Mechanisms of DNA Recognition by Transcription Factor c-Myb

Oda, Masayuki

(Degree)
博士（理学）

(Date of Degree)
1998-03-11

(Date of Publication)
2012-07-12

(Resource Type)
doctoral thesis

(Report Number)
乙2218

(JaLCDOI)
https://doi.org/10.11501/3141263

(URL)
https://hdl.handle.net/20.500.14094/D2002218

神戸大学博士論文

Mechanisms of DNA Recognition by Transcription Factor c-Myb

平成10年1月

織田昌幸

神戸大学博士論文

Mechanisms of DNA Recognition by Transcription Factor c-Myb

（転写因子 c-MybによるDNA認識機構）

January, 1998

Masayuki Oda

CONTENTS

General Introduction

Specific molecular recognition characterizes individual biological macromolecules in the mediation of many biological phenomena. Protein-DNA interactions are the origin of complex genetic regulation. Recently, a number of complex structures of DNA duplexes and proteins determined at atomic resolution have revealed that nature uses a wide variety of readout mechanisms (Pabo and Sauer, 1992; Werner *et al.*, 1996). In DNA-bound proteins, the specific base sequences in the DNA are directly and indirectly recognized through the hydrogen bonds, the van der Waals attractions, and the hydrophobic interactions, in addition to the non-specific electrostatic interactions between the negatively charged phosphate backbone and the basic residues of the protein. Although structural studies have advanced our understanding of protein-DNA recognition, some of the most critical questions remain. They include the conformational changes of protein and DNA, which are needed for the formation of the complementary surfaces, and the coordination of water molecules, which fill cavities or mediate intermolecular hydrogen bonds.

These additional events take place which are more difficult to understand based on structural data alone. All these events affect the thermodynamics of the equilibrium. They also influence the sequence specificity, i.e. the difference in binding free energy for the formation of sequence-specific compared to non-specific complexes. All events make both enthalpic and entropic contributions to the binding free energy. Of these, entropic contributions are especially difficult to estimate based on structural information. Additionally, with site-directed mutagenesis on both protein and DNA, the individual contacts between them and the roles of the substituted amino acid and base could be investigated quantitatively. These facts motivate the use of protein engineering and thermodynamic studies to complement structural data for the elucidation of various molecular contributions to binding affinity and specificity.

In this thesis, both a filter binding assay and calorimetry were used as experimental methods to detect the binding affinity. A filter binding assay is very useful for strong binding ($K_a > 10^7\ M^{-1}$), and needs only a small amount of sample. On the other hand, isothermal titration calorimetry (ITC) can analyze the weak binding affinity ($10^2 < K_a < 10^8\ M^{-1}$), and also detect the binding enthalpy change directly. So far, only a small number of investigations for protein-DNA interactions have been reported, because their affinities are too high to be measured using ITC (Ladbury *et al.*, 1994; Hyre and Spicer, 1995; Merabet and Ackers, 1995; Lundbäck and Härd, 1996). Using these two methods, all the protein-DNA interactions can be analyzed thermodynamically.

In order to understand the DNA recognition mechanisms, I focused on the interactions between the c-*myb* gene product (c-Myb) and the DNA. Based on the analyses of the interaction between c-Myb and DNA, using both a filter binding assay and ITC, I would like to deepen our understanding of specific DNA recognition. What is the difference between specific recognition and non-specific recognition? What contacts with the bases and the DNA backbone allow specific recognition? Is there anything else that determines specific recognition?

c-Myb is a transcriptional activator that binds with a dissociation constant of about 10^{-9} M to the specific DNA sequence PyAAC$^{G}/_{T}$G, where Py indicates a pyrimidine (Biedenkapp *et al.*, 1988; Weston, 1992; Tanikawa *et al.*, 1993). The DNA-binding domain of c-Myb consists of three imperfect 52 residue repeats (designated R1, R2, and R3 from the N-terminus) (Gonda *et al.*, 1985; Klempnauer and Sippel, 1987; Sakura *et al.*, 1989). Analysis of deletion mutants has indicated that the last two repeats, R2 and R3, are the minimum unit for specific DNA-binding (Sakura *et al.*, 1989; Howe *et al.*, 1990). The solution structures of the specific DNA complex of the minimum DNA-binding domain (R2R3) and its free form were determined by multidimensional NMR techniques (Ogata *et al.*, 1992, 1994, 1995). Both R2 and R3 are composed of three helices with a helix-turn-helix variation motif, and each third helix in R2 and R3 is

engaged in direct and specific base recognition. The overall structures of these repeats are almost identical: the root-mean-square-deviation (r.m.s.d.) of the backbone heavy atoms is only 0.81 ± 0.11 Å between R2 and R3 in their free forms, although their sequence identity in the tandem repeats is only 31%. R2 and R3 are closely packed in the major groove, so that the two recognition helices directly contact each other to bind cooperatively to the specific base sequence.

In the complex of c-Myb R2R3 with the Myb-binding DNA sequence (MBS-I), the consensus A4, the counterpart guanine of C6, and the last G8 directly interact with Asn183, Lys182, and Lys128, respectively (Ogata *et al.*, 1994). There are several additional possible base contacts, that are less well-defined in the NMR structure, for example, at Py3 with Ser187, and at A5 and its counterpart thymine with Asn136, Asn179, and Asn186 (Ogata *et al.*, 1994). These interactions can also be explained by the available experimental data concerning the effects of both amino acid substitutions and base substitutions on DNA-binding (Saikumar *et al.*, 1990; Gabrielsen *et al.*, 1991; Frampton *et al.*, 1991; Tanikawa *et al.*, 1993). In addition, not only the residues in the recognition helix, but also the residues in the hydrophobic core, including three conserved tryptophans in each repeat, are indispensable for specific DNA-binding (Kanei-Ishii *et al.*, 1990). The strong cooperativity between R2 and R3 originates from the putative polar interactions between the side-chains of Glu132 and Asn179, and between those of Arg131 and Asp178 (Ogata *et al.*, 1994).

In chapter 1, in order to investigate the role of Ser187 and the origin of this pyrimidine preference at the initial base of the consensus sequence, both Ser187 in the c-Myb R2R3 and the third T-A base-pair in the 22-mer MBS-I fragment were substituted by other amino acids and other base-pairs, respectively, and the interactions between the R2R3 mutants and the base-pair substituted MBS-I fragments were analyzed using a filter binding assay. The results indicated that the principal origin of this base specificity should not only occur from the direct-readout mechanism, namely, the direct recognition of the

third base by the amino acid at 187, but also from the indirect-readout mechanism, namely, the overall shape of the DNA. Additionally, following the conventional chemical rules of the direct-readout mechanism, amino acid mutagenesis at position 187 yielded several new base preferences for the protein.

In chapter 2, in order to elucidate the reason for the imperfection of the tandem R2 and R3 repeats at amino acid positions other than the recognition helices, a series of R2R3 mutants was generated by swapping the N-terminus, the first helix (helix-1), and the second helix (helix-2) in R2 to those in R3, and the binding affinities of these R2R3 mutants to the 22-mer MBS-I fragment were analyzed using a filter binding assay. Further mutational studies revealed that the only two residues are indispensable, which are involved in the hydrophobic core of R2, and do not directly interact with the DNA. The results indicated that the characteristic packing of R2, which originates from the hydrophobic core, should be important for the specific DNA recognition.

In chapter 3, in order to explore the DNA recognition mechanisms by the c-Myb, the thermodynamics of c-Myb R2R3 binding to DNA was characterized, using ITC to measure both the binding enthalpy change and the corresponding binding affinity. The binding reactions of R2R3 to non-cognate DNAs, in which the base-pairs in the consensus sequence were substituted, were analyzed and compared to the binding to the cognate DNA. In addition, the interactions of the R2R3 mutant proteins with the various DNAs were also investigated using ITC, to reveal the roles of the substituted residues. Thermodynamic analyses advanced our understanding of the molecular recognition, such as the individual contacts between amino acids and bases, the roles of water molecules, and the local folding mechanisms. Furthermore, the calorimetric measurements provided us much information about the weak binding, including non-specific binding, which could not be analyzed using only a filter binding assay. The characteristic thermodynamic features of specific and non-specific DNA binding can also be discussed.

References

Biedenkapp, H., Borgmeyer, U., Sippel, A. E., and Klempnauer, K. -H. (1988) *Nature*, **335**, 835-837.

Frampton, J., Gibson, T. J., Ness, S. A., Doderlein, G., and Graf, T. (1991) *Protein Eng.*, **4**, 891-901.

Gabrielsen, O. S., Sentenac, A., and Fromageot, P. (1991) *Science*, **253**, 1140-1143.

Gonda, T. J., Gough, N. M., Dunn, A. R., and de Blaquiere, J. (1985) *EMBO J.*, **4,** 2003-2008.

Howe, K. M., Reakes, C. F. L., and Watson, R.J. (1990) *EMBO J.*, **9**, 161-169.

Hyre, D. E., and Spicer, L. D. (1995) *Biochemistry*, **34**, 3212-3221.

Kanei-Ishii, C., Sarai, A., Sawazaki, T., Nakagoshi, H., He, D. -N., Ogata, K., Nishimura, Y., and Ishii, S. (1990) *J. Biol. Chem.*, **265**, 19990-19995.

Klempnauer, K. -H., and Sippel, A. E. (1987) *EMBO J.*, **6,** 2719-2725.

Ladbury, J. E., Wright, J. G., Sturtevant, J. M., and Sigler, P. B. (1994) *J. Mol. Biol.*, **238**, 669-681.

Lundbäck, T., and Härd, T. (1996) *Proc. Natl. Acad. Sci., U.S.A*, **93**, 4754-4759.

Merabet, E., and Ackers, G. K. (1995) *Biochemistry*, **34**, 8554-8563.

Ogata, K., Hojo, H., Aimoto, S., Nakai, T., Nakamura, H., Sarai, A., Ishii, S., and Nishimura, Y. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 6428-6432.

Ogata, K., Morikawa, S., Nakamura, H., Sekikawa, A., Inoue, T., Kanai, H., Sarai, A., Ishii, S., and Nishimura, Y. (1994) *Cell*, **79**, 639-648.

Ogata, K., Morikawa, S., Nakamura, H., Hojo, H., Yoshimura, S., Zhang, R., Aimoto, S., Ametani, Y., Hirata, Z., Sarai, A., Ishii, S., and Nishimura, Y. (1995) *Nature Struct. Biol.*, **2**, 309-320.

Pabo, C. O., and Sauer, R. T. (1992) *Annu. Rev. Biochem.*, **61**, 1053-1095.

Saikumar, P., Murali, R., and Reddy, E. P. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 8452-8456.

Sakura, H., Kanei-Ishii, C., Nagase, T., Nakagoshi, H., Gonda, T. J., and Ishii, S. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 5758-5762.

Tanikawa, J., Yasukawa, T., Enari, M., Ogata, K., Nishimura, Y., Ishii, S., and Sarai, A. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 9320-9324.

Werner, M. H., Gronenborn, A. M., and Clore, G. M. (1996) *Science*, **271**, 778-784.

Weston, K. (1992) *Nucleic Acids Res.*, **20**, 3043-3049.

Chapter 1

Investigation of the Pyrimidine Preference by the c-Myb DNA-binding Domain at the Initial Base of the Consensus Sequence

1.1 Introduction

Specific interactions between proteins and DNA are critical to gene expression and regulation, so a general readout mechanism of the information encoded in DNA has been sought (Seeman *et al.*, 1976; Lehming *et al.*, 1990; Pabo and Sauer, 1992). However, a number of complex structures of DNA duplexes and proteins determined at atomic resolution have revealed that nature uses a great variety of readout mechanisms (Otwinowski *et al.*, 1988; Lawson and Carey, 1993; Travers, 1992; Winkler *et al.*, 1993; Kim *et al.*, 1993; Schumacher *et al.*, 1994; Klimasauskas *et al.*, 1994; Vassylyev *et al.*, 1995).

In most complex structures, the direct-readout mechanism is mediated by intermolecular hydrogen bond networks and hydrophobic interactions between DNA duplexes and proteins. The interaction modes have been classified into (i) the intrinsic chemical features of bases and amino acids (Seeman *et al.*, 1976; Pabo and Sauer, 1992; Suzuki and Yagi, 1994), and (ii) the stereochemical relations between the amino acids and the bases inside the DNA major grooves (Suzuki and Yagi, 1994; Suzuki, 1993). In contrast, the indirect-readout mechanism works in several systems, such as the *trp* repressor/operator (Otwinowski *et al.*, 1988; Lawson and Carey, 1993), where the DNA bases are specifically recognized by proteins without the use of particular hydrogen bonds or non-polar contacts. Instead, each sequence-dependent deformation of the DNA conformation stabilizes the characteristic geometry of the phosphate backbone, which directly interacts with the protein through polar contacts (Travers, 1992). Water molecules are often observed to mediate the specific interaction through additional hydrogen bonds (Otwinowski *et al.*, 1988; Lawson and Carey, 1993). One common type of DNA deformation is a steep kink of the duplex (Winkler *et al.*, 1993; Kim *et al.*, 1993; Schumacher *et al.*), which substantially contributes to readout of the minor groove (Kim *et al.*, 1993; Schumacher *et al.*).

In general, a combination of the direct- and indirect-readout mechanisms results in specific base-pair recognition. In other words, both the specific binding affinity and the DNA bending contribute to the free energy of complex formation (Travers, 1992; Pontiggia *et al.*, 1994). This situation has made it difficult to determine how each consensus base sequence is recognized by the corresponding protein, even when the precise complex structure is known.

In the complex of c-Myb R2R3 with the Myb-binding DNA sequence (MBS-I), the consensus A4, the counterpart guanine of C6, and the last G8 directly interact with Asn183 in R3, Lys182 in R3, and Lys128 in R2, respectively (Figure 1-1) (Ogata *et al.*, 1994). The strong cooperativity between R2 and R3 originates from the putative polar interactions between the side-chains of Glu132 and Asn179, and between those of Arg131 and Asp178. However, it is not clear why the initial Py corresponding to the third base position in the MBS-I fragment is preferred by c-Myb R2R3, although this Py3 is less specific than the other A4, A5, C6, and G8 sites in the consensus DNA sequence (Tanikawa *et al.*, 1993). In the NMR structure shown in Figure 1-1, Ser187 is the only candidate that interacts with the T3 base, and this ability was suggested in the previous paper (Ogata *et al.*, 1994). The hydroxyl group in the Ser side-chain could form a hydrogen bond with the O_4 oxygen of the T3 base, either directly or through water molecules.

Thus far, the Myb-homologous DNA-binding domain (DBD) has been found in over 30 proteins from many species. An alignment of the DBDs shows that the Ser at position 187 is highly conserved in the animal sequences, whereas it is variable in the plant sequences (Ogata *et al.*, 1994; Bilaud *et al.*, 1996).

Here, in order to investigate the role of Ser187 and the origin of this pyrimidine preference at the third base position, both Ser187 in the c-Myb R2R3 and the third T-A base-pair in the 22-mer MBS-I fragment containing the Myb-binding site were substituted by other amino acids and other base-pairs, respectively. The interactions between them

were examined using a filter binding assay, whose efficiency has already been shown (Tanikawa *et al.*, 1993; Takeda *et al.*, 1989; Sarai and Takeda, 1989). The recognition mechanism will be discussed.

1.2 Experimental Procedures

Plasmids and site-directed mutagenesis

A DNA fragment encompassing R2R3 (Leu90-Val193) in the DNA binding domain of c-Myb was amplified by PCR, using pact-c-*myb* (Nishina *et al.*, 1989) as the template and two synthetic primers, to generate an *Nco*I site and a *Bam*HI site at the 5'- and the 3'-end of the amplified fragment, respectively. After digestion with *Nco*I and *Bam*HI, the DNA fragment was cloned into pAR2156NcoI (Tanikawa *et al.*, 1993) to yield the expression plasmid, pRP23. An additional Met-Glu- sequence was introduced at the N-terminus of R2R3. Site-directed mutagenesis was performed by two-step PCR, as described by Higuchi (Higuchi, 1989). Here the name of each mutant protein is indicated as, for example, C130I/S187G for the simultaneous mutations that replace Cys130 with Ile and Ser187 with Gly.

Protein expression and purification

Escherichia coli BL21(DE3) was transformed with the wild-type and mutant plasmids (Studier, 1990). Freshly precultivated cells were inoculated into growth medium containing 100 μg/ml ampicillin and were grown at 37 °C. When the culture reached an OD_{600} of about 0.4, isopropyl-1-thio-ß-D-galactopyranoside was added to a final concentration of 0.5 mM. The cells were cultured at 22 °C for another 12 h. The harvested cells were suspended in 50 mM Tris-HCl buffer (pH 7.8) containing 5 mM $MgCl_2$, and were lysed by sonication at 4 °C. After the cell debris was removed by centrifugation, ammonium sulfate was added to the supernatant to 50 % saturation. After an incubation at 4 °C for 1 h, the supernatant was dialyzed against 50 mM potassium phosphate buffer (pH 7.5) containing 200 mM NaCl, and was then applied to a phosphocellulose column (Whatman, P11). The purified fractions were pooled, and the buffer was exchanged to 100 mM potassium phosphate buffer (pH 7.5) containing 20 mM KCl. The protein concentrations were determined from UV absorption at 280 nm and

were calculated by using the molar absorption coefficient of 3.7×10^4 M^{-1} cm^{-1} (Tanikawa *et al.*, 1993).

CD measurements

Circular dichroism (CD) spectra were measured at 20 °C on a Jasco J-600 spectropolarimeter equipped with a water-circulating cell holder. The spectra were obtained in 100 mM potassium phosphate buffer (pH 7.5) containing 20 mM KCl, using a 0.2 cm optical path length cell. The protein concentration was 0.1 mg/ml. CD spectra between 200 and 250 nm were obtained using a scanning speed of 20 nm per min, a time response of 1 s, a bandwidth of 1 nm, and an average over 8 scans.

Preparation of oligonucleotides

The 22-mer oligonucleotide CACCCTAACTGACACACATTCT, containing the Myb-binding site in the simian virus 40 enhancer sequence (MBS-I) (Nakagoshi *et al.*, 1990), and the third base substituted variants were synthesized and purified by HPLC with a C_{18} reverse-phase column (Figure 1-2). The purified DNA was suspended in STE (10 mM Tris-HCl (pH 8.0), 100 mM NaCl, 1 mM EDTA), and complementary strands were annealed and end-labeled with [γ-^{32}P]ATP (Amersham) using T4 polynucleotide kinase (Toyobo, Osaka). The labeled DNAs were purified by passage through spin columns (Pharmacia, HR-300). Here the name of each variant DNA is indicated as, for example, [C3]MBS-I, for the substitution of the T-A base-pair at the third position by a C-G base-pair.

Filter binding assay

All filter binding assays for the protein-DNA binding were carried out essentially as described (Riggs *et al.*, 1970a, 1970b; Kim *et al.*, 1987). [^{32}P]DNA and various amounts of the c-Myb R2R3 mutant proteins were incubated in 100 µl of binding buffer

(100 mM potassium phosphate buffer (pH 7.5), 20 mM KCl, 0.1 mM EDTA, 500 μg/ml bovine serum albumin, and 5% (vol/vol) glycerol) on ice for 30 min. The final concentration of the [^{32}P]DNA in binding buffer was 0.4 nM, which was always a lower concentration than the K_d value. The incubated samples were filtered through a nitrocellulose membrane (Schleicher and Schuell, BA-85, 0.45 μm) in approximately 10 sec with suction. The filters were dried and counted by a liquid scintillation counter. The equilibrium dissociation constants K_d were obtained from the binding titration curve, based on the least square fitting to the normalized bound DNA (y) with the protein concentration (x) using the formula, $y = x/(x+ K_d)$

1.3 Results

Prior to the mutational analyses of Ser187, the Cys130 in R2, which is the only cysteine residue in the c-Myb R2R3 and is located at a position equivalent to an isoleucine in R3, was replaced with Ile, to facilitate the protein purification and the DNA-binding assay (Guehmann *et al.*, 1992). It was reported that this mutation has little effect on DNA-binding (Myrset *et al.*, 1993). The affinity of the C130I mutant was also measured in our own assay system, and it was shown to be almost equal to that of the wild-type, and to maintain the pyrimidine preference at the third base position (Table 1-1).

A series of 10 amino acids, Gly, Ala, Thr, Asn, Gln, Val, Leu, Lys Arg, and Asp, were introduced into position 187 of the c-Myb R2R3, which is a Ser residue in the wild-type. The purity of each mutant protein was about 95%, as monitored by SDS/PAGE. All of the mutant proteins have secondary structure contents similar to the wild-type, as confirmed by the CD spectra at the far UV region (Figure 1-3). The perfect coincidence of all the spectra suggests that the global tertiary structures of the mutant proteins were not deformed.

The binding affinities of the mutants to the cognate 22-mer MBS-I fragments and the third base-pair substituted variants were analyzed using the filter binding assay, and the results are summarized in Table 1-1. All measurements were repeated at least twice, and typical experimental errors for the K_d value were less than 10%. Although the retention efficiency was 20 ± 10 % depending on the experimental conditions, that had little effect on the K_d value. The methylation interference experiments (Tanikawa *et al.*, 1993) and the NMR analyses (Ogata *et al.*, 1994) have suggested that the stoichiometry of binding was one to one within the concentration used in this assay. As already indicated in the previous experiments (Tanikawa *et al.*, 1993; Takeda *et al.*, 1989; Sarai and Takeda, 1989), the filter binding assay was validated for this investigation.

The C130I/S187G mutant protein binds about one-third less strongly to the cognate MBS-I than the standard C130I mutant. The relative binding free energy change

for the replacement of Ser with Gly, calculated from the K_d values, is 0.65 kcal/mol. It should correspond to the free energy derived from the interaction between the Ser side-chain and the T3 base. This Gly mutant preferentially binds to both the cognate [T3]MBS-I and the substituted [C3]MBS-I. That is, even when residue 187 has no side-chain, the mutant protein prefers the third pyrimidine as well as the wild-type and the C130I mutant proteins.

The substitutions of Ser187 by Ala (C130I/S187A), Thr (C130I/S187T), or Val (C130I/S187V), reveal slightly reduced binding affinities, although the sequence specificities are retained like the standard C130I. In contrast, the C130I/S187N mutant preferentially binds to the [A3]MBS-I. The affinity for the A3 base is similar to that of the wild-type, although those for the other three bases (T, C, and G) are greatly reduced, by approximately one-half to one-sixth. The specific interaction between the Asn residue and the A3 base closely follows the intrinsic chemical features. Interestingly, for the substitution by Gln, which is one methylene group longer than Asn, the C130I/S187Q mutant loses the preference for the A3 base. Also, in the case of the C130I/S187L mutant, in which Leu is one methylene group longer than Val, the binding affinity is greatly reduced.

The mutant proteins C130I/S187K and C130I/S187R, which introduced basic amino acids into position 187, specifically prefer to bind to the [G3]MBS-I and [C3]MBS-I variants. In contrast, for the substitution of Ser187 by acidic Asp (C130I/S187D), the binding affinity is completely reduced and is no longer sequence specific.

1.4 Discussion

Thus far, many amino acid replacements in the c-Myb R2R3 have been created and assayed by specific DNA binding (Saikumar *et al.*, 1990; Gabrielsen *et al.*, 1991; Frampton *et al.*, 1991), and almost all of their effects have been explained by the specific polar contacts between the R2R3 and the DNA in the three-dimensional structure of the R2R3-DNA complex (Ogata *et al.*, 1994). The current mutational study clearly indicates that residue 187 in R2R3 is also able to interact with the T3 base, as estimated from the geometry of Ser187 in the NMR complex structure (Ogata *et al.*, 1994). This specific DNA-binding mode is very different from the telomeric DNA recognition by the yeast RAP1-DBD (König *et al.*, 1996), whose amino acid sequence is weakly homologous to that of the c-Myb R2R3.

However, the substitution of Ser187 with Gly, Ala, or Val unexpectedly resulted in only about a three-fold decrease in the binding affinity toward any base, which would be a consequence of a direct-readout mechanism, while the pyrimidine base preference at the third position in the MBS-I fragment was retained. Ser is thought to have weak specificity, because its side chain can act as either a hydrogen bond donor or an acceptor, and thus can bind to any base. Nevertheless, Ser187 of the R2R3 preferentially binds to the pyrimidine bases. If this interaction were attributable only to the direct-readout mechanism, then the substitution of Ser187 should have resulted in an over 100-fold reduction of the binding affinity and a loss of the sequence specificity, like the substitution of Lys128 by Ala (Ogata *et al.*, 1994), and those of Asn136 and Asn186 by Ala (Gabrielsen *et al.*, 1991). These results suggest that the preference of the pyrimidine bases at the third position of MBS-I should occur primarily by an indirect-readout mechanism.

In the previous structural study, no distinct deformation of the global DNA conformation was observed (Ogata *et al.*, 1994). However, when the local bending of the DNA duplex was carefully analyzed in 25 NMR complex structures and the refined

average structure (PDB codes 1MSF and 1MSE, respectively), significantly positive roll angles were always observed between the third pyrimidine and the fourth purine, as indicated by an arrow in Figure 1-4. Characteristic negative slides (-1.1 ± 0.3 Å) were also observed at the same pyrimidine-purine step, corresponding to positive rolling, while the twist angles at this step were 34.1 ± 2.3 °, nearly equal to the twist angle in standard B-form DNA.

Similar significant, positive rolls at pyrimidine-purine steps are general phenomena (Babcock and Olson, 1994), observed in many complex crystal structures of repressors and homeodomains with the helix-turn-helix motif, as summarized in Table 1-2. In every case, as a part of the consensus base sequence, the base-pair roll bends the DNA so that the recognition helix is wrapped by the DNA duplex in the major groove (Pabo and Sauer, 1992; Suzuki *et al.*, 1995). Consequently, a large contact area is created between the recognition helix and the DNA major groove, facilitating the preferable polar contacts between the protein side-chains and the DNA phosphate backbone. The local roll in the MBS-I fragment may be associated with the small magnitude of observed bending in long DNA duplexes bound with the c-Myb R2R3 (Saikumar *et al.*, 1994). This bending may be enhanced by other regions in the protein, like the transactivation domain.

Due to the intrinsic propeller-twist of the DNA base-pairs, the pyrimidine-purine step has two stable conformations, with rolling of 0° and around 10° (Calladine and Drew, 1984; Nelson *et al.*, 1987), from the physical requirements of the base stacking (Calladine, 1982). There is negligible additional free energy cost required for the 10° rolling at the pyrimidine-purine step, even for a free DNA duplex without a protein. This is the physical origin of the so called 'bendability' of kinked DNA duplexes, commonly observed in the minor groove readout mechanism (Kim *et al.*, 1993; Schumacher, 1994). At the other pyrimidine-pyrimidine, purine-purine, and purine-pyrimidine steps, no such tendency toward a strongly bistable step is observed (Calladine and Drew, 1984). In fact, the binding free energy differences between the pyrimidine bases and the purine bases at

the third base position for the current Gly187, Ala187, and Val187 mutants are 0.4 ± 0.1 kcal/mol, as calculated from the dissociation constants in Table 1-1.

Figure 1-5 shows the results of the relative binding free energy changes $\Delta\Delta G$ toward the C130I/S187G mutant: $\Delta\Delta G = \Delta G_{bind}$ (mutant against the third N base) - ΔG_{bind} (C130I/S187G against the same third N base), where $\Delta G_{bind} = RT \ln K_d$. Here, the difference was calculated while keeping the same third position base-pair. We can now separate the bendability effect from the total binding free energies between the c-Myb R2R3 mutants and the variety of DNA sequences, unless the binding modes vary from the wild-type. Each positive and negative free energy corresponds to a decrease and an increase of the binding affinity, depending upon the intrinsic chemical features of the amino acids and the bases, and subtracting the DNA bending effect.

For the Ala substitution, the binding affinity is increased as compared to Gly187, independent of the bases at the third position, probably due to the hydrophobic contacts. When the side-chain volume is larger in the Val substitution, a similar binding affinity to the pyrimidines remains, but the affinity becomes neutral to the purines. Therefore, the volume of space created between residue 187 and the third base may allow at most the Val-pyrimidine pair, but the Val-purine pair would be slightly too large for the space. In fact, other amino acids, such as Leu and Gln, with larger side-chain volumes than Val, significantly lack binding affinity, as indicated in Figure 1-5. Moreover, the Val, Leu, and Gln substituted mutants always have lower affinities for adenine than for guanine. This is also supported by the fact that the amino N_6 of adenine occupies a larger volume than the oxygen O_6 of guanine, which should be located at the position nearest to the side-chain of residue 187.

From this consideration of the space volume around residue 187 and the third base, the native and the optimum interaction between Ser187 and T3 should be mediated by water molecules, as long as the binding mode is assumed to be the same in all of the mutant proteins and DNAs. In the Thr mutant, the disposition of the water molecules

could be different from that in the wild-type, thus yielding a slight decrease in the binding affinity. Since there is no possible conformation on the helix in which the methyl group of the Thr side-chain would be able to access the methyl group in T3, as shown in a modeling study, a specific non-polar contact between the Thr mutant and T3 is not expected.

Following the conventional chemical rules for specific binding between amino acids and bases (Seeman *et al.*, 1976; Pabo and Sauer, 1992; Suzuki and Yagi, 1994), the current Asn mutant specifically binds to the A3 base relative to the other bases, as indicated in Figure 1-5. The Asn side-chain size is less than that of Val, and there should be enough space for the Asn-adenine pair, resulting in the formation of direct hydrogen bonds with a free energy gain of about 0.5 kcal/mol. In addition, the Lys and Arg mutant proteins prefer to bind to the G3 base. From their intrinsic chemical nature, both basic amino acids can bind to the guanine base almost exclusively by electrostatic interaction. In contrast, these mutant proteins bind to the [A3]MBS-I and [T3]MBS-I bases with only weak affinity, probably because of the bulky side-chains of the amino acids, like the Leu mutant. It is interesting that their long side chains seem to interact with the guanine base on the opposite side of C3. The acidic Asp substitution results in a severe reduction of its DNA-binding, which is much lower than the Gly substitution, suggesting that the Asp side chain cannot interact with any base, including cytosine, in this geometry. Rather, the negative ionic charge may disturb other specific hydrogen bonds between the protein and the DNA.

