**Solution Structure and DNA-binding Mode of the Matrix Attachment Region-binding Domain of the Transcription Factor SATB1 That Regulates the T-cell Maturation**

Received for publication, October 6, 2005, and in revised form, December 15, 2005. Published, JBC Papers in Press, December 21, 2005, DOI 10.1074/jbc.M510933200

Hirosi Yamaguchi †§, Masaru Tateno †‡, and Kazuhiro Yamasaki †

From the † Age Dimension Research Center, National Institute of Advanced Industrial Science and Technology, 1-1-1 Higashi, Tsukuba 305-8566, the ‡ Department of Materials Science, Faculty of Engineering, Tottori University, 4-101 Koyama-minami, Tottori 680-8552, and the § Institute of Physics and Center for Computational Sciences, Tsukuba University, Tsukuba 305-8571, Japan

SATB1 is a transcriptional regulator controlling the gene expression that is essential in the maturation of the immune T-cell. SATB1 binds to the nuclear matrix attachment regions of DNA, where it recruits histone deacetylase and represses transcription through a local chromatin remodeling. Here we determined the solution structure of the matrix attachment region-binding domain, possessing similarity to the DNA-binding domain of human SATB1 by NMR spectroscopy. The structure consists of five α-helices, in which the N-terminal four are arranged similarly to the four-helix CUT structure of the CUT domain of hepatocyte nuclear factor 6α. By an NMR chemical shift perturbation analysis and by surface plasmon resonance analyses of SATB1 mutant proteins, an interface for DNA binding was revealed to be located at the third helix and the surrounding regions. Surface plasmon resonance experiments using groove-specific binding drugs and methylated DNAs indicated that the domain recognizes DNA from the major groove side. These observations suggested that SATB1 possesses a DNA-binding mode similar to that of the POU-specific DNA-binding domain, which is known to share structural similarity to the four-helix CUT domain.

The nuclear matrix is a structural component inside the nucleus, to which chromatin binds via matrix (or scaffold) attachment regions (MARs) of the DNA, forming looped chromatin structures (1). MARs are frequently located at the boundaries of transcription units where they are likely to delimit the ends of the active chromatin domains in terms of transcription as well as replication (1). The MAR sequences commonly contain regions where base pairs tend to break under an unwinding stress (base-unpairing region (BUR)), which is important in binding to the nuclear matrix (2).

SATB1 (from special AT-rich sequence-binding protein) was initially identified as a factor that specifically binds to the BUR sequence, which is predominantly expressed in thymus (3). In thymocytes, SATB1 recruits histone deacetylase complex to the MAR site inside the transcription of the interleukin-2 receptor α gene, in order to repress their expression in the premature T-cells (4). Indeed, SATB1-null mice exhibit irregular expression of interleukin-2 receptor α gene and related genes in the premature CD4+/CD8− T-cells, which caused small thymi and spleens and death at the age of 3 weeks (5). SATB1 also regulates the expression of fetal globin genes in the erythroid progenitor cells by directly binding to MARs in the locus control region and the e-globin promoter region in the β-globin cluster (6). Inside the cells, SATB1 is localized at nuclei and surrounds heterochromatin, forming a cage-like network structure (7).

SATB1 is ~800 amino acids in length, in which a region of ~150 amino acids located nearly in the middle (Tyr346−Asn495) is originally reported to be relevant to binding to the MAR DNA (MAR-binding domain (MBD)) (8). This protein also contains a homeodomain located at a more C-terminal region, which alone does not show significant DNA binding activity but enhances the activity of MBD (9). The MBD sequence possesses a similarity to the CUT DNA-binding motif that is found in a group of DNA-binding proteins also containing a homeodomain at a more C-terminal region (10). Recently, a four-helix structure of the CUT domain of hepatocyte nuclear factor 6α has been determined (11). Although this structure showed similarity to those of the POU-specific domains of POU-homologous proteins (12, 13), the DNA-binding mechanism of the CUT domain is yet unknown.

In this study, a three-dimensional structure of SATB1-MBD was determined by NMR spectroscopy. The structure consists of five α-helices packing together into a globular shape, which possesses an additional helix as compared with the HNF6α CUT domain structure. In addition, the DNA-binding mode was elucidated by NMR titration, surface plasmon resonance (SPR), and point mutation experiments and was proposed to be similar to that by POU-specific domains.

**MATERIALS AND METHODS**

Sample Preparation—The cDNA that codes for the SATB1 MAR-binding region (Tyr346−Asn495) was amplified from Quick Clone™ human thymus cDNA mixture (Clontech), by PCR using Pyrobest DNA polymerase (Takara), and was subcloned into the Ndel-BamHI cloning site of pET15b vector (Novagen). Shorter fragments, Val353−Asn490, Asn368−Lys477, and Asn368−Ala455, were also produced by PCR from the above vector. The isotope-labeled and unlabeled proteins were expressed in *Escherichia coli* cells, as described previously (14). The

---

*This work was supported in part by the Ministry of Education, Science, Sports and Culture Grant-in-aid for Scientific Research (C), 17570103, 2005. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

1. To whom correspondence should be addressed: Age Dimension Research Center, National Institute of Advanced Industrial Science and Technology, 1-1-1 Higashi, Tsukuba 305-8566, Japan. Tel.: 81-29-8619473; Fax: 81-29-8612706; E-mail: k-yamasaki@aist.go.jp.

2. The abbreviations used are: MAR, matrix attachment region; BUR, base-unpairing region; m²A, N²-methyladenine; MBD, MAR-binding domain; SPR, surface plasmon resonance; PDB, Protein Data Bank; HSQC, heteronuclear single quantum coherence.

