Protein Structure and Folding:
A Structural Analysis of DNA Binding by Myelin Transcription Factor 1 Double Zinc Fingers


doi: 10.1074/jbc.M113.482075 originally published online October 6, 2013

Access the most updated version of this article at doi: 10.1074/jbc.M113.482075

Find articles, minireviews, Reflections and Classics on similar topics on the JBC Affinity Sites.

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC’s e-mail alerts

This article cites 35 references, 10 of which can be accessed free at http://www.jbc.org/content/288/49/35180.full.html#ref-list-1
A Structural Analysis of DNA Binding by Myelin Transcription Factor 1 Double Zinc Fingers*

Roland Gamsjaeger,1† Michell R. O’Connell,‡,§ Liza Cubeddu,‡ Nicholas E. Shepherd,† Jason A. Lowry,† Ann H. Kwan,† Marylene Van De Venn€e,†,§ Michael K. Swanton,† Jacqueline M. Matthews,†,‡ and Joel P. Mackay,†3,4

Background: Myelin transcription factor 1 (MyT1) contains seven similar zinc finger domains that bind DNA specifically. DNA-binding residues are conserved among all MyT1 zinc fingers, suggesting an identical DNA binding mode. Significance: Determination of the molecular details of DNA interaction will be crucial in understanding MyT1 function.

Results: A three-dimensional structural model explains how a double zinc finger unit is able to recognize DNA. A three-dimensional structural model explains how a double zinc finger unit is able to recognize DNA. Myelin transcription factor 1 (MyT1/NZF2), a member of the neural zinc-finger (NZF) protein family, is a transcription factor that plays a central role in the developing central nervous system. It has also recently been shown that, in combination with two other transcription factors, the highly similar paralog MyT1L is able to direct the differentiation of murine and human stem cells into functional neurons. MyT1 contains seven zinc fingers (ZF$s$) that are highly conserved throughout the protein and throughout the NZF family. We recently presented a model for the interaction of the fifth ZF of MyT1 with a DNA sequence derived from the promoter of the retinoic acid receptor (RAR$€$E$€$)$€$ element (see Fig. 1), although a second isoform exists that lacks finger 1 (F1) (1). MyT1 was first discovered through its ability to bind to sites in the proteolipid protein promoter (2). This protein (3) plays a major role in the structure and compaction of the myelin sheath that is located around the axons of the central nervous system. MyT1 has also been shown to promote commitment to a neuronal fate in Xenopus laevis (4) and, more recently, has been found to interact with Sin3B, a transcriptional coregulator that mediates transcriptional repression by recruiting histone deacetylases (5).

In humans, there are two paralogs of MyT1: MyT1-like (MyT1L/NZF1) (6) and suppressor of tumorigenicity 18 (ST18/NZF3) (7). NZF3 has been shown to be a breast cancer tumor suppressor gene (8) and has also been implicated in the regulation of mRNA levels of proapoptotic and proinflammatory genes in fibroblasts (9). The other paralog, MyT1L, has recently attracted substantial interest due to its ability to act in concert with two other transcription factors (Ascl1 and Brnd2) to transform mouse as well as human stem cells directly into functional neurons (10–12). The resulting neuronal cells displayed functional properties such as the generation of trains of action potentials and synapse formation, properties that might enable them to be used for applications in neurological disease modeling or regenerative medicine. However, the molecular mechanisms through which these three transcription factors, in particular MyT1L, act are not well understood.

Several studies have shown that MyT1-type ZFs (from MyT1, MyT1L, and ST18) are able to recognize DNA in a sequence-specific manner (4, 7, 13). The motif AAAGTT (the retinoic acid receptor element, or RARE), which is found in the human proteolipid protein promoter and is the core sequence in the cis-regulatory element of the retinoic acid receptor gene, has been identified as the consensus-binding sequence for MyT1.