The wild-type protein and the C130I mutant with Ser187 bind to the cognate DNA most tightly among the mutant proteins, and their K_d values are in the nano-molar order. Generally, transcriptional regulator proteins bind to their target genes with greater affinity (Spolar and Record, 1994). These results are consistent with the conservation of Ser in position 187 of c-Myb among animal species (Ogata *et al.*, 1994). In contrast, among plant species, the amino acid in this position varies (Bilaud *et al.*, 1996). This suggests

that the recognition mode in the plant Myb homologues may be different from that of the c-Myb DBD from animal species. In fact, in the case of the yeast RAP1 domain 1, the corresponding Val409 residue does not interact with the DNA in the complex structure (König *et al.*, 1996), although the free domain structure is similar to that of the c-Myb R3.

In conclusion, the current mutational analysis revealed that the pyrimidine preference of the native c-Myb DBD for the initial base of the consensus sequence originates principally in the intrinsic positive roll at the pyrimidine-purine step of the DNA duplex. For the purine-purine step, as much as 0.4 kcal/mol of additional free energy would be necessary, corresponding to the bendability. When these bending energies are separated, the conventional chemical rules between the amino acids and the bases are distinctively observed in the c-Myb R2R3 mutants.

It is still difficult to extract a definite 'recognition code' from the variety of DNA information readout mechanisms. The situation becomes much more complicated when the DNA flexibility is considered. Only a screening technology, such as a phage display library (Choo and Klug, 1994a, 1994b; Reber and Pabo, 1994), would be expected to reveal a novel, specific form of DNA recognition, instead of an artificial molecular design. However, based upon the complex structure and the mutational analysis, one may be able to dissect the sequence specific affinity into the DNA bendability and the specific interaction between the amino acids and the bases. Without this kind of precise analysis, we may never reach a complete understanding of the readout mechanism, nor produce any novel devices for molecular readout.

1.5 References

Aggarwal, A. K., Rodgers, D. W., Drottar, M., Ptashne, M., and Harrison, S. C. (1988) *Science*, **242**, 899-907.

Babcock, M. S., and Olson, W. K. (1994) *J. Mol. Biol.*, **237**, 98-124.

Beamer, L. J., and Pabo, C. O. (1992) *J. Mol. Biol.*, **227**, 177-196.

Bilaud, T., Koering, C. E., Binet-Brasselet, E., Ancelin, K., Pollice, A., Gasser, S. M., and Gilson, E. (1996) *Nucleic Acids Res.*, **24,** 1294-1303

Calladine, C. R. (1982) *J. Mol. Biol.*, **161**, 343-352.

Calladine, C. R., and Drew, H. R. (1984) *J. Mol. Biol.*, **178**, 773-782.

Choo, Y., and Klug, A. (1994a) *Proc. Natl. Acad. Sci. USA*, **91**, 11163-11167.

Choo, Y., and Klug, A. (1994b) *Proc. Natl. Acad. Sci. USA*, **91**, 11168-11172.

Frampton, J., Gibson, T. J., Ness, S. A., Doderlein, G., and Graf, T. (1991) *Protein Eng.*, **4**, 891-901.

Gabrielsen, O. S., Sentenac, A., and Fromageot, P. (1991) *Science*, **253**, 1140-1143.

Guehmann, S., Vorbrueggen, G., Kalkbrenner, F., and Moelling, K. (1992) *Nucleic Acids Res.*, **20**, 2279-2286.

Higuchi, R. (1989) in *PCR technology* (Erlich, H.A., ed.) pp.61-88, Stockton Press, New York.

Kim, J. G., Takeda, Y., Matthews, B. W., and Anderson, W. F. (1987) *J. Mol. Biol.*, **196**, 149-158.

Kim, Y., Geiger, J. H., Hahn, S., and Sigler, P. B. (1993) *Nature*, **365**, 512-520.

Kissinger, C. R., Liu, B., Martin-Blanco, E., Kornberg, T. B., and Pabo, C. O. (1990) *Cell*, **63**, 579-590.

Klemm, J. D., Rould, M. A., Aurora, R., Herr, W., and Pabo, C. O. (1994) *Cell*, **77**, 21-32.

Klimasauskas, S., Kumar, S., Roberts, R. J., and Cheng, X. (1994) *Cell*, **76**, 357-369.

König, P., Giraldo, R., Chapman, L., and Rhodes, D. (1996) *Cell*, **85**, 125-136.

Lawson, C. L., and Carey, J. (1993) *Nature*, **366**, 178-182.

Lehming, N., Sartorius, J., Kisters-Woike, B., von Wilcken-Bergmann, B., and Müller-Hill, B. (1990) *EMBO J.*, **9**, 615-621.

Li, T., Stark, M. R., Johnson, A. D., and Wolberger, C. (1995) *Science*, **270**, 262-269.

Mondragón, A., and Harrison, S. C. (1991) *J. Mol. Biol.*, **219**, 321-334.

Myrset, A. H., Bostad, A., Jamin, N., Lirsac, P. -N., Toma, F., and Gabrielsen, O. S. (1993) *EMBO J.*, **12**, 4625-4633.

Nakagoshi, H., Nagase, T., Kanei-Ishii, C., Ueno, Y., and Ishii, S. (1990) *J. Biol. Chem.*, **265**, 3479-3483.

Nelson, H. C. M., Finch, J. T., Luisi, B. F., and Klug, A. (1987) *Nature*, **330**, 221-226.

Nishina, Y., Nakagoshi, H., Imamoto, F., Gonda, T. J., and Ishii, S. (1989) *Nucleic Acids Res.*, **17**, 107-117.

Ogata, K., Morikawa, S., Nakamura, H., Sekikawa, A., Inoue, T., Kanai, H., Sarai, A., Ishii, S., and Nishimura, Y. (1994) *Cell*, **79**, 639-648.

Otwinowski, Z., Schevitz, R. W., Zhang, R. -G., Lawson, C. L., Joachimiak, A., Marmorstein, R. Q., Luisi, B. F., and Sigler, P. B. (1988) *Nature*, **335**, 321-329.

Pabo, C. O., and Sauer, R. T. (1992) *Annu. Rev. Biochem.*, **61**, 1053-1095.

Pontiggia, A., Rimini, R., Harley, V. R., Goodfellow, P. N., Lovell-Badge, R., and Bianchi, M. E. (1994) *EMBO J.*, **13**, 6115-6124.

Reber, E. J., and Pabo, C. O. (1994) *Science*, **263**, 671-673.

Riggs, A. D., Suzuki, H., and Bourgeois, S. (1970a) *J. Mol. Biol.*, **48**, 67-83.

Riggs, A. D., Bourgeois, S., and Cohn, M. (1970b) *J. Mol. Biol.*, **53**, 401-417.

Rodgers, D. W., and Harrison, S. C. (1993) *Structure*, **1**, 227-240.

Saikumar, P., Murali, R., and Reddy, E. P. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 8452-8456.

Saikumar, P., Gabriel, J. L., and Reddy, E. P (1994) *Oncogene*, **9**, 1279-1287.

Sarai, A., and Takeda, Y. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 6513-6517.

Schumacher, M. A., Choi, K.Y., Zalkin, H., and Brennan, R. G. (1994) *Science*, **266**, 763-770.

Schultz, S. C., Shields, G. C., and Steitz, T. A. (1991) *Science*, **253**, 1001-1007.

Seeman, N. C., Rosenberg, J. M., and Rich, A. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 804-808.

Shimon, L. J. W., and Harrison, S. C. (1993) *J. Mol. Biol.*, **232**, 826-838.

Spolar, R. S., and Record, T. Jr. (1994) *Science*, **263**, 777-784.

Studier, F. W., Rosenberg, A. H., Dunn, J. J., and Dubendorff, J. W. (1990) *Methods Enzymol.*, **185**, 60-89.

Suzuki, M. (1993) *EMBO J.*, **12**, 3221-3226.

Suzuki, M., and Yagi, N. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 12357-12361.

Suzuki, M., Yagi, N., and Gerstein, M. (1995) *Protein Eng.*, **8**, 329-338.

Takeda, Y., Sarai, A, and Rivera, V. M. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 439-443.

Tanikawa, J., Yasukawa, T., Enari, M., Ogata, K., Nishimura, Y., Ishii, S., and Sarai, A. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 9320-9324.

Travers, A. A. (1992) *Curr. Opin. Struct. Biol.*, **2**, 71-77.

Vassylyev, D. G., Kashiwagi, T., Mikami, Y., Ariyoshi, M., Iwai, S., Ohtsuka, E. and Morikawa, K. (1995) *Cell*, **83**, 773-782.

Winkler, F. K., Banner, D. W., Oefner, C., Tsernoglou, D., Brown, R.S., Heathman, S. P., Bryan, R. K., Martin, P. D., Petratos, K., and Wilson, K. S. (1993) *EMBO J.*, **12**, 1781-1795.

Zamyatnin, A. A. (1984) *Annu. Rev. Biophys. Bioeng.*, **13**, 145-165.

Table 1-1. Dissociation constants for the cognate 22-mer MBS-I fragments and the third base-pair substituted variants with the Ser187 substituted mutants

	K_d, nM			
protein	T3	C3	A3	G3
wild-type[a]	3.2	3.7	8.7	25
C130I	5.5	5.7	12	27
C130I/S187G	18	17	33	37
C130I/S187A	9.3	12	22	24
C130I/S187T	15	11	21	36
C130I/S187N	26	37	15	53
C130I/S187Q	36	39	54	39
C130I/S187V	13	8.9	37	34
C130I/S187L	130	73	$\geq 10^3$	72
C130I/S187K	61	22	$\geq 10^3$	36
C130I/S187R	43	22	$\geq 10^3$	21
C130I/S187D	$\geq 10^3$	$\geq 10^3$	$\geq 10^3$	$\geq 10^3$

[a]An additional Met-Ala- sequence was introduced at the N-terminus of R2R3, which was used in the NMR experiment (Ogata *et al.*, 1994).

Table 1-2. Significant roll angles of DNA duplexes at the Pyrimidine-Purine Steps in the protein-DNA crystal structures with the helix-turn-helix motif

Protein-DNA (PDB / resolution)	Pyrimidine-Purine step[a] (degree)	roll[b]	references
CAP[c]-DNA_{30} (1CGP / 3.0Å)	$C_{5C}A_{6C}/T_{26C}G_{27C}$ $C_{5D}A_{6D}/T_{26D}G_{27D}$	38.9 30.4	(Schultz *et al.*, 1991)
434Cro[d]-O_R1 (3CRO / 2.5Å)	$C_{6A}A_{7A}/T_{15B}G_{16B}$ $T_{16A}G_{17A}/C_{5B}A_{6B}$	7.8 13.8	(Mondragon and Harrison, 1991)
434R[e]-O_R1 (2OR1 / 2.5Å)	$C_{6A}A_{7A}/T_{15B}G_{16B}$ $T_{16A}G_{17A}/C_{5B}A_{6B}$	6.1 5.4	(Aggarwal *et al.*, 1988)
434R-O_R2 (1RPE / 2.5Å)	$C_{26A}A_{27A}/T_{15B}G_{16B}$ $T_{36A}G_{37A}/C_{5B}A_{6B}$	7.9 6.5	(Shimon and Harrison, 1993)
434R-O_R3 (1PER / 2.5Å)	$C_{6A}A_{7A}/T_{15B}G_{16B}$ $T_{16A}G_{17A}/C_{5B}A_{6B}$	1.8 9.0	(Rodgers and Harrison,)
λR[f]-O_L1 (1LMB / 1.8Å)	$C_{6(1)}A_{7(1)}/T_{35(2)}G_{36(2)}$ $T_{15(1)}G_{16(1)}/C_{26(2)}A_{27(2)}$	9.2 11.3	(Beamer and Pabo, 1992)
trpR[g]-*trp O* (1TRO / 1.9Å)	$T_{6I}A_{7I}/T_{14J}A_{15J}$ $T_{14I}A_{15I}/T_{6J}A_{7J}$ $T_{6K}A_{7K}/T_{14L}A_{15L}$ $T_{14K}A_{15K}/T_{6L}A_{7L}$	8.5 9.0 11.8 14.7	(Otwinowski *et al.*, 1988)
trpR[g]-*trp O* (1TRR / 2.4Å)	$T_{9C}A_{10C}/T_{9I}A_{10I}$ $T_{9F}A_{10F}/T_{9L}A_{10L}$	10.0 9.7	(Lawson and Carey, 1993)
MAT**a1**/α2[h]-DNA_{21} (1YRN / 2.5Å)	$T_{5C}G_{6C}/C_{39C}A_{40C}$ $T_{15C}A_{16C}/T_{29C}A_{30C}$ $C_{17C}A_{18C}/T_{27C}G_{28C}$	12.0 13.3 14.4	(Li *et al.*, 1995)
Oct-1 POU[i]-DNA_{15} (1OCT / 3.0Å)	$T_{205A}G_{206A}/C_{226B}A_{227B}$ $C_{207A}A_{208A}/T_{224B}G_{225B}$	8.2 11.5	(Klemm *et al.*, 1994)
Engrailed HD[j]-DNA_{21} (1HDD / 2.8Å)	$T_{11A}A_{12A}/T_{32B}A_{33B}$ $T_{15A}A_{16A}/T_{28B}A_{29B}$	3.7 9.6	(Kissinger *et al.*, 1990)

[a]The residue and chain identifiers are those as registered in PDB.

[b]The roll angles were calculated using a program rna (Babcock and Olson, 1994).

[c]*Escherichia coli* catabolite gene activator protein.

[d]phage 434 Cro protein. [e]phage 434 repressor. [f]λ represser.

[g]*Escherichia coli trp* repressor. [h]MAT**a1**/MATα2 homeodomain heterodimer.

[i]Oct-1 POU domain. [j]engrailed homeodomain.

Figure 1-1. Specific interaction between the c-Myb R2R3 and the MBS-I fragment (Ogata *et al.*, 1994). The side-chains of Lys128, Lys182, Asn183, and Ser187 are indicated by the thick black lines, and the consensus bases Py3(T3), A4, A5, C6, and G8 are indicated by the thick white lines. The black and white thin lines are the DNA double strands. The backbone of the protein is shown by a pipe model. The figure was drawn with the program InsightII (Molecular Simulations Inc., San Diego).

Figure 1-2. Sequences of the cognate 22-mer MBS-I and the 3rd base-pair substituted MBS-I fragments. The base numbering follows Ogata *et al.* (1994). The consensus base sequence is underlined in MBS-I, and the substituted bases in the non-cognate DNAs are indicated in italics.

		-3 1 5 10 15 19	
MBS-I	5′-	CACCCTAACTGACACACATTCT	-3′
[C3]MBS-I		CACCC*C*AACTGACACACATTCT	
[A3]MBS-I		CACCC*A*AACTGACACACATTCT	
[T3]MBS-I		CACCC*T*AACTGACACACATTCT	

Figure 1-3. Far-UV CD spectra of the wild-type c-Myb R2R3 and the 11 Ser187 substituted mutants. All of them are superimposed. The vertical scale is normalized by the molar concentration.

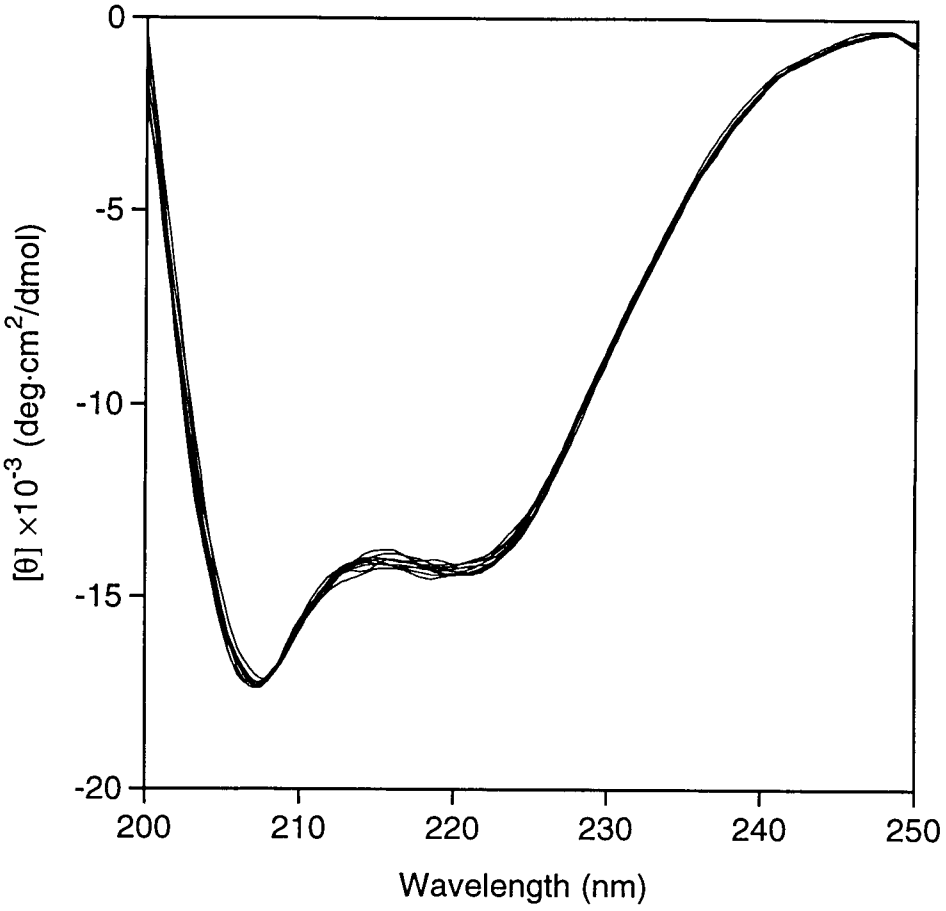


Figure 1-4. The roll angles of individual base-pair steps of the MBS-I DNA fragment complexed with the c-Myb R2R3 (Ogata *et al.*, 1994), calculated using a program developed by Babcock *et al.* (1994) Each filled circle with a straight line and an error bar is the average and the standard deviation for the 25 NMR structures (PDB code 1MSF), respectively. Each open circle with a dashed line indicates the rolling of the refined average structure (1MSE). The arrow indicates the pyrimidine-purine step between T3 and A4. The consensus bases are indicated in bold letters.

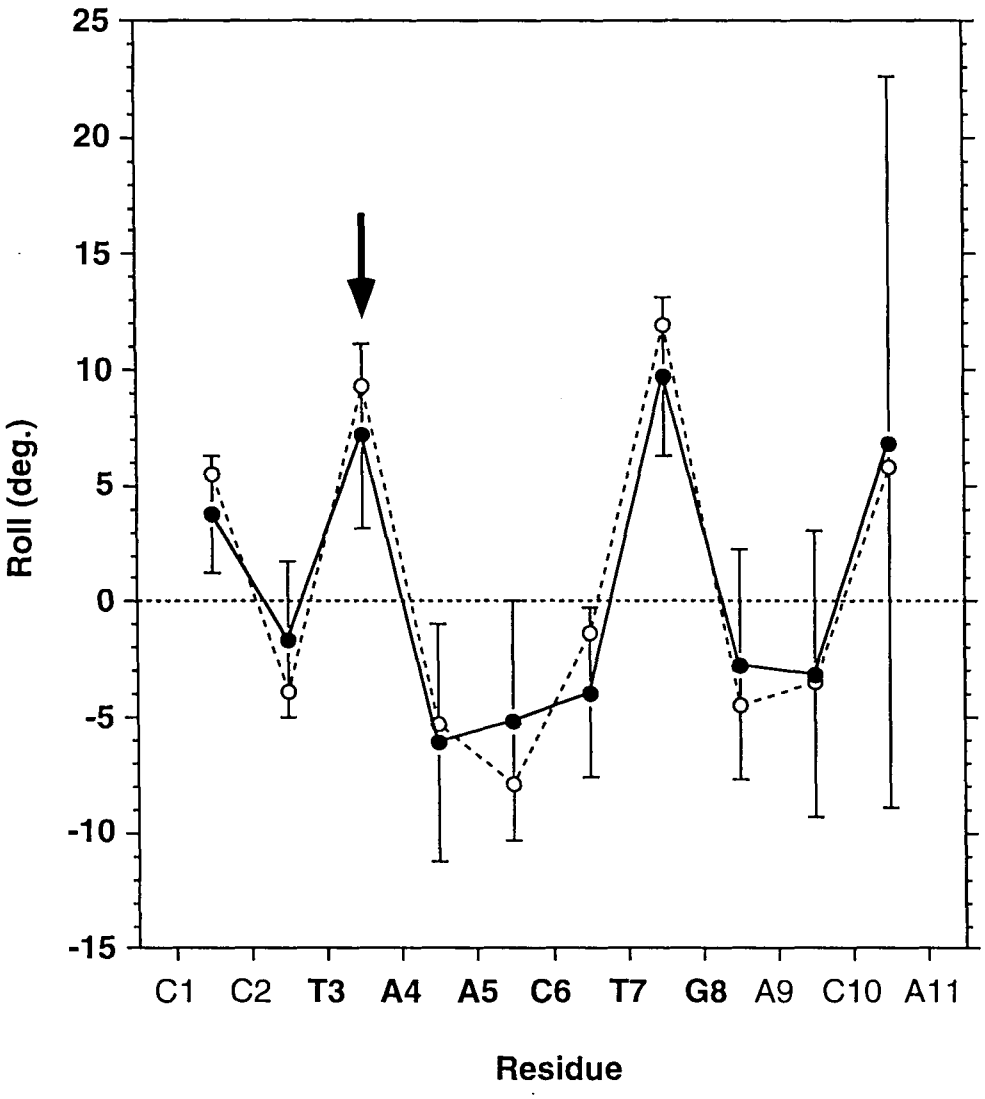
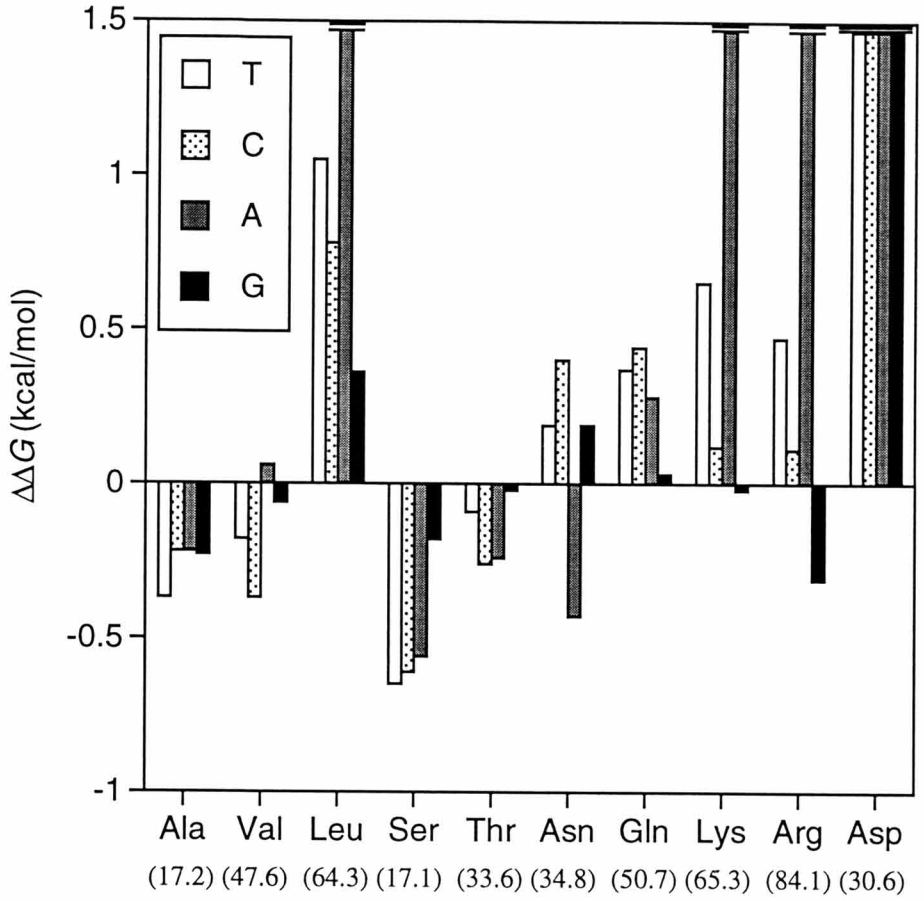


Figure 1-5. Relative binding free energy changes ($\Delta\Delta G$) toward the C130I/S187G mutant protein binding of the Ser187 substituted R2R3 mutant proteins to the cognate MBS-I and its variants. Each $\Delta\Delta G$ for the particular base-pair was calculated from the K_d values, as described in the text. The side-chain volumes (Å^3) of the substituted amino acids (Zamyatnin, 1984) are indicated in the parentheses below the individual residues.



Chapter 2

Identification of Indispensable Residues for Specific DNA-binding in the Imperfect Tandem Repeats of c-Myb R2R3

2.1 Introduction

Nature frequently uses tandem repeats of functional modules in a protein to attain accurate molecular recognition (Bork *et al.*, 1996). Indeed, transcription factors involved in specific DNA-binding often have tandem repeats of structural and functional modules, such as zinc fingers (Pavletich and Pabo, 1991, 1993) and the compact three helical structure represented by the homeodomain (Ogata *et al.*, 1994; Klemm *et al.*, 1994; Xu *et al.*, 1995; König *et al.*, 1996). In most cases, more than one repeat binds DNA simultaneously, with large binding constants from 10^9 to 10^{12} M^{-1}, while the affinity to DNA of a single repeat is as weak as the non-specific binding due to electrostatic attraction.

Once this cooperative mechanism in the tandem repeats is understood, it is possible to create an artificial molecule, which recognizes a novel base sequence, by fusion of the different DNA-binding modules. In fact, Pomerantz *et al.* (1995) succeeded in producing a new transcription factors by fusing two zinc fingers and a homeodomain. Recently, Kim *et al.* (1997) fused a TATA box-binding protein with three zinc fingers, yielding a site-specific repressor.

However, a direct interaction between the individual repeats has not been observed, except for the DNA-binding domain (DBD) of the c-*myb* protooncogene product (c-Myb). The linkers between the two repeat structures are usually long and flexible, such as between the POU-specific domain and the POU-homeodomain (Klemm *et al.*, 1994), between the N- and C-terminal domains of the prd paired domain (Xu *et al.*, 1995), and within the DBD of yeast RAP-1 (König *et al.*, 1996). In contrast, the linker within the c-Myb DBD is so short that strong cooperativity between the adjacent repeats was expected (Ogata *et al.*, 1994).

The solution structures of the specific DNA complex of the minimum DBD of c-Myb (R2R3) and its free form were determined by multidimensional NMR techniques (Ogata *et al.*, 1992, 1994, 1995). Both R2 and R3 are composed of three helices with a

helix-turn-helix variation motif, and each third helix (helix-3) in R2 and R3 is engaged in direct and specific base recognition. The overall structures of these repeats are almost identical: the root-mean-square-deviation (r.m.s.d.) of the backbone heavy atoms is only 0.81 ± 0.11 Å between R2 and R3 in their free forms (Ogata *et al.*, 1995). Previous mutational studies (Saikumar *et al.*, 1990; Gabrielsen *et al.*, 1991; Frampton *et al.*, 1991) indicated that the amino acid residues in recognition helix-3 and the three conserved tryptophans in each R2 and R3 have important roles in specific DNA-binding.

Thus, if the tandem sequence repeats are necessary for the overall backbone conformation of R2, and the imperfection within only the third helix is required for the characteristic recognition helix, then R2 can be engineered to be almost identical to R3, with only slight modifications for specific base recognition. Namely, the first helix (helix-1) or the second helix (helix-2) in R2 would be exchangeable with those helices in R3, respectively, because of their remarkably similar backbone structures.

However, the sequence identity between R2 and R3 is only 31 %, and the amino acids composing each repeat are highly conserved among animal species, even for the residues in helix-1 and helix-2, which do not seem to be in direct contact with the DNA bases. This suggests that these conserved amino acids have some particular role in specific DNA-binding, although they are located far from the interface between R2R3 and the DNA.

Here we prepared a series of R2R3 mutants, in which the helices in R2 were replaced by the corresponding helices in R3, and examined the affinities for specific DNA-binding, monitoring the overall folds by CD spectra. In addition, more localized mutations were introduced to identify the indispensable residues in R2, which should be different from those in R3. The roles of the identified residues in specific DNA recognition are discussed.

2.2 Experimental Procedures

Plasmid construction and protein purification

A DNA fragment encompassing R2R3 (Leu90-Val193) in the c-Myb DBD was amplified by PCR, using pact-c-*myb* (Nishina *et al.*, 1989) as the template and two synthetic primers, to generate an *Nco*I site and a *Bam*HI site at the 5'-end and the 3'-end of the amplified fragment, respectively. After digestion with *Nco*I and *Bam*HI, the DNA fragment was cloned into an expression vector, pAR2156NcoI_ΔEH, in which the *Eco*RI and *Hind*III sites of pAR2156NcoI (Tanikawa *et al.*, 1993) were deleted by site-directed mutagenesis, to yield the expression plasmid, pRP23. An additional Met-Glu- sequence was introduced at the N-terminus of R2R3 to generate an *Nco*I site. To replace a helix in R2 by the corresponding R3 sequence, the DNA fragment of R3 was amplified by PCR, using primers designed to allow the amplified DNA fragments to have suitable endonuclease restriction sites at both the 5'- and 3'- ends (Figure 2-1); for example, the replacement of helix-1 of R2 was performed using the restriction sites *Nco*I and *Apa*I. To obtain the site specific mutants, the two-step PCR method (Higuchi, 1989) was used. Protein expression and purification were performed as described previously (Oda *et al.*, 1997).

CD measurements

CD spectra were measured at 20 ℃ on Jasco J-600 and J-720 spectropolarimeter equipped with water-circulating cell holders. The spectra were obtained in 100 mM potassium phosphate buffer (pH 7.5) containing 20 mM KCl. The protein concentration was 0.1 mg/ml and 1.0 mg/ml for the far- and near-UV ranges, respectively. The optical path length was 0.2 cm for the far-UV range, and 0.5 cm for the near UV range. Spectra for CD between 200 and 250 nm were obtained using a scanning speed of 20 nm min^{-1}, a time response of 1 s, a bandwidth of 1 nm, and an average over 8 scans. For CD spectra between 250 and 320 nm, a scanning speed of 10 nm min^{-1} and a time response of 8 s

were used. Thermal denaturation curves were determined by monitoring the CD value at 222 nm as the temperature was increased by 1.0 °C per min. The temperature of the sample solution was directly measured by a thermistor (Takara, D641).