3. In the Protein Data Bank, the solution structure of the N-terminal CUT domain of KIAA1034 protein (SATB2) appears (PDB code 1WIZ), which is highly similar to that of SATB1-MBD.
proteins were purified by nickel-nitritriacetic acid Superflow (Qiagen) and Sephadex G-75 (Amersham Biosciences) column chromatography. The buffers used were 50 mM sodium phosphate (pH 8.0), 500 mM NaCl, 10–250 mM imidazole for nickel-nitritriacetic acid Superflow chromatography and 50 mM sodium phosphate (pH 5.5) for Sephadex G-75 chromatography. Protein concentration was determined by $A_{280}$ values, and molar absorption coefficients were calculated from the amino acid sequences. For NMR measurements, ~1.3–1.7 mg proteins were dissolved in 50 mM sodium phosphate buffer (pH 5.5) containing 0.5 mM sodium 2,2-dimethyl-2-silapentane-5-sulfonate, and 5% D$_2$O, unless otherwise stated.

NMR Measurements, Resonance Assignments, and Structural Calculation—Typical homonuclear and heteronuclear NMR spectra (15, 16) of the Val$^{353}$–Asn$^{490}$ fragment were recorded on Bruker DMX-750 (750.13 MHz for $^1$H and 76.02 MHz for $^{15}$N) and DMX-500 (500.13 MHz for $^1$H, 125.76 MHz for $^{13}$C, and 50.68 MHz for $^{15}$N) spectrometers at 308 K, essentially as described previously (17). The recorded spectra were analyzed for the backbone and side chain resonances assignment, also as described previously (17). HSQC spectra of samples containing 100% D$_2$O were recorded at 293 K, by which 43 hydrogen bond donors were identified. By using the HMBC-J (experiment) (18), 80 $^3$J$_{H\beta}$C values in the calculated region (see below) were obtained. By analyzing the nuclear Overhauser effect spectroscopy, total correlation spectroscopy, and double quantum-filtered correlation spectroscopy spectra, 25 and 4 pairs of H-$\beta$ and valine H-$\gamma$ resonances, respectively, in the calculated region were assigned stereospecifically.

The distance constraints derived from the nuclear Overhauser effect spectra and those maintaining hydrogen bonds were imposed as described (17, 19). The $\phi$ angle constraints were classified into three categories, $-120 \pm 50^\circ$, $-60 \pm 30^\circ$, and $-100 \pm 70^\circ$, corresponding to the $^3$J$_{H\alpha}$C coupling values of $>8.5$, $<6.0$, and $>6.0$–8.5 Hz, respectively. For stereospecifically assigned residues, $\gamma$ torsion angle constraints, classified into three categories, 60 $\pm$ 40$^\circ$, 180 $\pm$ 40$^\circ$, and $-60 \pm 40^\circ$, were imposed.

Structure was calculated for the Asn$^{368}$–Ala$^{455}$ region, because more N-terminal and C-terminal regions appeared to be unstructured in the NMR spectra, and resonance assignments were not completed mainly because of the cross-peak overlapping. Indeed, NMR spectra of the Asn$^{368}$–Ala$^{455}$ fragment were very similar to those of the fragment Val$^{353}$–Asn$^{490}$ as described under “Results.” Random simulated annealing (20) was carried out by using the program CNS (21), essentially as described (19).

SPR—Experiments were carried out at 293 or 288 K using a BiaCORE X apparatus (BIAcore). The running buffer was 50 mM sodium phosphate buffer (pH 5.5) containing 0.005% Tween 20. A total of 338 and 522 resonance units of two double-stranded DNAs (5’-bio-GCTTTCTAATATATGC-3’/5’-GCATAATTm$^{A}$AGAACG-3’) and 5’-bio-GCTTTCTAATATATGC-3’/5’-GCATAATTm$^{A}$ATTAGAAACG-3’ or in the presence of distamycin (Sigma), methyl green (Funakoshi, Japan), or spermine (Wako, Japan). Because methylation at the N$^m$H$_2$ group of adenine base destabilizes the DNA double strands (23), experiments using methylated DNAs were performed at 288 K. Experiments were also performed for the Val$^{353}$–Asn$^{490}$ fragment in 50 mM sodium acetate buffer (pH 5.5), which confirmed that the phosphate ions in the phosphate buffer do not interfere with the DNA binding (data not shown).

NMR Titration Analysis—$^1$H, $^{15}$N HSQC spectra of the protein at the initial concentration of 0.31 mM were recorded at 293 K by adding increasing amounts of 7.8 mM of the double-stranded 16-mer DNA (5’-CGTTTCTAATATATGC-3’/5’-GCATAATTAGAAACG-3’) (the BUR nucleation sequence (22) is underlined). The concentration of the double-stranded DNA was determined by using an extinction coefficient calculated after digestion of the strands with phosphodiesterase I (Worthington). Heteronuclear three-dimensional spectra of SATB1-MBD in the complex with DNA were recorded in order to confirm assignments of backbone $^1$H and $^{15}$N resonances.

