Structural and Biophysical Analysis of the DNA Binding Properties of Myelin Transcription Factor 1*

Roland Gamsjäger1,2, Michael K. Swanton1,3, Felix J. Kobus, Eija Lehtomaki, Jason A. Lowry, Ann H. Kwan, Jacqueline M. Matthews4, and Joel P. Mackay5

From the School of Molecular and Microbial Biosciences, University of Sydney, Building G08, New South Wales, Sydney 2006, Australia

Zinc binding domains, or zinc fingers (ZnFs), form one of the most numerous and most diverse superclasses of protein structural motifs in eukaryotes. Although our understanding of the functions of several classes of these domains is relatively well developed, we know much less about the molecular mechanisms of action of many others. Myelin transcription factor 1 (MyT1) type ZnFs are found in organisms as diverse as nematodes and mammals and are found in a range of sequence contexts. MyT1, one of the early transcription factors expressed in the developing central nervous system, contains seven MyT1 ZnFs that are very highly conserved both within the protein and between species. We have used a range of biophysical techniques, including NMR spectroscopy and data-driven macromolecular docking, to investigate the structural basis for the interaction between MyT1 ZnFs and DNA. Our data indicate that MyT1 ZnFs recognize the major groove of DNA in a way that appears to differ from other known zinc binding domains.

It is predicted that there are at least 15,000 zinc fingers (ZnFs)6 in over 1,000 different proteins in humans (1). These domains fall into more than twenty subclasses, based on their fold and zinc-ligation topology, and different members can mediate interactions with DNA, RNA, proteins, and other molecules (2). The structures and functions of several classes of ZnFs, including classical and GATA-type domains, have been analyzed in detail; in contrast, some classes of ZnFs are not well understood.

Myelin transcription factor 1 (MyT1, or neural zinc finger 2 (NZF2)) is a transcription factor that contains seven copies of a ZnF with a C2HC consensus sequence. The seven ZnFs are arranged in a 1 + 2 + 4 topology (see Fig. 1), although a second isoform exists that lacks finger 1 (F1) (3). MyT1 was first isolated on the basis of its interaction with DNA elements in the promoter of the proteolipid protein (PLP) gene (4). MyT1 acts during central nervous system development to induce differentiation of early neuronal precursors and to commit glial cell precursors to the glial cell lineage (5). After birth, MyT1 activates the PLP gene in Schwann cells, leading to the production of the myelin sheath around axons in the central nervous system (6, 7). Orthologs of MyT1 are found in organisms ranging from nematodes to mammals, and functional data indicate that MyT1 is essential for neural development in Xenopus laevis (5). Humans contain two other likely paralogs of MyT1: MyT1-like (MyT1L/NZF1) and Suppressor of Tumorigenicity 18 (ST18/NZF3). MyT1L appears to be involved in neural development (7), whereas ST18 was reported to be a breast cancer tumor suppressor (8). At least three additional vertebrate proteins that contain one MyT1-type ZnF have been identified (L3MBTL1, L3MBTL2, and L3MBTL4), and one of these, L3MBTL1, is a possible tumor suppressor (9).

Several biochemical studies have demonstrated that MyT1-type ZnFs from MyT1, MyT1L, and ST18 can recognize DNA in a sequence-specific manner (5, 10, 11). The consensus sequence, AAAGT, is the core element of the cis-regulatory element of the β-retinoic acid receptor gene (β-RARE) and is also found in the human PLP promoter. Until now, the molecular details of DNA recognition by MyT1-type zinc fingers have not been addressed.

In this work, we used surface plasmon resonance (SPR) and isothermal titration calorimetry (ITC) to delineate the relative importance of each ZnF in MyT1 for DNA binding. Finger 5 (F5) has the highest affinity for the β-RARE site. We went on to use a combination of NMR and mutagenesis to probe the structural basis for the MyT1-DNA interaction. A data-driven model of the interaction, calculated using HADDOCK (12–14), reveals that a whole MyT1-type ZnF can sit snugly in the major groove of double-stranded DNA, forming a number of specific electrostatic and hydrophobic contacts. This mode of recognition is distinct from that observed for other classes of DNA-binding ZnFs.