DNA-binding assay

The 22-mer oligonucleotide, CACCCTAACTGACACACATTCT, which is the fragment of cognate DNA containing the Myb-binding site in the simian virus 40 enhancer sequence (MBS-I) (Nakagoshi *et al.*, 1990), and its base substituted mutant, CACCCTAACTAACACACATTCT ([G8A]MBS-I), which is the substrate for the non-specific binding of c-Myb (Tanikawa *et al.*, 1993), were prepared. The complementary strands were annealed and end-labeled with [γ-^{32}P]ATP (Amersham) using T4 polynucleotide kinase (Toyobo, Osaka). The labeled DNAs were purified by passage through spin columns (Pharmacia, HR-300). The binding of each R2R3 mutant to the DNA was assayed using a filter binding method, as described previously (Oda *et al.*, 1997). The efficiency of this assay has already been shown (Takeda *et al.*, 1989; Sarai and Takeda, 1989; Tanikawa *et al.*, 1993).

2.3 Results

Prior to the current mutational analyses, the Cys130 in R2, which is the only cysteine residue in the c-Myb R2R3 and is located at a position equivalent to an isoleucine in R3, was replaced with Ile, to facilitate the protein purification and the DNA-binding assay (Guehmann *et al.*, 1992). It has been shown that the affinity and the specificity of the C130I mutant are almost equal to those of the wild-type (Myrset *et al.*, 1993; Oda *et al.*, 1997).

The constructs and the observed dissociation constants of the current R2R3 mutant proteins to MBS-I are summarized in Table 2-1, with the standard C130I mutant protein. The naming rules for each mutant protein are as follows: Repeat R2 is divided into six regions: the N-terminus (n or N), the three helices (h_1, h_2, and h_3, or H_1, H_2, and H_3), and the two turns between the first and the second helices (t_{12} or T_{12}) and between the second and the third helices (t_{23} or T_{23}). Here, residues 137 and 138 just after helix-3 are included in h_3 or H_3 for simplicity. When the sequence of each region is that of R2, small letters, n, h, and t, are used. In contrast, when the sequence is that of R3, capital letters, N, H, and T, are used. When further local mutations are introduced in the helices, italicized letters, *h* and *H*, are used with alphabetic designators after the helix number. In addition, a mutant with completely tandem repeats of R3 including the C-termini was also constructed with the linker composed of three residues, -Asn-Pro-Glu-. This mutant is separately named as R3-NPE-R3.

In Table 2-1, the precise amino acid sequence of each mutant is indicated, using red and blue one-letter amino acid codes, which correspond to the residues in R2 and in R3, respectively. The black one-letter codes are the amino acid residues conserved in both of the R2 and R3 repeats of the C130I standard.

The dissociation constants to the [G8A]MBS-I were also determined for all of the mutant proteins in Table 2-1, and they were all above μM concentrations, except for R2($NH_1T_{12}H_2T_{23}H_3$)R3, R2($nH_1t_{12}h_2t_{23}h_3$)R3, and R2($nH_1t_{12}H_2t_{23}h_3$)R3 (see below).

Group A mutations in R2

At first, we tried to make R2 as similar to R3 as possible. Unfortunately, when R2 was completely replaced with R3, except for the linker from Asn139 to Glu141, the expressed mutant protein, R2($NH_1T_{12}H_2T_{23}H_3$)R3, had poor solubility and formed inclusion bodies. The basic amino acids just after helix-3 of R3 may be required for the formation of the correct tertiary structure (Ogata *et al.*, 1992). Therefore, as a reference, we constructed the other mutant with completely tandem repeats, R3-NPE-R3. This peptide was soluble, and the far-UV CD spectrum indicated a typical α-helical structure, which is similar to that of the C130I standard mutant (Figure 2-2). The profile of R3-NPE-R3 spectrum is very similar to R3 alone, which was previously reported by Sarai *et al.* (1993). However, the affinity to MBS-I was very low, with a dissociation constant on the μM scale. It had a similar low binding affinity to [G8A]MBS-I, suggesting that the binding was non-specific. The result agrees with the previous observation that R3 alone did not show any specific DNA-binding (Sakura *et al.*, 1989).

Next, when the helix-3 from the original R2 was used, the expressed peptide, R2($NH_1T_{12}H_2T_{23}h_3$)R3, was produced in the soluble fraction of *Escherichia coli* BL21, and it was purified using the same protocol as that for the wild-type. The affinity to MBS-I was also very low, similar to that of R3-NPE-R3. The far-UV CD spectrum of the mutant protein, indicated in Figure 2-2, is also similar to that of R3-NPE-R3, and the secondary structure was stable, at least below 30 °C. In addition, slight modifications in the hydrophobic core of helix-3, h_{3a} and h_{3b} and restoration of the original turn of t_{23} resulted in the same low affinities to MBS-I. As indicated in Figure 2-2, R2($NH_1T_{12}H_2T_{23}h_{3b}$)R3, whose all core residues are derived from R3, has the similar α-helical structure to that of R3-NPE-R3.

Therefore, all of the group A mutants bound to DNA only with affinities as low as those of non-specific binding, although the overall helical conformations were stable and similar to the C130I standard.

Group B mutations in R2

In order to investigate the residues in R2 that are necessary for specific DNA-binding, we made a series of group B mutations by systematically exchanging the regions. Here, the N-terminus, the first helix, and the second helix were individually or simultaneously replaced with the corresponding R3 sequences to generate ten R2R3 mutant proteins, as indicated in Table 2-1. The third helix (helix-3), as well as the two turns, t_{12} and t_{23}, of the original R2 were retained. Since only few amounts of R2(n$H_1t_{12}h_2t_{23}h_3$)R3 and R2(n$H_1t_{12}H_2t_{23}h_3$)R3 could be expressed, Arg153, Ile154, and Tyr156 were substituted with the corresponding amino acids in R2, Gln, Arg, and Ile, respectively, to produce a modified helix-1, H_{1a}, instead of the original H_1. The mutant proteins, R2(n$H_{1a}t_{12}h_2t_{23}h_3$)R3 and R2(n$H_{1a}t_{12}H_2t_{23}h_3$)R3 were well expressed and soluble. When the N-terminus was simultaneously replaced with the R3 sequence, R2(N$H_1t_{12}h_2t_{23}h_3$)R3 could be expressed, and the affinities to the DNA and the CD spectra were measured with those of R2(N$H_{1a}t_{12}h_2t_{23}h_3$)R3.

In Table 2-1, it is clear that three mutant proteins with h_1, R2(N$h_1t_{12}h_2t_{23}h_3$)R3, R2(n$h_1t_{12}H_2t_{23}h_3$)R3, and R2(N$h_1t_{12}H_2t_{23}h_3$)R3, had similar K_d values to that of the C130I standard. Thus, the sequences in the N-terminus and helix-2 are exchangeable with those in R3, respectively. On the contrary, once helix-1 in R2 was replaced with the corresponding sequence H_1 or the slightly modified H_{1a} in R3, specific binding to the MBS-I was never recovered, irrespective of the other N-terminus or helix-2. The results clearly indicate that helix-1 of R2 plays an important role in specific DNA-binding.

Figure 2-3a shows the far-UV CD spectra of the C130I standard, R2(N$h_1t_{12}h_2t_{23}h_3$)R3, R2(n$H_{1a}t_{12}h_2t_{23}h_3$)R3, R2(n$h_1t_{12}H_2t_{23}h_3$)R3, and R2(n$H_{1a}t_{12}H_2t_{23}h_3$)R3, at 20 °C. All of them show typical α-helical conformations. In R2(n$H_{1a}t_{12}h_2t_{23}h_3$)R3, the helical content seems to be slightly reduced, as well as R2(N$H_1t_{12}h_2t_{23}h_3$)R3 in Figure 2-4a. However, the combination of H_{1a} and H_2 recovered it, as indicated in R2(n$H_{1a}t_{12}H_2t_{23}h_3$)R3.

In addition, reversible thermal denaturation of the mutant proteins indicated in Figure 2-3a were observed by monitoring the CD value at 222 nm, and each fraction unfolded is plotted as a function of the temperature in Figure 2-3b. The melting temperature (T_m) at 50 % fraction unfolded of each mutant protein is around 45 to 52 °C, and the unfolding is as cooperative as that of the C130I standard except R2(nH_{1a}t$_{12}$H$_2$t$_{23}$h$_3$)R3. It is interesting that only the mutant proteins with the H_2 sequence have slightly lower T_m values than that of C130I.

Thus, the reduced binding affinity with the H_1 or H_{1a} sequence is not due to an overall deformation of the mutant protein structures.

Group C mutations in R2

The above conclusion made in the group B mutations was also confirmed by introducing individual point mutations within helix-2 and the first turn in the group C mutations, while leaving all the other residues as those of R2. As long as helix-1 retained the original sequence in R2, the mutant proteins bound to the MBS-I in a specific manner, when each residue in helix-2 was exchanged individually with that in R3. Even when the turn between helix-1 and helix-2 was exchanged with that in R3, which is one residue shorten, the mutant protein with the h_1 sequence had a high affinity to MBS-I, and the far-UV CD spectrum is very similar to that of the C130I standard (data not shown).

Group D mutations in R2

We made detailed investigations of helix-1 of R2, by producing a series of group D mutant proteins with modified helices from H_{1b} to H_{1i}, in addition to H_{1a} in the group B mutations. The helix-1 mutant proteins that have almost the same affinity to MBS-I as that of the C130I standard are H_{1b}, H_{1c}, H_{1e}, H_{1f}, and H_{1g} (see Table 2-1). Among them, R2(NH_{1g}t$_{12}$h$_2$t$_{23}$h$_3$)R3 has only two amino acid residues, Val103 and Val107, originating from R2, with all the other residues in helix-1 replaced with those of R3. This high

affinity should be compared with the low affinity of R2($NH_1t_{12}h_2t_{23}h_3$)R3 in the group B mutations. This highlights the fact that the two Val residues are indispensable for specific DNA-binding.

In contrast, when Val107 was replaced with the corresponding His in the group D mutations, the DNA affinities of the mutant proteins always decreased significantly. When Val103 was replaced with the corresponding Ile residue, the effect was not as striking as that with the replacement at position 107, but the DNA affinity also greatly decreased.

The finding that Val103 and Val107 are essential residues for specific DNA-binding was also supported by the small K_d values of R2($NH_{1g}t_{12}H_2t_{23}h_3$)R3, in which half of the original sequence of R2 was exchanged with the corresponding sequence in R3. This specific affinity should be compared with the low affinity of R2($NH_{1a}t_{12}H_2t_{23}h_3$)R3 in the group B mutations.

Figures 2-4a and 4b indicate the CD spectra of the mutant proteins, R2($NH_{1g}t_{12}h_2t_{23}h_3$)R3, R2($NH_{1h}t_{12}h_2t_{23}h_3$)R3, and R2($NH_{1i}t_{12}h_2t_{23}h_3$)R3, with R2($NH_1t_{12}h_2t_{23}h_3$)R3 in the group B mutations and the C130I standard, at the far- and the near-UV regions, respectively. It is interesting that the introduction of Val107 always makes the CD spectrum similar to that of the C130I standard at both the far- and near-UV ranges.

Group E mutations in R2

The above finding was further examined by the group E mutations, in which a few amino acid residues in helix-1 in the standard C130I were locally replaced with the residues corresponding to those in R3. In fact, the introduction of Ile103 caused a distinct reduction in MBS-I binding, as indicated in Table 2-1. However, the reduction in the DNA affinity by the introduction of His107 was not so remarkable as that by the introduction of Ile103.

Figures 2-5a and 5b indicate the CD spectra of the mutant proteins, R2(nh_{1a}t$_{12}$h$_2$t$_{23}$h$_3$)R3, R2(nh_{1b}t$_{12}$h$_2$t$_{23}$h$_3$)R3, and R2(nh_{1c}t$_{12}$h$_2$t$_{23}$h$_3$)R3, with the C130I standard, at the far- and the near-UV ranges, respectively. At the far-UV range, all three mutant proteins show almost identical CD spectra. However, at the near-UV range, the individual spectra differed slightly, and R2(nh_{1c}t$_{12}$h$_2$t$_{23}$h$_3$)R3 displayed the CD spectrum most similar to that of the C130I standard.

2.4 Discussion

The group A mutants did not exhibit DNA-binding as strong as either that of the C130I standard or the wild-type R2R3. Therefore, an engineering strategy that simply fuses the repeats to engender new base recognition is not feasible for the c-Myb DNA-binding domain, although the strategy was successful with the zinc-fingers and other DBDs (Pomerantz *et al.*, 1995; Kim *et al.*, 1997). This is probably because the linker between the repeats is so short in the c-Myb DBD, that the cooperativity between R2 and R3 is too strong to be tailored. In fact, the unique features of the DNA-binding mode of c-Myb are that the recognition helices of R2 and R3 are closely packed together, and that they directly contact each other to bind cooperatively to the specific base sequence (Ogata *et al.*, 1994). A recent mutational study indicated that the amino acid residues located at the linker between R2 and R3 of the c-Myb DBD are also associated with specific recognition of DNA (Hegvold and Gabrielsen, 1996).

In the group B and C mutations, which swapped the corresponding helices and turns between R2 and R3, it was clearly shown that the helix-1 sequence in R2 is non-exchangeable, in addition to the recognition helix-3, while the other N-terminus, turns, and helix-2 are exchangeable into the corresponding sequences in R3. Every mutant protein forms the stable tertiary structure, which has a similar or a slightly reduced α-helical content to that of the C130I standard.

The mutant proteins, R2($NH_{1h}t_{12}h_2t_{23}h_3$)R3 and R2($NH_{1i}t_{12}h_2t_{23}h_3$)R3, which introduce Val at either position 103 or 107, recovered the DNA-binding to some extent. Moreover, the R2($NH_{1g}t_{12}h_2t_{23}h_3$)R3 mutant, with Vals at positions 103 and 107, specifically bound to the cognate DNA as strongly as the wild-type. These findings in the group D mutations indicate that only the two inner residues, Val103 and Val107 in the R2 helix-1, are indispensable for the specific binding of the MBS-I.

Previously, it was found that the replacement of Val103 with Leu reduced the binding affinity by one-third, although it stabilized the R2 structure significantly (Ogata *et*

al., 1996). It was then concluded that a cavity-filling mutation in R2 reduced the specific DNA-binding, by hampering the reorientation of the R2 N-terminal region, especially of Trp95, upon DNA-binding. In this study, a similar effect can explain the essential role of Val103 for DNA-binding, in both the group D and E mutations.

In addition, Val107 contributes to the specific binding of R2R3. The difference between the far-UV CD spectra of R2($NH_{1h}t_{12}h_2t_{23}h_3$)R3 and R2($NH_{1g}t_{12}h_2t_{23}h_3$)R3 in Figure 2-4a suggests that Val107 is necessary for the correct formation of helix-1. Moreover, the near-UV CD spectrum is remarkably affected by the amino acid at residue 107, as indicated in Figure 2-4b, suggesting that this residue switches the helix packing from the R2-type to that of the R3-type. The Val103 and Val107 residues probably function cooperatively to construct the characteristic core of R2.

The individual repeats, R2 and R3, have almost identical folds, with an entire backbone r.m.s.d. less than 1.3 Å, irrespective of the presence of MBS-I (Ogata *et al.*, 1994, 1995). However, when the individual structures were analyzed more precisely, the angle between the axes of helix-1 and -3 in R2 was found to be slightly different from that in R3. In Figure 2-6, the deviations in the helix axes can be seen, when the backbone heavy atoms of helix-3 of R3 in the minimized averaged structure of the MBS-I complexed form of R2R3 (Ogata *et al.*, 1994; PDB code, 1MBE) were superimposed onto those of R2.

The axes of helix-1 and -3 were determined in detail, following the algorithm of Barlow and Thornton (1988), for the 25 NMR complex structures (Ogata *et al.*, 1994; PDB code, 1MSF), and the 50 free structures of both R2 (Ogata *et al.*, 1995; 1MBH) and R3 (Ogata *et al.*, 1995; 1MBK). Then, the angles and the deviations between the helix axes as the vectors were calculated. Consequently, in the free form, the angles in R2 and R3 were 94.7° ± 2.0° and 98.7° ± 2.8°, respectively. In the complex form, they were 89.9° ± 4.9° and 97.5° ± 3.2°, respectively. The angle in R3 is remarkably larger than that in R2, irrespective of DNA-binding. It is interesting that the angles in R3 are almost

the same in both the free and the MBS-I complex structures. In contrast, the angle in R2 significantly decreased with DNA-binding. Therefore, in the DNA complex structure, the angle between the helix axes in R2 is 8 degrees less than that in R3. This difference could be accompanied by the local conformational change of Trp95 in R2, which could correspond to the changes in the far- and near-UV CD spectra in the mutant proteins.

Thus, the precise correlation of helix-1 and recognition helix-3 characterizes repeat R2 in specific DNA recognition, and the two residues, Val103 and 107, within the hydrophobic core, play an essential role in its formation, although they are located on helix-1, far from the direct interface with the DNA (see Figure 2-6).

The results of the group E mutations also confirmed the current finding of the importance of residues 103 and 107. When helix-1 is mostly made up of the sequence in R3, Val at position 107 is more effective for DNA-binding than Val at position 103. In contrast, when helix-1 in R2 is mostly composed of the sequence in the original R2, Val103 is more effective than Val107. This difference should be caused by the different contribution from the environment of residue 107. Therefore, the group E mutants, from R2(nh_{1c}T$_{12}$h$_2$t$_{23}$h$_3$)R3 to R2(nh_{1h}t$_{12}$h$_2$t$_{23}$h$_3$)R3, were constructed and the affinities were measured. As indicated in Table 2-1, the more amino acid residues were changed to the corresponding residues in R3, the more weakly the mutant proteins bind the MBS-I. However, no significant cooperativity was observed between His107 and any particular residue in its neighborhood. These results indicate that the importance of Val107 in the group E mutants would be influenced by most of the surrounding residues additively, not by only one or a few residues.

Although the simple fusion engineering was not successful, it was possible to engineer an R2R3 to be as similar to the complete R3 tandem as possible, revealing Val103 and Val107 in R2 are required for specific DNA-binding. In the mutant protein, R2(NH_{1g}t$_{12}$H$_2$t$_{23}$h$_3$)R3, only the residues Val103, Val107, and recognition helix-3 with the two short turns originate from the R2 sequence, while all the other residues are those

of R3. This mutant protein specifically binds to the MBS-I sequence, with a K_d value of 24 nM, while it binds to the [G8A]MBS-I only weakly, with a K_d value larger than μM.

In conclusion, from the combination of mutational analyses with filter binding assays, we have determined the two essential residues in R2, that are characteristic of the R2 fold, and are required for specific DNA-binding in addition to recognition helix-3. These residues are involved in the interdigitation between helix-1 and helix-3 in the hydrophobic core of R2, form the correct helix-1 structure, and characterize the packing of the R2 fold, which is slightly different from that of R3, but meaningful for the biological function.

2.5 References

Barlow, D. J., and Thornton, J. M. (1988) *J. Mol. Biol.*, **201**, 601-619.

Bork, P., Downing, A. K., Kieffer, B., and Campbell, I. D. (1996) *Quart. Rev. Biophys.*, **29**, 119-167.

Frampton, J., Gibson, T. J., Ness, S. A., Doderlein, G., and Graf, T. (1991) *Protein Eng.*, **4**, 891-901.

Gabrielsen, O. S., Sentenac, A., and Fromageot, P. (1991) *Science*, **253**, 1140-1143.

Guehmann, S., Vorbrueggen, G., Kalkbrenner, F., and Moelling, K. (1992) *Nucleic Acids Res.*, **20**, 2279-2286.

Hegvold, A. B., and Gabrielsen, O. S. (1996) *Nucleic Acids Res.*, **24**, 3990-3995.

Higuchi, R. (1989) in *PCR Technology* (Erlich, H.A., ed.), pp.61-88, Stockton Press, New York.

Kim, J. -S., Kim, J., Cepek, K. L., Sharp, P. A., and Pabo, C. O. (1997) *Proc. Natl. Acad. Sci. USA*, **94**, 3616-3620.

Klemm, J. D., Rould, M. A., Aurora, R., Herr, W., and Pabo, C. O. (1994) *Cell*, **77**, 21-32.

König, P., Giraldo, R., Chapman, L., and Rhodes, D. (1996) *Cell*, **85**, 125-136.

Myrset, A. H., Bostad, A., Jamin, N., Lirsac, P. -N., Toma, F., and Gabrielsen, O. S. (1993) *EMBO J.*, **12,** 4625-4633.

Nakagoshi, H., Nagase, T., Kanei-Ishii, C., Ueno, Y., and Ishii, S. (1990) *J. Biol. Chem.*, **265,** 3479-3483.

Nishina, Y., Nakagoshi, H., Imamoto, F., Gonda, T. J., and Ishii, S. (1989) *Nucleic Acids Res.*, **17,** 107-117.

Oda, M., Furukawa, K., Ogata, K., Sarai, A., Ishii, S., Nishimura, Y., and Nakamura, H. (1997) *J. Biol. Chem.*, **272,** 17966-17971.

Ogata, K., Hojo, H., Aimoto, S., Nakai, T., Nakamura, H., Sarai, A., Ishii, S., and Nishimura, Y. (1992) *Proc. Natl. Acad. Sci. USA*, **89,** 6428-6432.

Ogata, K., Morikawa, S., Nakamura, H., Sekikawa, A., Inoue, T., Kanai, H., Sarai, A., Ishii, S., and Nishimura, Y. (1994) *Cell*, **79,** 639-648.

Ogata, K., Morikawa, S., Nakamura, H., Hojo, H., Yoshimura, S., Zhang, R., Aimoto, S., Ametani, Y., Hirata, Z., Sarai, A., Ishii, S., and Nishimura, Y. (1995) *Nature Struct. Biol.*, **2,** 309-320.

Ogata, K., Kanei-Ishii, C., Sasaki, M., Hatanaka, H., Nagadoi, A., Enari, M., Nakamura, H., Nishimura, Y, Ishii, S., and Sarai, A. (1996) *Nature Struct. Biol.*, **3,** 178-187.

Pavletich, N. P., and Pabo, C. O. (1991) *Science*, **252**, 809-817.

Pavletich, N. P., and Pabo, C. O. (1993) *Science*, **261**, 1701-1707.

Pomerantz, J. L., Sharp, P. A., and Pabo, C. O. (1995) *Science*, **267**, 93-96.

Saikumar, P., Murali, R., and Reddy, E. P. (1990) *Proc. Natl. Acad. Sci. USA*, **87,** 8452-8456.

Sakura, H., Kanei-Ishii, C., Nagase, T., Nakagoshi, H., Gonda, T. J., and Ishii, S. (1989) *Proc. Natl. Acad. Sci. USA*, **86,** 5758-5762.

Sarai, A., and Takeda, Y. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 6513-6517.

Sarai, A., Uedaira, H., Morii, H., Yasukawa, T., Ogata, K., Nishimura, Y., and Ishii, S. (1993) *Biochemistry*, **32**, 7759-7764.

Takeda, Y., Sarai, A., and Rivera, V. M. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 439-443.

Tanikawa, J., Yasukawa, T., Enari, M., Ogata, K., Nishimura, Y., Ishii, S., and Sarai, A. (1993) *Proc. Natl. Acad. Sci. USA*, **90,** 9320-9324.

Xu, W., Rould, M. A., Jun, S., Desplan, C., and Pabo, C. O. (1995) *Cell*, **80**, 639-650.

Table 2-1. Sequences of R2R3 mutant proteins and the dissociation constants for the 22-mer MBS-I fragment

protein		amino acid sequence[a]	K_d (nM)
		n h1 t12 h2 t23 h3	
		90 100 110 120 130 140	
R2R3 (C130I)	R2	LIKGPWTKEEDQRVIKLVQKYGPKRWSVIAKHLKGRIGKQIRERWHNHLNPE	5.5
		150 160 170 180 190	
	R3	VKKTSWTEEEDRIIYQAHKRLG-NRWAEIAKLLPGRTDNAIKNHWNSTMRRKV	
(i) group A mutations in R2[b]			
R2 ($NH_1T_{12}H_2T_{23}H_3$) R3		VKKTSWTEEEDRIIYQAHKRLG-NRWAEIAKLLPGRTDNAIKNHWNSTMNPE	I.B.[c]
R3-NPE-R3		VKKTSWTEEEDRIIYQAHKRLG-NRWAEIAKLLPGRTDNAIKNHWNSTMRRKVNPE	>1000
R2 ($NH_1T_{12}H_2T_{23}h_3$) R3		VKKTSWTEEEDRIIYQAHKRLG-NRWAEIAKLLPGRTGKQIRERWHNHLNPE	>1000
R2 ($NH_1T_{12}H_2T_{23}h_{3a}$) R3		VKKTSWTEEEDRIIYQAHKRLG-NRWAEIAKLLPGRTGKQIRERWHNHMNPE	>1000
R2 ($NH_1T_{12}H_2T_{23}h_{3b}$) R3		VKKTSWTEEEDRIIYQAHKRLG-NRWAEIAKLLPGRTGKQIREHWHNHMNPE	>1000
R2 ($NH_1T_{12}H_2t_{23}h_3$) R3		VKKTSWTEEEDRIIYQAHKRLG-NRWAEIAKLLKGRIGKQIRERWHNHLNPE	>1000
(ii) group B mutations in R2[b]			
R2 ($Nh_1t_{12}h_2t_{23}h_3$) R3		VKKTSWTKEEDQRVIKLVQKYGPKRWSVIAKHLKGRIGKQIRERWHNHLNPE	2.2
R2 ($nH_1t_{12}h_2t_{23}h_3$) R3		LIKGPWTEEEDRIIYQAHKRLGPKRWSVIAKHLKGRIGKQIRERWHNHLNPE	N.E.[d]
R2 ($nH_{1a}t_{12}h_2t_{23}h_3$) R3		LIKGPWTEEEDQRIIQAHKRLGPKRWSVIAKHLKGRIGKQIRERWHNHLNPE	>1000
R2 ($nh_1t_{12}H_2t_{23}h_3$) R3		LIKGPWTKEEDQRVIKLVQKYGPKRWAEIAKLLKGRIGKQIRERWHNHLNPE	11
R2 ($NH_1t_{12}h_2t_{23}h_3$) R3		VKKTSWTEEEDRIIYQAHKRLGPKRWSVIAKHLKGRIGKQIRERWHNHLNPE	>1000
R2 ($NH_{1a}t_{12}h_2t_{23}h_3$) R3		VKKTSWTEEEDQRIIQAHKRLGPKRWSVIAKHLKGRIGKQIRERWHNHLNPE	>1000
R2 ($Nh_1t_{12}H_2t_{23}h_3$) R3		VKKTSWTKEEDQRVIKLVQKYGPKRWAEIAKLLKGRIGKQIRERWHNHLNPE	6.3
R2 ($nH_1t_{12}H_2t_{23}h_3$) R3		LIKGPWTEEEDRIIYQAHKRLGPKRWAEIAKLLKGRIGKQIRERWHNHLNPE	N.E.[d]
R2 ($nH_{1a}t_{12}H_2t_{23}h_3$) R3		LIKGPWTEEEDQRIIQAHKRLGPKRWAEIAKLLKGRIGKQIRERWHNHLNPE	>1000
R2 ($NH_{1a}t_{12}H_2t_{23}h_3$) R3		VKKTSWTEEEDQRIIQAHKRLGPKRWAEIAKLLKGRIGKQIRERWHNHLNPE	>1000

(continued)

(iii) group C mutations in R2[b)]		
R2 ($nh_1T_{12}h_2t_{23}h_3$) R3	LIKGPWTKEEDQRVIKLVQKYG-NRWSVIAKHLKGRIGKQIRERWHNHLNPE	11
R2 ($nh_1t_{12}h_{2a}t_{23}h_3$) R3	LIKGPWTKEEDQRVIKLVQKYGPKRWAVIAKHLKGRIGKQIRERWHNHLNPE	15
R2 ($nh_1t_{12}h_{2b}t_{23}h_3$) R3	LIKGPWTKEEDQRVIKLVQKYGPKRWSEIAKHLKGRIGKQIRERWHNHLNPE	10
R2 ($nh_1t_{12}h_{2c}t_{23}h_3$) R3	LIKGPWTKEEDQRVIKLVQKYGPKRWSVIAKLLKGRIGKQIRERWHNHLNPE	5.2
(iv) group D mutations in R2[b)]		
R2 ($NH_{1b}t_{12}h_2t_{23}h_3$) R3	VKKTSWTKEEDQRVIKAVKRLGPKRWSVIAKHLKGRIGKQIRERWHNHLNPE	4.5
R2 ($NH_{1c}t_{12}h_2t_{23}h_3$) R3	VKKTSWTKEEDQRVIQAVKRLGPKRWSVIAKHLKGRIGKQIRERWHNHLNPE	6.0
R2 ($NH_{1d}t_{12}h_2t_{23}h_3$) R3	VKKTSWTKEEDQRVIQAHKRLGPKRWSVIAKHLKGRIGKQIRERWHNHLNPE	100
R2 ($NH_{1e}t_{12}h_2t_{23}h_3$) R3	VKKTSWTEEEDQRVIQAVKRLGPKRWSVIAKHLKGRIGKQIRERWHNHLNPE	6.9
R2 ($NH_{1f}t_{12}h_2t_{23}h_3$) R3	VKKTSWTEEEDRIVIQAVKRLGPKRWSVIAKHLKGRIGKQIRERWHNHLNPE	5.7
R2 ($NH_{1g}t_{12}h_2t_{23}h_3$) R3	VKKTSWTEEEDRIVYQAVKRLGPKRWSVIAKHLKGRIGKQIRERWHNHLNPE	7.3
R2 ($NH_{1h}t_{12}h_2t_{23}h_3$) R3	VKKTSWTEEEDRIVYQAHKRLGPKRWSVIAKHLKGRIGKQIRERWHNHLNPE	160
R2 ($NH_{1i}t_{12}h_2t_{23}h_3$) R3	VKKTSWTEEEDRIIYQAVKRLGPKRWSVIAKHLKGRIGKQIRERWHNHLNPE	75
R2 ($NH_{1g}t_{12}H_2t_{23}h_3$) R3	VKKTSWTEEEDRIVYQAVKRLGPKRWAEIAKLLKGRIGKQIRERWHNHLNPE	24
(v) group E mutations in R2[b)]		
R2 ($nh_{1a}t_{12}h_2t_{23}h_3$) R3	LIKGPWTKEEDQRIIKLHQKYGPKRWSVIAKHLKGRIGKQIRERWHNHLNPE	90
R2 ($nh_{1b}t_{12}h_2t_{23}h_3$) R3	LIKGPWTKEEDQRIIKLVQKYGPKRWSVIAKHLKGRIGKQIRERWHNHLNPE	30
R2 ($nh_{1c}t_{12}h_2t_{23}h_3$) R3	LIKGPWTKEEDQRVIKLHQKYGPKRWSVIAKHLKGRIGKQIRERWHNHLNPE	12
R2 ($nh_{1c}T_{12}h_2t_{23}h_3$) R3	LIKGPWTKEEDQRVIKLHQKYG-NRWSVIAKHLKGRIGKQIRERWHNHLNPE	6.9
R2 ($nh_{1d}t_{12}h_2t_{23}h_3$) R3	LIKGPWTKEEDQRVIKLHQKLGPKRWSVIAKHLKGRIGKQIRERWHNHLNPE	17
R2 ($nh_{1e}t_{12}h_2t_{23}h_3$) R3	LIKGPWTKEEDQRVIKLHKKLGPKRWSVIAKHLKGRIGKQIRERWHNHLNPE	34
R2 ($nh_{1f}t_{12}h_2t_{23}h_3$) R3	LIKGPWTKEEDQRVIKLHKRYGPKRWSVIAKHLKGRIGKQIRERWHNHLNPE	15
R2 ($nh_{1g}t_{12}h_2t_{23}h_3$) R3	LIKGPWTKEEDQRVIKLHQRLGPKRWSVIAKHLKGRIGKQIRERWHNHLNPE	28
R2 ($nh_{1h}t_{12}h_2t_{23}h_3$) R3	LIKGPWTKEEDQRVIQAHQKYGPKRWSVIAKHLKGRIGKQIRERWHNHLNPE	23

(continued)

[a)] Red and blue one-letter amino acid codes are used, corresponding to the residues in R2 and in R3, respectively. The black one-letter codes are the amino acid residues conserved in both of the R2 and R3 repeats of the C130I standard.