Site-directed Mutagenesis—A series of point mutations (Table 2) were introduced on the SATB1-MBD (Val$^{553}$–Asn$^{599}$) expression plasmid, using QuickChange™ XL site-directed mutagenesis kit (Stratagene). After annealing of primers of 30 bases containing mutated triple codons, DNA strands were elongated by KOD Plus DNA polymerase (Toyobo). The mutant proteins were prepared essentially as described above, and SPR experiments were performed also as described above. The error levels in the ratios of the binding constants of the wild-type and mutant proteins ($K_{AWT}/K_{Amut}$) were evaluated as $(\sigma_{WT}^2K_{AWT}^2 + \sigma_{mut}^2K_{AWT}^2)^{-1/2}/K_{AWT}^2$, where $K_{AWT}$ and $K_{Amut}$ represent the binding constants of wild-type and mutant proteins, respectively, and $\sigma_{WT}$ and $\sigma_{mut}$ represent the error levels in $K_{AWT}$ and $K_{Amut}$, respectively.

Modeling of Complex of SATB1-MBD and DNA—Starting from an initial model in which the side of the protein containing the residues with largely affected chemical shifts in the NMR titration experiment was oriented to the major groove side of a standard B-DNA, the most energetically favored structure was selected by a careful and systematic search, essentially as described previously (17).

RESULTS

MAR-binding Domain and CUT Domain—MAR-binding domain region was originally assigned to be Tyr$^{346}$–Asn$^{455}$ for murine SATB1 protein, where the human counterpart possesses a nearly identical sequence (8) (Fig. 1). We have initially expressed and purified a fragment corresponding to this region, and we found that the fragment contained significantly unfolded region(s) (data not shown). Therefore, we prepared shorter fragments Val$^{353}$–Asn$^{490}$, Asn$^{368}$–Lys$^{475}$, and Asn$^{368}$–Ala$^{455}$. All of them were expressed well and successfully purified. Comparison of the NMR spectra suggested that the shortest fragment contains most of the region with folded property and the longer ones possess additional unstructured region(s) (data not shown).

The abilities of the above fragments to bind to a DNA containing a MAR sequence (the BUR nucleation sequence; see Ref. 22) were evaluated by SPR experiments (Fig. 2, a and b). The original MBD region fragment (Tyr$^{346}$–Asn$^{455}$) and a slightly shorter fragment (Val$^{353}$–
Asn⁴⁹⁰ were demonstrated to bind to the DNA with similar binding constants (3.7 ± 0.4) × 10⁻⁹ M⁻¹ and 4.8 ± 0.6) × 10⁻⁹ M⁻¹, respectively (Fig. 2b). The response ratio of the maximal binding of the protein to the DNA immobilized on the sensor chip suggests that SATB1-MBD binds to this DNA at a stoichiometry of 1:1. Nonspecific binding to a DNA with an unrelated sequence was shown to be very weak, which indicates to this DNA is indicated by a 

This is probably because the deletion reduced the net charges of the fragments at neutral pH (difference between the numbers of basic and acidic residues), which need to be positive enough to stably bind to the negatively charged DNA, i.e., the net charges are +2, +2, +1, and -1 for fragments Tyr⁴⁶⁶-Asn⁴⁹⁰, Val⁴⁵⁵-Asn⁴⁹⁰, Asn⁴⁸⁶-Asn⁴⁹⁰, and Asn⁴⁸⁶-Ala⁴⁵⁵, respectively. Indeed, decreasing the pH value to 4.0 partially recovered the binding activity of fragment Asn⁴⁸⁶-Ala⁴⁵⁵ (data not shown). Also, another fragment containing Asn⁴⁸⁶-Leu⁴⁵² and additional four basic residues at the C terminus showed a strong sequence-specific binding (data not shown). Therefore, it is likely that the CUT domain region is mainly responsible for the MAR-DNA binding, although the other unstructured regions contribute to the strong binding at least by increasing the net charges. Considering the above, we used the fragment Val⁴⁵⁵-Asn⁴⁹⁰ as SATB1-MBD for all the experiments hereafter unless otherwise stated.

**Structural Description**—Although the NMR analyses were performed on the Val⁴⁵⁵-Asn⁴⁹⁰ fragment, the structure was calculated for region Asn⁴⁸⁶-Ala⁴⁵⁵, because the N-terminal and C-terminal regions appear to be largely unfolded as described above. The experimental constraints and stereoechemical properties of the NMR solution structure of SATB1-MBD are shown in Table 1. The secondary structure elements are 5 α-helices (α₁, Glu³⁷⁶-Ala³⁸⁶; α₂, Gln⁴³⁹-Ala⁴⁶⁷; α₃, Leu⁴⁰⁴-Lys⁴¹⁴; α₄, Gln⁴²⁰-Gln⁴³⁴; and α₅, Glu⁴³⁷-Gln⁴⁵⁵), as identified by the program Procheck-NMR (24) (Figs. 1 and 3). These helices pack against one another, forming a globular shape as a whole. The packing of the helices was achieved by hydrophobic interactions between side chains, many of which are those of aliphatic or aromatic residues, i.e., Leu³⁷⁵, Tyr³⁷⁶, Trp³⁷⁸, Val³⁷⁹, Glu³⁸², and Leu³⁸³ of α₁; Gln³⁹⁰, Phe³⁹³, Ala³⁹⁷; Val³⁹⁰, and Ala⁴⁰⁷ of α₂; Leu⁴⁰⁴, Leu⁴⁰⁵, Ile⁴⁰⁶, and Leu⁴⁰⁸ of α₃; Leu⁴ⁱ⁷, Leu⁴²⁶, Met⁴⁲⁹, Gln⁴³⁰, Phe⁴³¹, and Leu⁴⁵³ of α₄; and Arg⁴⁴⁶, Ile⁴⁴⁸, and Tyr⁴⁶⁰ of α₅, as defined by average C-C distances of less than 4.5 Å in the structural ensemble. In addition, Val⁴⁵³, Ile⁴⁶⁸, and Phe⁴⁷⁰ in