EXPERIMENTAL PROCEDURES

Subcloning, Expression, and Purification of MyT1 Constructs—The original plasmid encoding mouse 6-ZnF myelin transcription factor 1 (mMyT1) was a generous gift of Dr. Lynn Hudson (National Institutes of Health). Ten different con-
structs of MyT1 were cloned from the original plasmid and mutants were constructed using overlap extension PCR (see Fig. 1). All constructs were overexpressed as fusions with glutathione S-transferase at 25 °C under standard conditions; iso-
topically labeled F5 was overexpressed using the protocol described in a previous study (15). Proteins were purified using GSH affinity chromatography, thrombin cleavage, and gel fil-
tration (Superdex-75 in SPR buffer: 50 mM NaCl, 10 mM HEPES, 1 mM dithiothreitol, pH 7.4). Protein identities were
confirmed by electrospray mass spectrometry, and concentrations were determined by absorbance at 215, 225, and 280 nm (16). Fractions were stored at −20 °C until required.

**β-RARE DNA and Mutant Oligonucleotides—**Single-
stranded β-RARE DNA (5′-ACCGAAAGTTCAC and 5′-GTG-
AACTTTCCGGT), mutant oligonucleotides and biotinylated
DNA for SPR experiments were obtained from Sigma-Aldrich, annealed in SPR buffer without dithiothreitol (heated to 95 °C
for 5 min, then cooled to room temperature over the course of
1–2 h) and purified using gel filtration (Superdex-75). Concent-
trations were calculated by absorbance at 260 nm.

**SPR—**All experiments were performed on a Biacore 3000 sys-
tem (Biacore AB) at flow rates of 5–10 μl/min in SPR buffer. Biotinylated DNA (~10–100 nm) was immobilized on strepta-
vidin-coated Biacore SA chips. MyT1 (1–110 M) was injected
in SPR buffer and binding was monitored. To compare binding
affinities for different single- or double-finger constructs in a
semi-quantitative manner, equal concentrations of each con-
struct were added (10 or 20 μM) and the SPR responses were
compared. The system was washed with 1 M NaCl (1 min) after
each experiment. For kinetics studies (F5 and F7), curves were
fitted using the standard 1:1 Langmuir model. In the competi-
tion experiments, either F5 (12 μM) or F45 (5 μM) was added
to pre-bound DNA in the presence of 2 or 5 molar equivalents
of competitor DNA oligonucleotides, respectively.

**ITC—**Protein and DNA samples were dialyzed overnight
against the same reservoir of SPR buffer. β-RARE DNA (60–
130 μM) was titrated into F5 or F45 (8 μM). Titrations were
carried out on a MicroCal VP-ITC microcalorimeter at 25 °C.
For each titration, 25–30 injections of 5–10 μl of titrant were
made at 5-min intervals. Data were corrected for heats of dilu-
tion from control experiments of DNA into buffer, and ana-
lyzed using Origin ITC Analysis software (MicroCal Software,
Northampton, MA), as described previously (17).

**NMR Spectroscopy—**F5 (unlabeled, 15N- or 15N/13C-labeled)
was dialyzed into NMR buffer (2.5 mM deuterated MES, pH 6.5,
10 mM NaCl, 1 mM dithiothreitol or Tris(2-chloroethyl)phos-
phate) and concentrated in Microsep 1-K cutoff units to 200–
700 μM. Standard homonuclear and 15N-separated experi-
ments were acquired at 4 °C on a Bruker Avance 600 and
Avance 800 NMR spectrometers equipped with cryoprobes.
15N-separated NOESY and two-dimensional NOESY data were
used to derive distance constraints. Dihedral restraints were
obtained by analysis of the HNHA spectrum (18), and stereo-
specific assignments were made from short mixing time
NOESY (50 ms) and TOCSY (35 ms) data. 15N-HSQC titrations
of F5 with β-RARE DNA were carried out in 10 mM sodium
phosphate buffer with 50 mM NaCl (pH 7.4) in the presence of 1
mM dithiothreitol at 25 °C. HNCA, HNCOCA, CBCA(CO)NH,
and HNCO experiments were recorded for both 15N,13C-F5
alone, and the protein-DNA complex, together with single and
double half-filtered NOESY experiments and an HCCH-
TOCSY for the complex. Chemical shift changes were calcu-
lated as a weighted average of HN, N, Ca, and C′ changes, using
a to previously reported equation (19, 20). Assignments of
the DNA alone were obtained from two-dimensional NOESY spec-
tra. NMR data were processed using XWINNMR (Bruker, Karlsruhe) and analyzed with SPARKY (T. D. Goddard and
D. G. Kneller, University of California at San Francisco).

