociation constants of mutant *tv*Myb2$_{40–156}$ (K_D^m) and the wild-type *tv*Myb2$_{40–156}$ (K_D). It represents the factor in reduced binding affinity.
[c]No binding of R84A to MRE-1-20 can be detected by ITC.

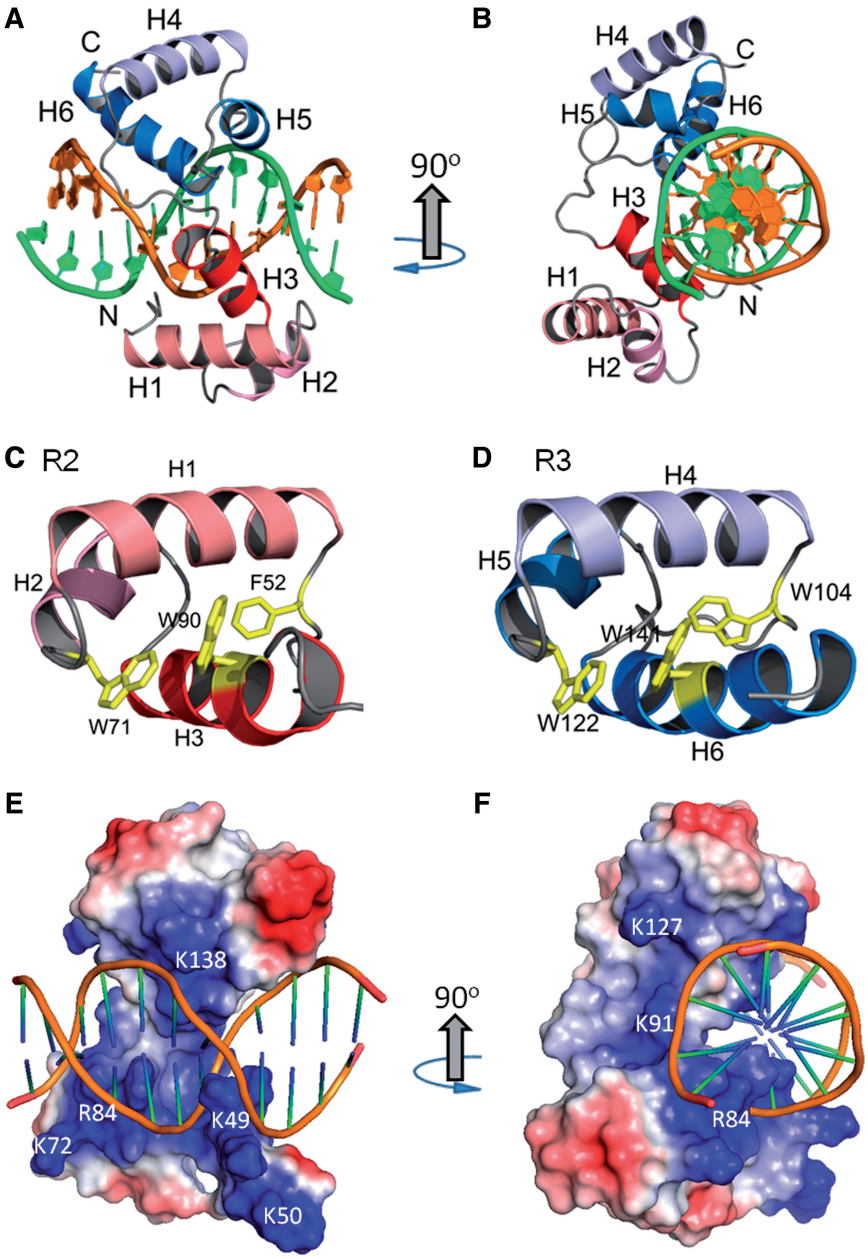

Figure 4. (**A** and **B**) Schematic representation of the crystal structure of *tv*Myb2$_{40–156}$ in complex with MRE-1-12 in two orientations. Helices are represented by colored ribbons. The two strands of DNA backbone and bases are represented by orange- and green-colored ropes and bases, respectively. (**C** and **D**) Schematic representation of the structures of the R2 (C) and R3 (D) motifs showing the spatial locations of the conserved aromatic residues. (**E** and (**F**) Surface charge representations of the *tv*Myb2$_{40–156}$/MRE-1-12 complex in two orientations to highlight the arrangement of the positively charged residues and the DNA backbone. Positively charged surface is colored blue and negatively charged surface is in red.

(Trp^{71}–Thr^{77}), H3 (Ala^{83}–Leu^{94}), H4 (Ala^{105}–Tyr^{118}), H5 (Trp^{122}–Phe^{128}) and H6 (Asp^{134}–Leu^{148}). The secondary structure assignments are almost identical to those determined by NMR in solution, with the exception of H6, which starts at Ile^{137} in contrast to Asp^{134} in solution. The six helices are arranged in two triple-helix bundles, with H1–H3 in the R2 motif forming the first bundle and H4–H6 within the R3 motif forming the second bundle. The two triple-helix bundles connected by a 10-residue linker comprising residues Ala^{95}–Thr^{104} (designated the R2–R3 linker) assume a similar structure fold. The conserved aromatic groups are buried inside the hydrophobic cores of the helix bundles (Figure 4C and D). *tv*Myb2$_{40–156}$ interacts with DNA bases by positioning the third long helix of each motif, H3 and H6, on the major groove in a HTH fashion. The DNA binding pocket is lined with numerous positively charged residues, including residues Arg^{120}–Asn^{124} at the N-terminal side of H5 (Figure 4E and F). Helices H1, H2 and H4 make no contact with the DNA molecule

Figure 5A depicts molecular interactions between *tv*Myb2$_{40–156}$ and MRE-1-12. The interactions can be categorized into four types. First, there are the direct base interactions: Lys^{49} (NZ)-T$_{3'}$(O2) and T$_5$(O2), Arg^{84}-

Table 2. Statistics for data collection and structural refinement

Data collection	Myb2/MRE-1-12	Myb2/MRE-2-13	Se-Met-Myb2/MRE-2-13		
			Peak	Edge	Remote (high energy)
Wavelength (Å)	0.97622	0.97622	0.97884	0.97906	0.96859
Space group	$P2_1$	$P2_12_12_1$	$P2_12_12_1$		
Unit cell (Å)	40.1, 127.5, 41.4 $\beta = 100.2$	73.9, 77.3, 84.4		73.4, 77.1, 84.1	
Resolution range (Å)	30–2.0 (2.07–2.0)[a]	50–2.03 (2.07–2.03)[a]	30–2.24 (2.32–2.24)[a]	30–2.1 (2.18–2.1)[a]	30–2.17 (2.25–2.17)[a]
Unique reflections	26 911 (2556)[a]	31 586 (1442)[a]	23 625 (2241)[a]	28 343 (2739)[a]	25 835 (2519)[a]
Redundancy	3.5 (3.2)[a]	4.9 (4.7)[a]	4.7 (4.5)[a]	4.7 (4.5)[a]	4.7 (4.5)[a]
Completeness (%)	97.3 (93.7)[a]	99.1 (93.1)[a]	98.3 (95.7)[a]	98.5 (97.2)[a]	98.6 (98.0)[a]
$I/\sigma<I>$	22.8 (3.2)[a]	30.3 (4.7)[a]	17.7 (2.2)[a]	19.5 (2.4)[a]	19.9 (2.7)[a]
R_{merge} (%)[b]	5.0 (37.8)[a]	4.7 (26.9)[a]	6.7 (53.3)[a]	6.2 (47.4)[a]	6.3 (43.1)[a]
Refinement					
Resolution range (Å)	30–2.0 (2.07–2.0)[a]	50–2.03 (2.1–2.03)[a]			
Reflections ($F > 0$ σ_F)	26 037 (2191)[a]	31 435 (2910)[a]			
R_{cryst} (%) for 95% data	19.0 (23.9)[a]	20.6 (25.7)[a]			
R_{free} (%) for 5% data	23.3 (32.3)[a]	25.8 (31.6)[a]			
RMSD					
Bond lengths (Å)	0.008	0.007			
Bond angles (°)	1.40	1.39			
Average *B*-factors (Å^2)					
Protein	39.9	38.0			
DNA	29.8	37.9			
Water	36.7	42.2			
Isopropanol	–	42.9			

[a]Values in parentheses are for the highest resolution shell.
[b]$R_{merge} = \Sigma|I-<I>|/\Sigma|I|$.