[b)] The naming rules for each mutant protein are described in text. Each red and blue colored helix or turn corresponds to that in R2 and R3, respectively. Green helices have farther local mutations.

[c)] I.B. means that the expressed mutant protein formed an inclusion body.

[d)] N.E. means that the mutant protein could not be expressed enough for observation of the K_d value.

Figure 2-1. Strategy to introduce novel restriction enzyme cleavage sites in pRP23 for construction of the R2R3 mutant genes. Helices-1, -2, and -3 in R2 are abbreviated by h_1, h_2, and h_3, respectively, while those in R3 are H_1, H_2, and H_3, respectively. The *Nco*I and *Bam*HI sites are located at the 5'-end and the 3'-end of the amplified R2R3 fragment, respectively. The *Hind*III site in R2 and the *Eco*RI site in R3 are within the original DNA sequence of c-*myb*. The *Apa*I site was introduced into the first turn (t_{12}) between h_1 and h_2, and the *Spe*I site was introduced into the N-terminal region of R3. These mutations were engineered without affecting the amino acid sequence.

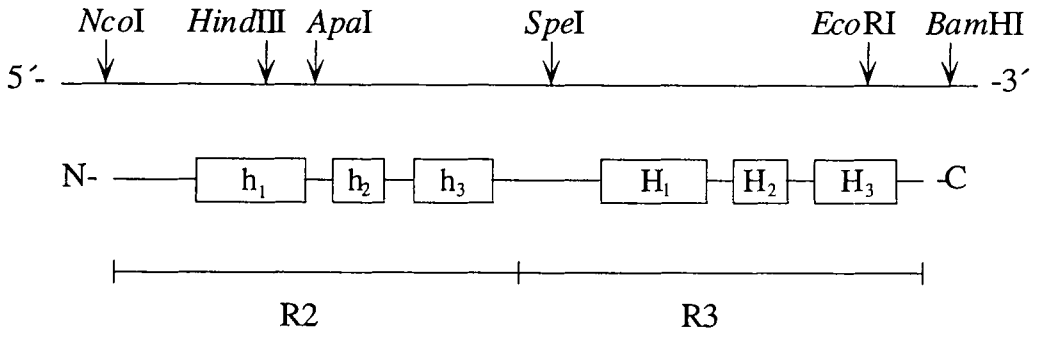


Figure 2-2. Far-UV CD spectra of group A mutant proteins. C130I standard (thick solid line), R3-NPE-R3 (thin solid line), R2($NH_1T_{12}H_2T_{23}h_3$)R3 (broken line), and R2($NH_1T_{12}H_2T_{23}h_{3b}$)R3 (dotted line). The vertical scale is normalized by the mole concentration.

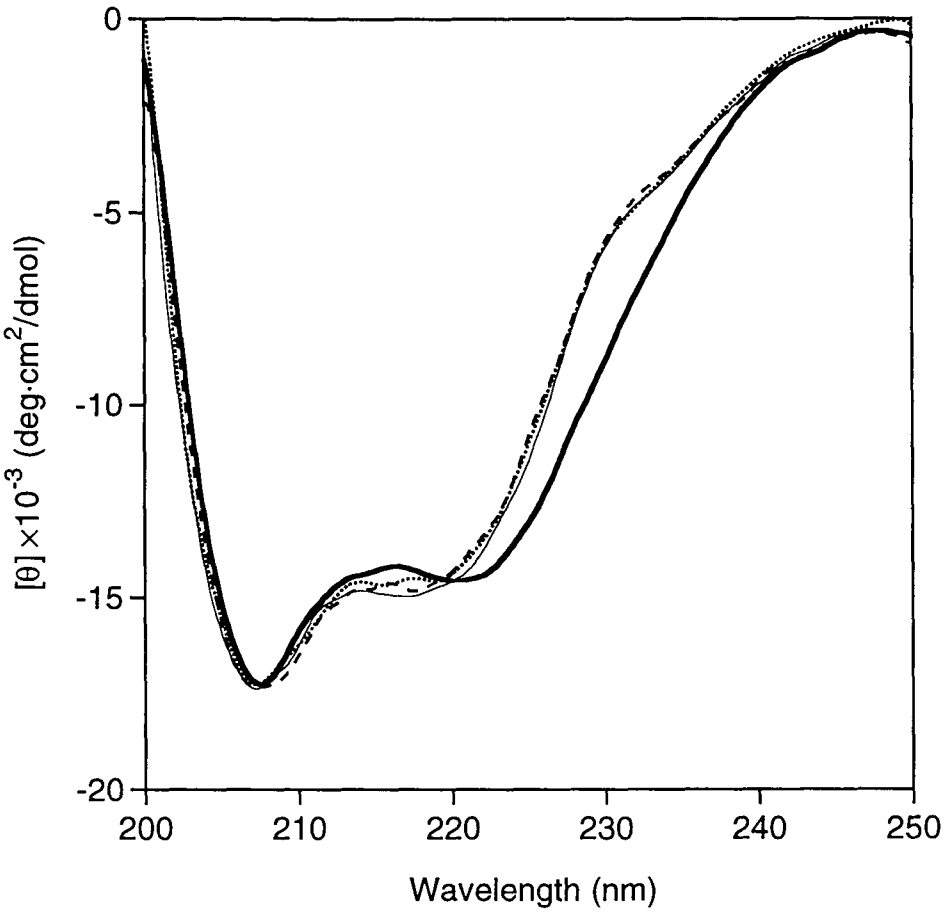


Figure 2-3. Far-UV CD spectra and thermal stabilities of group B mutant proteins. a) Far-UV spectra of C130I standard (thick solid line), R2($Nh_1t_{12}h_2t_{23}h_3$)R3 (thin solid line), R2($nH_{1a}t_{12}h_2t_{23}h_3$)R3 (broken line), R2($nh_1t_{12}H_2t_{23}h_3$)R3 (dotted line), and R2($nH_{1a}t_{12}H_2t_{23}h_3$)R3 (dot-dashed line). The vertical scale is normalized by the mole concentration. b) Thermal denaturation curves of C130I standard (●), R2($Nh_1t_{12}h_2t_{23}h_3$)R3 (O), R2($nH_{1a}t_{12}h_2t_{23}h_3$)R3 (×), R2($nh_1t_{12}H_2t_{23}h_3$)R3 (□), and R2($nH_{1a}t_{12}H_2t_{23}h_3$)R3 (△). The apparent fraction of folded protein by monitoring the CD value at 222 nm is shown as a function of the temperature.

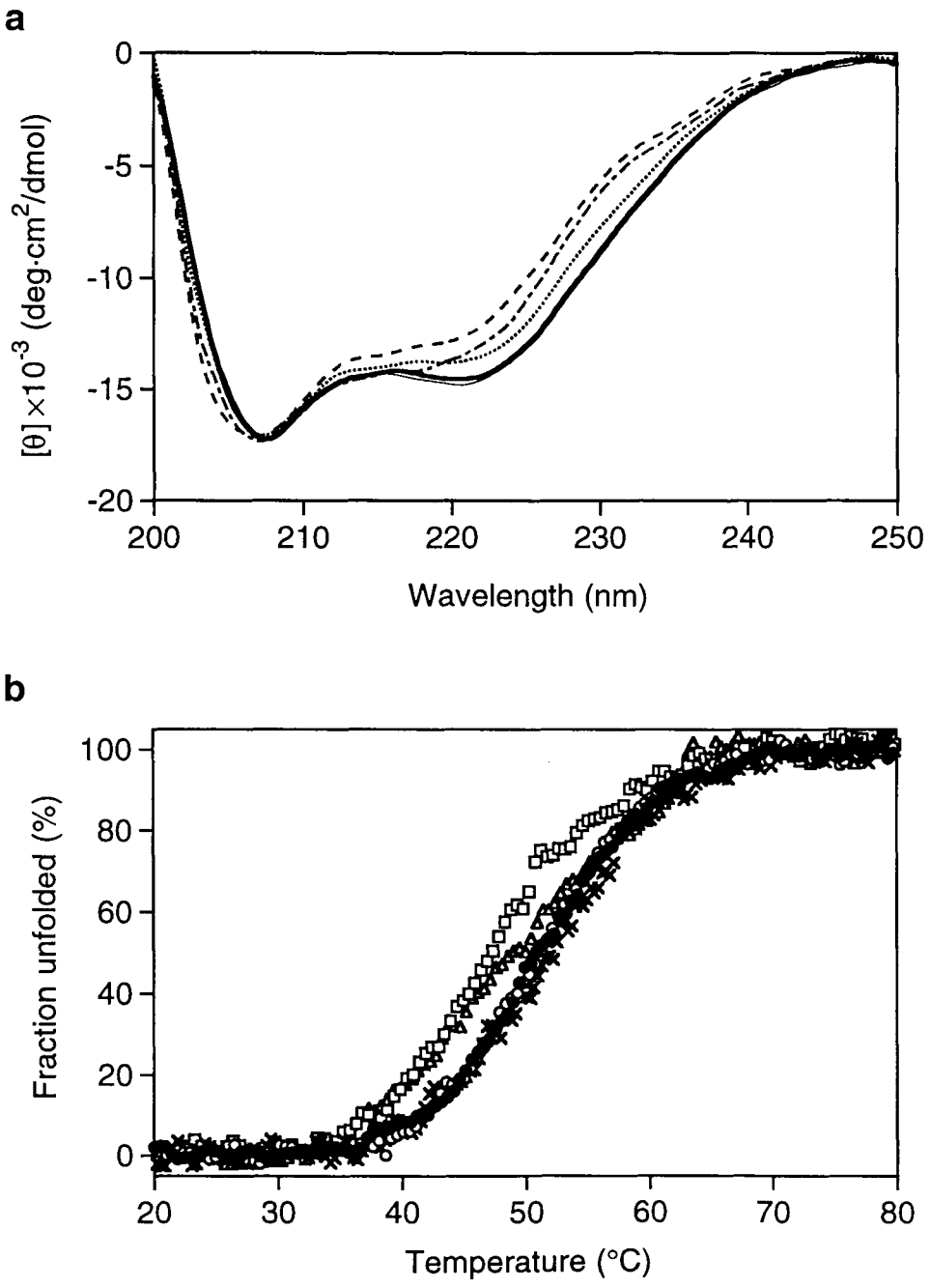


Figure 2-4. CD spectra of group D mutant proteins. The vertical scale is normalized by the mole concentration. a) CD spectra at the far-UV range: C130I standard (thick solid line), R2($NH_1t_{12}h_2t_{23}h_3$)R3 (thin solid line), R2($NH_{1g}t_{12}h_2t_{23}h_3$)R3 (broken line), R2($NH_{1h}t_{12}h_2t_{23}h_3$)R3 (dotted line), and R2($NH_{1i}t_{12}h_2t_{23}h_3$)R3 (dot-dashed line). b) at the near-UV range: The identities of the lines are the same as those in (a).

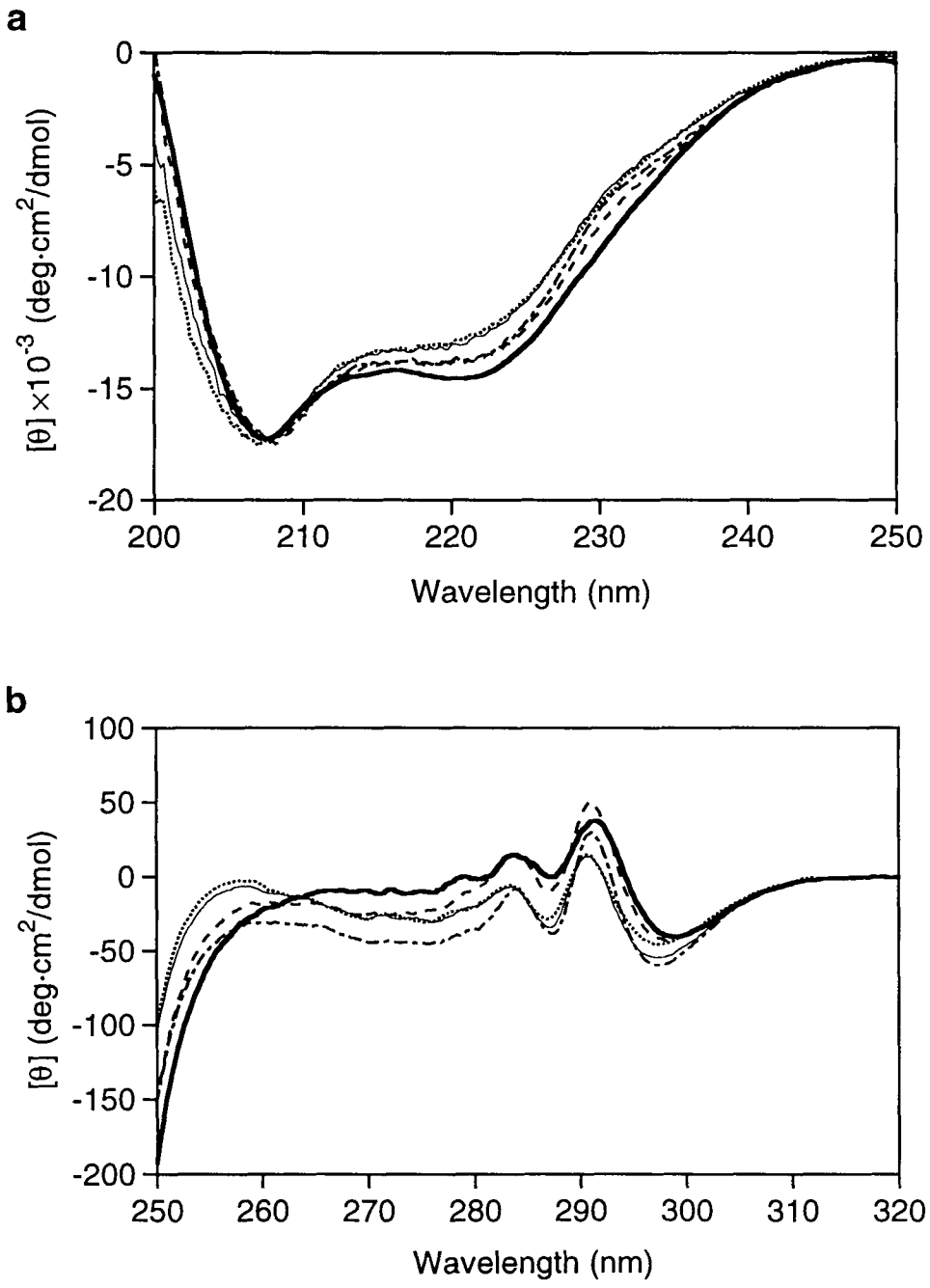


Figure 2-5. CD spectra of group E mutant proteins. The vertical scale is normalized by the mole concentration. a) CD spectra at the far-UV range: C130I standard (thick solid line), R2(nh_{1a}t$_{12}$h$_2$t$_{23}$h$_3$)R3 (thin solid line), R2(nh_{1b}t$_{12}$h$_2$t$_{23}$h$_3$)R3 (broken line), and R2(nh_{1c}t$_{12}$h$_2$t$_{23}$h$_3$)R3 (dotted line). b) at the near-UV range: The identities of the lines are the same as those in (a).

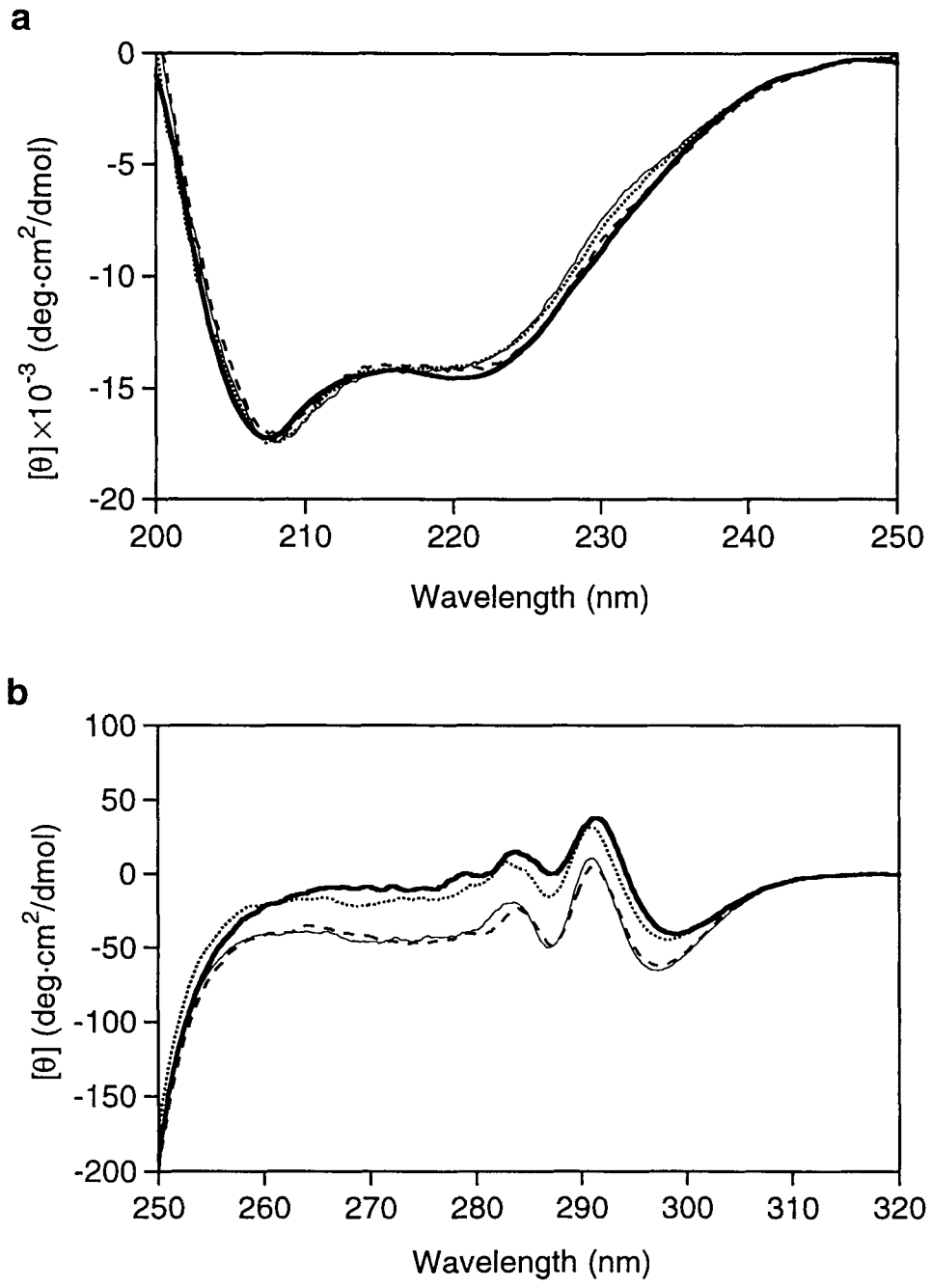
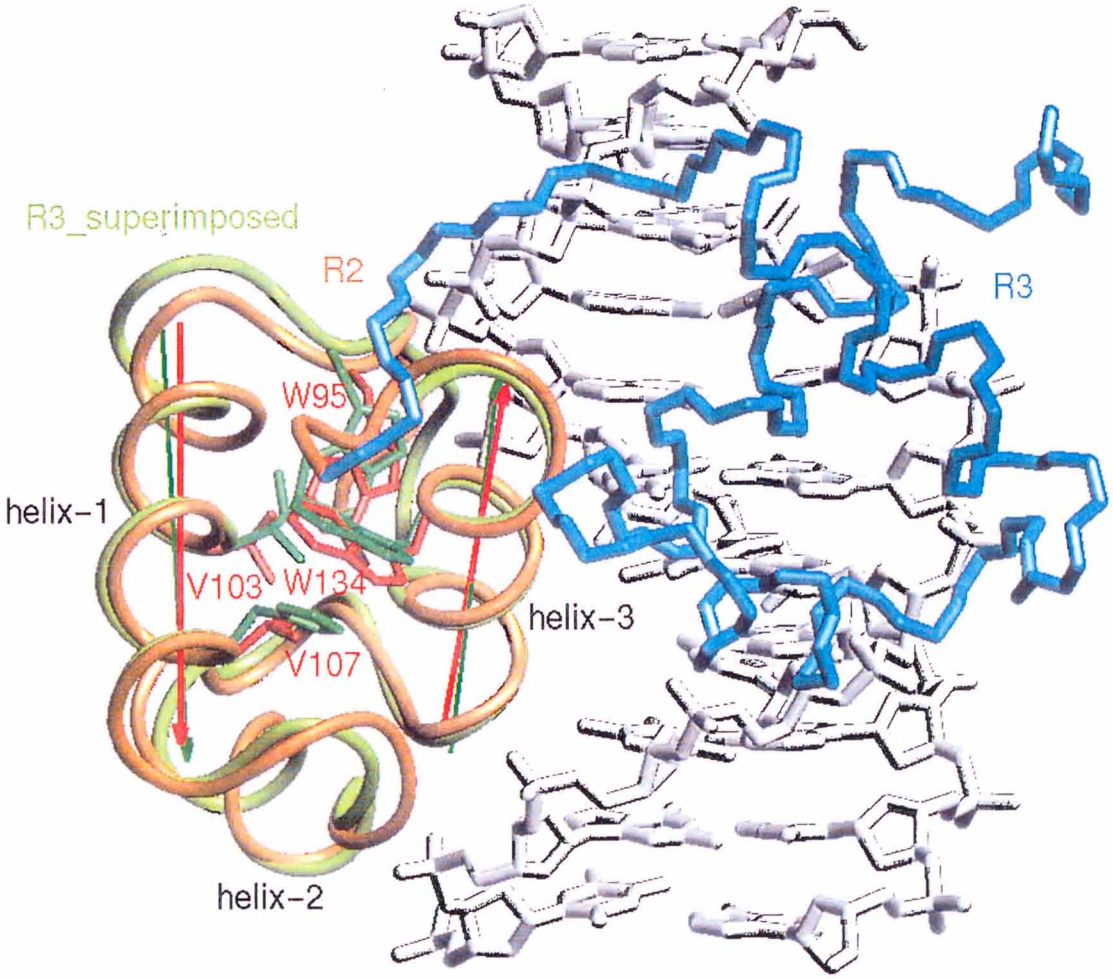


Figure 2-6. Complex structure of the c-Myb R2R3 with the MBS-I fragment (Ogata *et al.*, 1994). The orange and blue lines indicate the backbones of R2 and R3, respectively. The MBS-I DNA fragment is indicated by thick white lines. The backbone heavy atoms of helix-3 in R3 of the MBS-I complexed structure (PDB code, 1MSE) were superimposed onto those in R2, and are shown by a green pipe model. The individual axes of helix-1 and helix-3 of R2 and the superimposed R3 were calculated following Barlow and Thornton (1988), and are indicated by arrows with red and green colors, respectively. The side-chains of indispensable residues in R2, Val103 and Val107, are also shown by red lines, with the corresponding side-chains of Ile155 and His159 in the superimposed R3 by green lines. The figure was drawn with the program InsightII (Molecular Simulations Inc., San Diego).



Chapter 3

Thermodynamics of Specific and Non-specific DNA Binding by the c-Myb DNA-binding Domain

3.1 Introduction

Specific molecular recognition characterizes individual biological macromolecules in the mediation of many biological phenomena. Protein-nucleic acid interactions are the origin of complex genetic regulation. In DNA-bound proteins, the specific base sequences in the DNA are directly and indirectly recognized through the hydrogen bonds, the van der Waals attractions, and the hydrophobic interactions, in addition to the non-specific electrostatic interactions between the negatively charged phosphate backbone and the basic residues of the protein. These specific interactions are observed in the complex structures, yielded from either the rigid body association or the local folding mechanism. The protein-DNA interaction is not a simple phenomenon that can be revealed only from structural analyses, but is a much more intricate one in which different thermodynamic factors stabilize the complex form against the free form. Here, we describe extensive thermodynamic studies of the binding of the c-*myb* protooncogene product (c-Myb) to cognate and non-cognate DNA duplexes, and the dissection of the thermodynamics of specific molecular recognition, based on the known tertiary structures.

In the complex of the c-Myb R2R3 with the Myb-binding DNA sequence (MBS-I), shown in Figure 3-1, the consensus A4, the counterpart guanine of C6, and the last G8 directly interact with Asn183, Lys182, and Lys128, respectively (Ogata *et al.*, 1994). There are several additional possible base contacts, that are less well-defined in the NMR structure, for example, at Py3 with Ser187, and at A5 and its counterpart thymine with Asn136, Asn179, and Asn186 (Ogata *et al.*, 1994). These interactions can also be explained by the available experimental data concerning the effects of both amino acid substitutions and base substitutions on DNA-binding (Saikumar *et al.*, 1990; Gabrielsen *et al.*, 1991; Frampton *et al.*, 1991; Tanikawa *et al.*, 1993; Oda *et al.*, 1997). In addition, not only the residues in the recognition helix, but also the residues in the hydrophobic core, including three conserved tryptophans in each repeat, are indispensable for specific DNA-binding (Kanei-Ishii *et al.*, 1990). Our previous studies have shown that, although

the overall structure of each repeat is almost identical, the two valines in R2, Val103 and Val107, which differ from the corresponding residues in R3, are necessary for DNA-binding (Oda *et al.*, 1998).

The NMR structure of the R2R3-DNA complex shows that there is strong cooperativity between R2 and R3, originating from the putative polar interactions between their recognition helices (Ogata *et al.*, 1994). One of the unique features of R2R3 is that the linker connecting the two repeats is quite short, while other DNA-binding proteins, such as the Oct1 POU-domain (Klemm *et al.*, 1994), the prd paired domain (Xu *et al.*, 1995), and the yeast RAP-1 (König *et al.*, 1996), usually have long and flexible linkers. The short R2R3 linker could be critical for specific DNA-binding, especially for the cooperativity between the adjacent repeats. Recent mutational studies have also shown that the substitutions of some residues in the linker of R2R3 lead to a loss of DNA-binding (Hegvold and Gabrielsen, 1996).

In this study, we characterized the thermodynamics of c-Myb R2R3 binding to DNA, using isothermal titration calorimetry (ITC) to measure both the enthalpy change associated with the formation of the complex and the corresponding binding affinity. With the recent improvements in calorimeter sensitivity and reliability (Wiseman *et al.*, 1989; Livingstone, 1996), ITC has become a powerful tool for direct measurements of the thermodynamic quantities of various biological interactions, including protein-DNA interactions (Ladbury *et al.*, 1994; Hyre and Spicer, 1995; Merabet and Ackers, 1995; Lundbäck and Härd, 1996). These calorimetric measurements reveal the contributions of enthalpy and entropy to the binding free energy. In many cases, the interpretation of the effects on enthalpy and entropy in structure-activity relationships is complicated by the involvement of solvent. To explore the DNA recognition mechanisms by the c-Myb DBD, the binding of R2R3 to non-cognate DNAs, in which the base-pairs in the consensus sequence were substituted, was analyzed and compared to the binding to the cognate DNA. In addition, the interactions of the R2R3 mutant proteins with the various

DNAs were also investigated, using both ITC and a filter binding assay, to reveal the roles of the substituted residues.

In every c-Myb R2R3 and mutant protein used in this study, the Cys130 in R2, which is the only cysteine residue in R2R3 and is located at a position equivalent to an isoleucine in R3, was replaced with Ile, to facilitate the protein purification and the DNA-binding assay (Guehmann *et al.*, 1992). As shown previously, the affinity and the specificity of the C130I mutant are almost equal to those of the wild-type protein (Myrset *et al.*, 1993; Oda *et al.*, 1997), so this mutant protein was used as the standard R2R3 in this study and is denoted as R2R3*.

3.2 Experimental Procedures

Protein preparation

The expression and purification methods of R2R3* and all of the R2R3* mutant proteins were slightly modified from the previous method, to produce a large amounts of proteins (Oda *et al.*, 1997). The plasmids were constructed by site-directed mutagenesis, and were overexpressed in *Escherichia coli* BL21 (DE3). Freshly precultivated cells were inoculated into growth medium containing 100 μg/ml ampicillin and were grown at 37 °C. When the culture reached an OD_{600} of about 0.6, isopropyl-1-thio-ß-D-galactopyranoside was added to a final concentration of 1.0 mM. After the cells were cooled down at 4 °C, they were cultured at 22 °C for another 12 h. The harvested cells were lysed by sonication at 4 °C, and the cell debris was removed by centrifugation. Ammonium sulfate was added to the supernatant to 50% saturation, and the solution was incubated with stirring at 4 °C for 1 h. The supernatant was dialyzed against 50 mM potassium phosphate buffer (pH 7.5) containing 200 mM NaCl, and was then applied to a heparin column (Pharmacia, HiTrap Heparin), connected to an FPLC system. The purified fractions were pooled, and the buffer was exchanged to 100 mM potassium phosphate buffer (pH 7.5) containing 20 mM KCl (buffer A). The protein concentrations were determined from UV absorption at 280 nm and were calculated by using the molar absorption coefficient of 3.7×10^4 M^{-1} cm^{-1} (Tanikawa *et al.*, 1993). The salt concentration was regulated by adding KCl to buffer A.