**Structure Calculations—**Initial structures of F5 were calcu-
lated in DYANA (21) from manually assigned unambiguous
NOEs and dihedral angle constraints obtained from both 1J_{HN,Hα}
coupling constants and the GRIDSEARCH module of DYANA.
Zinc-ligating atoms were identified by examination of the ini-
tial structures calculated in the absence of zinc. Final calcula-
tions were then carried out using ARIA 1.2 (22) implemented
in CNS 1.1 (23), using the standard protocols provided, with the
zinc geometry and the experimentally determined tautomeric
state of the histidine side chains fixed (24). Final assignments
made by ARIA 1.2 were checked manually and corrected where
necessary. In the final set of calculations, the 100 lowest energy
structures were refined in a 9-Å shell of water using standard
ARIA 1.2 water refinement modules (minimization and dyna-
mics steps, for details see Refs 25 and 26). The 20 conform-
ers with the lowest value of total energy were analyzed using
MOLMOL (27) and PROCHECK_NMR (28).

**HADDOCK Docking—**F5 was docked to the DNA using the
program HADDOCK 1.3 (12–14). The starting structures for
the docking were a B-form model of the double helix DNA
fragment (5′-ACCGAAAGTTCAC) constructed with the
Nucleic Acid Builder package (29) and the 20 lowest energy
structures of F5. Based on NMR titration data, disordered
sequences consisting of the ten N-terminal and seven C-termi-
 nal residues of F5 (and the Zn atom) were excluded from the
calculation. Active and passive residues for both the protein
and the DNA were chosen based on experimental data and
solvent accessibility (>50%) determined by the program MOL-
MOL. A 2-Å distance was used to define the ambiguous inter-
action restraints (AIRs). For the DNA fragment, bases 6–9 and
18–20 located within the consensus sequence were selected as
“active” based on the chemical shift changes observed upon
complex formation. AIRs were defined solely from the unique
base atoms of bases 8, 9, 19, and 20 to suitable atoms (i.e. from H
to N or O and vice versa) of all active and passive residues of F5,
whereas for bases 6, 7, and 18 the whole base was selected
because of the lower specificity of these nucleotides (30). The
four flanking nucleotides (bases 5, 10, 17, and 21) were chosen
as “passive.” For F5 AIRs were defined between all atoms of all
active residues (Ile-857, Asn-860, Tyr-861, Ser-863, Arg-865,
and Ser-868) and all atoms of all active and passive DNA bases.
Passive residues of the protein were defined as the solvent-
accessible surface neighbors of active residues (Ser-854 and
Ala-862). A total of 18 AIRs resulted from all definitions
described above and were used as input in HADDOCK. Addi-
tional restraints to maintain base planarity and Watson Crick
bonds were introduced for the DNA. During the rigid body
energy minimization, 1000 structures were calculated, and the
RESULTS

The DNA-binding Properties of MyT1 ZnF Combinations—To understand the DNA-binding properties of MyT1 ZnFs, we first expressed and purified poly-Lysines corresponding to different combinations of the MyT1 ZnFs (Fig. 1, A and B) and tested their ability to bind a 13-bp oligonucleotide containing the β-RARE sequence (5'-ACCGAAAGTTCACT, Fig. 1C) by SPR. Both two-dimensional NMR and circular dichroism melting data confirmed that this oligonucleotide is a well ordered B-form double helix in the SPR buffer (data not shown). The B-form nature of the oligonucleotide was confirmed by analyzing intra- and internucleotide NOE intensities (31).