$G_{5'}$(O6 and N7), Lys^{138}(NZ)-G_3(O6) and Asn^{139}(ND2)-$A_{2'}$(N7) (Figure 5B). These interactions are responsible for the sequence-specific recognition between *tv*Myb2$_{40–156}$ and MRE-1/MRE-2r. The interactions of Lys^{49} (NZ) with $T_{3'}$(O2) and T_5(O2), as well as Lys^{49}(N) and Gln^{50}(N) with the backbone phosphate groups stabilize the N-terminal loop. These interactions involving residues at the N-terminal side of H1 are unique to the *tv*Myb2 protein. Second are the solvent-mediated base interaction: Arg^{87}(NH2)-A_{-1}(N7), Asp^{88}(OD2)-$C_{4'}$(N4) and Lys^{138}(NZ)-$T_{3'}$(O4). These types of interaction may also contribute to specific base recognition. Third are the direct Coulomb interactions with backbone phosphate. Residues involved include Lys^{48}(N), Lys^{49}(N), Gln^{50}(N), Phe^{52}(N), Gln^{85}(NE2), Arg^{87}(NH1), Arg^{89}(NH1), Tyr^{93}(OH), Arg^{120}(NE, NH2), Trp^{122}(NE1), Ala^{123}(N) and Asn^{146}(ND2). These interactions do not contribute to sequence-specific recognition but may account for a significant fraction of the enthalpy gain. Fourth are the solvent-mediated interactions with backbone phosphate. Residues involved include Phe^{52}(O), Ser^{69}(O), Arg^{81}(NH1,2), Arg^{84}(NH1,2), Asp^{134}(OD1) and Asn^{139}(ND2). These interactions do not contribute to sequence-specific recognition either, they mainly contribute to binding energy, particularly the unfavorable entropy lost due to immobilization of bound water molecules.

Crystal structure of *tv*Myb2$_{40–156}$ in complex with MRE-2-13

The crystal structure of Se-Met-*tv*Myb2$_{40–156}$ in complex with MRE-2-13 was determined by the multiple-wavelength anomalous diffraction method and refining it to 2.03 Å resolution (Table 2). The protein was crystallized in $P2_12_12_1$ symmetry with each asymmetric unit also consisting of two *tv*Myb2$_{40–156}$/MRE-2-13 complexes with indistinguishable structures. Although the recognition sequence of MRE-2f (5′-TATCGT-3′) differs from that of MRE-1/MRE-2r (5′-ACGATA-3′), its complementary strand is identical to that of MRE-1/MRE-2r, and vice versa. Thus, the structures of *tv*Myb2$_{40–156}$ in the two complexes are practically identical, with an rmsd of 0.322 Å. Furthermore, the residues involved in specific DNA sequence recognition, including Lys^{49} are identical. However, there is an additional solvent-mediated base interaction, i.e. Lys^{51}(NZ)-C_7(O2) observed in this complex (Figure 5C and D), which together with the results of MRE-1-12 indicates that 5′-a/gACGAT-3′ is the specific base sequence recognized by *tv*Myb2$_{40–156}$.

Nonetheless, a closer inspection of the structures showed that *tv*Myb2$_{40–156}$ has more extensive interactions with MRE-2-13 than with MRE-1-12, consistent with the observed binding affinity by ITC. Three solvent-mediated interactions beyond the promoter region, namely Lys^{51}(NZ)-C_7(O2), Arg^{87}(NH2)-A_{-2}(N7) and Arg^{87}(NH2)-G_{-1}(N7) were observed in the MRE-2-13 complex, whereas only the Arg^{87}(NH2)-A_{-1}(N7) interaction was observed in MRE-1-12/*tv*Myb2$_{40–156}$ complex. The Asn^{139}(OD1)-$T_{3'}$(O4) interaction was replaced by the Lys^{138}(NZ)-$T_{3'}$(O4) interaction. Moreover, we observed five protein–DNA backbone phosphate interactions in the MRE-2-13 complex but only three in the MRE-1-12 complex.