Myelin transcription factor 1 (MyT1,$5$ or neural zinc finger 2 (NZF2)) is a transcription factor that contains seven zinc finger (ZF) modules. These ZFs all contain a C2HC arrangement of zinc ligands and are located in the protein in a 1 + 2 + 4 topology (see Fig. 1), although a second isoform exists that lacks finger 1 (F1) (1). MyT1 was first discovered through its ability to bind to sites in the proteolipid protein promoter (2). This protein (3) plays a major role in the structure and compaction of the myelin sheath that is located around the axons of the central nervous system. MyT1 has also been shown to promote commitment to a neuronal fate in Xenopus laevis (4) and, more recently, has been found to interact with Sin3B, a transcriptional coregulator that mediates transcriptional repression by recruiting histone deacetylases (5).

In humans, there are two paralogs of MyT1: MyT1-like (MyT1L/NZF1) (6) and suppressor of tumorigenicity 18 (ST18/NZF3) (7). NZF3 has been shown to be a breast cancer tumor suppressor gene (8) and has also been implicated in the regulation of mRNA levels of proapoptotic and proinflammatory genes in fibroblasts (9). The other paralog, MyT1L, has recently attracted substantial interest due to its ability to act in concert with two other transcription factors (Ascl1 and Brnd2) to transform mouse as well as human stem cells directly into functional neurons (10–12). The resulting neuronal cells displayed functional properties such as the generation of trains of action potentials and synapse formation, properties that might enable them to be used for applications in neurological disease modeling or regenerative medicine. However, the molecular mechanisms through which these three transcription factors, in particular MyT1L, act are not well understood.

Several studies have shown that MyT1-type ZFs (from MyT1, MyT1L, and ST18) are able to recognize DNA in a sequence-specific manner (4, 7, 13). The motif AAAGTT (the retinoic acid receptor element, or RARE), which is found in the human proteolipid protein promoter and is the core sequence in the cis-regulatory element of the retinoic acid receptor gene, has been identified as the consensus-binding sequence for MyT1.
We recently assessed the binding of MyT1 to DNA and showed that the fifth zinc finger can fit into the DNA major groove and make contacts with the central AGT of the consensus sequence (14). Our data also indicated that double-finger constructs of both the two-ZF and the four-ZF clusters (F1F2, F4F5, F5F6, and F6F7) interact with the full AAGTTC consensus site with a higher affinity than does a single finger.

In this study, we have examined the DNA binding properties of a double-finger polypeptide consisting of MyT1 fingers 4 and 5 and calculated a data-driven structural model using a combination of NMR and SPR affinity data. We reveal that the full AAGTTC site can accommodate both fingers and that finger 4 contacts the DNA sequence in an orientation that differs by 180° from that observed for finger 5, consistent with the partially palindromic nature of the DNA site. Our NMR data also show that this binding mode is conserved among other MyT1 double ZF constructs, indicating that the full-length protein might be able to recognize two or three distinct AAGTTC sites in vivo.

**EXPERIMENTAL PROCEDURES**

**Subcloning, Expression, and Purification of MyT1 Constructs—**
The original plasmid encoding mouse 6-ZF myelin transcription factor 1 (mMyT1) was a gift of Dr Lynn Hudson (National Institutes of Health). Both F4F5 and F5F6 constructs of MyT1 (see Fig. 1) were cloned from the original plasmid (residues 18–904), and mutants were constructed using either overlap extension PCR or site-directed mutagenesis. All constructs were cloned into the pGEX-6P vector and overexpressed as GST fusions at 37 °C under standard conditions; isotopically labeled proteins were overexpressed using the protocol described previously (15). Proteins were purified using GSH affinity chromatography, HRV-3C cleavage, and gel filtration (Superdex-75 in SPR buffer: 50 mM NaCl, 10 mM HEPES, 1 mM DTT, pH 7.2). Protein concentrations were determined by absorbance at 215, 225, and 280 nm (16). Fractions were stored in the presence of protease inhibitors at −20 °C until required.