We first measured the affinity of individual ZnFs for the β-RARE DNA. As shown in Fig. 2A, significant differences in the binding properties were observed between fingers. The data for finger 5 (F5) fitted well to a 1:1 binding isotherm, yielding a $K_D$ of 200 μM (see Fig. 2G). Reduction of the salt concentration from 150 to 50 mM increased the binding affinity to 11 μM (Fig. 2H), indicating a significant electrostatic component to the F5-DNA interaction. In comparison, the binding of F7 was slightly weaker ($K_D = 51 \mu M$ at 50 mM NaCl) and F3, F4, and F6 bound so weakly that an affinity constant could not be reliably measured. As an alternative (albeit semi-quantitative) means of gauging the relative affinities of each ZnF for DNA, we compared the SPR response obtained following the injection of equal amounts of protein onto identically prepared chips (e.g. see inset in Fig. 2A for data on F5 and F7). This approach gives a reliable indication of the relative propensities for binding, provided the protein concentration used is of the order of the dissociation constants (as it was in these experiments). As shown in Fig. 2A, the different ZnFs clearly have different DNA-binding properties in isolation. Overall, F2 and F5 appeared to bind with similar affinities, whereas F3, F4, and F6 bound relatively poorly. F7 exhibits ~50% of F5 binding using this approach (Fig. 2A and inset), in good agreement with the affinities determined above. Interestingly, the affinities of double finger constructs displayed less variation (Fig. 2B).

ITC data revealed that F5 binds to β-RARE with a 1:1 stoichiometry and an affinity of 27 μM (Fig. 2C). Note that these data were recorded under conditions where the so-called “c-value,” the product of the association constant and the receptor (in this case the protein) concentration, was ~0.3. ITC data are typically recorded with $c$ in the range 10–300, under which conditions the data are sigmoidal in shape. However, it has been rigorously established that accurate $K_D$ values can still be determined for weak interactions where $c$ values are as low as 0.01 or less (32). The binding constant obtained for the F5-DNA interaction agrees well with that obtained by SPR (11 versus 27 μM, see above).

A double finger construct containing both F4 and F5 (F45) also binds in a 1:1 fashion, but two orders of magnitude more tightly ($K_D = 0.27 \mu M$, Fig. 2D). Thus, binding of the β-RARE probe is dominated by a subset of the MyT1 ZnFs (F2 and F5, and probably also F7), but the other fingers can make a significant contribution to binding at a GAAAGTT site.

Analysis of ZnF-DNA Interactions—We next sought to examine the specificity of DNA binding by MyT1-type ZnFs in more detail. Preliminary NMR data (not shown) indicated that ZnFs 2 and 4, while folded, existed in solution as multiple species that were in slow exchange on the chemical shift timescale. Additionally, SPR binding curves for the F2-DNA interaction could not be fitted by any available kinetic model. Both of these

![Schematics of MyT1 and DNA. A, schematic representation of mouse MyT1, showing the constructs used in this work. Gray boxes indicate zinc fingers. B, sequence alignment of ZnF domains in mMyT1. Zinc-ligating residues are marked by asterisks, and regions of very high sequence similarity are indicated by gray shading, DNA-binding residues are boxed, and point mutants are indicated by black arrows. Numbers refer to the full-length proteins. C, sequence of the β-RARE oligonucleotide used in this work. Boxed bases are part of the consensus sequence.](image-url)
findings are consistent with oligomerization of the proteins; a conclusion that was subsequently confirmed by multiangle laser light scattering (not shown). We therefore focused on the interaction of F5 with DNA.

To determine the importance of each DNA base for the interaction, SPR competition experiments were performed. β-RARE DNA was immobilized and F5 was injected in the presence of a 2-fold molar excess of a competitor oligonucleotide (Fig. 2E). This series of competitors was based on wild-type β-RARE, with each base mutated in turn (A to C, G to T, and vice versa). Under these conditions, wild-type β-RARE reduced the SPR signal by ~80% and all other competition experiments were compared with this value. The mutants A7C, G8T, T9G, and T10G were significantly less effective competitors for F5 binding than the other oligonucleotides (p < 0.01), indicating that the central AGTT sequence is likely to represent the target site for F5. The same experiment was repeated for the double finger construct F45 (Fig. 2F), whereupon three additional positions (GAA) upstream of the central AGTT motif were demonstrated to contribute to F45 binding.

Notably, the changes in relative competition at the AGTT motif were significantly larger for F45 compared with F5 because of the different conditions used (which were restricted by the amounts of protein and DNA available and the binding affinities for the different interactions).