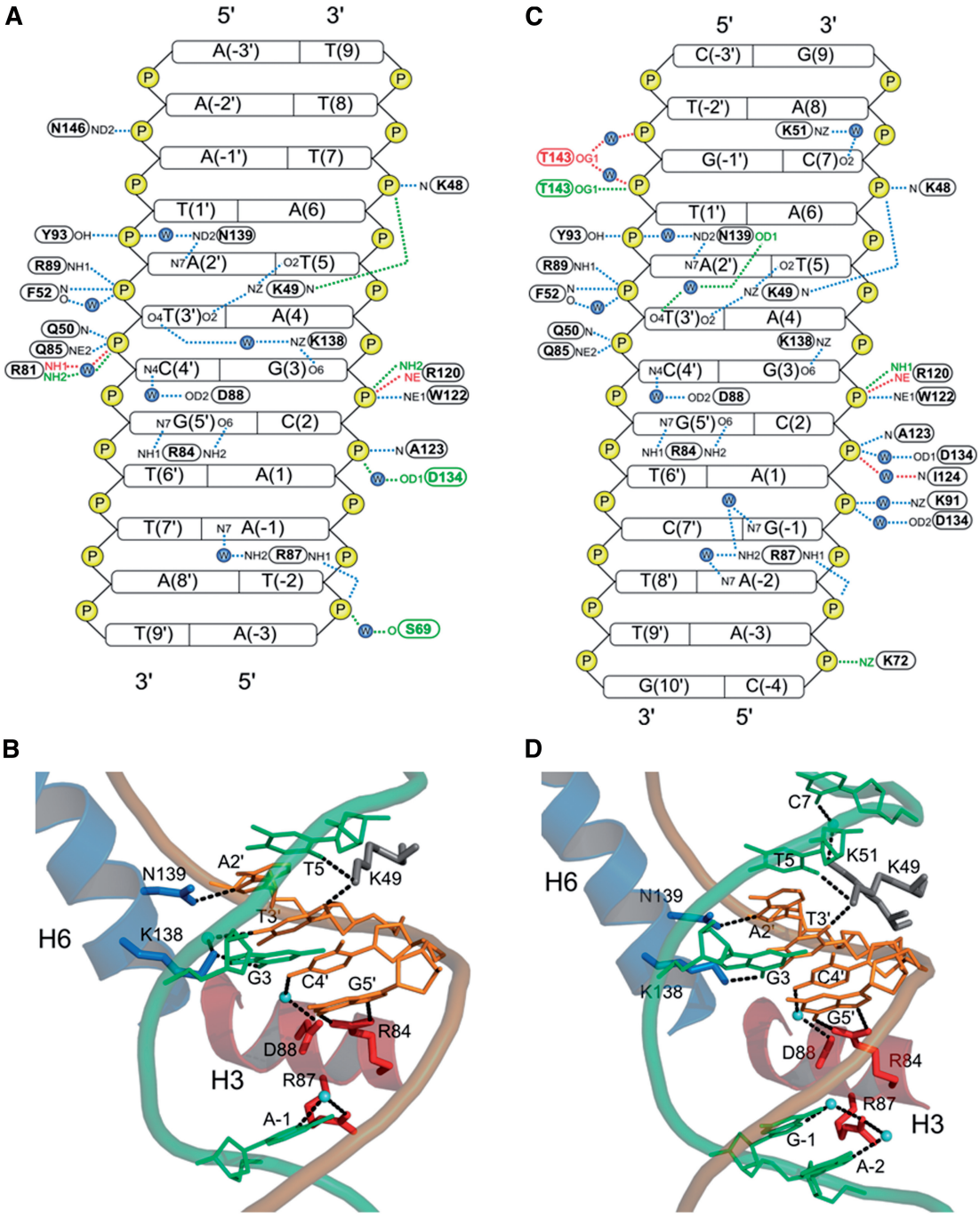

Figure 5. Schematic representation of the detailed interactions between *tv*Myb2$_{40–156}$ and MRE-1-12 (**A**) or MRE-2-13 (**C**). The corresponding pair-wise interactions of *tv*Myb2$_{40–156}$ with MRE-1-12 and MRE-2-13 are shown in (**B**) and (**D**), respectively. There are two monomers (a and b) in an asymmetric unit. The blue, red and green dotted lines indicate interactions observed in both monomers, monomer a and monomer b, respectively. W indicates a water molecule.

Energetic contribution to specific DNA sequence recognition of the base-recognition residues

To corroborate the structural studies and to further determine the energetic contribution of residues responsible for specific DNA sequence recognition we generated six alanine mutants (K49A, K51A, R84A, R87A, K138A and R139A) and determined their promoter DNA-binding affinities by ITC (Table 1). The mutation of any of these residues to alanine drastically reduced the binding affinity and in general the mutation has greater effect on the binding to MRE-1-20 than it does to MRE-2-20. The reduction in the binding affinity of mutant *tv*Myb2$_{40–156}$ follows the following order: R84A > K49A > R87A > N139A > K138A > K51A. In particular, the mutation of Arg^{84} to alanine completely abolished protein binding to MRE-1-20 and reduced the

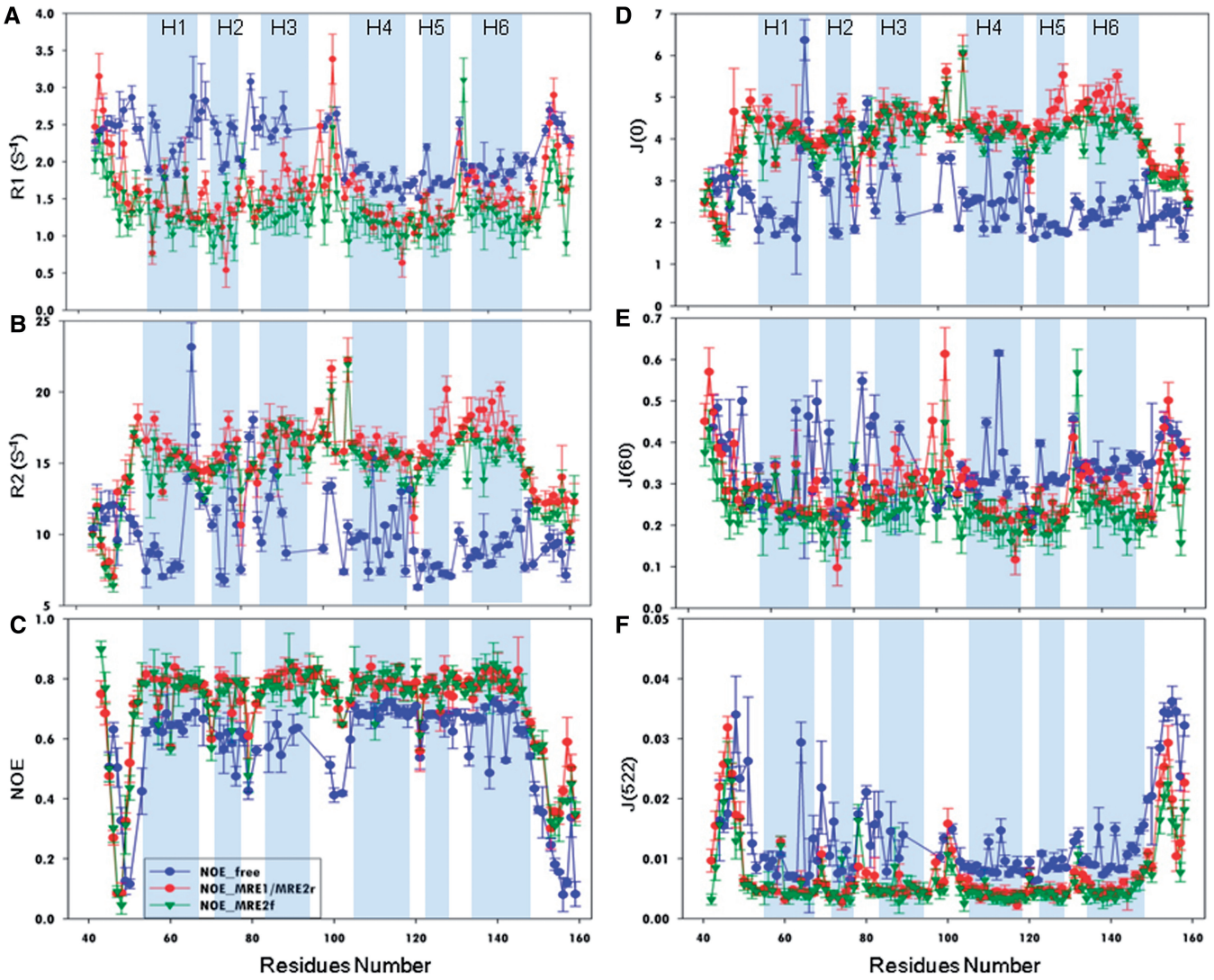

Figure 6. Sequence variation of ^{15}N-R1 (**A**), ^{15}N-R2 (**B**) and [^{1}H-^{15}N] NOE (**C**) of *tv*Myb2$_{40–156}$ (blue), the *tv*Myb2$_{40–156}$/MRE-1-12 complex (red) and the *tv*Myb2$_{40–156}$/MRE-2-13 complex (green). The calculated reduced spectral density functions J(0), J(60) and J(522) are shown in (**D–F**), respectively.

binding affinity to MRE-2-20 by a factor of 654. Arg^{84} interacts with the $T_{6'}$-A_1 and $G_{5'}$-C_2 base pairs. The corresponding residue, Lys^{128} in c-Myb is also the most important residue in promoter recognition as the K128A mutant has lost the ability to bind cognate DNA (49). The significant role of the N-terminal residue Lys^{49} not located on the third helix of each motif is surprising and is not previously reported in other Myb proteins (49).