---

**Figure 1:** Amino acid sequence alignment of the MAR-binding domains of the SATB proteins and CUT domains produced by the ClustalX program (35). Sequences of eight SATB proteins (hsSATB1 and hsSATB2 from Homo sapiens, mmSATB1 and mmSATB2 from Mus musculus, ggSATB1 and ggSATB2 from Gallus galus, and tnSATB1 and tnSATB2 from Tetradon nigroviridis), four ONECUT group proteins (human hepatocyte nuclear factor 6a (hsHNF6a)), human ONECUT2 (hsONECUT2), Drosophila melanogaster ONECUT (dmONECUT), and Danio rerio ONECUT (drONECUT), and two CDP/Cux proteins (D. melanogaster CUT protein (dmCUT)) and human CAAAT displacement protein (hsCDP), each containing three repeated CUT domains, were obtained from the NCBI data base (www.ncbi.nlm.nih.gov). Entry codes are AAH01744 for hsSATB1, AAI73772 for mmSATB1, XP_418746 for ggSATB1, CAB93466 for dmSATB1, and AAD00826 for hsHNF6a. AAL37172 for mmSATB2, XP_421919 for ggSATB2, CAF93666 for tnSATB2, AAD00826 for hsHNF6a, CAB38253 for hsONECUT2, AAL13705 for dmONECUT, AAH66466 for drONECUT, CAA30794 for mmSATB2, and AAB26579 for hsCUT. Numbers above the sequences are for hsSATB1, although those of the first residues of the aligned sequences of the individual proteins are indicated beside the sequences. Basic (Arg and Lys) residues conserved in eight or more of the proteins presented here are colored cyan, although aliphatic (Ile, Leu, Met, and Val) and aromatic (His, Phe, Trp, and Tyr) residues conserved in 17 or more are colored red. Colored boxes above the sequence alignment indicate regions of the helices of SATB1-MBD. Below the sequence, identical and similar residues are marked as produced by the ClustalX program (35). CUT domain region is indicated by a horizontal bar. The arrows indicate the introduced mutational sites (see Table 2).
Solution Structure of SATB1 MAR-binding Domain

The loop regions contribute to formation of the structural core, through hydrophobic interactions. Also, an electrostatic interaction between Glu\(^{382}\) and Arg\(^{440}\) side chains, with an average O\(^-\)–H\(^+\) distance of 3.2 Å, is likely to contribute to the packing of α-helices 1 and 5.

It should be noted that many of the above residues, except those in α-helix 5, are conserved among the CUT domains (Fig. 1, top), indicating that the domains of these proteins share a common structural architecture except for α5. Indeed, structure of the CUT domain of hepatocyte nuclear factor 6 (α5) has been reported to possess four helices (11), which corresponds to α1–4 of SATB1-MBD. More will be described on the structural comparisons below under “Discussion.”

Binding to the Major Groove of DNA—To test to which groove of DNA SATB1-MBD binds, SPR experiments employing DNAs methylated on the major groove side and groove-specific binding drugs, such as methyl green (major groove binding; Ref. 25), spermine (major groove binding; Ref. 26), and distamycin (minor groove binding; Ref. 27), were carried out. When a DNA containing an m\(^6\)A base outside the BUR nucleation sequence (5′-CGTTTCTAATATATGC-3′/5′-GCATATTTAGAAAAGC-3′) (the BUR nucleation sequence (22) is underlined), the binding ability of SATB1-MBD is only slightly affected (Fig. 2c). In contrast, binding to another DNA containing an m\(^6\)A base in the BUR nucleation sequence (5′-CGTTTCTAATATATGC-3′/5′-GCATATTTAGAAAAGC-3′) is weaker by ~40-fold than that of the DNA without modification (Fig. 2c). Because the N\(^6\)H\(_2\) group of adenine base is exposed to the major groove side of DNA, the above results indicate that SATB1-MBD binds to the BUR nucleation sequence from the major groove side. Consistently, experiments using the groove-specific DNA-binding drugs showed the major
groove binding of SATB1-MBD, namely the presence of 50 μM dita-
mycin did not affect the DNA binding of SATB1-MBD, although
presence of 50 μM methyl green or spermine significantly reduced
the apparent affinity of the protein (Fig. 2c). Therefore, it was con-
cluded that SATB1-MBD binds to the DNA from the major groove
side.

Interface for DNA Binding—The protein surface involved in the DNA
binding tends to be positively charged to bind to negatively charged
DNA. On the surface of the SATB1-MBD structure, a characteristic
array of positively charged amino acid side chains, i.e. Lys411–Arg410–
Arg380–Lys401–Arg385, is observed (Fig. 3c). These residues are con-
served in the SATB2 proteins and either of CUT domains of ONECUT
group or Cux/CDP proteins and are considered to be candidate residues
that form the interface for DNA binding (Fig. 1). As described later, the
first four residues of this array are suggested to be involved in the DNA
binding.