Structure of F5—Next, we used standard 1H- and 15N-separated NMR experiments to determine the solution conformation of F5. Initial structures calculated in DYANA (21) using unambiguous restraint data indicated that Cys-486, Cys-851, His-864, and Cys-870 were the zinc ligands. The structure was then refined in ARIA (22, 33), with the inclusion of restraints to define the zinc geometry (34). The 20 lowest energy structures from a total of 700 were used to represent the structure.
MyT1 Interaction with DNA

![Image](https://example.com/image.png)

FIGURE 3. Solution structure of F5. A, stereo view of the backbone of the 20 lowest energy structures of F5 (residues 845–874). Residues 837–844 and 875–880 are disordered and are not shown. Heavy side-chain atoms of the zinc ligating residues (Cys-846, Cys-851, His-864, and Cys-870) are shown in green, and zinc atoms are colored yellow. The dimensions of the domain are 21 Å (top to bottom) by 22 Å at the widest point. B, electrostatic surface representation of F5. The structure on the left has been rotated by 45° clockwise about the vertical axis when looking from above relative to A. Important residues have been labeled. (The coordinates for the structure were deposited to the Protein Data Bank with the accession code 2JYD.)

of F5 (Fig. 3A, PDB code 2JYD). These structures display good covalent geometry and no restraint violations >0.5 Å or 5° (Table 1). The structure is well defined, with an r.m.s.d. for backbone atoms of 0.41 Å (residues 845–874), but lacks any elements of repeating secondary structure (i.e. α-helices and β-sheets). Rather, it comprises several irregular loops with a central zinc atom (Fig. 3A). The four zinc-ligating residues, together with Thr-848, Gly-855, Gly-859, and Gly-869, make up a small hydrophobic core. Fig. 3B shows an electrostatic surface representation of F5, revealing a distinct groove on one side and a mixture of positive and negative potential on the opposing surface. A small electrostatic network also appears to exist on one surface of F5, comprising Lys-845, Asp-852, and Arg-872 (Fig. 3B). Mutation of the surface exposed Arg-872 to alanine disrupts the structure of F5, as determined by one-dimensional NMR spectroscopy (data not shown). Fig. 3C shows a section of the F5-DNA complex also allowed chemical shift changes to be evaluated for protons in the DNA (Fig. 4, D and E). Overall, the largest chemical shift changes were observed for the central AGT portion of the duplex, consistent with the SPR competition experiments.

A Model of the F5-DNA Complex—Although the F5-DNA complex was ~99% populated under the conditions of our NMR experiments ([F5] = [β-RARE] = 1 mM), no intermolecular NOEs could be unambiguously identified from either double half-filtered or a combination of two-dimensional [F1,F2] and two-dimensional [F1]15C,15N-filtered NOESY spectra (35, 36). Notably, resonances from a number of DNA base protons (e.g. GUA8 and CYT19) as well as F5 side-chain protons (including those of Tyr-861) that are likely to form the protein-DNA interface could not be located in spectra of the complex. These signals were most likely experiencing intermediate exchange.

Therefore, to gain insight into the likely DNA-binding mode of F5, we used our chemical shift perturbation and mutagenesis data to create a model of the F5-DNA complex using the data-driven docking program HADDOCK (12–14). HADDOCK comprises a series of Python scripts that run on top of the structure determination programs ARIA (22) and CNS (23); it is designed to use biochemical and/or biophysical interaction data to generate AIRs and to then carry out docking calculations to generate models of a complex that satisfy the experimental data (37–39).

In the current case, active protein and DNA residues were defined as those that both underwent significant changes in chemical shift and were >50% surface exposed (Fig. 4). Six AIRs...
forms hydrogen bonds with GUA8, CYT19, and THY20, as well as phobic contacts with GUA8 and ADE18, whereas the latter is seen in Fig. 5 structures (Table 2), and the base-specific interactions can be ligands (Fig. 5, Table 1). The DNA-binding surface of F5 comprises mainly and 14 hydrophobic contacts are observed in 4 bp and makes contacts over mainly four base pairs: the AAGT of the DNA complex.