The amide resonance of Phe^{52} experienced the largest chemical shift perturbation upon binding to the promoter elements with its amide nitrogen and carbonyl group contacting the phosphate group between A(2′) and T(3′). In c-Myb, the indole ring of the corresponding Trp^{95} significantly shifted toward the cavity of the R2 motif on DNA binding (49). We found that the F52A mutant reduced the binding affinity toward MRE-1-20 and MRE-2-20 by a factor of 279 and 123, respectively, corresponding to ΔΔGs of 3.68 and 2.82 kcal/mol. The drastic reduction in binding energy illustrates the pivotal role of this hydrophobic residue in the structural integrity and DNA binding of the Myb family of proteins.

Dynamics of tvMyb2$_{40–156}$

Figure 6A–C show the backbone amide ^{15}N-T_1, ^{15}N-T_2 and [^{1}H-^{15}N]-NOE of the DNA-free and DNA-bound *tv*Myb2$_{40–156}$ at 600 MHz. The spectral density functions, J(0), J(60) and J(522), extracted from the relaxation data represent the extent of motion at 0, 60 and 522 MHz, respectively (Figure 6D–F) (29,30,32). As expected, the N- and C-termini of the protein in all three forms are the most flexible regions. Surprisingly, the extreme terminal residues have J values close to those of the structured regions even in the DNA-free form, suggesting that these residues are partially immobilized, likely by

folding back to interact with the structured part. This is confirmed in the structure of *tv*Myb2$_{40–156}$/MRE-1-12 complex where the segment Val^{40}–Asn^{43} loops back, placing Val^{40} very close to Gln^{48}. However, the structure of the N-terminal segment cannot be traced in the structure of *tv*Myb2$_{40–156}$/MRE-2-13, nor are the C-terminal residues in both complexes, probably due to disorder as reflected in the low NOE values. For the DNA-free form the structured core region between Gln^{50}–Ile^{150} showed considerable variation in all J values. This is most prominent for residues in the region between Ala^{63}–Thr^{104}, which encompasses the C-terminal end of H1 in R2 to the beginning of H4 in R3. Furthermore, amide resonances of several residues in the H3 and the R2–R3 linker region were not observed, which further reflects the dynamic nature of this region. In comparison, residues in the R3 motif have lower but more uniform J values. Binding of the protein to DNA completely immobilizes the R2 motif such that the structured R2R3 region is now fully immobilized except for some of the residues in the loop regions connecting the helices.

For a globular protein, the rotational correlation time can be predicted from the empirical Stokes–Einstein relation, $\tau_c^{-1} = \kappa_B T/8\pi\eta r^3$, where κ_B is the Bolzmann's constant, T is the absolute temperature and η is solvent viscosity. The predicted τ_c values are ~8.0 and ~4.1 ns for *tv*Myb2$_{40–156}$ and the isolated R2 or R3 motif, respectively (50). For a rigid globular molecule, the overall rotational correlation time, τ_c, can be obtained from spectral density functions, $\tau_c^2 = \omega_N^{-2}[J(0) - J(\omega_N)]/J(\omega_N)$, as previously described (29,32). For DNA-free *tv*Myb2$_{40–156}$, we obtained a mean τ_c = 5.3 ns for the more rigid region, Thr^{53} to Ala^{63}, of the R2 motif and τ_c = 6.0 ns for the R3 motif. These values indicate that the R2 and R3 motifs move with a correlation time of a globular protein of ~78 and 87 residues, respectively. Thus, the two motifs move with a considerable degree of independence. In contrast, we obtained τ_c of 10.3 ns and 10.4 ns for *tv*Myb2$_{40–156}$/MRE-1-12 and *tv*Myb2$_{40–156}$/MRE-2-13, respectively, consistent with the predicted correlation times of their sizes. In addition, the two motifs in the DNA complexes move with the same rotational correlation times. Thus, *tv*Myb2$_{40–156}$/MRE-1-12 and *tv*Myb2$_{40–156}$/MRE-2-13 move as monomeric rigid bodies.

DISCUSSION

Solution structure of DNA-free Myb proteins

*tv*Myb2$_{40–156}$ shares 26% sequence identity with the R2R3 regions of vertebrate c-Myb and 40% sequence identity with the R2R3 region of *tv*Myb1 (13) (Figure 1B). The solved structures of DNA-free c-Myb (PDB ID: 1GV2) and DNA-free *tv*Myb1$_{35–141}$ (PDB ID: 2K9N) are highly homologous (an rmsd of 2.813 Å) (15,51). DNA binding induced a significant structural rearrangement in the relative orientation of R2 and R3 with little effect on the folding of the two motifs. In fact, when superimposed, the R2 domains of the DNA free and DNA-bound structures of c-Myb match quite well, but the H6 on R3 showed a ~90° rotation about the H3 (51). A similar change was observed in the DNA-free structure versus and the HADDOCK modeled DNA-bound structure of *tv*Myb1$_{35–141}$, but with a 50° reorientation of the two motifs (15).

We were unable to solve the solution structure of DNA-free *tv*Myb2$_{40–156}$ to high resolution by NMR due to insufficient NOE data, likely due to the high flexibility of the protein. To gain further insights on the structure transition upon DNA binding to *tv*Myb2$_{40–156}$, we have measured the ^{1}H-^{15}N RDC of DNA-free (u-^{15}N) *tv*Myb2$_{40–156}$ and (u-^{15}N)*tv*Myb2$_{40–156}$/MRE-1-12 complex. Totally, 59 measured $^1D_{NH}$ RDCs for the helical region of the *tv*Myb2$_{40–156}$/MRE-1-12 complex were fit against the calculated values from the crystal structure of the *tv*Myb2$_{40–156}$/MRE-1-12 complex (Figure 7A). A low value of 0.21 RDC quality factor (Q-factor) was obtained, suggesting that the solution structure of the DNA-bound *tv*Myb2$_{40–156}$ is very similar to that in the crystal state. For the DNA-free *tv*Myb2$_{40–156}$, the measured $^1D_{NH}$ values for the R2 motif (25 $^1D_{NH}$ RDCs) and R3 motif (32 $^1D_{NH}$ RDCs) individually fit well with the calculated values from the crystal structure of the individual motifs in the *tv*Myb2$_{40–156}$/MRE-1-12 complex (Q-factors of 0.23 and 0.32 for R2 and R3, respectively) (Figure 7B), suggesting that the structures of R2 and R3 are similar in the presence and absence of DNA. On the other hand, the measured $^1D_{NH}$ RDCs from R2 and R3 of *tv*Myb2$_{40–156}$ together fit poorly with the calculated RDCs from the crystal structure of the *tv*Myb2$_{40–156}$/MRE-1-12 complex (Q-factor = 1.04, Figure 7C). This suggests that the relative orientation of R2 and R3 in *tv*Myb2$_{40–156}$ changes substantially upon DNA binding. We also compared the measured RDC values with those predicted from two homology models of DNA-free *tv*Myb2$_{40–156}$, based on DNA-free c-Myb or DNA-free *tv*Myb1$_{35–141}$ (Figure 7D). The fit is also rather poor, suggesting that the structure of DNA-free *tv*Myb2$_{40–156}$ is very different from that of the other two DNA-free Myb proteins. Alternatively, because the DNA-free *tv*Myb2$_{40–156}$ is highly flexible, it is possible that the R2 and R3 motifs in DNA-free *tv*Myb2$_{40–156}$ are undergoing significant structural fluctuation so that the RDC values are further averaged.