**β-RARE DNA and Mutant Oligonucleotides—**Single-stranded β-RARE DNA (5'-ACCGAAGTTCAC and 5'-GTGAACCTTT-CGGT), mutant oligonucleotides and biotinylated DNA for SPR experiments were obtained from Sigma-Aldrich, annealed in 20 °C water until required, and purified using gel filtration (Superdex-75). Concentrations were calculated from absorbance at 260 nm.

**Surface Plasmon Resonance—**All experiments were performed on a Biacore 3000 system (Biacore AB) at flow rates of 20 μl/min in SPR buffer to which was added 0.01% polyborate 20 (P20) detergent. Biotinylated DNA (~10–100 nM) was immobilized on streptavidin-coated Biacore SA chips (50–100 resonance units). MyT1 and MyT1 mutants (0.2–10 μM) were injected in SPR buffer, and binding was monitored. The system was washed with 1 M NaCl (1 min) after each experiment. For kinetics studies, the Biacore BiaEvaluation software was utilized to calculate affinity constants using global fitting algorithms. In the competition experiments, F5F6 (5 μM) was added to prebound DNA in the presence of 5 nM eq of competitor DNA oligonucleotides.

**NMR Spectroscopy—**F4F5 or F5F6 (unlabeled, 15N-labeled, or 15N/13C-labeled) were exchanged into NMR buffer (50 mM NaCl, 10 mM phosphate, 1 mM DTT, pH 7.2) with 1 mM DTT and concentrated in Microspec 3K cutoff filters to 200–1000 μM. Resonance assignments were made from standard triple-resonance experiments that were acquired at 25 °C on Bruker Avance 600 and 800 NMR spectrometers equipped with cryoprobes. 15N HSQC titrations as well as two-dimensional NOESY experiments of proteins with β-RARE DNA were carried out in NMR buffer at 25 °C. Chemical shift changes were calculated as a weighted average of HN, N, and Ca changes, using a previously reported equation (17, 18). Assignments of the DNA alone were obtained from our previous study (14). One-bond HN residual dipolar couplings (RDCs) were recorded for the F4F5-DNA complex in NMR buffer containing 22.2 mg/ml Pf1 phage (ASLA Biotech), using the in-phase/anti-phase pulse sequence (19). Alignment was assessed by measuring the D2O splitting (19 Hz). The program PALES (20) was used for the calculation of the magnitude and orientation of the sterically induced alignment tensor (see below for details). NMR data were processed using Topspin (Bruker, Karlsruhe) and analyzed with SPARKY 3 (37).

**HADDOCK Docking—**F4F5 was docked to the DNA using the program HADDOCK 1.3 (21–23). The starting structure for the DNA was a B-form model of the double helix DNA fragment (5'-ACCGAAGTTCAC and 5'-GTGAACCTTT-CGGT) constructed with the Nucleic Acid Builder package (24). Based on our NMR data (see Figs. 2 and 3), a starting structure of F4F5 was made in silico by fusing two individual ZF domains and the native linker sequence (see Fig. 1) together using the calculated NMR structure of F5 (14) (Protein Data Bank (PDB) ID 2JYD) as a template. A total of 10 different starting orientations between F4F5 and the DNA were chosen as starting structures for the docking. Sequences shown to be disordered in our previous NMR analysis, namely residues 799–800 (N-terminal) and 872–873 (C-terminal), were defined as fully flexible during the calculations, as was the internal linker (828–845) (14). Ambiguous interaction restraints for both the protein and the DNA were chosen based both on our NMR data from Figs. 2 and 3 and on solvent accessibility (>30%, determined by the program MOLombok) and fixed at 2 Å. For the DNA fragment, ambiguous interaction restraints were defined solely from the unique base atoms of bases Ade6, Ade7, Thy20, and Thy21, whereas for F5, DNA bases Thy9, Thy10, Ade17, and Ade18 were selected. For the protein, restraints between unique side-chain atoms of F4 (residues His-812, Tyr-817, Ser-819, Arg-821, Ser-822, Leu-823, Ser-824) as well as corresponding residues in F5 (His-856, Tyr-861, Ser-863, Arg-865, Ser-866, Leu-867, Ser-868) were chosen. A total of 48 ambiguous interaction restraints resulted from these definitions and were used as input into HADDOCK for all 10 different F4F5-DNA starting configurations. Additional restraints to maintain base planarity and Watson-Crick bonds for the DNA, intramolecular non-crystallographic symmetry restraints between F4F5 and DNA (F4+Ade6/7 = F5+Ade17/18 and F4+Thy20/21 = F5+Thy9/10), and zinc-coordinating restraints for F4F5 were introduced. During the rigid body energy minimization, 1000 structures were calculated, and the 200 best solutions based on the inter-
molecular energy were used for the semiflexible, simulated annealing. 10 different runs were carried out with the 10 F4F5-DNA starting orientations, respectively. The best 10 structures of each run were part of the lowest energy cluster (cut-off of 0.5 Å root mean square deviation (RMSD) based on the pairwise backbone RMSD matrix).