Following a two-stage docking, simulated annealing, and water refinement protocol (30), the ten lowest energy structures were selected to represent a model of the complex structure (Fig. 5A, PDB code 2IX1). These structures overlay with an r.m.s.d. of 0.7 Å over all backbone atoms of the complex. No restraint violations >0.5 Å were observed in the final structures (Table 1). The DNA-binding surface of F5 comprises mainly residues in the long irregular loop between the two pairs of zinc ligands (Fig. 5, B and C); this loop inserts into the major groove and makes contacts to DNA by mainly four base pairs: the AAGT of the consensus sequence. Ten intermolecular hydrogen bonds and 14 hydrophobic contacts are observed in >50% of the structures (Table 2), and the base-specific interactions can be seen in Fig. 5C. The side chains of both Tyr-861 and Ser-868 play a dominant role in DNA recognition: the former makes two hydrogen bonds with THY9 and CYT19, as well as hydrophobic contacts with GUA8 and ADE18, whereas the latter forms hydrogen bonds with GUA8, CYT19, and THY20, as well as van der Waals’ contacts to bases ADE7, GUAY, and CYT19. Other hydrogen bonds to base protons are formed by the backbone O of Leu-867, whereas Arg-865 and Ser-866 form hydrogen bonds with atoms in the phosphate backbone. Notably, the chemical shift of the carbonyl carbon of Leu-867 (180.2 ppm) is significantly high compared with other residues, which agrees well with the observation that the carbonyl oxygen of Leu-867 forms a hydrogen bond with base protons of both ADE6 and ADE7.

The conformation of the DNA after the docking process was analyzed using the program 3DNA (40). Overall, the DNA remained double-stranded throughout the calculation, although some deviations from canonical base-pairing geometry were observed in the protein-binding region (ADE6–THY9, Fig. 6). The width of the major groove of F5-bound DNA is at a maximum between ADE7–GUA8 and GUAY–THY9 indicating the weak intercalation of residues Ser-868 and Tyr-861 (Fig. 5B), respectively.

In an effort to validate our model, we constructed point mutants of F5, targeting residues Tyr-861 and Ser-868, which form the majority of specific hydrogen bonds with the DNA (Table 2) as well as Thr-858, Asn-860, Arg-865, and Arg-872. One-dimensional 1H NMR spectra of alanine mutants T858A, Y861A, S863A, R865A, S868A, and R872A revealed that they were unfolded; however, the more conservative mutations Y861F, S868V, N860A, and S866D, which all target the proposed DNA-binding face of F5, folded correctly according to one-dimensional 1H NMR spectra and all resulted in significant decreases in DNA binding (Fig. 7).

Finally, we noted that, although the DNA binding ability of F4 and F5 is significantly different (Fig. 2A), the amino acid sequence of the two domains is conserved at the DNA binding interface (Fig. 1). In an attempt to explore the reason leading to this difference, Lys-845 was mutated to alanine. This residue is one of the few in the F5 structure that is not present in F4 (the corresponding residue is Thr-801). The mutation resulted in a decrease of DNA binding compared with F5 (Fig. 7) suggesting that the electrostatic network between Lys-845, Asp-852, and Arg-872 plays a role in stabilizing the protein-DNA complex.

### DISCUSSION

**The DNA Binding Mode of MyT1 Zinc Fingers**—In this study, we have sought to understand the molecular basis for DNA recognition by MyT1-type ZnFs. Our data define the footprint made on the DNA by a single MyT1 ZnF and suggest the molecular basis for the DNA-binding properties of these domains.

The model that we have constructed from a combination of NMR and SPR binding data indicates that MyT1 ZnFs form small, compact structures that can sit entirely within the major groove of DNA. The side chains of a small subset of residues (primarily Tyr-861 and Ser-868) make specific contacts with DNA bases in the 4-bp AAGT portion of the MyT1 consensus sequence, and these interactions most likely contribute to specificity as well as binding affinity. The weak intercalation of Tyr-861, through hydrogen bonding of the terminal OH group and van der Waals’ interactions of the aromatic ring, probably plays a significant role in the recognition of the DNA (Fig. 5 and Table 2). Tyrosine intercalation during DNA binding has been observed with other non-ZnF proteins (41).
Our model reveals that there is good shape complementarity between MyT1 ZnFs and the DNA major groove (Fig. 5). The positioning of the domain in the major groove corroborates published methylation interference data obtained on the homologous rat NZF1 protein, which also indicated an interaction with the major groove (10, 42).

During HADDOCK calculations, we often observed an alternative cluster of structures in which the same protein residues made interactions with the DNA (predominantly Tyr-861, Ser-866, and Ser-868), but where the orientation of F5 in the major groove differed by 180°. In these complexes, the general fit of the domain into the major groove was good, but a statistical analysis of calculated CNS energies (for details see Ref. 23) revealed a significantly higher (p < 0.01) energy compared with our model (Δ ~ 70 kcal/mol). In addition, distance violations were observed in all of these structures (>0.5 Å), arguing against this alternative structure being correct.