To further corroborate the structural information, we conducted small angle X-ray scattering (SAXS) experiments to assess the structure of DNA-free *tv*Myb2$_{40–156}$. The radius of gyration (Rg) values, estimated from experimental curves using Guinier analysis (52), are 22.5 ± 0.09 Å and 18.6 ± 0.05 Å for DNA-free *tv*Myb2$_{40–156}$ and the *tv*Myb2$_{40–156}$/MRE-1-12 complex, respectively (Supplementary Figure S4). Thus, the R_g value of the DNA-free *tv*Myb2$_{40–156}$ is considerably larger than that in the DNA-bound form, suggesting that the DNA-free *tv*Myb2$_{40–156}$ assumes a much more open conformation. This conclusion is consistent with the poor fit of RDC data of DNA-free *tv*Myb2$_{40–156}$ with that predicted from the structures of *tv*Myb2$_{40–156}$/MRE-1-12.

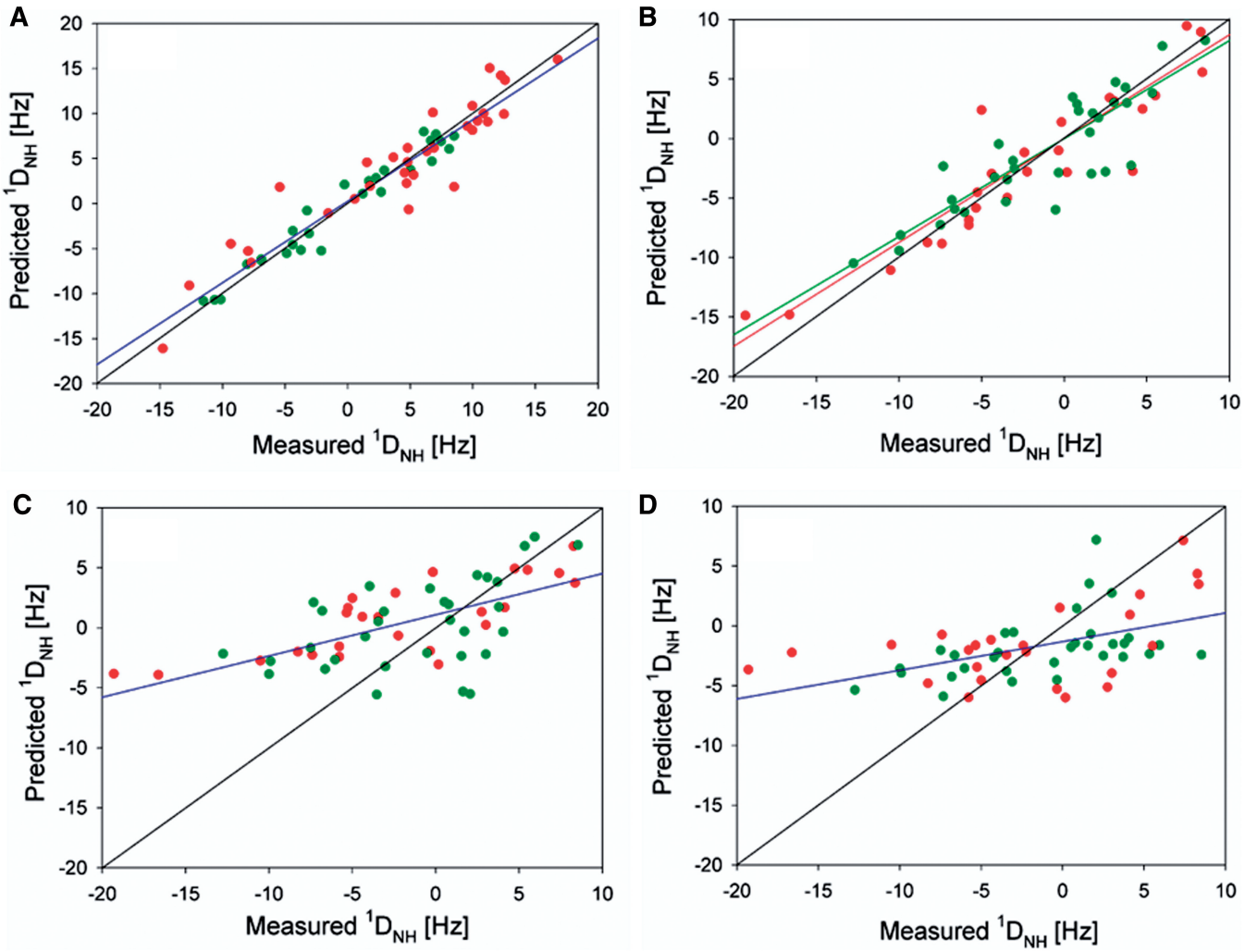

Figure 7. Structural assessment by ^1H-^{15}N RDC, $^1D_{NH}$. In all figures, the diagonal line represents the case for a perfect agreement between the experimental data and the predicted values. (**A**) Correlation between measured $^1D_{NH}$ values of u-^{15}N-*tv*Myb2$_{40–156}$/MRE-1-12 and values predicted from the X-ray structures of *tv*Myb2$_{40–156}$/MRE-1-12. The regression line is shown in blue (Q-factor = 0.21). (**B**) Correlation between measured $^1D_{NH}$ values of DNA-free u-^{15}N-*tv*Myb2$_{40–156}$ and values predicted from the X-ray structures of the R2 or R3 motifs in the X-ray structure of *tv*Myb2$_{40–156}$/MRE-1-12. Q-factor values are 0.23 and 0.32 for R2 and R3, respectively. (**C**) Correlation between measured $^1D_{NH}$ values of DNA-free u-^{15}N-*tv*Myb2$_{40–156}$ and values predicted from the X-ray structures of *tv*Myb2$_{40–156}$/MRE-1-12 (Q-factor = 1.04). (**D**) Correlation between measured $^1D_{NH}$ values of DNA-free u-^{15}N-*tv*Myb2$_{40–156}$ and values predicted from the homology structure of DNA-free *tv*Myb2$_{40–156}$ modeled with DNA-free *tv*Myb1$_{35–141}$ (PDB ID: 2K9N) (Q-factor = 1.03). The predicted $^1D_{NH}$ values were obtained by the PALES software (38). Data points are colored red for the residues located in the helices of the R2 motif and green for residues in the R3 motif.

Roles of protein flexibility in specific DNA sequence recognition in Myb proteins

Sarai *et al.* (53) have shown that the R2 motif of c-Myb is the least stable motifs with a thermal melting temperature of 43°C compared with 57°C for R3. The flexibility of the isolated motifs was further confirmed by NMR relaxation studies (3,51) (For a lack of better term, the word 'flexibility' here indicate a protein, or a portion of it, is not rigid and shows motion at various time scales as indicated by NMR relaxation parameters). However, the dynamics of the R2–R3 linker was not studied in c-Myb. Recently Lou *et al.* (15) also reported the presence of considerable flexibility in DNA-free *tv*Myb1$_{35–141}$, especially the R2 motif and the linker region. Here, we demonstrated that in the DNA-free state *tv*Myb2$_{40–156}$ is even more flexible than that of either c-Myb or *tv*Myb1$_{35–141}$, as indicated by the lower ^1H-^{15}NOE for a larger number of residues. Consistently, the linker region is the most flexible part of the molecule, and R2 is more flexible than R3. Upon binding to the cognate DNA, all three Myb proteins lost their flexibility.