To elucidate the protein-DNA interface, an NMR titration experi-
ment was carried out (Fig. 4a). By adding increasing amounts of the
DNA, chemical shift perturbations were observed in the HSQC spectra
as follows. The positions of some cross-peaks did not change at all, or
they shift only slightly so that the chemical shift changes were easy to
follow (e.g. Ser372, Ser373, Phe398, Glu437, and Gln430 H-ε in Fig. 4a).
For some cross-peaks, even though the differences were significant, the
change could be followed as the fast-exchange manner and/or reso-
nances were assigned by analyses of three-dimensional spectra (e.g.
Asp381, Asp414, Leu409, and Arg427 in Fig. 4a). For others, however, the
changes could not be followed, for which distances to the nearest unas-
signed cross-peaks of the complex were temporally considered as the
chemical shift difference or classification into the highest category was
applied if there were no such candidate peaks within the classification
limit (Δδαrms(2 ± 4)/4.5)1/2 ≥ 0.2 ppm) (e.g. Gln390, Thr401, Gln402, Gln420,
Glu430, and Gln390(H-ε) in Fig. 4a). This treatment may cause underes-
terimation of the difference but causes no overestimation except for the

case where a strong exchange broadening occurs without large change
in the chemical shift. The chemical shift perturbations were completed
when the concentration ratio of DNA to protein reached ~1.0, which
suggested a 1:1 binding stoichiometry of the protein-DNA interaction.

After classifying the residues according to their chemical shift differ-
ences, it became apparent that a relatively limited region of the structure
was largely affected by the binding of DNA (Fig. 4b). These regions are
loop between α1 and α2 (loop 12), α3, loop between α3 and α4 (loop 34),
and α4. In contrast, most residues of α1 and α5 are not significantly
affected. Considering the above, a model of the complex of SATB1-
MBD and DNA was built using a computational approach (Fig. 4b).
In this model, α-helix 3 deeply enters the major groove of the DNA in such
a way that the axis of the α-helix is nearly perpendicular to that of
the DNA double helix. Accordingly, the side chains of Ser380, Glu384, Arg410,
and Lys411 in α3 contact the DNA bases of the ~4-bp region, probably
forming hydrogen bonds. The side chains of Arg380, Lys384, Ser389,
Glu390, Arg400, Arg410, Ser419, and Ser421, and the backbone amide of
Gln390 form intramolecular hydrogen bonds to the DNA phosphate
groups. The NMR titration experiment is highly consistent with these
interactions, e.g. cross-peaks of both the side chain and the backbone
signals of Gln390 are significantly affected upon adding DNA, as shown
in Fig. 4a. It should be noted that Ala328 and neighboring residues in the
middle of α4 are also mainly affected in the NMR titration experiment,
although the residues are distant from the interface to DNA. This sug-
gests a possibility that α4 slightly kinks upon binding to DNA, in order
to achieve better intramolecular fitting.

To evaluate the reliability of the proposed model, point mutations at
8 Arg/Lys residues and 3 Ser residues were introduced, and their effects
on the DNA binding ability were evaluated by SPR (Fig. 2d and Table 2).
It should be noted that all the Arg/Lys to Asn mutations decreased the
affinity constants at least by ~3-fold. This is likely to be at least partially
because of the effect of reducing net charge (from +2 to +1) by elimi-
nating a basic residue. However, three mutations at Lys384, Arg410, and

|$a$ Values calculated with the force field for NMR structure determination in the CNS package (21).

|$b$ Values calculated with Val371–Gln445.
Lys416 decreased affinity constants more significantly (>10-fold) (Fig. 2 and Table 2). The first two residues contact the base and/or phosphate of DNA in the proposed model, as described above, although Lys416 is significantly distant from the interface to the DNA. In the structure, the side chain of this residue is capable of forming a salt bridge with those of Glu370 and Asp414. It is likely that the mutation of Lys416 affected structure at least locally, which is not suitable for DNA binding. It should be noted that Glu370, Asp414, and Lys416 are conserved only in the SATB proteins (SATB1 and related SATB2 proteins; Fig. 1), indicating that the above possible salt bridge network is specific for the SATB proteins. Thus, although a possibility cannot be excluded that Lys416 is directly involved in DNA binding through relatively large conformational changes of loop 34, this is not the case with the CUT domains of the other families.

Four of the other Arg/Lys residues with relatively small mutational effects on the DNA binding (Arg380, Arg395, Arg427, and Arg442; Table 2) are distant from the protein-DNA interface in the model, except for Arg380 in α1. Arg380 forms a hydrogen bond with a DNA backbone phosphate, as described above. Although the reason why the mutation at this site does not have a large effect is unknown, it is possible that an Asn residue at this site also forms a hydrogen bond to the DNA phosphate. Indeed, Thr or Ser residues, which are also capable of hydrogen bonding, are conserved at this position in the ONECUT group of proteins (Fig. 1). Lys416 in the C-terminal unstructured loop also showed a relatively small mutational effect. Although the basic residues in the unstructured regions were considered to be important in increasing the net charge, as described above, a possibility that their side chains directly associate with the DNA phosphate cannot be excluded.

Among the three mutants of Ser residues, only that of Ser406 in α3 induced a large decrease in the affinity to DNA (Fig. 2 and Table 2). This residue contacts the DNA bases possibly forming hydrogen bonds, as described above. The other two Ser residues in the N- or C-terminal loop are likely to be distant from the interface. Therefore, the model proposed in the present study is largely consistent with the mutational effects on the DNA binding activity and is likely to be reliable at least to define the framework of the protein-DNA binding. To reveal the
FIGURE 4. DNA-recognition mode of SATB1-MBD. a, an overlay of selected regions of $^{1}H-^{15}N$ HSQC spectra of SATB1-MBD in the absence (black) or the presence of 0.25 (blue), 0.5 (cyan), 0.75 (green), 1.0 (magenta), 1.25 (brown), and 1.5 (red) times molar concentrations of the 16-mer DNA (5'–CGTTTCTAATATATGC–3' / 3'–GCATATATTAGAAACG–5') (the BUR nucleation sequence (22) is underlined). The continuous arrows indicate changes that could be followed, and the dashed arrows indicate temporarily assumed changes to the nearest unassigned peaks (see text). Spectra were recorded at the proton frequency of 750 MHz. 