DNA preparation

The 22-mer oligonucleotides indicated in Figure 3-2 were synthesized and purified by high performance liquid chromatography with a C_{18} reverse-phase column. The complementary strands were annealed, and were dialyzed against buffer A. The DNA concentrations were determined from UV absorption at 260 nm using an average extinction coefficient of 1.32×10^4 M base $pair^{-1}$ cm^{-1} (Mahler *et al.*, 1964).

Isothermal titration calorimetry

Isothermal titration calorimetric experiments were carried out on a Microcal MCS calorimeter interfaced with a microcomputer (Wiseman *et al.*, 1989). All solutions were carefully degassed before the titrations, using the equipment provided with the instrument. The DNA solution (500 μM) was titrated into the protein solution (≈17 μM) using a 100 μl syringe. Each titration consisted of a preliminary 2-μl injection followed by 19 subsequent 5-μl additions. The heat for each injection was subtracted from the heat of dilution of the injectant, which was measured by injecting the DNA solution into the experimental buffer. Each corrected heat was divided by the moles of DNA injected and was analyzed with the Microcal Origin software supplied by the manufacturer.

Circular dichroism

The far- and near- UV CD spectra were measured at 20 °C on Jasco J-600 and J-720 spectropolarimeters, as described previously (Oda *et al.*, 1998). The protein concentration was 0.1 mg/ml and 1.0 mg/ml for the far- and near-UV ranges, respectively. Thermal denaturation curves were determined by monitoring the change in the CD value at 222 nm.

Differential scanning calorimetry

Differential scanning calorimetric experiments were carried out on a Microcal MCS DSC calorimeter. All solutions were carefully degassed before the measurements. Data were collected in the temperature range between 10 °C and 110 °C at a heating rate of 1 °C min^{-1}. The protein concentration was 3.0 mg ml^{-1}. The deconvolutions were done by a non-linear least-square method, as described in the DSC analysis software Microcal Origin.

Filter binding assay

All filter binding assays for the protein-DNA binding were carried out as previously described (Oda *et al.*, 1997). [^{32}P]DNA and various amounts of the c-Myb R2R3 mutant proteins were incubated on ice for 30 min. The final concentration of the [^{32}P]DNA in the binding buffer was 0.4 nM, which was always a lower concentration than the K_a^{-1} value. The incubated samples were filtered through a nitrocellulose membrane (Schleicher and Schuell, BA-85, 0.45 μm) in approximately 10 sec with suction. The filters were dried and were counted by a liquid scintillation counter.

3.3 Results

Determination of thermodynamic parameters

Figure 3-3a shows a typical ITC profile at 20.2 °C for the interaction between the c-Myb R2R3* and its cognate 22-mer DNA containing the Myb-binding site in the simian virus 40 enhancer (MBS-I), indicated in Figure 3-2. An exothermic heat pulse was observed after each injection of MBS-I into R2R3*. Each area of this exothermic peak was integrated, and the heat of dilution of MBS-I was subtracted from the integrated values. The corrected heat was divided by the moles of MBS-I injected, and the resulting values were plotted as a function of the molar ratio, as shown in Figure 3-3b. The resultant data was best-fit according to a model for one binding site, by using a nonlinear least-squares method. The stoichiometry of binding, n, the binding constant, K_a, and the enthalpy change, ΔH, were obtained from the fitted curve. The Gibbs free energy change, ΔG, and the entropy change, ΔS, were calculated from the equation, $\Delta G = -RT \ln K_a = \Delta H - T\Delta S$. The thermodynamic parameters obtained for the interaction of R2R3* and MBS-I at 20.1 (± 0.1) °C are the following: n = 1.00 (± 0.02), K_a = 1.8 (± 0.2) x 10^7 M^{-1}, ΔH = –12.5 (± 0.1) kcal mol^{-1}, ΔS = –9.5 (± 0.5) cal mol^{-1} K^{-1}, and ΔG = –9.7 (± 0.1) kcal mol^{-1}. The experimental errors were estimated from four independent measurements, and they were almost equal to the fitting errors in each measurement.

The magnitudes of the ΔS and ΔG values depend on the concentration units for the standard state. Thus, the unitary entropy change, $\Delta S°$, and the unitary Gibbs free energy change, $\Delta G°$, should be used after subtraction of the cratic part (Gurney, 1953; Torigoe *et al.*, 1995). In the binding of R2R3* to MBS-I, $\Delta S°$ and $\Delta G°$ are –1.3 (± 0.5) cal mol^{-1} K^{-1} and –12.1 (± 0.1) kcal mol^{-1}, respectively. Hereafter, the unitary Gibbs free energy and the unitary entropy are always indicated.

Interaction between the c-Myb R2R3 and the cognate DNA*

ITC measurements of the interaction of R2R3* and MBS-I were performed at five different temperatures, ranging from 11 °C to 30 °C. This reaction showed a strong temperature dependence for both ΔH and $T\,\Delta S^{\circ}$, which both compensated each other to make ΔG° almost insensitive to temperature, as indicated in Figure 3-4. The temperature at which ΔH changes sign was estimated to be 0.3 °C. Obviously, the interaction of R2R3* and MBS-I is enthalpically driven throughout the physiological temperature range, under these conditions. The heat capacity change, ΔC_p, was –0.62 kcal mol^{-1} K^{-1}, calculated from the linear fitting to the ΔH values shown in Figure 3-4, on the assumption that ΔC_p is constant within the experimental temperatures. From the linear van't Hoff plot between 11 °C and 30 °C, the corresponding van't Hoff enthalpy change was calculated as –13.2 (± 4.0) kcal mol^{-1}, which did not deviate much from the calorimetric enthalpy change.

Using a filter binding assay, the K_a value of the interaction between R2R3* and MBS-I, preincubated on ice, was determined to be 1.8 x 10^8 M^{-1} (Oda *et al.*, 1997). In contrast, using the thermodynamic parameters observed from the ITC measurements, we calculated the K_a value as 1.3 x 10^8 M^{-1} at 0 °C. Therefore, the K_a value analyzed by our previous filter binding assay was slightly higher than that by the current ITC measurement, but both values are essentially consistent.

Interactions between the c-Myb R2R3 and the non-cognate DNAs*

For the non-cognate DNA duplexes, the base-pair substitutions were introduced at the positions from 2 to 10 in the 22-mer MBS-I (Figure 3-2). At the base positions 3, 4, 5, 6, 8, and 10, which are specifically recognized by the c-Myb DBD (Tanikawa *et al.*, 1993; Oda *et al.*, 1997), every original base-pair was substituted by the other purine or pyrimidine base to maintain the DNA structure (Calladine, 1982). In [NC-a]MBS-I, all of the bases at positions 4, 5, 6, and 8 were simultaneously exchanged. Since a part of the

consensus base sequence, AAC, remains in the complementary strand of [NC-a]MBS-I, another non-cognate DNA, [NC-b]MBS-I, was also prepared to eliminate any trace of the consensus sequence.

The binding of R2R3* to these 12 non-cognate DNAs was thermodynamically analyzed using ITC, and was compared with the cognate DNA interaction (Table 3-1). The decrease in the current binding affinity (K_a) for each non-cognate DNA duplex was similar to the corresponding K_a value that was previously obtained using the filter binding assay (Tanikawa *et al.*, 1993).

In the A4, C6, and G8 substituted MBS-I, the enthalpic contribution to the reduced affinity was larger than the entropic one at 20 °C, judging from the changes of ΔH ($\Delta\Delta H$) and $\Delta S°$ ($\Delta\Delta S°$) relative to those for the original MBS-I (Table 3-1). In contrast, the reductions of the binding affinities to the T3 and A5 substituted MBS-Is were mainly attributed to the entropy changes. As indicated in the previous report (Tanikawa *et al.*, 1993), the binding affinity to [C10G]MBS-I was significantly reduced, in spite of the facts that this base position is out of the consensus sequence and no residue in R2R3 seems to contact it in the NMR structure (Ogata *et al.*, 1994). The [C6T]MBS-I, [NC-a]MBS-I, and [NC-b]MBS-I binding affinities were reduced as much as 100-fold (Figure 3-3 and Table 3-1), and the stoichiometry number, n, remarkably deviated from 1.0, while the n values were nearly equal to 1.0 in the binding to all of the other non-cognate DNAs. For the three non-cognate DNAs, the increases of ΔH were sufficiently large, so that less than half of the individual ΔH values were compensated by a gain in the corresponding $\Delta S°$.

It should be noted that significantly negative ΔC_p values were always observed (Figure 3-4 and Table 3-1) upon R2R3* binding to the non-cognate DNAs, and even to [NC-b]MBS-I, with no trace of the consensus sequence, although most of the ΔC_p values were smaller than that obtained in the binding to the cognate DNA.

Salt dependence of the interactions between the c-Myb R2R3 and the cognate and the non-cognate DNAs*

In general, the electrostatic force should strongly govern the intermolecular interaction between DNA, with the negatively charged phosphate backbone, and the basic residues in a protein. In order to reveal the electrostatic contributions to DNA-binding, we changed the concentration of the cationic counterions in the solvent and made ITC measurements.

Figures 3-5a and 5b show the salt dependencies of the log of the binding affinity, log K_a, and ΔH of R2R3*, respectively, in binding to the cognate MBS-I and the non-cognate [NC-b]MBS-I at 20 °C. The log of the binding affinity changed linearly with the log of the K^+ concentration, and the slopes of the linear least-squares fits to the data are –5.2 and –5.5 for the cognate and the non-cognate DNAs, respectively. The salt dependency of ΔH in the binding of R2R3* to the cognate DNA is smaller than that to the non-cognate DNA.

The ΔC_p values were also measured in the binding of R2R3* to the cognate and the non-cognate MBS-I at different salt concentrations, and they are shown in Figure 3-5c. It clearly indicates that the ΔC_p associated with both specific and non-specific DNA binding significantly depends upon the salt conditions.

*Effects of mutations in the recognition sites and the hydrophobic core of R2R3**

In order to reveal the thermodynamic roles of the individual factors in specific DNA recognition by the protein, the thermodynamic parameters for the binding of several mutated c-Myb DBD to the cognate and the non-cognate DNAs were measured.

First, the K128M R2R3* mutant was prepared, in which Lys128 was substituted with Met, to eliminate the direct read-out of the G8 base. Second, for the analysis of the recognition of Py3 as an indirect read-out mechanism, the Ser187 substituted R2R3* mutants with Gly and Ala, S187G and S187A, were prepared, as indicated by Oda *et al.*

(1997). Third, Val103 and Val107 in R2 were substituted with the corresponding residues in R3, Ile and His, respectively, and the mutant is designated as V103I/V107H. These two valines in the first helix of R2 were found to be essential for the tight binding of MBS-I besides the third helix of R2, from the systematic mutational study of R2 (Oda *et al.*, 1998). The binding affinities of these mutant proteins were observed by the filter binding assay, and are listed in Table 3-2. The K_a values of the interactions between the mutant proteins and the non-cognate DNAs were also measured, but they were all less than 10^6 M^{-1}.

Figure 3-6a shows the far-UV CD spectra of R2R3* and the R2R3* mutants, K128M, S187G, S187A, and V103I/V107H, at 20 °C. All of the spectra indicate that these mutants have typical α-helical conformations similar to that of R2R3*. Figure 3-6b shows their near-UV CD spectra at 20 °C. Only the spectrum of the V103I/V107H mutant deviated from that of R2R3*, suggesting that the simultaneous substitutions of Val103 and Val107 slightly altered the local conformations around the Trp residues, all of which are in the hydrophobic core of R2, although the overall conformation is similar to R2R3* (Oda *et al.*, 1998). In addition, the reversible thermal denaturations of the mutant proteins were observed by monitoring the CD value at 222 nm, and the unfolded fractions are plotted as a function of the temperature in Figure 3-6c. The melting temperatures at which 50% of each protein was unfolded ranged from 48 °C to 52 °C, and the structures were stable at least below 30 °C. The thermal stability of the R2R3* was also examined using differential scanning calorimetry (DSC). Analysis of the melting curve in Figure 3-7 shows that it can be fitted to a three-state model with two components, consistent with the previous DSC analysis of the wild-type R2R3 where the two repeats behave independently in their thermal denaturation (Ogata *et al.*, 1996). As indicated in the previous analysis, the more stable component of the R2R3* curve (T_m = 55 °C) corresponds to the melting of R3. The less stable component (T_m = 50 °C) should correspond to the melting of R2 in R2R3*, which is slightly more stable than R2 in the

wild-type R2R3 with T_m = 47 °C (Ogata *et al.*, 1996). The current DSC experiment also indicated clearly that the change of the heat capacity occurs above 30 °C.

The binding of the R2R3* mutants to the cognate and the non-cognate DNAs was thermodynamically analyzed using ITC, and the results are indicated in Tables 3-3 and 3-4, respectively. The stoichiometry of binding, n, of every interaction was approximately 1.0, except for the binding to [NC-b]MBS-I, where no significant ΔH was observed.

The binding ΔG° of the K128M mutant was highly increased with both the cognate and the non-cognate DNAs. Apparently, the ΔG° was mainly derived from the ΔH in the binding of K128M to MBS-I, as well as in the binding of R2R3* to MBS-I. Although the low binding affinity of the K128M mutant to MBS-I was almost the same as that to [G8T]MBS-I, the entropy contribution increased significantly in the latter binding. In addition, the binding of K128M to [NC-b]MBS-I was accompanied by only a small $|\Delta H|$ (< 1 kcal mol^{-1}). In this interaction, ΔC_p could not be observed, as shown in Figure 3-8.

The binding affinities of both the S187G and S187A mutants were slightly reduced, but they retained the Py3 preference, as indicated in Tables 3-3 and 3-4, as well as in our previous investigation using the filter binding assay (Oda *et al*, 1997). However, the enthalpy and entropy contributions to the Py3 preference were not the same among R2R3* and the Ser187 substituted mutants. In R2R3*, the enthalpy changes were the same for the binding to MBS-I and [T3G]MBS-I, but more negative entropy changes were observed in the binding to the latter (Table 3-1). On the contrary, in the S187G and S187A mutants, the enthalpy and the entropy changes were both larger in the binding to [T3G]MBS-I than in the binding to MBS-I (Tables 3-3 and 3-4).

The reduction of the binding of the V103I/V107H mutant to MBS-I was attributed to both an increased ΔH and a decreased ΔS°. Thus, in this case, the enthalpy-entropy compensation mechanism did not work. The affinity of V103I/V107H to the non-cognate [G8A]MBS-I was much weaker than that to the cognate MBS-I. This means that the

V103I/V107H mutant retains the ability to discriminate between the cognate and non-cognate sequences. It is also interesting that the ΔC_p of V103I/V107H was more negative than that of R2R3* in the binding to MBS-I, while that of K128M was approximately only half of the R2R3* ΔC_p.

Effects of mutations in the linker region connecting R2 and R3

When three amino acid residues, Asn139, Pro140, and Glu141, in the first half of the linker region were simultaneously substituted with Gly residues, the binding affinity to the cognate DNA was completely lost (Table 3-2). To identify the residues in the linker region connecting R2 and R3 that are required for specific DNA-binding, glycine and alanine scanning mutants were made for the three residues to generate six R2R3* mutant proteins, N139G, P140G, E141G, N139A, P140A, and E141A. Analysis using the filter binding assay showed that the substitution of Asn139 with Gly and that of Pro140 with Gly or Ala resulted in a significant loss of the affinity to MBS-I (Table 3-2). In contrast, the affinities of all of these mutant proteins to the non-cognate [G8A]MBS-I were less than 10^6 M^{-1} (data not shown). Therefore, the affinities of these mutant proteins to the cognate MBS-I were still higher than that to [G8A]MBS-I.

Figures 3-9a and 9b indicate the CD spectra of the six mutant proteins and R2R3*, at the far- and the near-UV regions, respectively. The far-UV CD spectra were almost identical, indicating that the secondary structures of the mutants are the same as that of R2R3*. Although the near-UV CD spectra of the mutants, E141G, N139A, and E141A, were similar, those of the other mutants, N139G, P140G, and P140A, were slightly different from that of R2R3*. Thermal denaturation experiments indicated that the stabilities of all of the mutants were similar to that of R2R3* (Figure 3-9c).

The binding of the mutant proteins substituted with Gly or Ala in the linker region to MBS-I was thermodynamically analyzed using ITC, and was compared with the

R2R3* interaction (Table 3-3). The differences in the binding ΔG° obtained by the ITC measurements were similar to those revealed by the filter binding assay (Table 3-2).

The filter binding assay showed that three mutants, N139A, E141G, and E141A, had almost the same affinities as that of R2R3*, while those of the other three mutants, N139G, P140G, and P140A, were lower by approximately one order of magnitude. In the Glu141 substituted mutants, the slightly unfavorable ΔH values were compensated by the favorable ΔS° values, and as a result, their binding ΔG° values were similar to that of R2R3*. In the three mutants with lower affinities, the thermodynamic origins for the loss of the DNA-binding are apparently different, depending on the substituted site. Namely, the increase of ΔG° for the N139G mutant is attributed to an increase of ΔH, while those of the Pro140 substituted mutants, P140G and P140A, are attributed to a decrease of ΔS°.

The thermodynamic parameters of the interactions of the mutant proteins substituted in the linker region to the non-cognate DNA, [G8A]MBS-I, are shown in Table 3-4. Interestingly, their binding affinities and their ΔH values were not identical, although these differences could not be discriminated in the filter binding assay, as described above. The results of the ITC measurements also indicated that the stronger the mutant binds to the cognate MBS-I, the stronger it also binds to the non-cognate [G8A]MBS-I. In addition, the binding of the mutant proteins, N139G, P140G, and P140A, to [NC-b]MBS-I was accompanied by only small $|\Delta H|$ values (< 1 kcal mol^{-1}), and the ΔC_p values of these interactions were nearly zero, as seen in the binding of K128M to [NC-b]MBS-I.

3.4 Discussion

Recent studies have suggested that sequence specific protein-DNA association is not accomplished solely by a direct read-out mechanism of the bases at the rigid molecular interface. Rather, the specific complex is stabilized more than the non-specific molecular association by combinations of different factors: (i) the direct read-out mechanism, including the intermolecular hydrogen-bonds and the van der Waals interactions (Seeman *et al.*, 1976; Pabo and Sauer, 1992), (ii) the conformational changes of proteins upon DNA binding, known as the local folding (Spolar and Record, 1994), (iii) the conformational changes of DNA, such as duplex bending at a specific base sequence (Travers, 1992; Oda *et al.*, 1997), (iv) the change in the hydration, yielding the hydrophobic attraction (Ha *et al.*, 1989), (v) the electrostatic interaction between the protein and the counterion condensed DNA in an environment with many free counterions (Record *et al.*, 1991), and (vi) the entropic loss associated with fixing the translational and rotational movements of the mutual molecular disposition (Finkelstein and Janin, 1989). The current thermodynamic results dissect the individual roles in the formation of the specific complex of the c-Myb R2R3* and DNA, based upon the free and the DNA-complexed tertiary structures determined by the NMR-distance geometry method (Ogata *et al.*, 1994, 1995).

Electrostatic interaction as the thermodynamic driving force

The ionic strength dependence of the thermodynamic parameters reveals the role of the electrostatic terms upon the association of the c-Myb DBD and the DNA. From the polyelectrolyte theory (Manning, 1969, 1972), the sum of the direct condensation of the counterions and the associating Debye-Hückel type screening process, ψ, neutralizes the high charge density in the DNA duplex due to the phosphate groups, and stabilizes the duplex conformation. In typical B-form DNA, ψ equals 0.88. For protein-DNA binding, Record *et al.* (1976, 1991) derived the next equation,

$$\frac{\partial (\log K_a)}{\partial (\log [M^+])} = - m' \psi \qquad [1]$$

where the protein makes m' ion pairs with the DNA and $[M^+]$ is the molar concentration of the counterion.

The linear slopes in Figure 3-5a agree well with this theory, and m' is calculated as about 6 in the cognate MBS-I. In fact, in the complex structure, five and two phosphate groups evidently interact with R2 and R3, respectively (Ogata *et al.*, 1994). Although the slope for the binding to the non-cognate [NC-b]MBS-I is nearly the same as that to the cognate MBS-I, the binding modes should be very different, because of the small n values indicated in Table 3-1. Similar slopes in binding to both the cognate and the non-cognate DNA were also observed by Mossing and Record (1985) and Frank *et al.* (1997).

As shown in Figure 3-5b, the salt dependency of ΔH in the cognate MBS-I binding is not as sensitive as that in the non-cognate MBS-I binding. This is reasonable because the specific interaction is mediated by the hydrogen bonding and the van der Waals interactions, in addition to the electrostatic attraction. In contrast, only the electrostatic term should be dominant in the non-specific interaction, as discussed below.

The origin of the heat capacity change

The negative ΔC_p values associated with complex formation with the cognate and the non-cognate DNAs characterize the molecular association phenomenon. As described by Sturtevant (1977), among the possible sources of the ΔC_p, the hydration contribution or the hydrophobic effect by Sturtevant, ΔC_p^{hyd}, and the internal vibration effect, ΔC_p^{vib}, could be the major factors in the observed negative ΔC_p. Positive contributions might be expected from the formation of the contact ion pairs between the phosphate groups of the DNA backbone and the basic side-chains (Edsall and McKenzie, 1978; Ha *et al.*, 1989). The hydrogen bond contribution is also considered to be small (Sturtevant, 1977). In

fact, when the intermolecular hydrogen bonds between the protein and the DNA are deformed, alternative hydrogen bonds should be formed between those polar groups and the solvent water molecules.

The hydration changes of the protein and the DNA upon specific association, ΔC_p^{hyd}, were usually estimated from the corresponding change of the solvent accessible polar (ΔA^p) and non-polar (ΔA^{np}) surface areas of both molecules, based upon the empirical protein folding thermodynamics (Spolar *et al.*, 1992):

$$\Delta C_p^{hyd} \approx (0.32\pm0.04)\ \Delta A^{np} + (-0.14\pm0.04)\ \Delta A^p. \qquad [2]$$

Murphy and Freire (1992) derived the similar equation:

$$\Delta C_p^{hyd} \approx (0.45\pm0.02)\ \Delta A^{np} + (-0.26\pm0.03)\ \Delta A^p. \qquad [3]$$

Several protein-DNA associations correlate well with reductions of the solvent accessible non-polar surfaces, where induced conformational changes were also noted (Ha *et al.*, 1989; Spolar *et al.*, 1994; Lundbäck and Härd, 1996; Frank *et al.*, 1997).

Using the atomic coordinates of the c-Myb R2R3 complexed with the MBS-I (Ogata *et al.*, 1994; PDB code, 1MSE), the accessible surface area was calculated by the method of Shrake and Ruply (1973), using the van der Waals radii proposed by Richards (1977) for the proteins and those proposed by Alden and Kim (1979) for the DNA. For the free protein, the atomic coordinates of the free R2 and R3 structures (Ogata *et al.*, 1995; PDB codes, 1MBG and 1MBJ, respectively) were used. The extended conformations (Oobatake and Ooi, 1993) for the regions from Leu90 to Asp94, from Asn139 to Ser146, and from Arg190 to Val193 were used, because the root-mean-square-deviations (r.m.s.d.) of the backbone heavy atoms in these regions are over 2 Å in the 50 converged NMR structures of the free R2 (PDB code, 1MBH) and the free R3 (PDB code, 1MBK). For the free DNA, the atomic coordinates of typical B-form DNA (Arnott and Hukins, 1972) were used. Thus, we calculated $\Delta A^p = -2477$ Å^2 and $\Delta A^{np} = -1842$ Å^2, corresponding to $\Delta C_p^{hyd} \approx -0.24\ (\pm 0.17)$ kcal mol^{-1} K^{-1} from eq.[2] and to $\Delta C_p^{hyd} \approx -0.18\ (\pm 0.11)$ kcal mol^{-1} K^{-1} from eq.[3]. These values essentially did not

change, when the accessible surface area of each amino acid residue X in the flexible regions was calculated using the typical conformation of Gly-X-Gly, following the procedure of Livingstone *et al.* (1991). Obviously, these values do not coincide with the observed ΔC_p values, and it is impossible to interpret ΔC_p only by the hydration effect of the protein and the DNA even when their conformational changes upon the specific association were considered. This discrepancy has also been observed in several recent studies, suggesting that it is necessary to consider other contributions (Jin *et al.*, 1993; Lundbäck *et al.*, 1993; Ladbury *et al*, 1994; Merabet *et al.*, 1995; Frisch *et al.*, 1997).

The globular domains of R2 and R3, which are connected by a flexible linker as shown in Figure 3-1, can move almost independently in the free form, as revealed by NMR studies (Ogata *et al.*, 1995, 1996). The corresponding soft internal vibration modes with low frequencies should disappear upon DNA binding, even to the non-cognate MBS-I. Empirically, the effect of the structural tightening, ΔC_p^{vib}, contributes about 20% to the total ΔC_p (Sturtevant, 1977). However, in the association of the c-Myb DBD and the DNA, the contribution of the vibrational motion may be much larger. This may be one of the reasons why $|\Delta C_p|$ is large enough to be observed in almost all of the ITC measurements, for the binding of R2R3* and the mutants to the cognate and even the non-cognate DNAs.

However, the above two factors cannot satisfactorily explain the significant dependency of ΔC_p on the excess counterion concentration, as shown in Figure 3-5c. The linear dependencies of ΔG° and ΔH suggest that the complexed structure of the c-Myb DBD with DNA is not greatly deformed in this salt concentration range. Thus, we propose another source, which has never been clearly identified as a possible factor contributing to the negative ΔC_p. It is the hydration effect of the released cations from the DNA. As described above, the negative charges of the phosphates in the free DNA duplex are largely screened by the condensed counterions, which are then released into the solvent when the protein binds to the DNA, depending upon the excess counterion

concentration (Record *et al.*, 1976, 1991). The heat capacity change accompanied by the hydration of the ions, $\Delta C_{p,\ ion}^{\ hyd}$, is generally a large, negative value (Edsall and McKenzie, 1978; Makhatadze and Privalov, 1990), because the ionic interaction with the water promotes increased ordering of the water molecules. When the counterions are condensed on the free DNA, the positive ionic charges are almost neutralized by the negative phosphate groups of the DNA. Therefore, $|\Delta C_{p,\ ion}^{\ hyd}|$ due to the hydration of the condensed counterions should be much smaller than that of the free cations in the solvent. Although there is no reliable $\Delta C_{p,\ ion}^{\ hyd}$ value for this process, if we assume that it is about –0.01 ~ –0.03 kcal mol^{-1} K^{-1}, about half of the $\Delta C_{p,\ ion}^{\ hyd}$ in the ionization process (Edsall and McKenzie, 1978), then the effect could contribute to $-0.01\ m'\psi \sim -0.03 m'\psi \approx -0.1$ ~ –0.2 kcal mol^{-1} K^{-1}, at the low ionic strength limit. Makhatadze and Privalov (1990) used –0.01 kcal mol^{-1} K^{-1} as the heat capacity difference between a cationic and a neutral amino group. The sign, the amplitude, and the ionic strength dependency of this contribution to ΔC_p seem to agree with our observations in Figure 3-5c. If this proposed mechanism is correct, it can be another reason for the observed ΔC_p in the binding to the cognate and the non-cognate DNAs. In any case, without considering the salt effects, it would be meaningless to argue the coincidence between the experimental ΔC_p value and the calculated $\Delta C_p^{\ hyd}$ value from ΔA^p and ΔA^{np}.

Thermodynamic driving forces in sequence specific base recognition

The remarkable reduction of the binding affinities in the A4 and C6 substituted MBS-I is predominantly due to an increase of ΔH, together with a small amount of entropy compensation. This corresponds well to the observation in the complex structure that A4 and the counterpart guanine of C6 are well defined and directly interact with Asn183 and Lys182, respectively. When definite hydrogen bonds are broken, an increase of ΔH should be observed (Makhatadze and Privalov, 1993), and ΔS° should also be

increased, as a general characteristic of the enthalpy-entropy compensation in weak interactions (Dunitz, 1995).

In contrast, the base substitutions at A5 resulted in an unfavorable ΔS°, but the ΔH remained similar to that in the binding to MBS-I. In the complex structure, although A5 and its counterpart thymine seem to interact with Asn136, Asn179, and Asn186, the distances between them are relatively longer than those of the directly interacting pairs. One possible explanation for this significant decrease of ΔS° is the change of the hydration, because there is enough space for an immobilized water around the binding surface. The entropic penalty for the immobilization of a water molecule in ice or a crystalline salt has been estimated to be a maximum of 2 kcal mol^{-1} at 300 K (Dunitz, 1994). It corresponds well to the experimental values for the base substitution entropic penalties, which ranged from 0.6 to 3.3 kcal mol^{-1}.

The binding reductions of the G8 substitutions were not as large as those of the A4 or C6 substitutions. The reductions were mainly attributed to the increase of ΔH, with a small amount of ΔS° compensation. Since the G8 base at the edge of the consensus sequence is recognized only by Lys128 in R2, other DNA recognition modes mainly responsible in R3 may still be maintained, although the direct interaction between the 8'th base and Lys128 disappears in the G8 substituted MBS-Is. This assumption is further supported by the results that the ΔC_p values of the G8 substituted MBS-Is were almost identical to that for the cognate MBS-I, while those of the A4, A5, and C6 substituted MBS-Is were about half of that for the cognate MBS-I.

The binding reductions with the C10 substitutions, especially the replacement of the original C-G base-pair with the G-C pair, cannot be explained by the complex structure, because there do not seem to be any possible contacts between the c-Myb DBD and this base-pair. However, significant increases in ΔH and ΔS° were observed, as shown in Table 3-1, which are consistent with the previous report by Tanikawa *et al.* (1993). This suggests another DNA-bending at the C10-A11 pyrimidine-purine step or

other dynamic interactions between the C-G base-pair and the protein, which were not observed in the static NMR structure. In fact, a modeling study on computer graphics suggests that Lys128 can reach the counterpart guanine of C10, assuming an alternative rotamer structure of the side chain. This hypothesis that Lys128 can dynamically interact with both G8 and the counterpart guanine of C10 may be supported by the fact that the affinity was significantly reduced only to the DNA having G10 with its counterpart cytosine (Tanikawa *et al.*, 1993). Only cytosine among the four bases does not work as a hydrogen bond acceptor on the surface of the major groove. These interactions can also explain why the binding reductions of the G8 substitutions were not large as discussed above, because the interaction between Lys128 and the counterpart guanine of C10 could be mainteined.