The structure and dynamics information obtained so far suggest a unique molecular mechanism of Myb–DNA recognition. The DNA-binding pocket of the Myb proteins is lined with many positively charged residues that interact with numerous negatively charged backbone phosphate groups. Since Coulomb charge–charge interaction is a long range interaction, we suggest that the open and flexible structure of a Myb protein behaves much like an open net to catch the DNA molecule as the first step in DNA recognition. The intrinsic flexibility of the two

motifs and the linker region allow the protein to adapt to the optimal conformation in placing the H3 and H6 on the major groove in an induced-fit manner. Upon first contact, the properly folded protein could presumably slide along the DNA molecule until the conserved residues on the H3 and H6 make proper contact with the specific DNA bases without altering the conformation of the R2 and R3 motifs. Thus, the intrinsic flexibility of the Myb proteins is essential for DNA binding and sequence specific recognition in a manner resembling that proposed for the intrinsically disordered proteins (IDP) in molecular recognition (54–56). The importance of protein flexibility in DNA recognition has also been demonstrated recently in a papillomavirus E2 proteins (57).

Ogata attributed the source of the flexibility to the presence of a cavity in the center of the hydrophobic core of the R2 motif due to the presence of a small hydrophobic residue, Val^{103}. The presence of flexibility in R2 allows the side chain of Trp^{95} to rearrange its packing for optimal DNA binding (3,51,53). The flexibility can be greatly reduced upon mutating the small Val residue to a larger Leu. Sequence alignment shows that Val^{103} in c-Myb is substituted by a larger Leu in *tv*Myb2$_{40–156}$. However, such a substitution does not reduce the flexibility of the R2 motif in the DNA-free *tv*Myb2$_{40–156}$, perhaps due to the simultaneous substitution of Trp^{95} in c-Myb by Phe^{52} in *tv*Myb2$_{40–156}$. In c-Myb the indole ring is significantly shifted toward the cavity of R2 on DNA binding. Such a rearrangement appears to be hampered by the absence of the cavity in the V103L of c-Myb, but it may not be a problem when Trp is simultaneously mutated to a smaller Phe in *tv*Myb2. Nonetheless, the important role played by the hydrophobic aromatic residue, Trp^{95} in c-Myb, was substantiated by our F52A study in which mutation of Phe^{52} to Ala reduced the binding affinity by two orders of magnitude.

Specific DNA sequence recognition in Myb proteins

In spite of the low sequence identity, the tertiary folds of the DNA-bound forms of *tv*Myb2$_{40–156}$ and c-Myb (PDB ID: 1MSE) are very similar (an rmsd of 2.3 Å) (49). Here, we showed that *tv*Myb2$_{40–156}$ recognizes the a/gACGAT sequence, which does not fully conform to the MBS sequence recognized by c-Myb. Nonetheless, the four conserved residues in the H3 and H6, namely Arg^{84}, Asp^{88}, Lys^{138} and Asn^{139} (as the Lys^{128}, Glu^{132}, Lys^{182} and Asn^{183} in c-Myb) and two unique residues, the N-terminal Lys^{49} and Arg^{87} in H3, are the major contributors in *tv*Myb2$_{40–156}$ for recognizing the specific DNA sequence. In c-Myb, specific base sequence recognition involves four other non-conserved residues: Asn^{136}, Asn^{179}, Asn^{186} and Ser^{187}. The participation of residues outside of the R2R3 motifs in specific DNA sequence recognition is unique among the Myb proteins. It is likely that these unique base-recognition residues contribute to the differential recognition of the MBS by a particular Myb protein. The variation in recognition of MBS is conceivably important in transcription regulation of *T. vaginalis* since the parasite genome encodes more than 400 distinct Myb proteins (58).

Using the program HADDOCK (59,60), Lou *et al.* (15) also proposed a structural model of *tv*Myb1$_{35–141}$/MRE-1/MRE-2f complex, which revealed the binding region in *tv*Myb1$_{35–141}$ to be similar to that identified in the crystal structures of *tv*Myb2$_{40–156}$–DNA complex reported here as well as that in c-Myb (49). However, there are two significant differences. First, *tv*Myb1$_{35–141}$ and *tv*Myb2$_{40–156}$ bind to the promoter element in opposite orientations (Supplementary Figure S5). Specifically, in the *tv*Myb1$_{35–141}$–DNA complex, both the H3 and H6 align in parallel to the 5′-$A_1C_2G_3A_4T_5A_6$- 3′ orientation. In contrast, in *tv*Myb1$_{35–141}$/DNA complex the H3 and H6 are aligned anti-parallelly to the 5′-$A_1C_2G_3A_4$-3′ sequence orientation. As a result, *tv*Myb1$_{35–141}$ and *tv*Myb2$_{40–156}$ use different residues to recognize the same bases. Second, in *tv*Myb2$_{40–156}$ the N-terminal residues, Lys^{49} and Lys^{51}, participate in DNA-specific sequence recognition but no interaction between the corresponding Lys residues with promoter DNA was observed in the *tv*Myb1$_{35–141}$/DNA complex. The HADDOCK structure was generated based on RDC, chemical shift perturbation and DNA specificity data, but without specific distance constraints, such as NOEs. It is known that two pairs of ^{1}H-^{15}N dipoles in opposite orientation have the same RDC and that chemical shift perturbation can only locate the binding site, but not the specific interacting pairs. Thus, two HADDOCK structures that bind to DNA in opposite orientations may both satisfy the HADDOCK constraints. In spite of the 40% sequence homology between *tv*Myb1$_{35–141}$ and *tv*Myb2$_{40–156}$, it is probable that *tv*Myb1 and *tv*Myb2 may bind to the MRE-1/MRE-2r in opposite orientations (21). This difference may have biological implications that are yet to be exploited. Intriguingly, examination of the structures shown on Figure 5A and Supplementary Figure S5A, revealed that DNA sequence-specific recognition of *tv*Myb1 involves two Asn residues (Asn^{110} with A2′ and Asn^{122} with C4′) that are not conserved and not involved in DNA sequence-specific recognition in *tv*Myb2 (The corresponding residues in *tv*Myb2 are Ala^{123} and Ile^{135}, respectively) (Figure 1B). It is tempting to speculate that these specific recognition pairs might affect the binding orientation.

Biological implications

The genome of *T. vaginalis* encodes more than 400 Myb proteins, most of which share conserved R2R3 Myb domains with highly variable N- and C-termini (58). It is conceivable that they are the major regulators of gene transcription in *T. vaginalis*. To better understand the transcription regulation of the parasite, it is desirable to know how myriad *tv*Myb proteins select their target genes. Intriguingly, the DNA-binding specificity and nuclear translocation of *tv*Myb2 rely on a contiguous region spanning the entire R2R3 domain. A single point mutation at Ile^{74} to Ala, which slightly changes the helical content and ternary folding of the R2R3 domain resulted in the inhibition of *tv*Myb2 nuclear import and

loss of DNA-binding activity. This suggests that the R2R3 domain is likely a common module for myriad Myb proteins in the parasite to regulate both DNA-binding specificity and nuclear import (Tai, J.H., unpublished data). Moreover, the nuclear import of Myb2 could be greatly facilitated in the presence of hydrogen peroxide (Tai, J.H., unpublished observation), suggesting an important role of this transcription factor in protecting the parasite from oxidative stress produced from host-immune surveillance. Consistent with this speculation, the Myb2 recognition sequence, MRE-1/MRE-2r, could be found in the promoter regions of several antioxidant genes, such as the genes encoding superoxide dismutase (accession number Gi123485668) and thioredoxin peroxidase (accession number Gi123459140). It is therefore important to study the structural basis of *tv*Myb2 in both DNA-free form and DNA-bound forms.