b, a predicted model of the complex of SATB1-MBD (a ribbon (and wires for selected residues) of different colors described below) and a standard B-DNA molecule (orange wire) in stereo view. Amino acid residues with backbone or side chain (Asn, Gln, or Trp) chemical shifts affected by the DNA binding are colored red ($\Delta\delta H^2 + (\Delta\delta D/5.0)^2)^{1/2} < 0.2$ ppm: residues 384, 387, 390, 401, 404, 406, 407, 414, 417, 419, 421, 422, 428; or yellow ($\Delta\delta H^2 + (\Delta\delta D/5.0)^2)^{1/2} < 0.1$ ppm: residues 381, 392–394, 397, 399, 402, 405, 410, 411, 416, 420, 423, 424, 427, 429, 430, 431; although those only slightly affected ($\Delta\delta H^2 + (\Delta\delta D/5.0)^2)^{1/2} < 0.1$ ppm: residues 370–380, 382, 383, 385, 386, 389, 391, 395, 396, 398, 400, 408, 409, 412, 413, 418, 425, 426, 432–435, 437–446, 448–450 are shown in blue. Residues for which no meaningful information is available, i.e. those with unobserved backbone resonances in the free protein, including proline residues (residues 403, 415, 436, 447), are shown in white. The presented region of the protein is Glu370–Arg450. Mutated residues in the presented region, DNA-contacting residues in the present model, and Ala428 (see text) are indicated by the wire representation. The major and minor grooves of DNA are indicated by M and m, respectively. The figure was produced by the Insight II molecular display program (Accelrys).
Solution Structure of SATB1 MAR-binding Domain

sequence-specific DNA-recognition mechanism, however, future experiments on the structure determination of the protein-DNA complex are necessary.

DISCUSSION

Major Groove Binding—In contrast to the present results, a previous report suggested the minor groove binding of the full-length SATB1 protein (3). This is mainly based on the result that a minor groove-specific binding drug, distamycin, at high concentrations (>25 μM) competes out the DNA binding of the SATB1 protein in a gel retardation experiment, which requires more stable binding than the SPR experiment. In the present SPR experiment, distamycin at a higher concentration (50 μM) did not interfere with DNA binding by MBD (Fig. 2c). It should be noted that the full-length protein includes the homeodomain that significantly enhances the affinity of the protein and DNA (9), which is likely to ensure that the binding is stable enough to be detected in the gel retardation experiments. It was reported that distamycin at even lower concentrations (>5 μM) inhibits the DNA binding of the Antennapedia homeodomain (28) that is known to bind specifically to the major groove. Therefore, the previous observation of chemical competition is possibly because of its competition to homeodomain but not to MBD.

Comparison with the Structures of CUT, POU-specific and Related Helix-Turn-Helix DNA-binding Domains—The CUT domain of HNF6α was reported to possess a four-helix structure (11), as shown in Fig. 5a, although the SATB1-MBD region similar to the CUT domain possesses an additional C-terminal helix, α5. The HNF6α CUT domain structure was shown to be similar to the POU-specific DNA-binding domains of the POU-homologous homeodomain proteins (12, 13, 29–31). The common structure of the POU-specific domain contains four helices arranged similarly to the HNF6α CUT domain and the N-terminal four helices of SATB1-MBD, although an atypical variation possessing an additional N-terminal helix was reported for the HNF1α POU-specific domain (31) (Fig. 5, b and c). Furthermore, the POU-specific domain is known to show similarity to the helix-turn-helix DNA-binding domains of phage A and 434 repressors and 434 Cro proteins (13). Although these prokaryotic DNA-binding domains possess five helices (32), the relative position of the fifth helix, which is known to be required for the dimer formation, and the length of the fourth helix are substantially different from SATB1-MBD (Figs. 3b and 5d).

It is also clear that the DNA-binding mode of the POU-specific domain is very similar to the proposed model of the complex of SATB1-MBD and DNA (Fig. 4b, and Fig. 5, b and c), in which the third helix (of the common four-helix structure) deeply enters the major groove of DNA in order to contact bases. Also for λ/434 repressors and 434 Cro DNA-binding domains, the third helix acts as the DNA-recognition helix. Therefore, these groups of the DNA-binding domains are likely to possess a common DNA-binding framework. It is especially noted that at the N terminus of the second α-helix, a Gln residue is strictly conserved for all these proteins (residue 390 of SATB1 in Fig. 1) (13, 31). This residue forms a hydrogen bond to a DNA phosphate group and is likely to contribute to defining the DNA-binding framework. Therefore, it is concluded that SATB1-MBD is likely to be classified into the same helix-turn-helix group as the POU-specific domain, the CUT domain, and the λ/434 repressors/434Cro DNA-binding domains, commonly containing four helices, although it is atypical with regard to the C-terminal helix. The role of this additional helix of SATB1-MBD is highly likely to be stabilizing the structural core, mainly by hydrophobic interactions between Val396(α2)–Ile443(α5), Phe432(α4)–Arg440(α5), and Phe432(α4)–Tyr444(α5).

Despite the similarity in the DNA-binding modes of the above domains, the prokaryotic and eukaryotic proteins containing the above domains are different in the DNA-binding modes of the whole molecules, i.e. the prokaryotic proteins form homodimers of the above domains when bound with DNA, although the eukaryotic proteins possess homeodomains and bind to DNA without forming homodimeric protein-protein contacts. Furthermore, among the eukaryotic members, the POU-homologous proteins and CUT-like proteins, including the SATB proteins, are different in that the homeodomains in the former proteins are involved in the specific recognition of DNA (29–31), although the homeodomains of CUT-like proteins are not required in the specific recognition or only assist the CUT-like domains (9, 10, 33).

