Biochemical and Biophysical Research Communications **262,** 94–97 (1999)
Article ID bbrc.1999.1159, available online at http://www.idealibrary.com on IDEAL®

Construction of an Artificial Tandem Protein of the c-Myb DNA-Binding Domain and Analysis of Its DNA Binding Specificity

Masayuki Oda,*[,1] Koji Furukawa,*[,2] Akinori Sarai,† and Haruki Nakamura*[,3]
**Biomolecular Engineering Research Institute (BERI), Furuedai, Suita, Osaka 565-0874, Japan; and †Tsukuba Life Science Center, Institute of Physical and Chemical Research (RIKEN), Koyadai, Tsukuba, Ibaraki 305-0074, Japan*

Received June 24, 1999

An artificial tandem protein was generated using the third repeat of the c-Myb DNA-binding domain, and its DNA binding affinity and specificity were analyzed by a filter binding assay, isothermal titration calorimetry, and surface plasmon resonance. Although this artificial protein had the proper secondary structure, which is similar to the third repeat by itself, it could not bind to the expected base sequences specifically. Compared with the successful results of the zinc finger fusion proteins with novel sequence specificities, the cooperativity between the adjacent repeats, observed in the c-Myb-DNA complex, should also be required for the DNA recognition by the artificial tandem protein. Using the previous analyses of the DNA binding specificities by Myb homologous proteins, the differences in the DNA recognition mechanisms between the animal and plant Myb domains are also discussed.

c-Myb is a transcriptional activator that specifically binds to DNA fragments containing PyAAC$^{G}/_{T}$G, where Py indicates a pyrimidine, with a dissociation constant of about 10^{-9} M to the specific DNA sequence (1–3). The DNA-binding domain (DBD) of c-Myb consists of three imperfect 51- or 52- residue repeats (designated as R1, R2, and R3 from the N-terminus) (4–6). The solution structures of the minimum DBD (R2R3) of mouse c-Myb and the complex of R2R3 with the Myb-binding DNA sequence (MBS-I) were determined by multidimensional NMR techniques (7, 8). Both R2 and R3 are composed of three helices with a helix-turn-helix variation motif, and the overall structures of these repeats are almost identical. Additionally, it is noteworthy that the two recognition helices of R2 and R3 contact each other directly to bind cooperatively to the specific base sequence.

Thus far, Myb-homologous domain proteins have been found in many species (9). While the mouse and human c-Myb proteins have been shown to have roles as transcriptional regulators, a large number of Myb domain proteins in plants participate in a variety of important cellular functions (9, 10). In terms of the DNA consensus sequence, several studies have shown that the Myb domains in plants recognize different DNA sequences, for instance, the MYB.Ph3 protein from Petunia recognizes both AACNG and TAACTAACT, and the P protein from maize recognizes CC$^{T}/_{A}$ACC (11, 12). Solano *et al.* showed that a single residue substitution changes the DNA binding specificity between MYB.Ph3 and c-Myb, and Williams and Grotewold showed that a chimeric Myb domain, composed of R2 of P and R3 of v-Myb, has novel DNA binding specificity (9, 13). These results give rise to the question as to whether the cooperative interaction of c-Myb R2R3 is actually required for specific DNA binding.

We have recently characterized the c-Myb R2R3-DNA interactions, by a filter binding assay, isothermal titration calorimetry (ITC), and surface plasmon resonance (SPR) (14–17). In this study, we generated an artificial molecule by fusing the c-Myb R3 in tandem, and analyzed its DNA binding affinity and specificity. Based on these results, we discuss the importance of

[1] To whom correspondence should be addressed at present address: Research Institute for Biological Sciences (RIBS), Science University of Tokyo, Yamazaki, Noda, Chiba 278-0022, Japan. Fax: +81-471-24-1541. E-mail: oda@rs.noda.sut.ac.jp.
[2] Present address: Research Institute for Biological Sciences (RIBS), Science University of Tokyo, Yamazaki, Noda, Chiba 278-0022, Japan.
[3] Present address: Institute for Protein Research, Osaka University, Yamadaoka, Suita, Osaka 565-0871, Japan.