Thermodynamic cycles by amino acid mutations and base substitutions

Combining the thermodynamic parameters in Tables 3-1, 3-3, and 3-4 of the binding of R2R3* and the other mutant proteins to the cognate and the non-cognate DNAs, one can draw a thermodynamic cycle, analogous to the thermodynamic cycles from double mutations in protein stability studies (Horovitz *et al.*, 1990). By subtracting the common thermodynamic quantities between any two states, it is possible to dissect the DNA recognition mechanism of the protein more precisely, while referring to the NMR tertiary structure to visualize the set of specific contacts between them (Ogata *et al.*, 1994).

The thermodynamic cycle composed of R2R3*•MBS-I, R2R3*•[G8T]MBS-I, K128M•MBS-I, and K128M•[G8T]MBS-I is shown in Figure 3-10a. Since the G8 base is directly recognized by Lys128, this thermodynamic cycle would be specified for this interaction, as long as the Lys128 substitution does not disrupt the protein structure and the other binding modes remain. Both the enthalpy and the entropy increases from the cognate to the non-cognate binding of K128M are much larger than those of R2R3*,

respectively. Similarly, the enthalpy and entropy increases from R2R3* to K128M accompanied by [G8T]MBS-I binding are much larger than those by MBS-I. They suggest that the direct interaction between the protein and the DNA essentially disappeared between K128M and the DNA, and so the effect of the loss of one positive charge is more significant in the non-cognate DNA binding than in the cognate DNA binding. Therefore, the thermodynamic cycle in Figure 3-10a should not be specified for a single interaction between residue 128 in the protein and the 8'th base in the DNA, although the overall structure of K128M is not deformed, as confirmed by CD measurements of the DNA-free form (Figure 3-6).

The thermodynamic cycle composed of R2R3*•MBS-I, R2R3*•[T3G]MBS-I, S187G•MBS-I, and S187G•[T3G]MBS-I is shown in Figure 3-10b. The thermodynamic cycle of S187A is essentially the same as that of S187G, and it is not shown here. In the NMR structure, Ser187 is the only candidate that interacts with Py3 at the initial base of the consensus sequence (Ogata *et al.*, 1994). Our previous study showed that the intrinsic bendability of this pyrimidine-purine step, Py3-A4, is important for the indirect-readout mechanism (Oda *et al.*, 1997). In the current thermodynamic studies, both $\Delta\Delta H$ and $\Delta\Delta S^\circ$ from S187G•MBS-I to S187G•[T3G]MBS-I were increased. One possible reason for the $\Delta\Delta S^\circ$ increase is that the maintenance of the bent DNA conformation in S187G•MBS-I is relaxed in S187G•[T3G]MBS-I, and is accompanied by a lack of preferable molecular contacts, which may result in an increase in $\Delta\Delta H$. Given the geometrical space volume around Ser187 and the T3 base, water molecules may mediate the interaction between the side-chain of Ser187 and the T3 base (Oda *et al.*, 1997). In fact, increases in both $\Delta\Delta H$ and $\Delta\Delta S^\circ$ from R2R3*•MBS-I to S187G•MBS-I were observed, and they may be attributed to the liberation of an immobilized water molecule, as described for the interaction between R2R3* and the A5 base. From R2R3*•MBS-I to R2R3*•[T3G]MBS-I, $\Delta\Delta S^\circ$ is negative, but $\Delta\Delta H$ is zero. This $\Delta\Delta S^\circ$ is in contrast to the cases of S187G and S187A, and is not readily understood. When both MBS-I and

[T3G]MBS-I bind to R2R3*, the DNAs may bend at the step between the third and the fourth base to optimize the interactions with the side-chain of Ser187. This may not impose a large difference in ΔH, but the ΔS° cost may appear in the bending of [T3G]MBS-I.

Our previous studies (Ogata *et al.*, 1996; Oda *et al.*, 1998) clearly indicated that not only the amino acids in the DNA contact sites, but also those in the hydrophobic core are critical to recognize the target DNA, although the overall protein structure does not change greatly. From the systematic mutations of many different amino acids in R2, we have ascertained that Val103 and Val107, located in the R2 core, are characteristic of the R2 fold, and that they are required for tight DNA-binding. When they are replaced with the corresponding residues in R3, Ile and His, respectively, the binding affinity is reduced by approximately one-order of magnitude. Since the overall conformation of V103I/V107H, monitored in the far-UV CD spectrum, remains the same as that of R2R3*, but the near-UV spectrum changed, as indicated in Figures 3-6a and 6b, the helix packing between the first and the third helices in R2 is considered to be deformed slightly, resulting in a significant decrease in the DNA affinity (Oda *et al.*, 1998).

Figure 3-10c shows the thermodynamic cycle composed of R2R3*•MBS-I, R2R3*•[G8A]MBS-I, V103I/V107H•MBS-I, and V103I/V107H•[G8A]MBS-I. It is obvious that the $\Delta\Delta G^\circ$ increase from R2R3*•MBS-I to V103I/V107H•MBS-I originates from both the unfavorable increase of $\Delta\Delta H$ and the decrease of $\Delta\Delta S^\circ$. The rearrangement of the helix packing probably requires more entropy loss, and some specific atom contacts may be disrupted, which would be reflected in an increase of ΔH. The large ΔC_p value of the V103I/V107H binding to MBS-I also suggests a structural rearrangement. The V103I/V107H mutant binds to the cognate MBS-I more than 10 times more tightly than to the non-cognate [G8A]MBS-I. This is due to the significant increase of $\Delta\Delta H$ and the compensating $\Delta\Delta S^\circ$ from V103I/V107H•MBS-I to V103I/V107H•[G8A]MBS-I, and

suggests the lack of some direct interactions between the mutant protein and the non-cognate DNA.

Local folding of the linker region between R2 and R3

It has been indicated that the linker connecting the two imperfect tandem repeats is important for the cooperativity in protein-DNA recognition (Choo and Klug, 1993; Klemm and Pabo, 1996). In the c-Myb R2R3, both Asn139 and Pro140 are involved in the high affinity of the MBS-I. In Figure 3-10d, the thermodynamic cycles composed of R2R3*•MBS-I, R2R3*•[G8A]MBS-I, N139G•MBS-I, N139G•[G8A]MBS-I, N139A•MBS-I, and N139A•[G8A]MBS-I are shown. In Figure 3-10e, those of R2R3*•MBS-I, R2R3*•[G8A]MBS-I, P140G•MBS-I, P140G•[G8A]MBS-I, P140A•MBS-I, and P140A•[G8A]MBS-I are also shown. Since the effects of the amino acid substitutions of Glu141 were not significant, the corresponding thermodynamic cycle is not shown. The role of Glu141 seems to be negligible in the cognate and the non-cognate DNA binding.

Since $\Delta\Delta H$ from R2R3*•MBS-I to N139G•MBS-I was increased, with only a small $\Delta\Delta S^\circ$ remaining, and both $\Delta\Delta H$ and $\Delta\Delta S^\circ$ were almost unchanged from R2R3*•MBS-I to N139A•MBS-I, the specific contact of the Cβ atom of residue 139 with the R2 core should be required for the tight binding of MBS-I. In fact, several NOE signals were observed between the Hβ atoms of Asn139 and the protons of other R2 residues in the free and the DNA-complexed structures (Ogata *et al.*, 1994, 1995). This is further supported by the near-UV CD spectra indicated in Figure 3-9b, in which the N139A spectrum is similar to that of R2R3*, but the N139G spectrum slightly deviates from that of R2R3*. The large increases of $\Delta\Delta H$ and $\Delta\Delta S^\circ$, both from R2R3*•[G8A]MBS-I to N139G•[G8A]MBS-I and from N139G•MBS-I to N139G•[G8A]MBS-I, suggest that the two changes, one in the linker of the protein and

one of a consensus base in the DNA, disrupt the cooperative binding that is characteristic of the c-Myb DBD and MBS-I (Ogata *et al.*, 1994).

On the contrary, the $\Delta\Delta S^\circ$ from R2R3*•MBS-I to P140G•MBS-I was significantly decreased, while the $\Delta\Delta H$ was unchanged. The $\Delta\Delta S^\circ$ from R2R3*•MBS-I to P140A•MBS-I was also decreased, with the $\Delta\Delta H$ unchanged, but the amplitude of $\Delta\Delta S^\circ$ is less than that in the case of P140G. These results are a clear evidence of the local folding of the linker, which accompanies specific DNA binding. In fact, in the protein folding problem, Matthews *et al.* (1987) pointed out that the stability of a protein can be increased by the introduction of a Pro residue, due to the decrease in the configurational entropy of the unfolded state. They enhanced of the thermal stability of T4 lysozyme with that substitution, which is expected to contribute about 1 kcal mol^{-1} to the folding free energy change. The role of Pro in the conformation and the stability was also studied in other proteins (Yutani *et al.*, 1991; Kimura *et al.*, 1992). A Pro substitution in this DNA-binding protein should follow the same mechanism. Namely, the DNA-free form of the protein corresponds to the unfolded state, and the DNA-bound form corresponds to the folded state in the protein folding problem. The increase in the entropy of the locally "unfolded" state of P140G and P140A as compared with the conformational entropy of R2R3*, with Pro140, should contribute to the stabilization of the R2R3*-DNA complex by increasing ΔG°. The role of Pro140 in the linker region is to restrict the conformational flexibility between R2 and R3 in the DNA-free state. The effect should be larger in P140G than in P140A, because of the larger freedom of the glycine backbone than that of alanine in the DNA-free form. In the cycles in Figure 3-10e, large increases of $\Delta\Delta H$ and $\Delta\Delta S^\circ$ were observed from R2R3*•[G8A]MBS-I to P140G•[G8A]MBS-I, from P140G•MBS-I to P140G•[G8A]MBS-I, from R2R3*•[G8A]MBS-I to P140A•[G8A]MBS-I, and from P140A•MBS-I to P140A•[G8A]MBS-I. Again, the cooperative binding may be disrupted by these changes of both an amino acid and a DNA base.

Recently, Hegvold and Gabrielsen (1996) also identified the importance of the linker connecting R2 and R3 in the c-Myb DBD for maintaining specific DNA-binding. However, their results are somewhat different from our results. Their gel-shift assay experiment indicated that the DNA-binding affinity of the P140A mutant was slightly lower than the wild-type R2R3, and that it was much higher than that of the P140G mutant. Our precise thermodynamic analysis revealed that the DNA-binding affinities of both the P140A and P140G mutants were similar, but they have the different thermodynamic roles in the local folding.

Thermodynamic parameters accompanying specific and non-specific protein-DNA interactions

In general, specific binding is defined as the molecular association, in which a particular molecule is tightly and exclusively bound, by forming a specific complex structure. Therefore, from structural and thermodynamic studies, specific molecular recognition could be characterized by several definite intermolecular interactions in an energetically stabilized complex structure, as a result of either rigid-body association or a local folding procedure. In contrast, non-specific binding has the broad meaning of any phenomena opposite to specific binding.

In the case of an interaction between a protein and DNA, their random association is always due to the electrostatic attractive force between them. Therefore, in living cells, any biologically meaningful, specific interaction should have a higher affinity than the average of such a random association, with $K_a \approx 10^6$ M^{-1}. Otherwise, particular DNA sequences might be covered with other DNA binding proteins due to random association, and biological regulation would cease. Therefore, in most conventional molecular biology experiments, such as gel-shift or filter binding assays, any protein-DNA interaction with much weaker affinity than that in the specific binding is usually considered as "non-specific" binding.

From *in vitro* thermodynamic studies, it has been found that those "non-specific" binding is accompanied by small ΔH and negligible ΔC_p values (Takeda *et al.*, 1992; Ladbury *et al.*, 1994; Merabet *et al.*, 1995; Frank *et al.*, 1997). They concluded that "non-specific" weak binding in protein-DNA associations should be dominated by entropically driven electrostatic interactions.

However, in the current studies using several mutant proteins, significant ΔH and ΔC_p values were observed, even when the affinity was greatly reduced. In addition, the contributions of ΔH and ΔS° to the affinity were not always the same. In fact, the binding of R2R3* to all of the DNAs used in this study, even to [NC-b]MBS-I, was accompanied by both detectable ΔH and ΔC_p. The binding affinities of the K128M mutant to all of the DNAs were very weak, with K_a values less than 10^6 M^{-1}, as determined by the filter binding assay. However, the current calorimetric analysis indicated that the ΔH values of the binding of the K128M mutant to three different DNAs, cognate MBS-I, [G8T]MBS-I, and [NC-b]MBS-I, were different. Moreover, the V103I/V107H mutant and the other mutant proteins with substitutions of Asn139 and Pro140 at the linker region bind to the cognate MBS-I exclusively, in spite of the reduced affinity, and discriminate the non-cognate [G8A]MBS-I by a 10-fold higher affinity.

Now, let us define "non-specific" binding as the procedure of random molecular association, and discriminate it from any other type of weak binding by assuming unique associations. In all of the current experiments, the binding of R2R3* to [C6T]MBS-I, [NC-a]MBS-I, and [NC-b]MBS-I is obviously "non-specific" binding in the limited sense, because their stoichiometries of binding, n, significantly deviated from 1.0. The binding of the other mutant proteins to [NC-b]MBS-I is also considered to be the "non-specific" binding, but no stoichiometry values were available, due to the greatly reduced ΔH values in the experiments. The characteristic features of "non-specific" binding are (i) positive ΔS° and (ii) small negative ΔH. From Figure 3-5b, (iii) the high sensitivity of ΔH against the salt concentration may be another feature. All of these features coincide

with the conventional view that "non-specific" binding in protein-DNA associations is entropically driven by electrostatic interactions, although ΔH was detectable. Theoretically, a smaller ΔC_p is expected in "non-specific" binding than in specific binding, because of the lower contributions of ΔC_p^{hyd} and ΔC_p^{vib} in "non-specific" binding, even if we assume the hydration effect of the released counterions from DNA. In fact, this tendency was found in the current experiments.

In protein-DNA interactions, long-range electrostatic forces bring the protein and the DNA into proximity at the first step, and at the second step, specific hydrogen bonds and van der Waals contacts are formed, along with the accompanying conformational changes (Takeda *et al.*, 1992; Wilson, 1996). The "non-specific" binding ends at the first step, and should be predominantly entropy driven and strongly dependent on the salt concentration. In contrast, in specific binding, the high binding affinity is achieved by the specific contacts between the native protein and the DNA. Several mutant proteins retain the ability of specific binding, although the affinities are reduced as much as those in the "non-specific" binding.

3.5 References

Alden, C. J., and Kim, S. -H. (1979) *J. Mol. Biol.*, **132**, 411-434.

Arnott, S., and Hukins, D. W. L. (1972) *Biochem. Biophys. Res. Commun.*, **47**, 1504-1510.

Calladine, C. R. (1982) *J. Mol. Biol.*, **161**, 343-352.

Choo, Y., and Klug, A. (1993) *Nucleic Acids Res.*, **21**, 3341-3346.

Dunitz, J. D. (1994) *Science*, **264**, 670.

Dunitz, J. D. (1995) *Chemistry and Biol.*, **2**, 709-712.

Edsall, J. T., and McKenzie, H. A. (1978) *Adv. Biophys.*, **10**, 137-207.

Finkelstein, A. V., and Janin, J. (1989) *Protein Eng.*, **3**, 1-3.

Frampton, J., Gibson, T. J., Ness, S. A., Doderlein, G., and Graf, T. (1991) *Protein Eng.*, **4,** 891-901.

Frank, D. E., Saecker, R. M., Bond, J. P., Capp, M. W., Tsodikov, O. V., Melcher, S. E., Levandoski, M. M., and Record, M. T., Jr. (1997) *J. Mol. Biol.*, **267**, 1186-1206.

Frisch, C., Schreiber, G., Johnson, C. M., and Fersht, A. R. (1997) *J. Mol. Biol.*, **267**, 696-706.

Gabrielsen, O. S., Sentenac, A., and Fromageot, P. (1991) *Science,* **253**, 1140-1143.

Guehmann, S., Vorbrueggen, G., Kalkbrenner, F., and Moelling, K. (1992) *Nucleic Acids Res.*, **20**, 2279-2286.

Gurney, R. W. (1953) in *Ionic Processes in Solution*, p. 89, McGraw-Hill, New York.

Ha, J. -H., Spolar, R. S., and Record, M. T., Jr. (1989) *J. Mol. Biol.*, **209**, 801-816.

Hegvold, A. B., and Gabrielsen, O. S. (1996) *Nucleic Acids Res.*, **24**, 3990-3995.

Horovitz, A., Serrano, L., Avron, B., Bycroft, M., and Fersht, A. R. (1990) *J. Mol. Biol.*, **216**, 1031-1044.

Hyre, D. E., and Spicer, L. D. (1995) *Biochemistry*, **34**, 3212-3221.

Jin, L., Yang, J., and Carey, J. (1993) *Biochemistry*, **32**, 7302-7309.

Kanei-Ishii, C., Sarai, A., Sawazaki, T., Nakagoshi, H., He, D. -N., Ogata, K., Nishimura, Y., and Ishii, S. (1990) *J. Biol. Chem.*, **265**, 19990-19995.

Kimura, S., Nakamura, H., Hashimoto, T., Oobatake, M., and Kanaya, S. (1992) *J. Biol. Chem.*, **267**, 21535-21542.

Klemm, J. D., Rould, M. A., Aurora, R., Herr, W., and Pabo, C. O. (1994) *Cell*, **77**, 21-32.

Klemm, J. D., and Pabo, C. O. (1996) *Genes and Dev.*, **10**, 27-36.

König, P., Giraldo, R., Chapman, L., and Rhodes, D. (1996) *Cell*, **85**, 125-136.

Ladbury, J. E., Wright, J. G., Sturtevant, J. M., and Sigler, P. B. (1994) *J. Mol. Biol.*, **238**, 669-681.

Livingstone, J. R. (1996) *Nature*, **384**, 491-492.

Livingstone, J. R., Spolar, R. S., and Record, M. T., Jr. (1991) *Biochemistry*, **30**, 4237-4244.

Lundbäck, T., Cairns, C., Gustafsson, J. -A., Carlstedt-Duke, J., and Härd, T. (1993) *Biochemistry*, **32**, 5074-5082.

Lundbäck, T., and Härd, T. (1996) *Proc. Natl. Acad. Sci., U.S.A*, **93**, 4754-4759.

Makhatadze, G. I., and Privalov, P. L. (1990) *J. Mol. Biol.*, **213**, 375-384.

Makhatadze, G. I., and Privalov, P. L. (1993) *J. Mol. Biol.*, **232**, 639-659.

Mahler, H. R., Kline, B., and Mehrota, B. D. (1964) *J. Mol. Biol.*, **9**, 801-811.

Manning, G. S. (1969) *J. Chem. Phys.*, **51**, 924-933.

Manning, G. S. (1972) *Biopolymers*, **11**, 937-949.

Matthews, B. W., Nicholson, H., and Becktel, W. J. (1987) *Proc. Natl. Acad. Sci., USA*, **84**, 6663-6667.

Merabet, E., and Ackers, G. K. (1995) *Biochemistry*, **34**, 8554-8563.

Mossing, M. C., and Record, M. T., Jr. (1985) *J. Mol. Biol.*, **186**, 295-305.

Murphy, K. P., and Freire, E. (1992) *Advan. Protein Chem.*, **43**, 313-361.

Myrset, A. H., Bostad, A., Jamin, N., Lirsac, P. -N., Toma, F., and Gabrielsen, O. S. (1993) *EMBO J.*, **12**, 4625-4633.

Oda, M., Furukawa, K., Ogata, K., Sarai, A., Ishii, S., Nishimura, Y., and Nakamura, H. (1997) *J. Biol. Chem.*, **272**, 17966-17971.

Oda, M., Furukawa, K., Ogata, K., Sarai, A., Ishii, S., Nishimura, Y., and Nakamura, H. (1998) *Protein Eng.*, **11**, in press.

Ogata, K., Morikawa, S., Nakamura, H., Sekikawa, A., Inoue, T., Kanai, H., Sarai, A., Ishii, S., and Nishimura, Y. (1994) *Cell*, **79**, 639-648.

Ogata, K., Morikawa, S., Nakamura, H., Hojo, H., Yoshimura, S., Zhang, R., Aimoto, S., Ametani, Y., Hirata, Z., Sarai, A., Ishii, S., and Nishimura, Y. (1995) *Nature Struct. Biol.*, **2**, 309-320.

Ogata, K., Kanei-Ishii, C., Sasaki, M., Hatanaka, H., Nagadoi, A., Enari, M., Nakamura, H., Nishimura, Y, Ishii, S., and Sarai, A. (1996) *Nature Struct. Biol.*, **3**, 178-187.

Oobatake, M., and Ooi, T. (1993) *Prog. Biophys. Mol. Biol.*, **59**, 237-284.

Pabo, C. O., and Sauer, R. T. (1992) *Annu. Rev. Biochem.*, **61**, 1053-1095.

Record, M. T., Jr., Lohman, T. M., and de Haseth, P. (1976) *J. Mol. Biol.*, **107**, 145-158.

Record, M. T., Jr., Ha, J. -H., and Fisher, M. A. (1991) *Methods Enzymol.*, **208**, 291-343.

Richards, F. M. (1977) *Annu. Rev. Biophys. Bioeng.*, **6**, 151-176.

Saikumar, P., Murali, R., and Reddy, E. P. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 8452-8456.

Seeman, N. C., Rosenberg, J. M., and Rich, A. (1976) *Proc. Natl. Acad. Sci. USA*, **73**, 804-808.

Shrake, A., and Rupley, J. A. (1973) *J. Mol. Biol.*, **79**, 351-371.

Spolar, R. S., Livingstone, J. R., and Record, M. T., Jr. (1992) *Biochemistry*, **31**, 3947-3955.

Spolar, R. S., and Record, M. T., Jr. (1994) *Science*, **263**, 777-784.

Sturtevant, J. M. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 2236-2240.

Takeda, Y., Ross, P. D., and Mudd, C. P. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 8180-8184.

Tanikawa, J., Yasukawa, T., Enari, M., Ogata, K., Nishimura, Y., Ishii, S., and Sarai, A. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 9320-9324.

Torigoe, H., Nakayama, T., Imazato, M., Shimada, I., Arata, Y., and Sarai, A. (1995) *J. Biol. Chem.*, **270**, 22218-22222.

Travers, A. A. (1992) *Curr. Opin. Struct. Biol.*, **2**, 71-77.

Wilson, W. D. (1996) in *Nucleic acids in chemistry and biology*, Second Ed. (Blackburn, G. M. and Gait, M. J. eds.), chapter 8, pp.331-374, Oxford Univ. press, Oxford.

Wiseman, T., Williston, S., Brandts, J. F., and Lin, L. (1989) *Anal. Biochem.*, **179**, 131-137.

Xu, W., Rould, M. A., Jun, S., Desplan, C., and Pabo, C. O. (1995) *Cell*, **80**, 639-650.

Yutani, K., Hayashi, S., Sugisaki, Y., and Ogasahara, K. (1991) *Proteins*, **9**, 90-98.

Table 3-1. Thermodynamic parameters of R2R3* binding to non-cognate DNAs

DNA	n	K_a (M^{-1})	$\Delta G°$ (kcal mol^{-1})	$\Delta\Delta G°$ [a]	ΔH (kcal mol^{-1})	$\Delta\Delta H$ [b]	$\Delta S°$ (cal mol^{-1} K^{-1})	$\Delta\Delta S°$ [c]	ΔC_p [d] (kcal mol^{-1} K^{-1})
MBS-I	1.01	2.0×10^7	−12.1	—	−12.5	—	−1.3	—	−0.62
[T3G]MBS-I	0.99	6.8×10^6	−11.5	0.6	−12.5	0.0	−3.3	−2.0	−0.56
[A4G]MBS-I	0.98	4.9×10^5	−10.0	2.1	−8.7	3.8	4.2	5.5	−0.34
[A5G]MBS-I	1.02	6.7×10^5	−10.2	1.9	−13.9	−1.4	−12.7	−11.4	−0.39
[A5T]MBS-I	1.01	1.0×10^6	−10.4	1.7	−12.2	0.3	−6.2	−4.9	−0.39
[A5C]MBS-I	1.00	4.3×10^6	−11.2	0.9	−12.2	0.3	−3.2	−1.9	−0.46
[C6T]MBS-I	0.87	1.4×10^5	−9.3	2.8	−8.0	4.5	4.4	5.7	−0.38
[G8A]MBS-I	0.96	1.3×10^6	−10.5	1.6	−10.6	1.9	−0.2	1.1	−0.65
[G8T]MBS-I	1.00	2.7×10^6	−11.0	1.1	−11.1	1.4	−0.5	0.8	−0.63
[C10T]MBS-I	1.02	8.4×10^6	−11.6	0.5	−10.2	2.3	4.8	6.1	−0.57
[C10G]MBS-I	0.99	1.6×10^6	−10.7	1.4	−8.8	3.7	6.5	7.8	−0.64
[NC-a]MBS-I	0.82	3.2×10^5	−9.7	2.4	−8.8	3.7	3.1	4.4	−0.55
[NC-b]MBS-I	0.81	2.0×10^5	−9.5	2.6	−9.2	3.3	1.0	2.3	−0.46

All ITC measurements were in 100 mM potassium phosphate buffer (pH 7.5) containing 20 mM KCl at 20.2 (±0.3)°C.

[a] The difference in unitary free energy change of binding between the interaction with MBS-I and with the non-cognate DNAs indicated.

[b] The difference in binding enthalpy change between the interaction with MBS-I and with the non-cognate DNAs indicated.

[c] The difference in unitary entropy change of binding between the interaction with MBS-I and with the non-cognate DNA indicated.

[d] Calculated from the slope of the regression line of the linear fit of the enthalpies measured at five (for MBS-I & [G8A]MBS-I) or three (for others) different temperatures between 11 and 30°C.

Table 3-2. The interaction of R2R3* mutants with MBS-I analyzed using filter binding assay

protein	K_a [a] (M^{-1})	$\Delta G°$ [b] (kcal mol^{-1})	$\Delta\Delta G°$ [c] (kcal mol^{-1})
R2R3*	1.8×10^8	−12.5	—
K128M	$\leq 10^6$		
S187G	5.5×10^7	−11.8	0.7
S187A	1.1×10^8	−12.2	0.3
V103I/V107H	1.1×10^7	−11.0	1.5
N139G/P140G/E141G	$\leq 10^6$		
N139G/E141G	1.5×10^7	−11.2	1.3
N139G	1.6×10^7	−11.2	1.3
P140G	1.6×10^7	−11.2	1.3
E141G	8.8×10^7	−12.1	0.4
N139A	6.0×10^7	−11.9	0.6
P140A	2.1×10^7	−11.3	1.2
E141A	2.2×10^8	−12.6	−0.1

All measurements were performed in 100 mM potassium phosphate buffer (pH 7.5) containing 20 mM KCl.

[a] Obtained by filter binding assay, in which samples were preincubated on ice.

[b] Calculated from K_a values obtained by filter binding assay.

[c] The difference in free energy change of binding between the interaction with R2R3* and with the R2R3* mutant indicated.

Table 3-3. Thermodynamic parameters of R2R3* mutants binding to cognate DNA

protein	n	K_a (M^{-1})	$\Delta G°$ (kcal mol^{-1})	$\Delta\Delta G°$ [a]	ΔH (kcal mol^{-1})	$\Delta\Delta H$ [b]	$\Delta S°$ (cal mol^{-1} K^{-1})	$\Delta\Delta S°$ [c]	ΔC_p [d] (kcal mol^{-1} K^{-1})
R2R3*	1.01	2.0×10^7	−12.1	—	−12.5	—	−1.3	—	−0.62
K128M	1.04	2.9×10^5	−9.7	2.4	−9.6	2.9	0.1	1.4	−0.28
S187G	0.96	1.3×10^7	−11.9	0.2	−11.5	1.0	1.4	2.7	−0.60
S187A	0.97	1.6×10^7	−12.0	0.1	−12.1	0.4	−0.4	0.9	−0.59
V103I/V107H	1.00	3.6×10^6	−11.2	0.9	−12.0	0.5	−2.8	−1.5	−0.74
N139G	1.00	3.3×10^6	−11.1	1.0	−11.3	1.2	−0.7	0.6	−0.59
P140G	0.99	3.3×10^6	−11.1	1.0	−12.4	0.1	−4.6	−3.3	−0.62
E141G	1.01	1.1×10^7	−11.8	0.3	−11.5	1.0	1.1	2.4	−0.64
N139A	0.99	1.3×10^7	−11.9	0.2	−12.4	0.1	−1.8	−0.5	−0.56
P140A	1.01	3.9×10^6	−11.2	0.9	−12.3	0.2	−3.8	−2.5	−0.58
E141A	1.00	2.1×10^7	−12.2	−0.1	−11.9	0.6	0.9	2.2	−0.62

All ITC measurements were in 100 mM potassium phosphate buffer (pH 7.5) containing 20 mM KCl at 20.2 (±0.3)°C.

[a] The difference in unitary free energy change of binding between the interaction with R2R3* and with the R2R3* mutant indicated.

[b] The difference in binding enthalpy change between the interaction with R2R3* and with the R2R3* mutant indicated.

[c] The difference in unitary entropy change of binding between the interaction with R2R3* and with the R2R3* mutant indicated.

[d] Calculated from the slope of the regression line of the linear fit of the enthalpies measured at three different temperatures between 15 and 25°C.