The 10 best structures from run I were subjected to a second round of semiflexible annealing following the inclusion of 43 HN RDCs as additional direct restraints (using the SANI statement); axial and rhombic components of the alignment tensor (Da and Dr) were calculated using the 10 run I structures and the software PALES (20). The alignment tensor was then recalculated based on the resulting best 10 of a total of 200 calculated structures (lowest SANI energies), and HADDOCK was run again (see above) using these new values. After this protocol, the final 10 structures were not significantly different from the ones calculated without the RDCs (RMSD over all atoms of the lowest energy structure /H11005/0.3 Å). These structures were analyzed using standard HADDOCK protocols.

PRE Measurements—To attach a paramagnetic moiety to the RARE oligonucleotide, we synthesized modified RARE DNA containing a phosphorothioate linkage at Thy15 (which is located next to the DNA-binding site) was resuspended in 100 mM phosphate buffer (pH 7) to a concentration of /H11011/400 M. This solution was then used to dissolve the complementary strand, and the DNA was annealed by heating to 95 °C for 5 min and then cooling to room temperature over a period of 1 h. A 100 molar excess of 3-(2-iodoacetamido-)proxyl radical (in 100% ethanol) was added to the annealed oligonucleotide and incubated in the dark for 20 h under shaking. To remove excess single-stranded DNA and unreacted 3-(2-iodoacetamido-) proxyl, the reaction mixture was subjected to size exclusion chromatography using NMR buffer. The progress of the reaction was monitored using UV spectroscopy and mass spectroscopy. The final product (containing about 50% labeled and 50% unlabeled double-stranded DNA) was used to carry out a semiquantitative paramagnetic resonance enhancement (PRE) analysis.

To acquire PRE data, two-time point HSQC experiments as described in Ref. 25 were performed (Ta = 0 ms; Tb = 14 ms) on a 1 mM F5-DNA complex. Spectra were first recorded for the paramagnetic (oxidized) sample and then for the diamagnetic (reduced) sample following the addition of a 50 molar excess of sodium dithionate. After normalizing all four 15N HSQC spectra, peak intensities were obtained for the assigned residues and PRE rates (25) were calculated. Distances shown in Fig. 8B are calculated from the phosphate of Thy15 to the HN backbone atom of the corresponding residue.

RESULTS

Identification of the DNA-binding Interface of the MyT1 Double-finger F4F5—In our recent study (14), we tested different combinations of MyT1 ZFs and their ability to bind a 13-bp oligonucleotide containing the RARE sequence (Fig. 1C). Our data showed that F2 and F5 bind with similar affinities (Ka = 1 × 10^6 M^-1 under the conditions of our binding experiments), whereas F3, F4, and F6 bind at least 10-fold less tightly (precise affinities were not measured). In contrast, the affinity of polypeptides comprising two ZF (F2F3, F4F5, F5F6, and F6F7) is significantly higher and displays less variation. Using isothermal titration calorimetry, we also showed that the F4F5-RARE complex forms with a 1:1 stoichiometry. Taken together, these data suggest that both fingers in the two-ZF polypeptides contribute to DNA binding.