Abbreviations used: DBD, DNA-binding domain; R2R3, minimum c-Myb DNA-binding domain composed of the second and the third repeat fragments; MBS, Myb-binding DNA sequence; ITC, isothermal titration calorimetry; SPR, surface plasmon resonance; CD, circular dichroism; k_{off}, dissociation rate constant; k_{on}, association rate constant; R2R3*, stable mutant of R2R3, in which Cys130 is substituted with Ile; \RU, response unit.

0006-291X/99 $30.00

cooperativity in the binding of c-Myb R2R3 to DNA and the differences between the Myb domains of different kingdoms in their specific DNA recognition.

MATERIALS AND METHODS

Preparation of c-Myb R2R3 mutant proteins and oligonucleotides. The expression and purification methods of the c-Myb R2R3 mutant proteins were described previously (14). The purified proteins were dialyzed against 100 mM potassium phosphate buffer (pH 7.5) containing 20 mM KCl (buffer A). Five kinds of ds oligonucleotides were obtained from BEX (Tokyo). The sequences of these fragments are shown in Fig. 1A. The complementary strands were annealed, and were dialyzed against buffer A.

Circular dichroism. Circular dichroism (CD) spectra were measured at 20°C on a Jasco J-600 spectropolarimeter, as described previously (14). The protein concentration was 0.1 mg/ml.

DNA bindng assay. Filter binding assays were carried out, as described previously (14). [^{32}P]DNA and various amounts of the c-Myb R2R3 mutant proteins were incubated on ice for 30 min. The incubated samples were filtered through a nitrocellulose membrane (Schleicher & Schuell, BA-85, 0.45 μm) in approximately 10 sec with suction. The filters were dried and were counted by a liquid scintillation counter.

ITC experiments were carried out on a Microcal MCS calorimeter, as described previously (16). The 0.5 mM DNA solution was titrated into the 0.017 mM protein solution with a preliminary 2 μl injection, followed by 19 subsequent 5 μl additions. The data were analyzed with the Microcal Origin software supplied by the manufacturer.

SPR measurements were carried out on a BIAcore biosensor system, BIAcore 2000 (Biacore AB), as described previously (17). The obtained data were analyzed with the BIA evaluation program 2.1, using a nonlinear least squares method. The dissociation rate constants (k_{off}) and the association rate constants (k_{on}) were calculated from the dissociation phases and the association phases of the sensorgrams, respectively. The association constants (K_a) were calculated from the two rate constants, $K_a = k_{on}/k_{off}$. The equilibrium association constants ($K_{a,eq}$) can also be determined from a Scatchard analysis.

A

		-3 1 5 10 15 19
MBS-I	5′-	CACCCTAACTGACACACATTCT -3′
	3′-	GTGGGATTGACTGTGTGTAAGA -5′
[G8A]MBS-I		CACCCTAACT*A*ACACACATTCT
[G8A(s−1)]MBS-I		CACCCTAAC*A*ACACACATTCT
[G8A(s+2)]MBS-I		CACCCTAACT<u>AT</u>AACACACATTCT
[G8A(s+4)]MBS-I		CACCCTAACT<u>ATAT</u>*A*ACACACATTCT

B

M-E-V-K·····K-V-G-G-G-G-G-G-G-V-K·····K-V

FIG. 1. Sequences of the DNAs used in the present study and the R3-G7-R3 mutant. (A) The base numbering follows that suggested by Ogata *et al.* (7). The consensus base sequence is boxed in MBS-I. In the non-cognate DNAs, the sequence of one strand is indicated. The substituted bases are indicated in italics, and the additional introduced bases are underlined. The names of the base-pair substituted DNAs follow those suggested by Oda *et al.* (16). The names of the space changed [G8A]MBS-I are, for example, [G8A(s + 2)]MBS-I, for the introduction of two bases into the consensus sequence. (B) The R3 sequence (residues 142-193) is indicated by the box. An additional Met-Glu- sequence was introduced at the N-terminus, and seven Gly residues were introduced as the linker connecting the two R3 repeats.