Table 3-4. Thermodynamic parameters of R2R3* mutants binding to non-cognate DNAs

protein	DNA	n	K_a (M^{-1})	$\Delta G°$ (kcal mol^{-1})	ΔH (kcal mol^{-1})	$\Delta S°$ (cal mol^{-1} K^{-1})	ΔC_p [a] (kcal mol^{-1} K^{-1})
K128M	G8T	1.03	2.6×10^5	−9.6	−6.6	10.2	−0.27
K128M	NC-b	n.d.[b]	n.d.[b]	n.d.[b]	n.d.[b]	n.d.[b]	≈ 0
S187G	T3G	0.99	3.0×10^6	−11.0	−10.1	3.1	−0.59
S187A	T3G	1.01	4.4×10^6	−11.3	−10.1	4.1	−0.52
V103I/V107H	G8A	0.98	2.8×10^5	−9.6	−8.4	4.3	−0.73
N139G	G8A	1.02	3.9×10^5	−9.8	−7.0	9.6	−0.52
N139G	NC-b	n.d.[b]	n.d.[b]	n.d.[b]	n.d.[b]	n.d.[b]	≈ 0
P140G	G8A	0.98	3.0×10^5	−9.7	−8.1	5.4	−0.62
P140G	NC-b	n.d.[b]	n.d.[b]	n.d.[b]	n.d.[b]	n.d.[b]	≈ 0
E141G	G8A	0.97	1.4×10^6	−10.6	−9.1	5.3	−0.59
N139A	G8A	1.03	6.0×10^5	−10.1	−10.6	−1.9	−0.58
P140A	G8A	0.97	4.2×10^5	−9.9	−8.8	3.9	−0.52
P140A	NC-b	n.d.[b]	n.d.[b]	n.d.[b]	n.d.[b]	n.d.[b]	≈ 0
E141A	G8A	1.03	1.3×10^6	−10.5	−9.9	2.0	−0.63

All ITC measurements were in 100 mM potassium phosphate buffer (pH 7.5) containing 20 mM KCl at 20.2 (±0.3)°C.

[a] Calculated from the slope of the regression line of the linear fit of the enthalpies measured at three different temperatures between 15 and 25°C.

[b] Not determined from the fit to the data, because the amplitude of ΔH was too small ($|\Delta H| < 1$ kcal mol^{-1}).

Figure 3-1. The DNA-complexed structure of the c-Myb DBD (Ogata *et al.*, 1994; PDB code, 1MSE). The backbones of R2, R3, and the linker between them are shown by pink, blue, and light yellow pipe models, respectively. The side-chains of Lys128, Lys182, Asn183, and Ser187 are indicated by thick purple lines. The thin yellow and green lines are the DNA double strands, and the consensus bases Py3 (T3), A4, A5, C6, and G8 are indicated by thick red lines. The Cα positions of the amino acid residues that were substituted with other amino acids in the current study are indicated by spheres with the one-letter amino acid codes following the residue numbers. The figure was drawn with the program InsightII (Molecular Simulations Inc., San Diego).

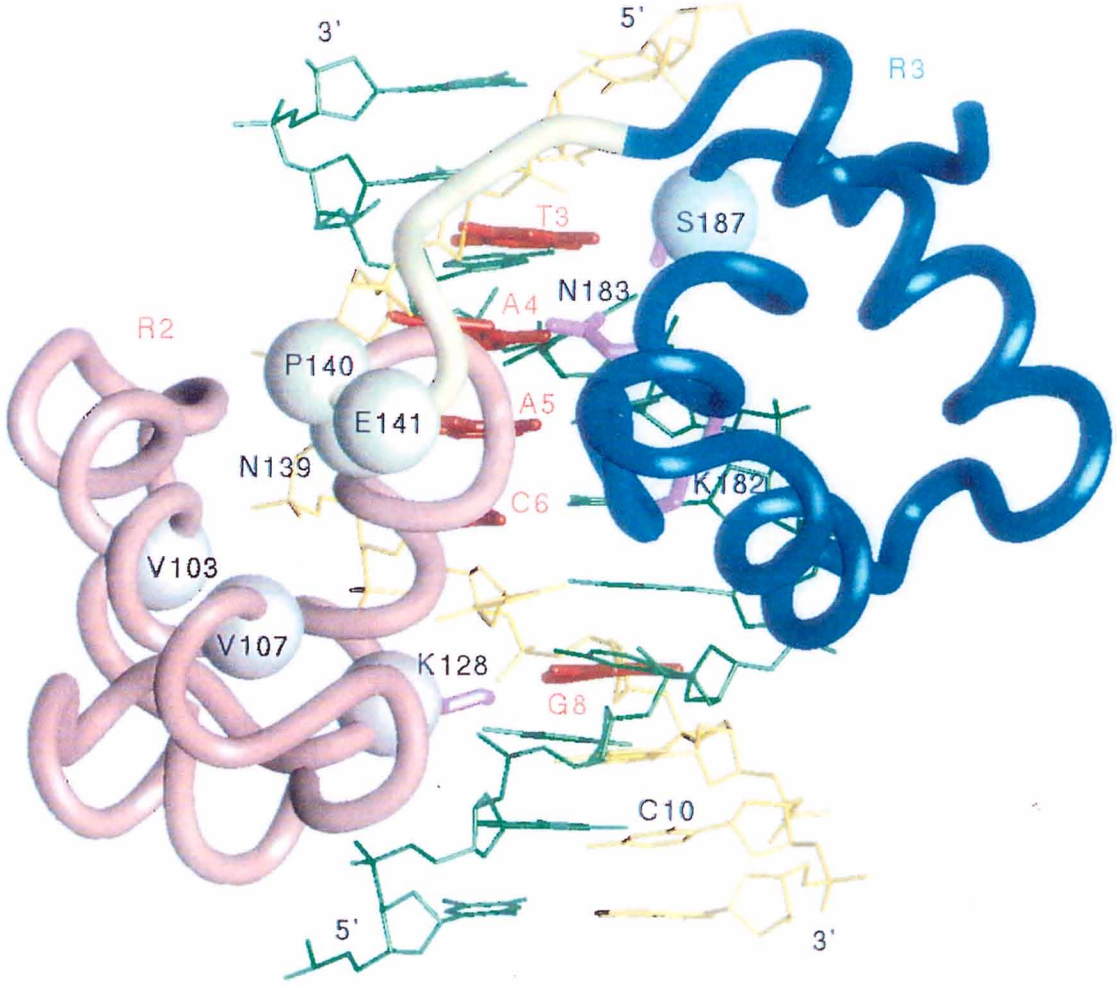


Figure 3-2. Sequences of DNAs used in this study. The base numbering follows that suggested by Ogata *et al.* (1994). The consensus base sequence is underlined in MBS-I, and the substituted bases in the non-cognate DNAs are indicated in italics. The names of the single base-pair substituted DNAs are, for example, [T3G]MBS-I, for the substitution of the T-A base-pair at the third position by a G-C base-pair. In the [NC-a]MBS-I, the base sequence, AAC, which is part of the consensus sequence, is included in the complementary strand. In contrast, the [NC-b]MBS-I has almost no trace of the consensus sequence.

	-3 1 5 10 15 19
MBS-I	5′- CACCCTAACTGACACACATTCT -3′
[T3G]MBS-I	CACCC*G*AACTGACACACATTCT
[A4G]MBS-I	CACCCT*G*ACTGACACACATTCT
[A5G]MBS-I	CACCCTA*G*CTGACACACATTCT
[A5T]MBS-I	CACCCTA*T*CTGACACACATTCT
[A5C]MBS-I	CACCCTA*C*CTGACACACATTCT
[C6T]MBS-I	CACCCTAA*T*TGACACACATTCT
[G8A]MBS-I	CACCCTAACT*A*ACACACATTCT
[G8T]MBS-I	CACCCTAACT*T*ACACACATTCT
[C10T]MBS-I	CACCCTAACTGA*T*ACACATTCT
[C10G]MBS-I	CACCCTAACTGA*G*ACACATTCT
[NC-a]MBS-I	CACCCT*GGT*TAACACACATTCT
[NC-b]MBS-I	CACC*TTGCT*TGACACACATTCT

Figure 3-3. Typical isothermal titration calorimetric profiles of the interaction between c-Myb R2R3* and the cognate or the non-cognate DNAs at 20.2 °C. (a) (*lower*) The 0.5 mM MBS-I (Figure 3-2) solution was injected 20 times in 5-μl increments into the 0.017 mM R2R3* solution. *(middle)* The 0.5 mM [NC-b]MBS-I (Figure 3-2) solution was injected into the 0.017 mM R2R3* solution. (*upper*) The MBS-I solution was injected into the experimental buffer. All samples were in 100 mM potassium phosphate buffer (pH 7.5) containing 20 mM KCl. Injections were performed over 10 s at 270 s intervals. (b) The data points were obtained by integration of the above peaks, corrected for the dilution heat, and plotted against the molar ratio (MBS-I / R2R3* and [NC-b]MBS-I / R2R3*). The data were fitted using a nonlinear least-squares method. Filled circles are for R2R3* with MBS-I, and open circles are for R2R3* with [NC-b]MBS-I.

Figure 3-3

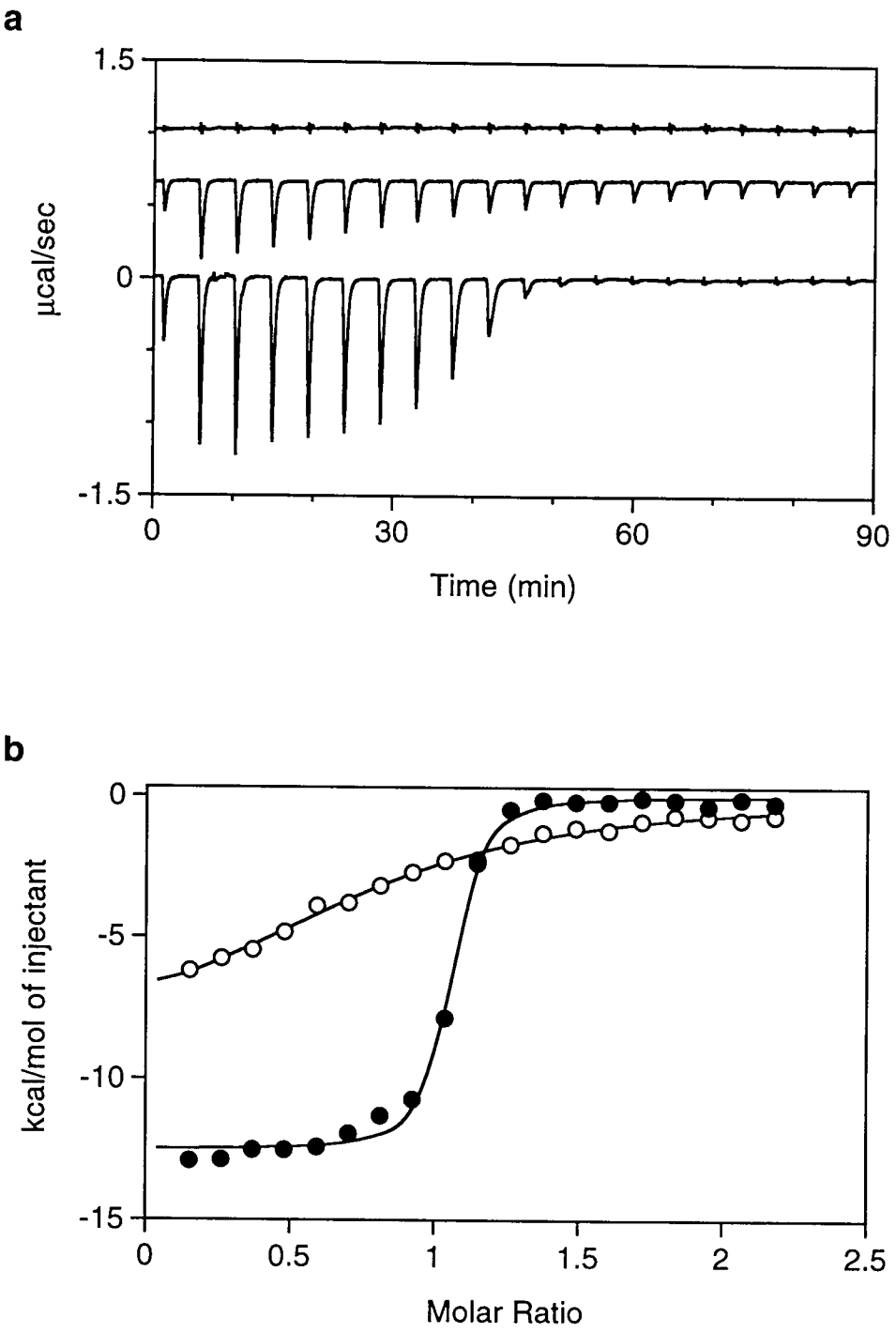

a
1.5
0
-1.5
µcal/sec
0
30
60
90
Time (min)
b
0
-5
-10
-15
kcal/mol of injectant
0
0.5
1
1.5
2
2.5
Molar Ratio

Figure 3-4. Diagram representing the thermodynamics of binding of the c-Myb R2R3* to the cognate MBS-I (filled marks with solid lines) and to the non-cognate [NC-b]MBS-I (open marks with broken lines). Values of ΔH (● and ○), $T\,\Delta S^\circ$ (■ and □), and ΔG° (▲ and △) are plotted as a function of temperature. Each line was fitted to a linear function.

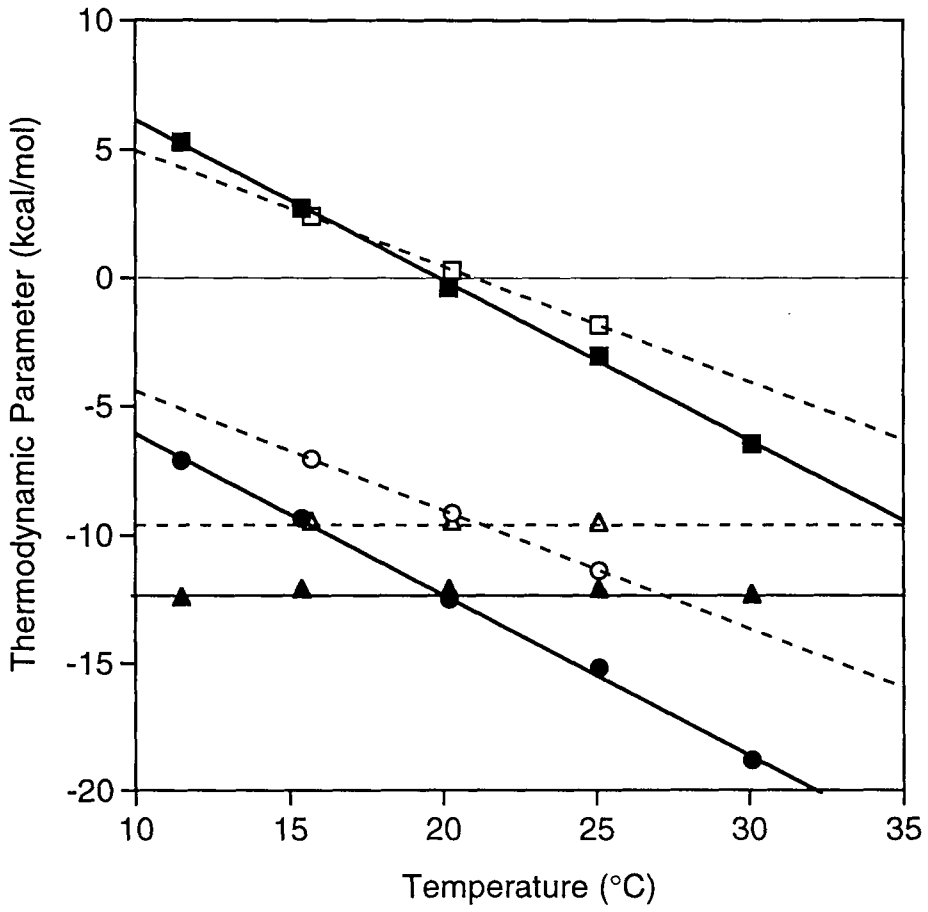


Figure 3-5. Salt dependence of the R2R3* interactions with the cognate and the non-cognate DNA. (a) Plot of the log of the binding affinity, log K_a, (b) ΔH, and (c) ΔC_p, versus the log of the K^+ concentration. Filled circles are for binding to the cognate MBS-I, and open circles are for binding to the non-cognate [NC-b]MBS-I. Error bars are indicated in the figures.

Figure 3-5

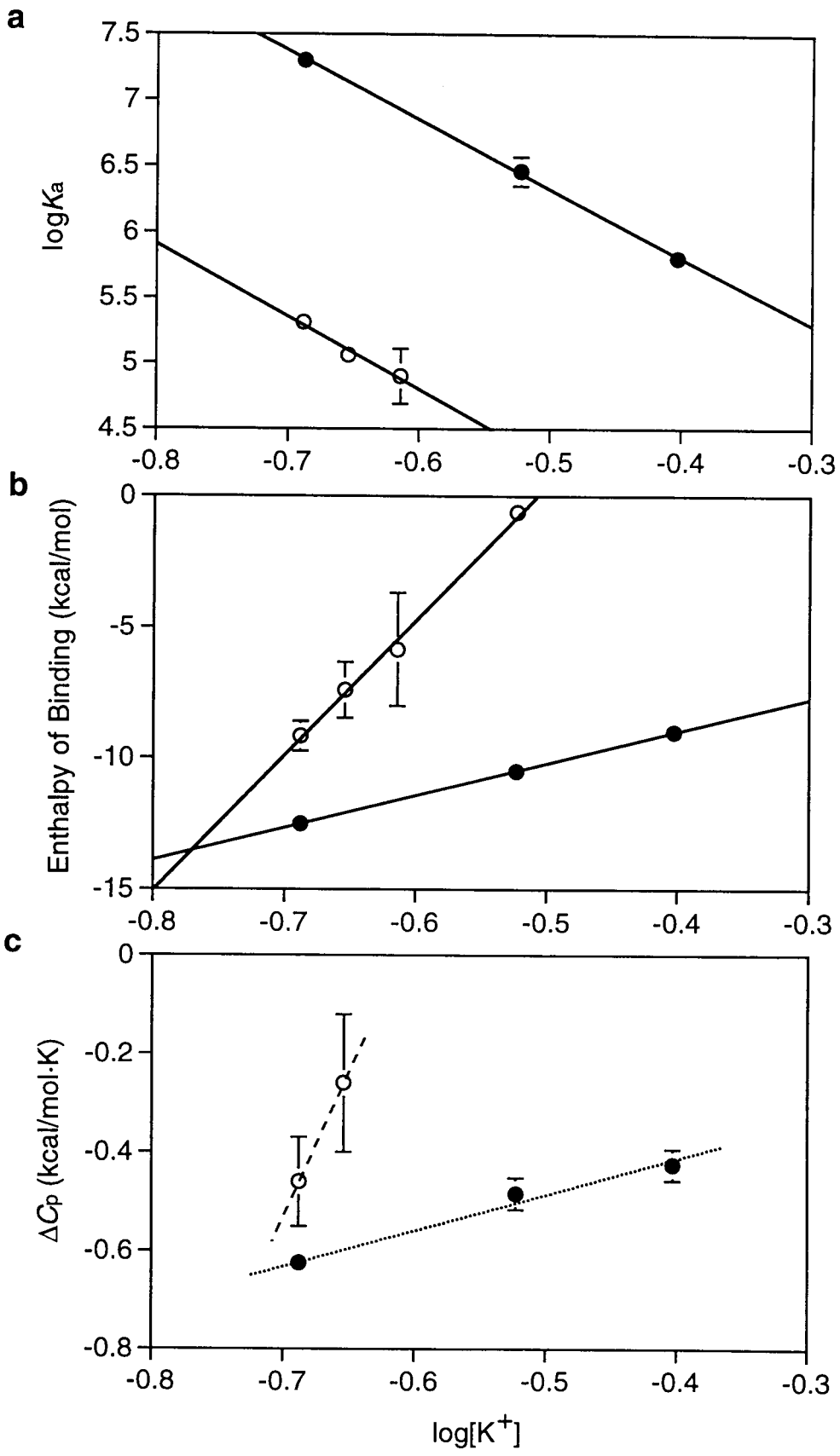

a
7.5
7
6.5
6
5.5
5
4.5
logKa
-0.8
-0.7
-0.6
-0.5
-0.4
-0.3
b
0
-5
-10
-15
Enthalpy of Binding (kcal/mol)
c
0
-0.2
-0.4
-0.6
-0.8
ΔCp (kcal/mol·K)
log[K+]

Figure 3-6. CD spectra and thermal stabilities of R2R3* mutants substituted in the recognition sites and the hydrophobic core. (a) Far-UV spectra of R2R3* (thick solid line), K128M (thin solid line), S187G (broken line), S187A (dotted line), and V103I/V107H (dot-dashed line). All of them are superimposed. The vertical scale is normalized by the mole concentration. (b) At the near-UV range. (c) Thermal denaturation curves of R2R3* and the four mutant proteins. The apparent fraction of folded protein, obtained by monitoring the CD value at 222 nm, is shown as a function of the temperature.

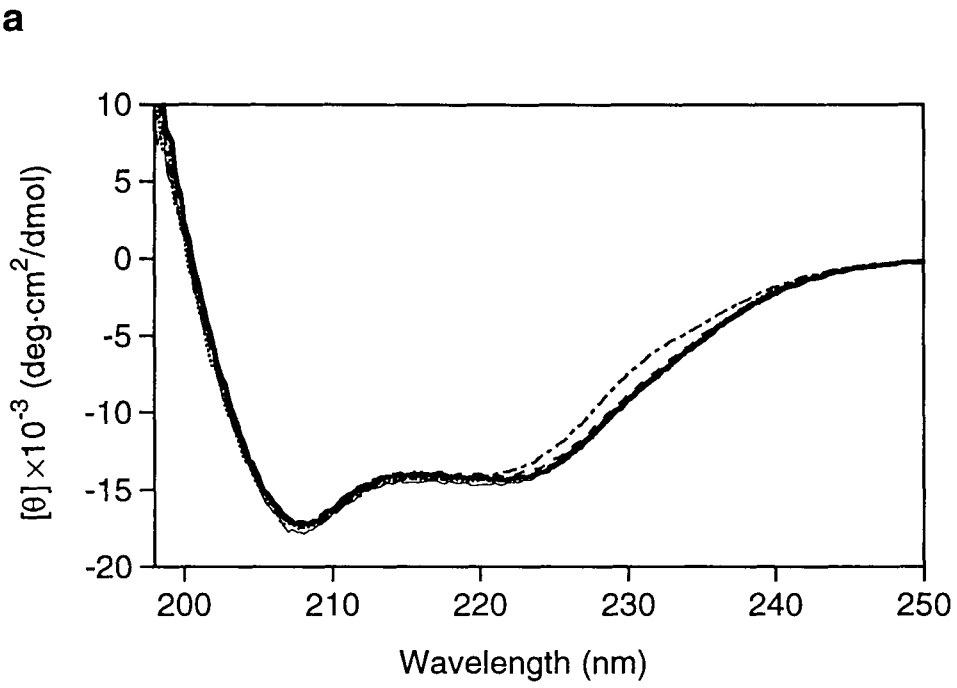


Figure 3-6

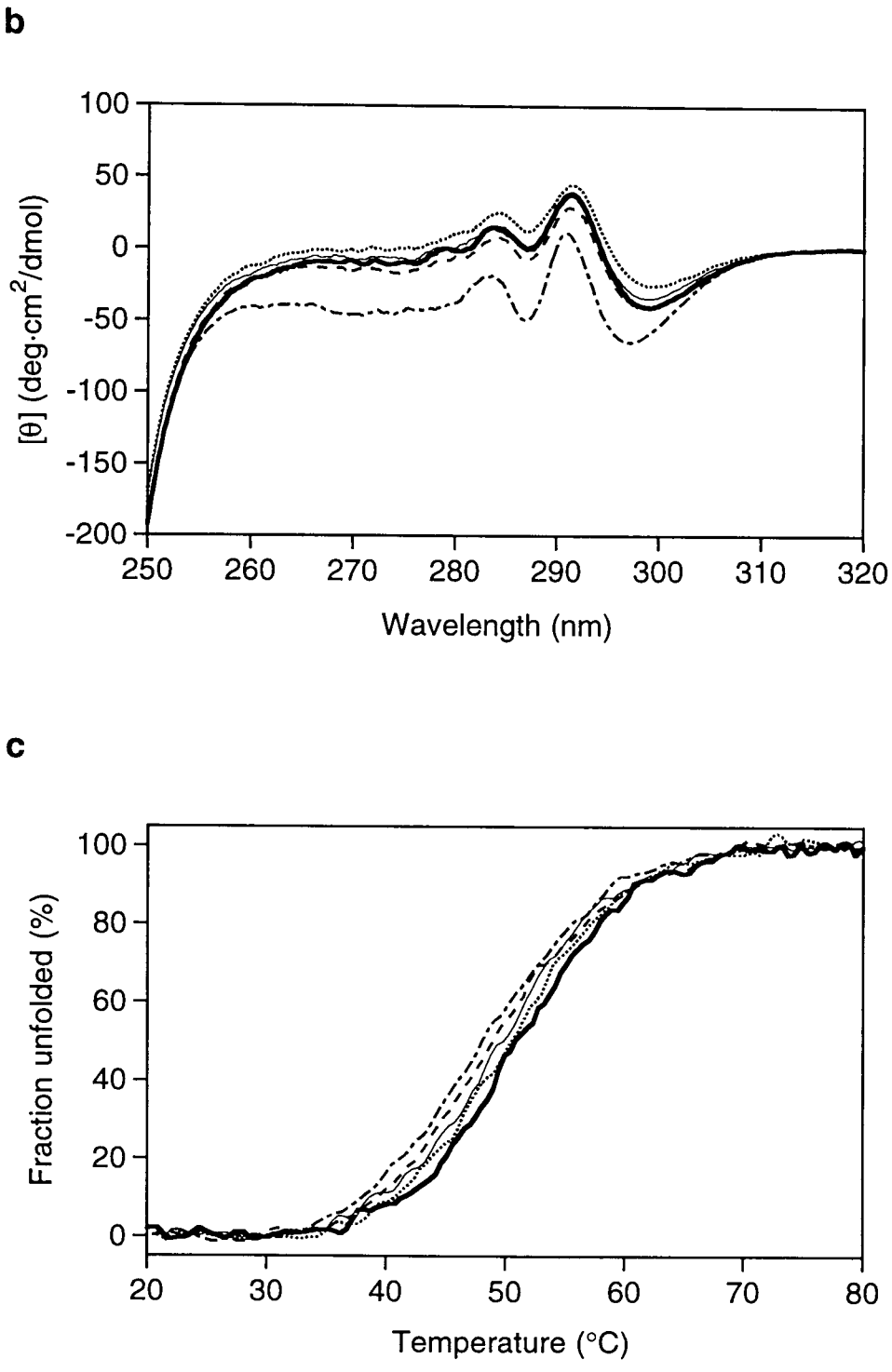

b
100
50
0
-50
-100
-150
-200
[θ] (deg·cm²/dmol)
250
260
270
280
290
300
310
320
Wavelength (nm)
c
100
80
60
40
20
0
Fraction unfolded (%)
20
30
40
50
60
70
80
Temperature (°C)

Figure 3-7. Thermal stability analysis for R2R3* by DSC. The melting curve (solid line) cannot be fitted to a two-state model, but can be fitted to a three-state model with two components. T_ms for the first (broken line) and second (dotted line) components are 50 °C and 55 °C, respectively.

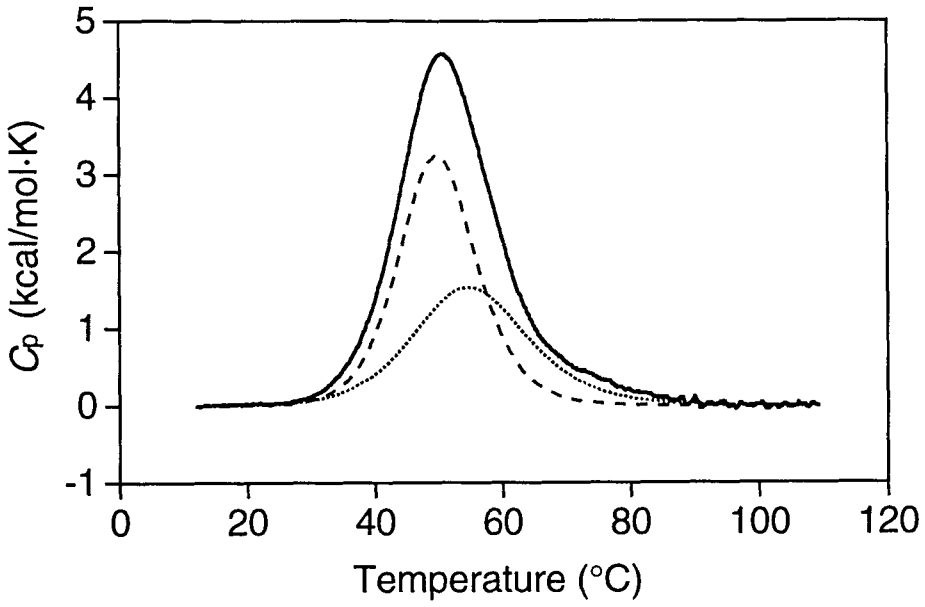


Figure 3-8. Isothermal titration calorimetric profiles of the binding of the K128M mutant at 20 °C. (*lower*) The 0.5 mM MBS-I solution was injected 20 times in 5-μl increments into the 0.017 mM K128M solution. (*middle*) The 0.5 mM [NC-b]MBS-I solution was injected into the 0.017 mM K128M solution. (*upper*) The 0.5 mM [NC-b]MBS-I solution was injected into the experimental buffer. All samples were in 100 mM potassium phosphate buffer (pH 7.5) containing 20 mM KCl. Injections were performed over 10 s at 270 s intervals.

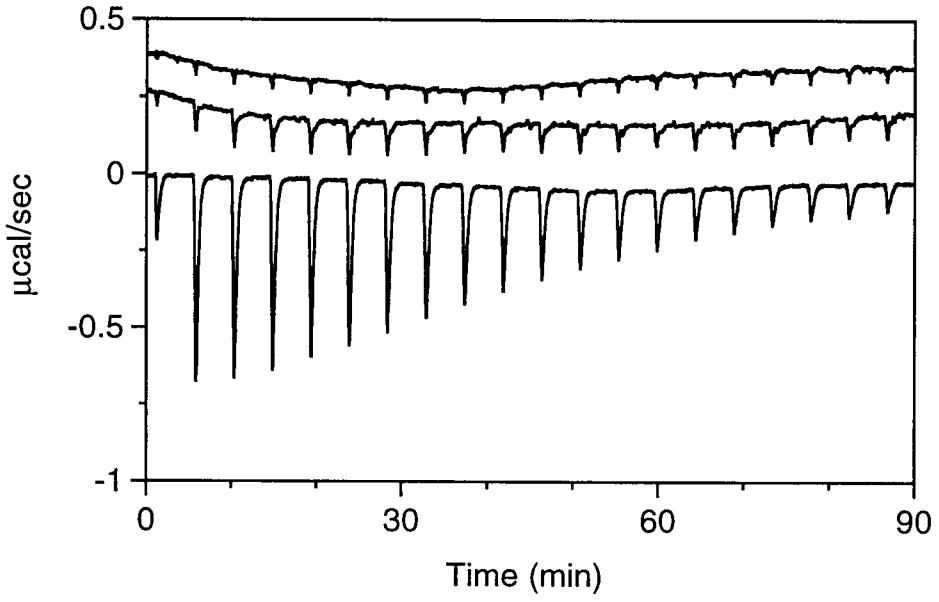


Figure 3-9. CD spectra and thermal stabilities of R2R3* mutants substituted in the linker region. (a) Far-UV spectra of R2R3* and the six mutant proteins (N139G, P140G, E141G, N139A, P140A, and E141A). All of them are superimposed. The vertical scale is normalized by the mole concentration. (b) Near-UV spectra of R2R3* (thick solid line in both upper and lower plates), N139G (thin solid line in upper plate), P140G (broken line in upper plate), E141G (dotted line in upper plate), N139A (thin solid line in lower plate), P140A (broken line in lower plate), and E141A (dotted line in lower plate). (c) Thermal denaturation curves of R2R3* and the six mutant proteins at the near-UV range. The identities of the lines are the same as those in (b). The apparent fraction of folded protein, obtained by monitoring the CD value at 222 nm, is shown as a function of the temperature.