Although many zinc fingers recognize the major groove of DNA, the mechanism through which MyT1 achieves this end, insertion of the entire domain in the groove, contrasts sharply with the classical, GATA-type, and steroid hormone receptor ZnFs. These latter domains all utilize an α-helix to position DNA-contacting residues appropriately and in fact many other DNA-recognition domains, including homeodomains and other helix-loop-helix proteins, take advantage of the size match between the α-helix and the DNA major groove. MyT1-type ZnFs are found in far fewer proteins than classical ZnFs, and in this regard are more similar to folds such as GATA-type, TAZ, and RanBP-type zinc finger domains, which are also comparatively rare.

Weak versus Nonspecific Interactions—Our data show that the binding of either single or double MyT1 zinc finger constructs to DNA is significantly weaker than that typically observed for sequence-specific DNA-binding proteins (~1–10 nm). However, it is important to emphasize that, despite its low affinity, the MyT1-DNA interaction is sequence specific: our SPR competition experiments showed that even single base mutations significantly reduced the binding of F5 (or F45) to β-RARE. Interestingly, a recent analysis of genome-wide ChIP-on-chip data from yeast (43) suggested that weak protein-DNA interactions are extremely abundant and are likely to be functional.

More generally, it is often tempting to use the terms “weak” and “nonspecific” interchangeably when discussing bimolecular interactions, but it is important to point out that these descriptors are definitely not equivalent. The interaction between MyT1 and DNA is relatively weak, but a wide variety of experimental data demonstrates that it is highly specific. For
example, the F5-DNA NMR titration was saturable, point mutations to the DNA sequence disrupted binding, and the SPR response reached equilibrium following each of the protein injections (Fig. 2G).

A related point is that it is often presumed that weak interactions are unlikely to be physiologically relevant because the relevant species will not be present in the cell at the required concentrations. However, there is an abundance of data in the literature that contradicts this assumption. For example, protein-protein interactions, such as those involving either NZF zinc-finger domains and ubiquitin (44), the transcription factors GATA-1 and FOG (45), or modified histone tails and their recognition domains (e.g. plant homeo domain (PHD) and chromo- and bromo-domains (46–48)), frequently have dissociation constants in the 1–10 \( \mu M \) range.

Given that interactions described above are known to be important biologically, the simplest conclusion that can be drawn from these data is that binding affinities measured \textit{in vitro} do not fully reflect the situation inside a cellular compartment, where for example, local concentrations might be drastically altered by compartmentalization, subcellular targeting processes, or the formation of large multiprotein complexes. Thus, \textit{in vitro} binding affinities are a useful relative measure (when comparing different domains or a panel of mutants, for example), but the relevance of their absolute values is difficult to gauge.

\textbf{Multi-ZnF Proteins and DNA Recognition}—The current study has improved our understanding of how the multiple zinc fingers of MyT1 combine to recognize DNA. SPR analysis shows that, although neither F4 nor F6 bind well to DNA, they can each combine with F5 to increase the affinity of MyT1 for DNA. Our data indicate that the double finger construct F45 is likely to make base-specific contacts across the full length of a GAAAGTT site (Fig. 2F), indicating that a two-finger unit is both necessary and sufficient to obtain full recognition of the \( \beta \)-RARE site. Given the high similarity between the sequences of F4 and F5 and the partially palindromic nature of the DNA site, it is likely that F4 of the double finger construct contacts the GAA sequence in an orientation that differs by 180\(^\circ\) from that observed for F5 binding to AGTT. This idea is consistent with our model of the F5-DNA complex, wherein the N terminus of F5 points...
toward the 5'-end of the \( \beta \)-RARE site (Fig. 5), as well as with the measured 1:1 binding stoichiometry of F45 (Fig. 2D) and the SPR competition experiments (Fig. 2F).

The DNA recognition site identified for MyT1 (4) in the human PLP promoter contains a single copy of an AGTTT sequence and is preceded by three purines (GGA). Earlier studies also suggested that purines upstream of the AGTT site are crucial for DNA binding (10). This site is located 256 bases upstream of the coding region of the PLP1 and is 100% conserved in the chimp. Despite the fact that mouse and rat MyT1 share a high homology with the human protein (5), particularly in the DNA-binding region identified in this study, a corresponding target site could not be located at the same position in these genomes. However, a very similar site was found 777 bases upstream of the PLP1 start site, comprising an AGTTT site preceded by three purines (AAG). It is quite possible that this represents the target site for MyT1.