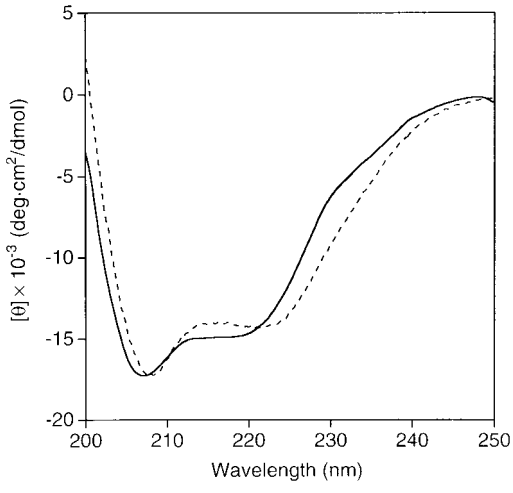

FIG. 2. Far-UV CD spectra of R3-G7-R3 (solid line), and R2R3* (broken line). The vertical scale is normalized by the mole concentration.

RESULTS

The NMR structure of the c-Myb R2R3-DNA complex shows that the recognition helix of R3 specifically binds to the AAC subsite of the consensus DNA sequence, while that of R2 recognizes the eighth G-C base pair, and both repeats also interact directly, which is unique among protein-DNA complexes (7). We surmised it would be possible to create an artificial tandem protein, c-Myb R3R3, which could recognize a novel base sequence, like other successful cases (9, 18–20). That is, the engineered R3R3 protein might specifically recognize two AAC subsites. This R3R3 mutant was designed as shown in Fig. 1B, and is named R3-G7-R3. Seven Gly residues were introduced as the linker connecting the two R3 repeats, so as not to cause steric interference between them. Our previous studies have shown that the R3-NPE-R3 mutant, which has a three residue linker, cannot specifically bind to [G8A]MBS-I (15). As shown in Fig. 2, the far-UV CD spectrum of R3-G7-R3 indicated a typical α-helical structure, and its profile is very similar to both R3 alone and R3-NPE-R3 (15, 21), while it is slightly different from the C130I mutant, denoted as R2R3*. The affinity and the

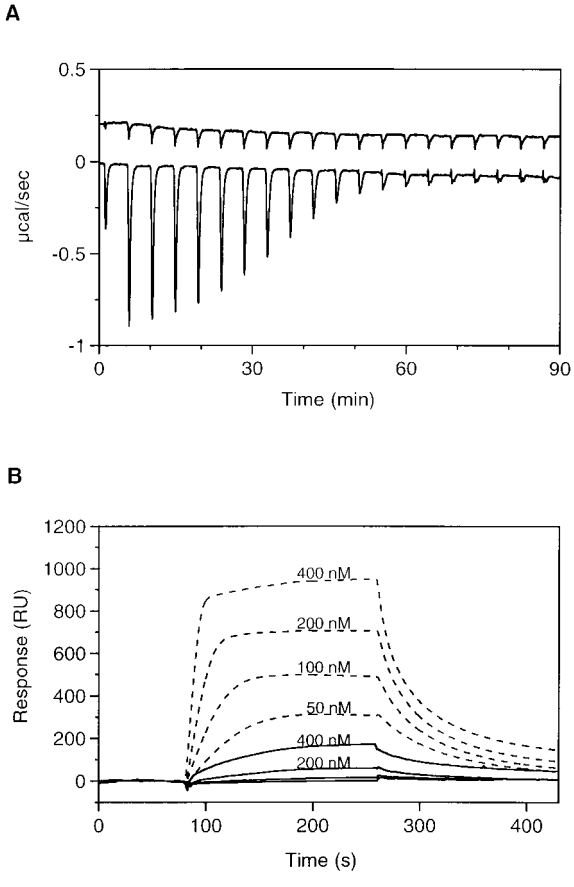

FIG. 3. DNA binding analyses of the R3-G7-R3 mutant using ITC and SPR. (A) Typical ITC profiles at 20°C. (upper) The 0.5 mM [G8A]MBS-I solution was injected into the 0.017 mM R3-G7-R3 solution. (lower) The 0.5 mM [G8A]MBS-I solution was injected 20 times in 5 μl increments into the 0.017 mM R2R3* solution. Injections were performed over 10 s at 270 s intervals. (B) A 60 μl aliquot of a 50, 100, 200, or 400 nM solution of R3-G7-R3 (solid line) or R2R3* (broken line) was passed over the immobilized [G8A]MBS-I, respectively, at 20°C with subsequent washing with the experimental buffer only.