Previously, we experienced problems with degradation and intermediate chemical exchange during NMR studies of the double-finger proteins (14). We therefore optimized our experimental conditions (e.g. the addition of protease inhibitor and...
change in buffer components) to circumvent these issues and have consequently been able to record backbone triple resonance spectra for $^{15}$N MyT1 F4F5 and assign the $^{15}$N HSQC spectrum before and after the addition of 1 molar eq of RARE DNA (Fig. 2A). Comparison of the F4 sequence with F5 (Fig. 1) reveals that all residues between the first and last zinc-coordinating cysteines are conserved between the two domains. Consistent with this conservation, most signals occur in closely spaced pairs or as two superimposed peaks in $^{15}$N HSQC spectra (Fig. 2A), indicating that the fold of the two domains is essentially identical. Moreover, the direction and magnitude of chemical shift changes that occur for residues in F4 as compared with their counterparts in F5 upon the addition of 1 molar eq of RARE DNA are very similar (Fig. 2B). Comparison of all chemical shift changes for HN, N, and C atoms in F4 with those in F5 (Fig. 2C) shows clearly that the DNA-binding surface is conserved between the two domains.

We used our assignments of the RARE oligonucleotide (14) to assess which part of the DNA was involved in the interaction with F4F5 (Fig. 3). Analysis of amino-, imino-, and methyl-proton chemical shift changes (Fig. 3, A and B) reveals the sequence AAAGTTCA (palindromic sequence is underlined) as the F4F5 interaction surface. These data are in good agreement with our SPR and NMR analysis of the F5-DNA interaction (14), which showed that F5 has a footprint on the DNA that comprises the smaller AGT motif.

In an attempt to obtain intermolecular NOEs, we recorded two-dimensional NOESY spectra of F4F5 in the presence of RARE DNA (Fig. 3, C and D) at both 25 °C and 4 °C. Although F4F5 binds DNA with higher affinity than the single finger F5, no intermolecular NOEs could be unambiguously identified under these conditions. Fig. 3, C and D, show two different portions of the DNA imino-proton region of the 25 °C NOESY spectrum; no cross-peaks to any protein resonances are observed. Notably, resonances from a number of DNA base protons as well as F4F5 side-chain protons that are likely to form part of the protein-DNA interface could not be located in the spectra of the complex. These signals were most likely expe-
riencing intermediate exchange, preventing the observation of NOEs. In summary, our HSQC data indicate that the structures of both F4 and F5 are largely conserved and that both fingers use the same surface to bind the same target sequence on the DNA, explaining the partly palindromic nature of the RARE DNA motif.

**Data-driven HADDOCK Docking Yields a Structural Model of the F4F5-DNA Complex**—Due to the lack of any intermolecular NOEs, structure determination using the standard NOE-based approach was not possible. Thus, we used our NMR data to calculate a structural model of an F4F5-DNA complex (Figs. 4 and 6) using HADDOCK (23, 26). To achieve this, we first created an F4F5 polypeptide by linking two individual ZF domains and the native linker sequence (Fig. 1) *in silico*, using the NMR structure of F5 (14) as a template. We defined ambiguous restraints between DNA atoms and specific side-chain atoms of DNA-interacting residues of each ZF, based on our NMR data (Figs. 2 and 3). Restraints to keep the structure of the DNA in a standard B-form geometry were also included. Consistent with our chemical shift data, which indicate that both fingers interact in an identical manner, we additionally introduced intramolecular noncrystallographic symmetry constraints between F4F5 and the palindromic part of the DNA (AA
gTT).