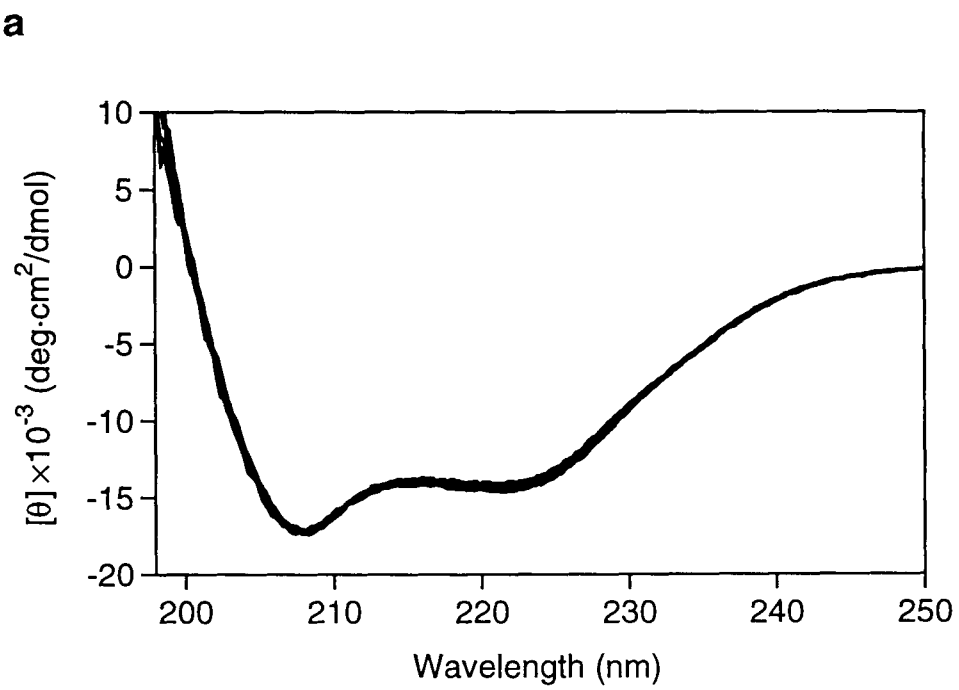


Figure 3-9

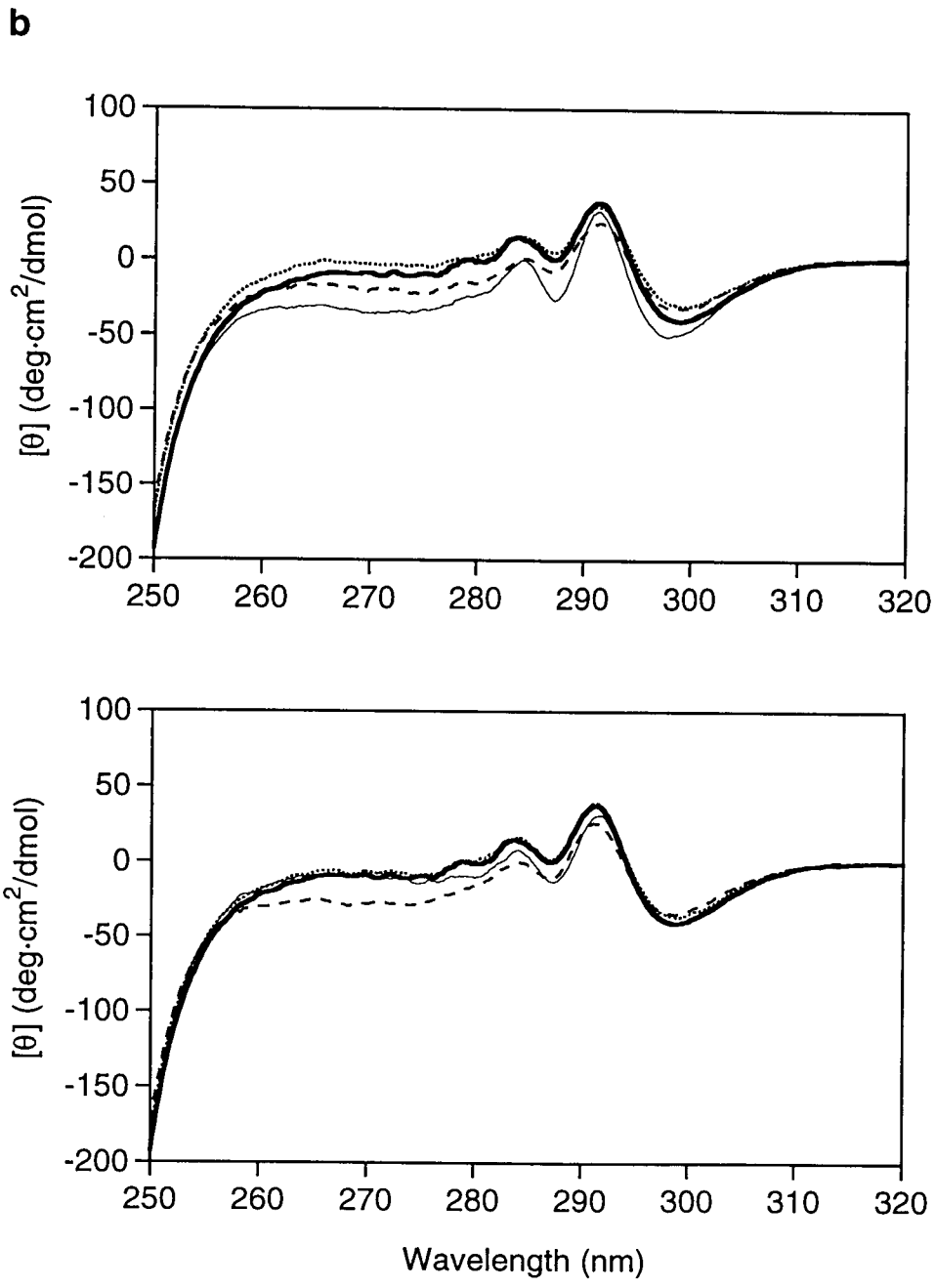

b
100
50
0
-50
-100
-150
-200
[θ] (deg·cm²/dmol)
250
260
270
280
290
300
310
320
Wavelength (nm)

Figure 3-9

c

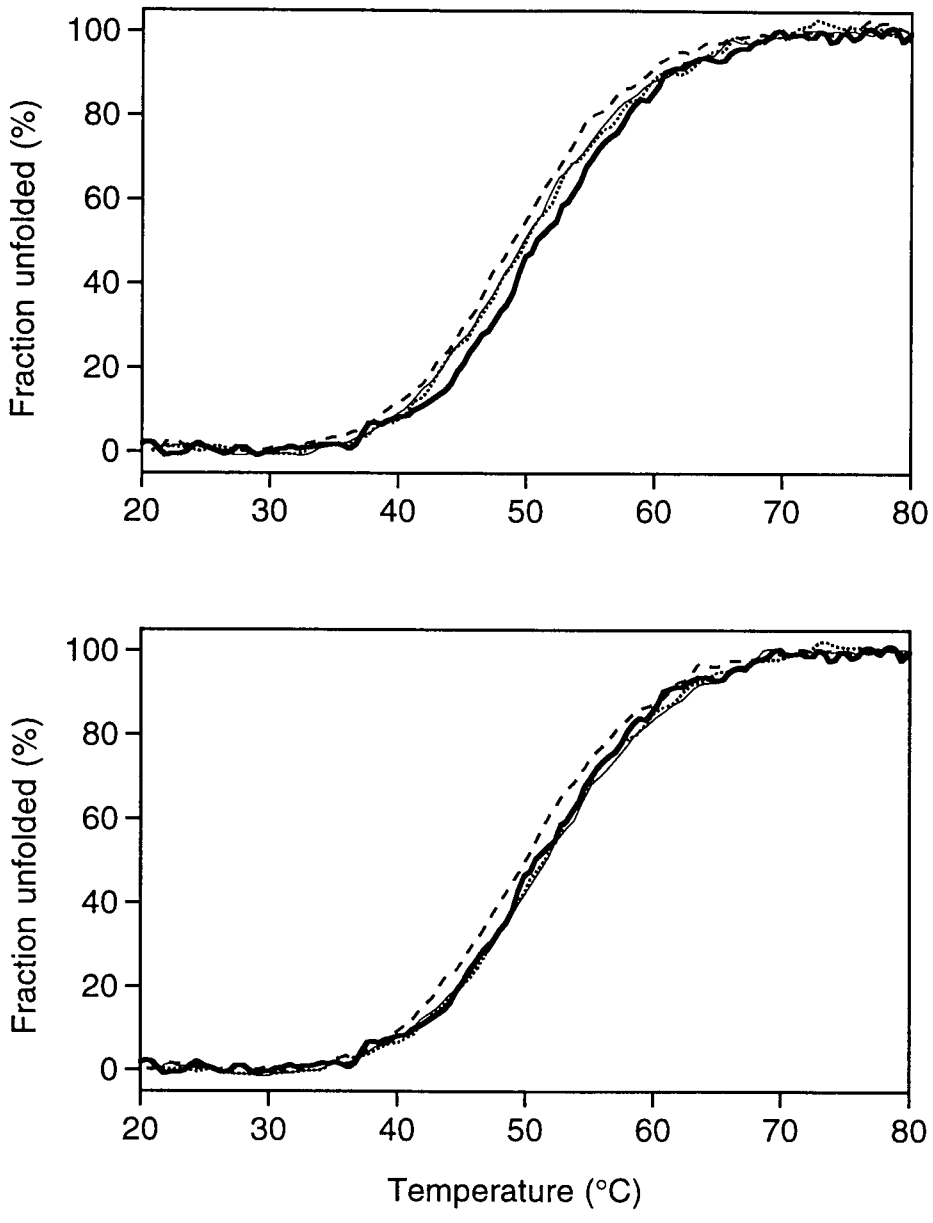

Fraction unfolded (%)
100
80
60
40
20
0
20
30
40
50
60
70
80
Fraction unfolded (%)
100
80
60
40
20
0
20
30
40
50
60
70
80
Temperature (°C)

Figure 3-10. The thermodynamic cycle with the substitutions of an amino acid in R2R3* and a base in MBS-I. The unit of the numbers in the figure is kcal mol^{-1}. (a) The thermodynamic cycle with the substitutions of Lys128 with Met and G8 with thymine. (b) The thermodynamic cycle with the substitutions of Ser187 with Gly and T3 with guanine. (c) The thermodynamic cycle of V103I/V107H with the substitution of G8 with adenine. (d) The thermodynamic cycles with the substitutions of Asn139 with Gly and Ala, and G8 with adenine. (e) The thermodynamic cycles with the substitutions of Pro140 with Gly and Ala, and G8 with adenine.

Figure 3-10

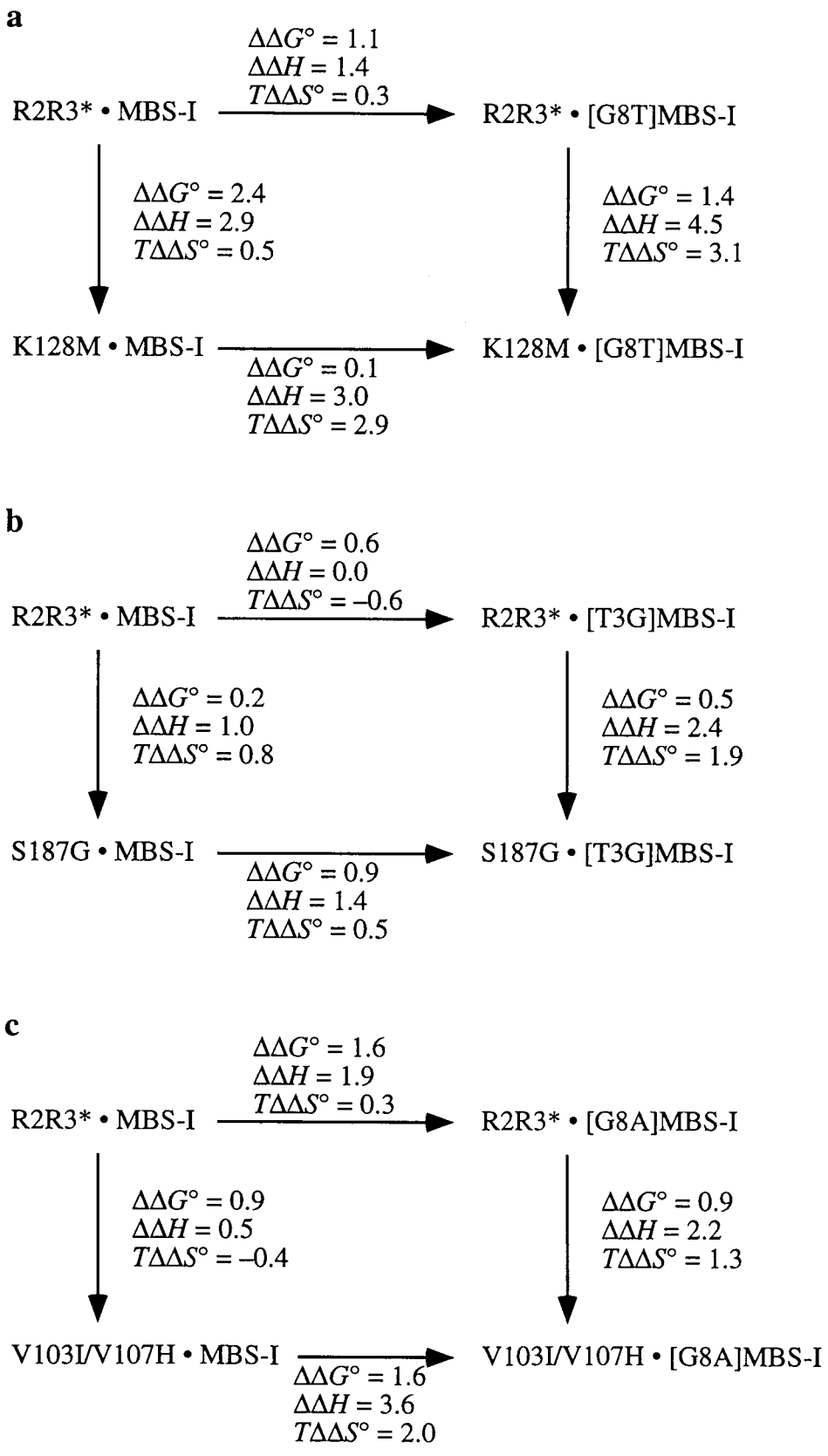
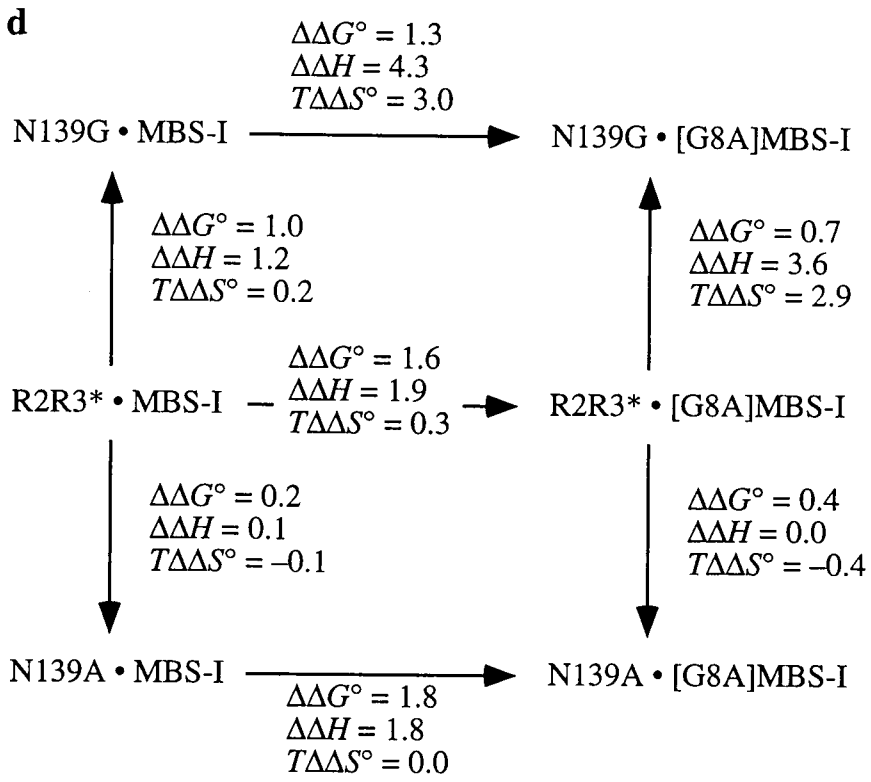
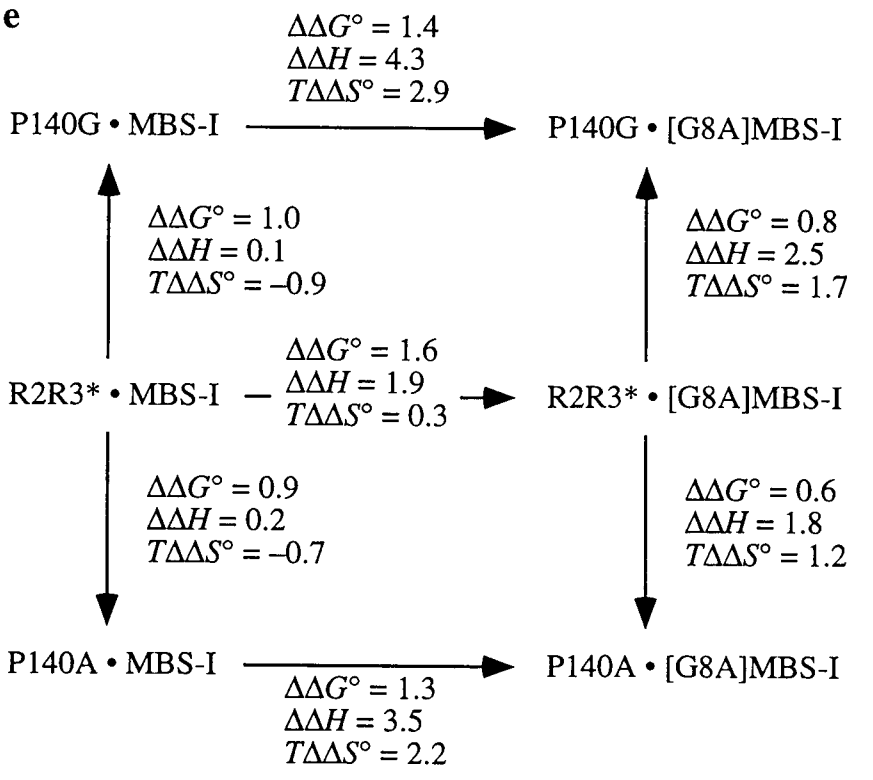

a
ΔΔG° = 1.1
ΔΔH = 1.4
TΔΔS° = 0.3
R2R3* • MBS-I
R2R3* • [G8T]MBS-I
ΔΔG° = 2.4
ΔΔH = 2.9
TΔΔS° = 0.5
ΔΔG° = 1.4
ΔΔH = 4.5
TΔΔS° = 3.1
K128M • MBS-I
K128M • [G8T]MBS-I
ΔΔG° = 0.1
ΔΔH = 3.0
TΔΔS° = 2.9
b
ΔΔG° = 0.6
ΔΔH = 0.0
TΔΔS° = –0.6
R2R3* • MBS-I
R2R3* • [T3G]MBS-I
ΔΔG° = 0.2
ΔΔH = 1.0
TΔΔS° = 0.8
ΔΔG° = 0.5
ΔΔH = 2.4
TΔΔS° = 1.9
S187G • MBS-I
S187G • [T3G]MBS-I
ΔΔG° = 0.9
ΔΔH = 1.4
TΔΔS° = 0.5
c
ΔΔG° = 1.6
ΔΔH = 1.9
TΔΔS° = 0.3
R2R3* • MBS-I
R2R3* • [G8A]MBS-I
ΔΔG° = 0.9
ΔΔH = 0.5
TΔΔS° = –0.4
ΔΔG° = 0.9
ΔΔH = 2.2
TΔΔS° = 1.3
V103I/V107H • MBS-I
V103I/V107H • [G8A]MBS-I
ΔΔG° = 1.6
ΔΔH = 3.6
TΔΔS° = 2.0

Figure 3-10


d
ΔΔG° = 1.3
ΔΔH = 4.3
TΔΔS° = 3.0
N139G • MBS-I
N139G • [G8A]MBS-I
ΔΔG° = 1.0
ΔΔH = 1.2
TΔΔS° = 0.2
ΔΔG° = 0.7
ΔΔH = 3.6
TΔΔS° = 2.9
ΔΔG° = 1.6
ΔΔH = 1.9
TΔΔS° = 0.3
R2R3* • MBS-I
R2R3* • [G8A]MBS-I
ΔΔG° = 0.2
ΔΔH = 0.1
TΔΔS° = –0.1
ΔΔG° = 0.4
ΔΔH = 0.0
TΔΔS° = –0.4
N139A • MBS-I
N139A • [G8A]MBS-I
ΔΔG° = 1.8
ΔΔH = 1.8
TΔΔS° = 0.0


e
ΔΔG° = 1.4
ΔΔH = 4.3
TΔΔS° = 2.9
P140G • MBS-I
P140G • [G8A]MBS-I
ΔΔG° = 1.0
ΔΔH = 0.1
TΔΔS° = –0.9
ΔΔG° = 0.8
ΔΔH = 2.5
TΔΔS° = 1.7
ΔΔG° = 1.6
ΔΔH = 1.9
TΔΔS° = 0.3
R2R3* • MBS-I
R2R3* • [G8A]MBS-I
ΔΔG° = 0.9
ΔΔH = 0.2
TΔΔS° = –0.7
ΔΔG° = 0.6
ΔΔH = 1.8
TΔΔS° = 1.2
P140A • MBS-I
P140A • [G8A]MBS-I
ΔΔG° = 1.3
ΔΔH = 3.5
TΔΔS° = 2.2

Summary

The formation of specific protein-DNA complexes would represent a delicate balance between the energy gained by formation of new hydrogen bonds, optimal buried surface area, and other stabilizing interactions. In protein-DNA interactions, long-range electrostatic forces bring the protein and the DNA into proximity at the first step, and at the second step, specific hydrogen bonds and van der Waals contacts are formed, along with the accompanying conformational changes. The non-specific binding ends at the first step, and should be predominantly entropy driven and strongly dependent on the salt concentration. In contrast, in specific binding, the high binding affinity is achieved by the specific contacts between the native protein and the DNA.

In order to attain specific DNA recognition, not only the specific binding between amino acids and bases, but also the structure and flexibility of protein and DNA are required. The current investigation described in this thesis has shown that all of them are interrelated and important for the specific DNA recognition. In the Chapter 1 of this thesis, the results showed the importance of specific contacts between amino acids and bases pyrimidine and the overall DNA structure, as follows. The preference by the c-Myb DNA-binding domain at the initial base of the consensus sequence should originates not only from the direct recognition by the amino acid at 187, but also from the indirect recognition of the overall shape of the DNA, that is, the intrinsic bendability of the pyrimidine-purine step of the DNA duplex. Additionally, following the conventional chemical rules of the direct-readout mechanism, the Asn187 mutant showed stronger affinity with adenine than with pyrimidines, and the Lys187 and the Arg187 mutants also changed the base specificity, preferring cytosine and guanine to adenine or thymine. In the Chapter 2 of this thesis, the results showed the requirements of the amino acid residues in the hydrophobic core, especially for the protein structure, as follows. The residues in the first helix of R2, Val103 and Val107, which are involved in the hydrophobic core and do not directly interact with the DNA, are indispensable for specific

binding, since they contribute to the correct formation of helix-1 and the characteristic packing of R2, which is slightly different from that of R3. In the Chapter 3 of this thesis, the results showed the dissection of the individual roles in the formation of the specific complex. Of these, the role of the linker connecting two repeats was clearly shown, as follows. The binding reduction of the substitution of Pro140 in the linker by Gly or Ala was mainly attributed to the decrease of ΔS°, with the ΔH unchanged. These results are a evidence of the local folding of the linker, which accompanies specific DNA binding. The role of Pro140 in the linker is to restrict the conformational flexibility between R2 and R3 in the DNA-free state.

In order to understand the difference between specific recognition and non-specific recognition, we have to understnad non-specific binding as the procedure of random molecular association. While structural analysis can give little information about non-specific binding, calorimetric analysis can provide much information. The characteristic features of non-specific binding are (i) positive ΔS° and (ii) small negative ΔH, (iii) the high sensitivity of ΔH against the salt concentration may be another feature. All of these features coincide with the conventional view that non-specific binding in protein-DNA associations is entropically driven by electrostatic interactions, although ΔH was detectable. Theoretically, a smaller ΔC_p is expected in non-specific binding than in specific binding, because of the lower contributions of the hydrophobic effect and the internal vibration effect in non-specific binding, even when the hydration effect of the released counterions from DNA is observed on the detectable ΔC_p. In fact, this tendency was first found in the current investigation.

In this study, the binding analyses using both a filter binding assay and ITC have revealed new principles of the c-Myb R2R3 binding to the DNA that had not clearly been shown by the NMR structure, such as the conformational changes of protein and DNA and the roles of water molecules for the recognition. Additionally, the thermodynamic analyses in this study contain the largest amount of data among the investigations

previously reported for protein-DNA interactions. These data, obtained with site-directed mutagenesis on both protein and DNA, can dissect the molecular interactions in a way that assign specific energetic contributions to individual contacts. They have expanded our understanding of protein-DNA interactions, along with structural analyses.

In order to really understand mechanisms of DNA recognition, much additional work is needed. There is much to learn from combinations of experiments addressing structure, dynamics, and thermodynamics complemented by computer simulations and statistical-mechanical considerations. Studies of DNA recognition are closely linked to issues of gene regulation. A careful analysis of the binding energies is important, because we would like to understand the physical basis for the differential regulation of gene expression. Progress in understanding DNA recognition will also allow us to design new DNA-binding proteins. This will provide a rigorous test of our understanding, and will be a key to the success of projects designed to target DNA regulatory elements with new therapeutic or diagnostic agents.

List of Publications

The contents of this thesis are composed of the following papers.

1. Oda, M., Furukawa, K., Ogata, K., Sarai, A., Ishii, S., Nishimura, Y., and Nakamura, H. (1997) Investigation of the pyrimidine preference by the c-Myb DNA-binding domain at the initial base of the consensus sequence. *J. Biol. Chem.*, **272**, 17966-17971.
2. Oda, M., Furukawa, K., Ogata, K., Sarai, A., Ishii, S., Nishimura, Y., and Nakamura, H. (1998) Identification of indispensable residues for specific DNA-binding in the imperfect tandem repeats of c-Myb R2R3. *Protein Eng.*, **11**, in press.
3. Oda, M., Furukawa, K., Ogata, K., Sarai, A., and Nakamura, H. (1998) Thermodynamics of specific and non-specific DNA binding by the c-Myb DNA-binding domain. *J. Mol. Biol.*, submitted.

The other supplementary paper.

1. Furukawa, K., Oda, M., and Nakamura, H. (1996) A small engineered protein lacks structural uniqueness by increasing the side-chain conformational entropy. *Proc. Natl. Acad. Sci. USA*, **93**, 13583-13588.

Acknowledgments

The present investigation was carried out under the guidance of Dr. Haruki Nakamura in Protein Engineering Research Institute (PERI) and Biomolecular Engineering Research Institute (BERI) from 1993 to 1997. The author wishes to express his thanks to him for his constant guidance and valuable discussions throughout this work.

The author is deeply indebted to Professor Kazuyuki Akasaka at the Division of Molecular Science, Graduate School of Science and Technology, Kobe University for his encouragement since 1988, when the author was a student of Kyoto University, and for affording the author a chance to obtain the Ph.D. degree.

The author wishes to make a grateful acknowledgment to Dr. Koji Furukawa of Research Institute for Biological Sciences (RIBS), Science University of Tokyo for his guidance, helpful discussions, and intimate encouragement throughout this investigation.

The author is deeply thankful to Dr. Kazuhiro Ogata of Kanagawa Academy of Science and Technology (KAST), and at the Department of Structural Biology, School of Medicine, Yokohama City University, Drs. Akinori Sarai and Shunsuke Ishii of Tsukuba Life Science Center, the Institute of Physical and Chemical Research (RIKEN), and Professor Yoshifumi Nishimura at the Graduate School of Integrated Science, Yokohama City University for their collaboration in this work and kind assistance.

The author is deeply thankful to Dr. Motohisa Oobatake of RIKEN, Mr. Soichi Morikawa of Kaneka Cooperation, Dr. Tadayasu Ohkubo of Japan Advanced Institute of Science and Technology, Hokuriku, Dr. Atsuo Tamura of Kobe University, Dr. Kenji Kanaori of Kyoto Institute of Technology, and Dr. Nobutoshi Ito of BERI for their helpful discussions.

The author also wishes to extend his hearty thanks to members of PERI and BERI for their kind help, and to members of Kyowa Hakko Kogyo Co., Ltd. for affording the author a chance to work in PERI and BERI.

Finally, the author greatly appreciates to his parents and his wife, Motoko, for encouragement and daily assistant.

January, 1998

Masayuki Oda