Yee and Yu (11) have found that NZF-3, the third member of the MyT1 family, binds substantially better to a DNA probe containing two repeated AAAGTTT motifs. However, the reported gel shift experiments were carried out using glutathione S-transferase fusions of MyT1 zinc finger domains, and we have observed in other contexts that the propensity of glutathione S-transferase to dimerize can lead to such results.7

Given that there are no tandem AAAGTT sites within the proximal promoter of the human or rodent PLP1 gene, the question arises as to why MyT1 carries three double-ZnF domains (F45, F67, and the distant F23). It is possible that distinct two-ZnF clusters might affect chromatin structure by binding simultaneously to highly separated DNA sites. This possible activity is interesting in light of the recent finding that MyT1 can recruit Sin3B (49) and could indicate a mechanism in which the binding of MyT1 to distant sites and the co-opting of Sin3B results in the reorganization of local chromatin structure. Mechanisms of this type have long been suggested for many transcriptional activators, including multi-ZnF proteins that can contain up to nearly 40 classic zinc fingers, but we still have very little direct evidence of such activities.

In conclusion, we have used a range of biophysical methods to delineate the mechanism of DNA recognition by the unusual MyT1 class of zinc fingers. Our results indicate that MyT1 ZnFs recognize 3–4 bp of double-stranded DNA in a sequence-specific manner by inserting the whole domain into the major groove of the DNA. These data further our understanding of protein-DNA recognition.

Acknowledgments—We thank Alexandre Bonvin and Marc van Dijk for assistance with the docking calculations and Lynn Hudson for the original mMyT1 construct.

7 F. Loughlin and J. P. Mackay, unpublished results.

### TABLE 2

| F5 atom | DNA atom | Distance (Å) | No. (of 10) |
|---------|----------|--------------|-------------|
| A) hydrogen bonds |
| Y861.OH | THY9.H3   | 2.19         | 10          |
| Y861.OH | CYT19.H42 | 2.29         | 7           |
| R865.HN | CYT19.O2P | 2.20         | 9           |
| S866.HG | CYT19.O2P | 1.95         | 8           |
| L867.O  | ADE6.H61  | 2.25         | 6           |
| L867.O  | ADE7.H61  | 2.21         | 9           |
| S868.HG | CYT19.N3  | 2.22         | 6           |
| S868.HG | THY20.OA  | 2.18         | 10          |
| S868.OG | AUA8.H1   | 2.19         | 10          |
| S868.OG | THY20.H3  | 2.21         | 10          |
| B) hydrophobic contacts |
| N860.CG | AUA8.C8   | 3.54         | 7           |
| Y861.C1 | ADE18.C5  | 3.38         | 10          |
| Y861.C1 | ADE18.C6  | 3.42         | 10          |
| Y861.C2 | AUA8.C5   | 3.38         | 10          |
| Y861.C2 | AUA8.C6   | 3.48         | 10          |
| Y861.CZ | ADE18.C6  | 3.68         | 7           |
| Y861.CZ | AUA8.C6   | 3.83         | 6           |
| R865.CG | ADE18.C3’ | 3.82         | 6           |
| S868.CA | ADE7.C5   | 3.68         | 10          |
| S868.CA | ADE7.C6   | 3.73         | 10          |
| S868.CB | CYT19.C4  | 3.22         | 10          |
| S868.CB | CYT19.C5  | 3.76         | 8           |
| S868.CB | AUA8.C6   | 3.83         | 10          |
| A873.CB | ADE5.C3’  | 3.76         | 7           |

**FIGURE 6.** Averaged helical parameters for \( \beta \)-RARE DNA bound to MyT1 F5 (solid lines) as calculated with 3DNA. Twist (squares), roll (dots), and width of the major groove (triangles) are shown. The corresponding values for the initial (computer generated) B-form DNA duplex are indicated for comparison (dotted lines).

**FIGURE 7.** Relative binding of different MyT1 ZnFs mutants as calculated from SPR resonance units 2 min after injection of protein onto a \( \beta \)-RARE derivatized sensor chip (\( n = 3–4 \)).