During the docking calculations, active residues were defined as semiflexible (side-chain atoms were allowed to move freely), whereas the linker region and the last two residues at both the N-terminal and the C-terminal ends were defined as fully flexible. We carried out 10 runs with different starting orientations of the F4F5 construct relative to the DNA in the presence and absence of noncrystallographic symmetry restraints; all other restraints were kept constant for all runs. In 2 of the 10 docking calculations, we observed significant convergence as seen from HADDOCK energy versus RMSD graphs (Fig. 4A, runs I and II, black and blue) following
Structural Analysis of MyT1 Double Zinc Fingers

two-stage docking and simulated annealing. Both total HADDOCK energy (Fig. 4A) and intermolecular HADDOCK energy (Fig. 4B) are significantly higher (less favorable) for run II as compared with run I. Based on these data, the lowest energy cluster of structures from run I was used to represent a model of the complex. Inspection of the structures with the lowest energies of each run (Fig. 4C) reveals that run II differs from run I by a rotation of each individual zinc finger of 90° around an axis running perpendicular to the major groove of the DNA.

Mutational Analysis and RDCs Were Used to Evaluate and Refine the Structural Model—Based on our 15N HSQC titration data and a visual examination of the structural model, we made a series of point mutations in both ZFs in an effort to assess the validity of our HADDOCK model. Residues were mutated in pairs (the two corresponding residues in each ZF), and seven double mutants were made, namely I813A/I857A, N816A/N860A, Y817F/Y861F, S819D/S863D, R821K/R865K, S822D/S866D, and S824D/S868D. Note that mutations of serine to alanine were avoided as such mutations had been previously demonstrated to disrupt the fold of these domains (14). All seven mutants were correctly folded as judged by their 15N HSQC spectra. The ability of each mutant to bind RARE DNA was then assessed by recording 15N HSQC spectra. For the final 10 structures, the correlation between the predicted and observed HN RDCs is good (r = 0.90; Fig. 5F). Fig. 6A shows the structural model calculated with (blue) and without (light green) RDCs, revealing no significant difference (RMSD over all atoms = 0.3 Å), supporting overall the validity of our structural model.

Polar Interactions Are Mostly Responsible for Specific DNA Binding—Both F4 and F5 contact the major groove of the DNA, together making contacts over the entire AAGTT motif (Fig. 6B). Base-specific hydrogen bonds observed in >50% of the structures were identified between base protons of Ade7/18 (that is, the symmetry-related bases from the AAGTT motif on the two DNA strands) and Arg-821/865 (from F4/F5), Ade6/17,
and Ser-824/868 as well as Thy9/20 and Ser-822/866, whereas Ser-819/Ser-863 contacts the DNA backbone. Multiple hydrophobic contacts made by residues Leu-823 and Leu-867 to methyl groups of Thy20 and Thy9, respectively, suggest that these residues might be important for binding. Taken together, these observations indicate that polar interactions are a strong determinant of binding affinity to MyT1 F4F5 and that several serines and arginines, which are conserved in any of the MyT1 ZF domains apart from finger F1 (Fig. 1), play an important role in recognition of the DNA. We have also analyzed the conformation of the DNA after the docking process using the software 3DNA (27) (Fig. 6 C). Consistent with the restraints applied during the calculation, the geometry of the DNA is conserved, and only minor deviations from the original B-form conformation are observed.

DNA Binding Properties of an F5F6 Construct—To determine whether DNA binding of other MyT1 double ZF domains is similar to F4F5, we expressed a 15N-labeled F5F6 polypeptide and recorded NMR 15N HSQC spectra. A portion of the spectrum in the presence and absence of 1 molar eq of RARE DNA (Fig. 2 A). Consistent with the restraints applied during the calculation, the geometry of the DNA is conserved, and only minor deviations from the original B-form conformation are observed.
is shown in Fig. 7A. As seen in the F4F5 construct, resonances from the corresponding residues in each ZF module occur in pairs, and furthermore, the direction and magnitude of the chemical shift changes upon binding of DNA are conserved, indicating that F6 has a similar fold and DNA-binding interface to F5. To determine the importance of each DNA base for the interaction, we carried out SPR competition experiments as described earlier (14) (Fig. 7B). Single base changes across the whole GAAAGTT motif reduced the ability of a RARE-based oligonucleotide to compete with WT RARE DNA for binding to F5. This profile obtained closely resembled that measured previously for WT RARE DNA (see “Experimental Procedures” for more details) further confirmed that the DNA binding orientation of F5 is equivalent to that proposed in our new F4F5-DNA model.