In summary, the structures of $tv\text{Myb2}_{40–156}$ in free form and in DNA-bound form were elucidated in this report, which reveals conserved and unique features of a particular *tv*Myb contributing to its recognition of specific DNA sequences. These characteristics are conceivably important in transcription regulation of *T. vaginalis* since the parasite genome encodes more than 400 distinct Myb proteins (58). Information derived from current study expands our understanding of the molecular basis of Myb protein function in general and for further study on the structure-functional relationship of Myb-regulated gene transcription in *T. vaginalis* in particular.

ACCESSION NUMBERS

3OSF, 3OSG, BMRB accession no. 17170.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Dr Yu-Shan Huang and Dr U-Ser Jeng of the National Synchrotron Radiation Research Center in Taiwan for helping in small angle X-ray diffraction data collection and analysis. We acknowledge the use of synchrotron radiation beamline 13B1 and 13C1 at the National Synchrotron Radiation Research Center, Taiwan and the staff of beamline BL44XU at SPring-8, Japan, for their help in X-ray crystal data collection. The NMR spectra were obtained at the High-Field Nuclear Magnetic Resonance Center, Taiwan, which is supported by the National Research Program for Genomic Medicine, the Republic of China.

FUNDING

The National Science Council (NSC grant 99-2119-M-001-004); National Health Research Institute of the Republic of China (grant NHRI-EX99-9933B) to T.H.H. Funding for open access charge: Institute of Biomedical Sciences, Academia Sinica and National Science Council, The Republic of China.

Conflict of interest statement. None declared.

REFERENCES

1. Ramsay,R.G. and Gonda,T.J. (2008) MYB function in normal and cancer cells. *Nat. Rev. Cancer*, **8**, 523–534.
2. 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.
3. Ogata,K., Kanei-Ishii,C., Sasaki,M., Hatanaka,H., Nagadoi,A., Enari,M., Nakamura,H., Nishimura,Y., Ishii,S. and Sarai,A. (1996) The cavity in the hydrophobic core of Myb DNA-binding domain is reserved for DNA recognition and trans-activation. *Nat. Struct. Mol. Biol.*, **3**, 178–187.
4. Biedenkapp,H., Borgmeyer,U., Sippel,A.E. and Klempnauer,K.-H. (1988) Viral myb oncogene encodes a sequence-specific DNA-binding activity. *Nature*, **335**, 835–837.
5. Jia,L., Clegg,M.T. and Jiang,T. (2004) Evolutionary dynamics of the DNA-binding domains in putative R2R3-MYB genes identified from rice subspecies indica and japonica genomes. *Plant Physiol.*, **134**, 575–585.
6. Sorvillo,F., Smith,L., Kerndt,P. and Ash,L. (2001) Trichomonas vaginalis, HIV, and African-Americans. *Emerg. Infect. Dis.*, **7**, 927–932.
7. Alderete,J.F. and Garza,G.E. (1988) Identification and properties of *Trichomonas vaginalis* proteins involved in cytoadherence. *Infect. Immun.*, **56**, 28–33.
8. 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.
9. Arroyo,R., Engbring,J. and Alderete,J.F. (1992) Molecular basis of host epithelial cell recognition by Trichomonas vaginalis. *Mol. Microbiol.*, **6**, 853–862.
10. Tsai,C.-D., Liu,H.-W. and Tai,J.-H. (2002) Characterization of an iron-responsive promoter in the protozoan pPathogen Trichomonas vaginalis. *J. Biol. Chem.*, **277**, 5153–5162.
11. 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.
12. 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.
13. Ong,S.J., Hsu,H.M., Liu,H.M., 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.
14. 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.
15. 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.
16. Gardner,K.H. and Kay,L.E. (1998) The use of ^{2}H, ^{13}C, ^{15}N multidimensional NMR to study the structure and dynamics of proteins. *Annu. Rev. Biophys. Biomol. Struct.*, **27**, 357–406.
17. Kanelis,V., Forman-Kay,J.D. and Kay,L.E. (2001) Multidimensional NMR methods for protein structure determination. *IUBMB Life*, **52**, 291–302.
18. Bax,A. (1994) Multidimensional nuclear magnetic resonance methods for protein studies. *Curr. Opin. Struct. Biol.*, **4**, 738–744.