It should be noted that the residues in the recognition helices of the helix-turn-helix DNA-binding domains described here are not conserved very much, which should cause differences in the DNA recognition.

---

**TABLE 2**

| Mutation | Location | $K_a$ | -Fold decrease
|----------|----------|-------|----------------|
| Wild type | a1 | 4.8E6 ± 6.0E5 | |
| R380N | a1 | 1.1E6 ± 1.7E5 | 4.5 ± 0.9 |
| R384N | a1 | 2.6E5 ± 2.4E5 | 18.7 ± 2.9 |
| R392N | a2 | 1.1E6 ± 8.7E4 | 4.6 ± 0.7 |
| R410N | α3 | 1.3E5 ± 1.9E4 | 36.8 ± 7.0 |
| K416N | loop 34$^a$ | 3.6E5 ± 7.6E4 | 13.4 ± 3.3 |
| R427N | α4 | 1.6E6 ± 1.2E5 | 3.1 ± 0.5 |
| R492N | α5 | 1.0E6 ± 7.0E4 | 4.7 ± 0.7 |
| K475N | C-loop$^b$ | 6.3E5 ± 5.1E4 | 7.7 ± 1.1 |
| S373A | N-loop$^c$ | 3.9E6 ± 3.8E5 | 1.2 ± 0.2 |
| S406A | α3 | 4.4E5 ± 1.4E5 | 11.0 ± 1.4 |
| S451A | C-loop | 3.6E6 ± 9.7E5 | 1.4 ± 0.4 |

$^a$ Error levels were estimated as described under "Materials and Methods."

$^b$ Loop is between α3 and α4.

$^c$ Loop is C-terminal to α5.

$^d$ Loop is N-terminal to α1.

---

**FIGURE 5.** Structures of HNF6α CUT domain (PDB code 1s7e) (a), POU-specific domains of Pit-1 (PDB code 1au7) (b), HNF1 α (PDB code 1ic8) in the complex with DNA (c), and N-terminal DNA-binding domain of phage 434 repressor (PDB code 1r69) (d), b and c, DNA strands are shown in red loops. Loops that lack coordinates in d are connected by dashed lines. The structures are viewed from similar orientation as the SATB1-MBD structure in the right panel of Fig. 3b. The equivalent secondary structures are shown in the same colors as used for SATB1-MBD in Fig. 3b. The figure was produced by Molscript (36).
sequence. Even among the three subgroups of CUT domain proteins, including the SATB family, the residues of the α3 region are similar but slightly different, i.e., LSSELIRK for the SATB proteins, TLSDDLRRN for the ONECUT group proteins, and (S/T)VS(D/E)L/I/M(K/R) for the CDP/Cux proteins (Fig. 1) (commonly conserved or similar residues are underlined). This is very likely to cause similarity and differences in the DNA recognition sequence. Indeed, the recognition sequences of SATB1 (BUR nucleation sequence), HNF6α, and CDP R2/R3 are ATATAT, ATCGAT, respectively (10, 22, 33).

Biological Implications of the SATB1/MAR-DNA Interaction—The BUR sequence is known to be essential to the binding of MAR-DNAs and the nuclear matrix, in which the tendency of the base unpairing appears to be important (2). Although the SATB1 protein is found as a factor that specifically binds to the BUR sequence, the present experimental results are consistent with a model in which SATB1-MBD binds to a double-stranded DNA in the standard B-form from the major groove side. This complex should stabilize the double-stranded MAR-DNA in the B-form and protect it from the base-unpairing force. Therefore, it is likely that the binding of SATB1 to MAR competes with the binding of MAR and the nuclear matrix, if the base-unpaired DNA is assumed to be involved in the binding to the nuclear matrix. By considering that SATB1 recruits the histone deacetylase from the base-unpairing force. Therefore, it is likely that the binding of SATB1 to MAR competes with the binding of MAR and the nuclear matrix, if the base-unpaired DNA is assumed to be involved in the binding to the nuclear matrix. By considering that SATB1 recruits the histone deacetylase complex in order to inactivate the gene expression through the chromatin remodeling (4), and that SATB1 itself possesses a nuclear matrix targeting sequence that is important in the transcriptional repression (34), the multistep gene-silencing reactions driven by SATB1 would be as follows: (i) translocation of SATB1 to the nuclear matrix, (ii) competing out the binding of the nuclear matrix and MAR-DNA, and (iii) chromatin remodeling by histone deacetylation.

Acknowledgments—We thank K. Hiratsu, P. Reay, T. Koyama, M. Kuwahara, T. Hamada, K. Yamamoto, K. Inoue, T. Tanabe, and T. Yamasaki at National Institute of Advanced Industrial Science and Technology (AIST) for helpful technical advice. We thank M. Ikemoto, S. Kurachi, K. Suzuki, and H. Shinsaki at AIST for use of the experimental facilities. We also thank K. Kurachi, M. Ohme-Takagi, and S. Nishikawa at AIST for support and encouragement.