### DISCUSSION

**Differences in the DNA Binding Mode between MyT1 Single and Double Fingers**—In this study, we have used a combination of NMR and SPR data to calculate a structural model of a MyT1 double ZF domain bound to DNA. We have shown that these domains interact with the major groove of the entire RARE motif and that both fingers bind in an identical manner with two-fold symmetry. Specificity is achieved through several key residues, namely Arg-821/865, Ser-822/866, and Ser-824/868, which make hydrogen bonds to base protons of adenesines and thymines, whereas the two leucines Leu-823 and Leu-867 make hydrophilic interactions with the methyl groups of the thymines in the consensus sequence.

To compare our double ZF F4F5-RARE model with our previous single ZF F5-RARE model (14), we compared the $^{15}$N HSQC spectra in the absence (gray) and presence (black) of RARE DNA (Fig. 8A). Both the magnitude and the direction of the chemical shift changes are very similar, suggesting that the DNA binding mode is conserved. However, comparison of our published F5-DNA model with the F4F5-DNA model reveals that the orientations of the ZF and the surface used for DNA recognition are different in the two cases. The high degree of similarity of the NMR spectra argues that the difference observed in the models does not reflect a real difference in solution binding. Indeed, a semiquantitative PRE analysis (Fig. 8B) carried out using F5 in the presence of a paramagnetic nitroxide spin-labeled RARE DNA (see "Experimental Procedures" for more details) further confirmed that the DNA binding orientation of F5 is equivalent to that proposed in our new F4F5-DNA model.

Substantially more data were used to derive and validate the model in the F4F5 case; in particular, the NMR measurements that were made on the F4F5 mutants clearly showed in significant detail which mutations did not appreciably affect the interaction. The close agreement of these mutational data with our F4F5-DNA model as well as the good correlation of the measured with the calculated RDCs reveals that our current model is more likely to represent the true binding mode for these ZFs. It is worth emphasizing the value of recording $^{15}$N HSQC spectra of complexes made with point mutants when validating docking models; the high information content of such experiments (compared for example with simple affinity measurements) can be very valuable in distinguishing different possible models (e.g. in our case between run I and II) of the same complex.

**Comparison with Members of the NZF Protein Family**—MyT1 (NZF2) belongs to the class of ZF transcription factors that is characterized by the presence of multiple Cys-X$_4$-Cys-X$_4$-His-X$_4$-His-X$_4$-Cys sequences. So far, apart from our solution structure of the fifth finger of MyT1, only two other structures have been determined: the first ZF domain of NZF1 (28)
(PDB ID 1PXE) and the fifth and sixth ZF of NZF3 (or ST18) (PDB ID 2CS8), which were solved by the RIKEN structural genomics initiative. Both NZF1 and NZF3 double ZF modules have been shown to be able to recognize the RARE DNA motif (7, 13, 29).

The structures of F5 and F6 of NZF3 are highly similar to our MyT1 F5 structure (RMSD /H11021 1.5 Å over folded regions), indicating that the ZF fold is conserved within the whole protein family. This is also in good agreement with our NMR data on the F5F6 construct (Fig. 7). As a consequence, the same loop that binds the DNA specifically in the MyT1 F4F5 construct is likely to be utilized in NZF3. Indeed, sequence alignment of these proteins (Fig. 1B, arrows) confirms that the subsets of residues that make specific contacts with the DNA (Fig. 1B, arrows) are either conserved or very similar (e.g. threonine instead of serine).