19. Cavanagh,J., Fairbrother,W.J., Palmer,A.G. and Skelton,N.J. (1996) *Protein NMR Spectroscopy - Principles and Practice*. Academic Press, San Diego.
20. Sattler,M., Schleucher,J. and Griesinger,C. (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. Nucl. Mag. Res. Sp.*, **34**, 93–158.
21. Bax,A., Vuister,G.W., Grzesiek,S., Delaglio,F., Wang,A.C., Tschudin,R. and Zhu,G. (1994) In Thomas,L. and James,N.J.O. (eds), *Methods in Enzymology*, Vol. 239. Elsevier, Waltham, MA, USA, pp. 79–105.
22. Lyons,B.A., Tashiro,M., Cedergren,L., Bilsson,B. and Montelione,G.T. (1993) An improved strategy for determining resonance assignments for isotopically enriched proteins and its application to an engineered domain of staphylococcal protein A. *Biochemistry*, **32**, 7839–7845.
23. Montelione,G.T., Lyons,B.A., Emerson,S.D. and Tashiro,M. (1992) An efficient triple resonance experiment using carbon-13 isotropic mixing for determining sequence-specific resonance assignments of isotopically-enriched proteins. *J. Am. Chem. Soc.*, **114**, 10974–10975.
24. Logan,T.M., Olejniczak,E.T., Xu,R.X. and Fesik,S.W. (1992) Side chain and backbone assignments in isotopically labeled proteins from two heteronuclear triple resonance experiments. *FEBS Lett.*, **314**, 413–418.
25. Yamazaki,T., Forman-Kay,J.D. and Kay,L.E. (1993) Two-dimensional NMR experiments for correlating carbon-13.beta. and proton.delta./.epsilon. chemical shifts of aromatic residues in 13C-labeled proteins via scalar couplings. *J. Am. Chem. Soc.*, **115**, 11054–11055.
26. Markley,J.L., Bax,A., Arata,Y., Hilbers,C.W., Kaptein,R., Sykes,B., Wright,P.E. and Wuthrich,K. (1998) Recommendation of the presentation of NMR stuctures of proteins and nucleic acids. IUPAC-IUBMB-IUPAB Inter-Union Task Group on Standardization of data bases of protein and nucleic acid structure determined by NMR spectroscopy. *Eur. J Biochem.*, **256**, 1–15.
27. Small,J. (1983) Sparky. *Geriatr. Nurs.*, **4**, 166.
28. Kay,L.E., Nicholson,L., Delaglio,F., Bax,A. and Torchia,D.A. (1992) Pulse sequences for removal of the effects of cross correlation between dipolar and chemical-shift anisotropy relaxation mechanisms on the measurement of heteronuclear T1 and T2 values in proteins. *J. Magn. Res.*, **97**, 359–375.
29. Chang,C.-F., Chou,H.-T., Chuang,J.L., Chuang,D.T. and Huang,T.-h. (2002) Solution structure and dynamics of the lipoic acid-bearing domain of human mitochondrial branched-chain alpha -keto acid dehydrogenase complex. *J. Biol. Chem.*, **277**, 15865–15873.
30. Chang,C.-F., Chou,H.-T., Lin,Y.-J., Lee,S.-J., Chuang,J.L., Chuang,D.T. and Huang,T.-h. (2006) Structure of the subunit binding domain and dynamics of the di-domain region from the core of human branched chain alpha-ketoacid dehydrogenase complex. *J. Biol. Chem.*, **281**, 28345–28353.
31. Farrow,N.A., Zhang,O., Szabo,A., Torchia,D.A. and Kay,L.E. (1995) Spectral density mapping using 15N relaxation data exclusively. *J. Biomol. NMR*, **6**, 153–162.
32. Lefevre,L.-F., Dayie,K.T., Peng,J.W. and Wagner,G. (1996) Internal mobility in the partially folded DNA binding and dimerization domains of GAL4: NMR analysis of the N-H spectral density functions. *Biochemistry*, **35**, 2674–2686.
33. Ishima,R. and Nagayama,K. (1995) Quasi-spectral density function analysis for nitrogen-15 nuclei inproteins. *J. Magn. Res.*, **B 108**, 73–76.
34. Hiyama,Y., Niu,C.H., Silverston,J.V., Bavavo,A. and Torchia,D.A. (1988) Determination of ^{15}N chemical shift tensor *via* ^{15}N-^{2}H dipolar coupling in Bocglycylglycyl[^{15}N] glycine benzyl ester. *J. Am. Chem. Soc.*, **110**, 2378–2383.
35. Bracken,C., Carr,P.A., Cavanagh,J. and Palmer,A.G. (1999) Temperature dependence of intramolecular dynamics of the basic leucine zipper of GCN4: implications for the entropy of association with DNA. *J. Mol. Biol.*, **285**, 2133–2146.
36. Rückert,M. and Otting,G. (2000) Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments. *J. Am. Chem. Soc.*, **122**, 7793–7797.
37. Cordier,F., Dingley,A.J. and Grzesiek,S. (1999) A doublet-separated sensitivity-enhanced HSQC for the determination of scalar and dipolar one-bond J-couplings. *J. Biomol. NMR*, **13**, 175–180.
38. Zweckstetter,M. (2008) NMR: prediction of molecular alignment from structure using the PALES software. *Nat. Protocols*, **3**, 679–690.
39. Otwinowski,Z. and Minor,W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Macromol. Crystallogr. PtA*, **276**, 307–326.
40. Terwilliger,T.C. (2000) Automated structure solution, density modification and model building. *Acta. Crystallogr. D Biol. Crystallogr.*, **58**, 1937–1940.
41. McCoy,A.J., Grosse-kunstleve,R.W., Adams,P.D., Winn,M.D., Storoni,L.C. and Read,R.J. (2007) Phaser crystallographic software. *J. Appl. Cryst.*, **40**, 658–674.
42. Adams,P.D., Afonine,P.V., Bunkoczi,G., Chen,V.B., Davis,I.W., Echols,N., Headd,J.J., Hung,L.W., Kapral,G.J., Grosse-Kunstleve,R.W. *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta. Crystallogr. D Biol. Crystallogr.*, **66**, 213–221.
43. Emsley,P. and Cowtan,K. (2004) Coot: model-building tools for molecular graphics. *Acta. Crystallogr. D Biol. Crystallogr.*, **60**, 2126–2132.
44. Davis,I.W., Leaver-Fay,A., Chen,V.B., Block,J.N., Kapral,G.J., Wang,X., Murray,L.W., Arendall,W.B.I., Snoeyink,J., Richardson,J.S. *et al.* (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.*, **35**, W375–W383.
45. Chen,V.B., Arendall,W.B.I., Headd,J.J., Keedy,D.A., Immormino,R.M., Kapral,G.J., Murray,L.W., Richardson,J.S. and Richardson,D.C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta. Crystallographica. D: Biol. Crystallogr.*, **D66**, 12–21.
46. Romero,P., Obradovic,Z. and Dunker,A.K. (1997) Identifying disordered regions in proteins from amino acid sequences. *Proc. I.E.E.E. Internat'l Conf. Neural Networks*, **1**, 90–95.
47. Wishart,D.S. and Sykes,B.D. (1994) Chemical shifts as a tool for structural determination. *Methods Enzymol.*, **239**, 363–392.
48. Ayed,A., Mulder,F.A.A., Yi,G.-S., Lu,Y., Kay,L.E. and Arrowsmith,C.H. (2001) Latent and active p53 are identical in conformation. *Nat. Struct. Mol. Biol.*, **8**, 756–760.
49. 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.
50. Bae,S.-H., Dyson,H.J. and Wright,P.E. (2009) Prediction of the rotational tumbling time for proteins with disordered segments. *J. Am. Chem. Soc.*, **131**, 6814–6821.
51. Ogata,K., Morikawa,S., Nakamura,H., Hojo,H., Yoshimura,S., Zhang,R., Aimoto,S., Ametani,Y., Hirata,Z., Sarai,A. *et al.* (1995) Comparison of the free and DNA-complexed forms of the DMA-binding domain from c-Myb. *Nat. Struct. Mol. Biol.*, **2**, 309–320.
52. Konarev,P.V., Volkov,V.V., Sokolova,A.V., Koch,M.H.J. and Svergun,D.I. (2003) PRIMUS - a Windows-PC based system for small-angle scattering data analysis. *J. Appl. Cryst.*, **36**, 1277–1282.
53. Sarai,A., Uedaira,H., Morii,H., Yasukawa,T., Ogata,K., Nishimura,Y. and Ishii,S. (1993) Thermal stability of the DNA-binding domain of the Myb oncoprotein. *Biochemistry*, **32**, 7759–7764.
54. Dunker,A.K., Cortese,M.S., Romero,P., Iakoucheva,L.M. and Uversky,V.N. (2005) Flexible nets. The roles of intrinsic disorder in protein interaction networks. *FEBS J.*, **272**, 5129–5148.
55. Dyson,H.J. and Wright,P.E. (2005) Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.*, **6**, 197–208.
56. Tompa,P. (2005) The interplay between structure and function in intrinsically unstructured proteins. *FEBS Lett.*, **579**, 3346–3354.
57. Brown,C., Campos-León,K., Strickland,M., Williams,C., Fairweather,V., Brady,R.L., Crump,M.P. and Gaston,K. (2011) Protein flexibility directs DNA recognition by the papillomavirus E2 proteins. *Nucleic Acids Res.*, **39**, 2969–2980.

58. Carlton,J.M., Hirt,R.P., Silva,J.C., Delcher,A.L., Schatz,M., Zhao,Q., Wortman,J.R., Shelby,L.B., Alsmark,U.C.M., Besteiro,S. *et al*. (2007) Draft genome sequence of the sexually transmitted pathogen Trichomonas vaginalis. *Science*, **315**, 207–212.
59. Dominguez,C., Boelens,R. and Bonvin,A.M. (2003) HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.*, **125**, 1731–1737.
60. van Dijk,M., van Dijk,A.D., Hsu,V., Boelens,R. and Bonvin,A.M. (2006) Information-driven protein-DNA docking using HADDOCK: it is a matter of flexibility. *Nucleic Acids Res.*, **34**, 3317–3325.