REFERENCES

1. Gasser, S. M., and Laemmli, U. K. (1987) Trends Genet. 3, 16–22
2. Bode, J., Kohwi, Y., Dickinson, L., Joh, T., Klehr, D., Malik, C., and Kohwi-Shigematsu, T. (1992) Science 255, 195–197
3. Dickinson, L. A., Joh, T., Kohwi, Y., and Kohwi-Shigematsu, T. (1992) Cell 70, 631–645
4. Yasui, D., Miyano, M., Cai, S., Varga-Weisz, P., and Kohwi-Shigematsu, T. (2002) Nature 419, 641–645
5. Alvarez, J. D., Yasui, D. H., Niida, H., Joh, T., Loh, D. Y., and Kohwi-Shigematsu, T. (2000) Genes Dev. 14, 521–535
6. Wen, J., Huang, S., Rogers, H., Dickinson, L. A., Kohwi-Shigematsu, T., and Noguchi, C. T. (2005) Blood 105, 3330–3339
7. Cai, S., Han, H. J., and Kohwi-Shigematsu, T. (2003) Nat. Genet. 34, 42–51
8. Nakagomi, K., Kohwi, Y., Dickinson, L. A., and Kohwi-Shigematsu, T. (1994) Mol. Cell. Biol. 14, 1852–1860
9. Dickinson, L. A., Dickinson, C. D., and Kohwi-Shigematsu, T. (1997) J. Biol. Chem. 272, 11463–11470
10. Lannoy, V. L., Brügin, T. R., Rousseau, G. G., and Lemaigre, F. P. (1998) J. Biol. Chem. 273, 13552–13562
11. Sheng, W., Yan, H., Rausa, F. M., III, Costa, R. H., and Liao, X. (2004) J. Biol. Chem. 279, 33928–33936
12. Dekker, N., Cox, M., Boelens, R., Verrijzer, C. P., van der Vliet, P., and Kaptein, R. (1993) Nature 362, 852–855
13. Assa-Munt, N., Mortishire-Smith, R. J., Aurora, R., Herr, W., and Wright, P. E. (1993) Cell 73, 193–205
14. Reay, P., Yamashita, K., Terada, T., Kuramitsu, S., Shirouzu, M., and Yokoyama, S. (2004) Proteins Struct. Funct. Bioinfo. 56, 40–51
15. Wüthrich, K. (1986) NMR of Proteins and Nucleic Acids, pp. 44–92, John Wiley & Sons, Inc., New York
16. Bax, A. (1994) Curr. Opin. Struct. Biol. 4, 738–744
17. Yamashita, K., Kigawa, T., Inoue, M., Tatemoto, M., Yamashita, T., Yabuki, T., Aoki, M., Seki, E., Matsuda, T., Nakanoue, E., Ishizuka, Y., Terada, T., Shirouzu, M., Osnari, T., Tanaka, A., Seki, M., Shinozaki, K., and Yokoyama, S. (2004) J. Mol. Biol. 337, 49–63
18. Kay, L. E., and Bax, A. (1989) J. Magn. Reson. 86, 110–126
19. Yamashita, K., Kigawa, T., Inoue, M., Tatemoto, M., Yamashita, T., Yabuki, T., Aoki, M., Seki, E., Matsuda, T., Tono, Y., Hayami, N., Terada, T., Shirouzu, M., Osnari, T., Tanaka, A., Seki, M., Shinozaki, K., and Yokoyama, S. (2004) Plant Cell 16, 3448–3459
20. Nilges, M., Clore, G. M., and Gronenborn, A. M. (1988) FEBS Lett. 239, 129–136
21. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Grosse-Kunstleve, R. W., Jiang, J. S., Kuzewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Acta Crystallogr. Sect. D. Biol. Crystallogr. 54, 905–921
22. Kohwi-Shigematsu, T., and Kohwi, Y. (1990) Biochemistry 29, 9551–9560
23. Yamashita, T., Yamashita, K., and Suzuki, M. (1998) Proc. Jpn. Acad. Ser. B Phys. Biol. Sci. 74, 210–215
24. Laskowski, R. A., Rollman, J. A. C., MacArthur, W. M., Kaptein, R., and Thornton, J. M. (1996) J. Biomol. NMR 8, 477–486
25. Norden, B., and Tjerneld, F. (1977) Chem. Phys. Lett. 50, 508–512
26. Drew, H. R., and Dickerson, R. E. (1981) J. Mol. Biol. 151, 535–556
27. Luck, G., Triebel, H., Waring, M., and Zimmer, C. (1974) Nucleic Acids Res. 1, 503–530
28. Dorn, A., Affolter, M., Müller, M., Gehring, W. J., and Leupin, W. (1992) EMBO J. 11, 279–286
29. Klemm, J. D., Rould, M. A., Aurora, R., Herr, W., and Pabo, C. O. (1994) Cell 77, 21–32
30. Jacobson, E. M., Li, P., Leon-del-Rio, A., Rosenfeld, M. G., and Aggarwal, A. K. (1997) Genes Dev. 11, 198–212
31. Chi, Y.-i., Frantz, J. D., Oh, B.-C., Hansen, L., Dhe-Paganon, S., and Shoehorn, S. E. (2002) Mol. Cell 10, 1129–1137
32. Mondragon, A., Soubia, S., Almo, S. C., Drottar, M., and Harrison, S. C. (1989) J. Mol. Biol. 205, 180–200
33. Truscott, M., Raynal, L., Wang, Y., Bérubé, G., Leduy, L., and Nepveu, A. (2004) J. Biol. Chem. 279, 49787–49794
34. Seo, J., Zhowno, N. M., and Dudley, J. P. (2005) J. Biol. Chem. 280, 24600–24609
35. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997) Nucleic Acids Res. 25, 4876–4882
36. Kraulis, P. J. (1991) J. Appl. Crystallogr. 24, 946–950
37. Koradi, R., Billeter, M., and Wüthrich, K. (1996) J. Mol. Graphics 14, 51–55

Solution Structure of SATB1 MAR-binding Domain