The DNA binding affinities of constructs containing MyT1-like zinc fingers appear to vary over a wide range depending on the number of individual ZF modules present. For example, although the affinity of all six ZF domains of NZF3 for a double RARE site was estimated to be around 2 nM by EMSA experiments (7), a double ZF polypeptide consisting of the second and third finger of NZF1 binds with >10 nM as determined by fluorescence methods (30). In contrast, our isothermal titration calorimetry (14) and SPR data (Fig. 5) revealed a RARE DNA binding affinity of between 0.3 and 1 μM for the F4F5 construct. However, in all cases, a double ZF module was found to be required and sufficient to recognize one single palindromic AANTT site, indicating that binding of four or more zinc fingers most likely occurs via interaction of multiples of these units with DNA.

Other Putative Target Sites of MyT1—To search for other putative MyT1 DNA-binding sites, we used our structural and biophysical data to define a minimal DNA-binding site. The majority of the specific contacts are made to the two bases flanking the central guanine of the AAGTT site (underlined), in agreement with our SPR competition data (14). We therefore initially searched the GenBank database for a triple RANTR (R = G/A, N = any base) site that would allow both the MyT1 F2F3 and the F4F5F6F7 ZF clusters to bind simultaneously. Due to the importance of MyT1 in neuronal development, we limited our search to neuronally expressed genes with promoters of mice and humans available in GenBank. No hits were found with this search sequence, but a double RANTR motif was located at three sites (Fig. 9).

The first putative site is located around 350 bp upstream of the promoter of NeuroD1, a basic helix-loop-helix transcription factor that when combined with Ascl1, Brnd2, and MyT1L can convert fetal and postnatal human fibroblasts into func-
Figure 8. Comparison to the single-finger F5-DNA interaction. A, portion of the $^{15}$N HSQC of $^{15}$N-labeled double ZF protein F4F5 (upper) and single ZF F5 (lower), respectively, in the absence (gray) and presence (black) of 1 molar eq of RARE DNA. Note the similarity in both the magnitude and the direction of the chemical shift changes. B, semiquantitative PRE analysis of F5 in the presence of nitroxide spin-labeled RARE DNA. The PRE rate calculated from a two time point $^{15}$N HSQC experiment (25) is plotted as a function of the amino acid (upper part) and compared with the actual distance between the phosphorus atom of Thy15 (on which the PRE label is located) and the corresponding backbone HN proton for the two different F5-DNA models, respectively (middle and lower parts). The PRE rate profile agrees better with the distance profile of the new F5-DNA model than with that of our previous model (14). Residues colored in green are those with distances to the Thy15 phosphorus (in green as well) that are below the average distance minus one standard deviation. These residues are indicated on the corresponding model (right part of figure).

Figure 9. Putative MyT1 target sites. Sequence alignment of portions of three human and mouse genes (International Nucleotide Sequence Database Collaboration [INSDC] database accession numbers in parentheses) that contain potential target sites for MyT1 double or quadruple ZF clusters are shown.
Structural Analysis of MyT1 Double Zinc Fingers

around 140 nucleotides upstream of the start of the transcription site of neurogenin1, a transcription factor essential for neuronal differentiation and subtype specification during embryogenesis (33, 34). Interestingly, when NZF3, a member of the same protein family to which MyT1 belongs, and Neurog1 are co-expressed in cells, the rate of neuronal differentiation is significantly increased, suggesting that these proteins act in concert to stimulate these processes (35).

In conclusion, we have used a combination of structural and biophysical methods to determine the molecular basis of the interaction of double ZF domains of myelin transcription factor 1 with DNA. Our work has revealed that the entire RARE motif, which has been identified previously as a relevant binding site, can accommodate a two-ZF module that occupies the major groove of the DNA (14). The orientation of the two fingers relative to each other differs by 180°, and both domains contact the same protein family to which MyT1 belongs, and Neurog1 embryogenesis (33, 34). Interestingly, when NZF3, a member of neuronal differentiation and subtype specification during tran-
scription site of neurogenin1, a transcription factor essential for nervous system and pituitary gland. J. Biol. Chem. 271, 10723–10730