specificity of this C130I mutant are almost equal to those of the wild-type protein (14).

The bindings of R3-G7-R3 to MBS-I and [G8A]MBS-I were analyzed using a filter binding assay and ITC. The filter binding assay experiments showed that the affinities to both DNAs were very low, with dissociation constants on the μM scale, which are similar to those of R3-NPE-R3 (15), and slightly lower than those of R3 alone (3). The ITC experiments showed that these interactions were accompanied by only a small $|\Delta H|$ (<1 kcal mol^{-1}), in contrast to the interaction between the R2R3* and [G8A]MBS-I, as shown in Fig. 3A. These results indicate that the bindings of R3-G7-R3 to both MBS-I and [G8A]MBS-I are non-specific and are entropically driven.

The interaction between R3-G7-R3 and [G8A]MBS-I was also analyzed using the BIAcore. As shown in Fig. 3B, a slight increase in the response units (RU) was observed, in which 400 nM of R3-G7-R3 binding approximately corresponds to 25 nM of R2R3* binding to the same DNA. In order to analyze the specificity of R3-G7-R3, three kinds of DNA fragments, [G8A(s − 1)]MBS-I, [G8A(s + 2)]MBS-I, and [G8A(s + 4)]MBS-I, all of which include two AAC subsites with different spacing, were prepared (Fig. 1A), and the bindings of R3-G7-R3 and R2R3* to these DNAs were analyzed using the BIAcore. The sensorgrams of the interactions between R3-G7-R3 and these three DNAs were similar to that of the interaction with [G8A]MBS-I, indicating that this R3-G7-R3 protein lacked the ability of specific DNA binding. The binding kinetics of these interactions, the k_{off} and k_{on} values, could not be determined accurately using the BIAcore, since their sensorgrams were poorly defined, due to their weak interaction. On the other hand, the binding affinities between R2R3* and these DNAs were reduced, as a result of the increased k_{off} values (Table 1), similar to the non-specific DNA binding of R2R3* observed previously (17).

DISCUSSION

We investigated the possibility of creating a new DNA-binding protein, using only the R3 repeat, and the requirement for cooperativity in specific DNA recognition by c-Myb. If the cooperativity were not essential and only the effective concentration of the linked two domains were required for the specific DNA bind-

TABLE 1

Kinetic and Association Constants of R2R3* Binding to DNAs

DNA	k_{off} (s^{-1})	k_{on} (M^{-1} s^{-1})	K_a (M^{-1})	$K_{a,eq}$ (M^{-1})
MBS-I[a]	$2.6 \pm 0.1 \times 10^{-3}$	$2.3 \pm 0.5 \times 10^{5}$	8.9×10^{7}	1.0×10^{8}
[G8A]MBS-I[a]	$2.0 \pm 0.2 \times 10^{-2}$	$2.2 \pm 0.2 \times 10^{5}$	1.1×10^{7}	6.6×10^{6}
[G8A(s − 1)]MBS-I[b]	$1.8 \pm 0.2 \times 10^{-2}$	$2.3 \pm 0.2 \times 10^{5}$	1.3×10^{7}	5.0×10^{6}
[G8A(s + 2)]MBS-I[b]	$1.8 \pm 0.2 \times 10^{-2}$	$2.3 \pm 0.2 \times 10^{5}$	1.3×10^{7}	5.0×10^{6}
[G8A(s + 4)]MBS-I[b]	$2.4 \pm 0.3 \times 10^{-2}$	$2.3 \pm 0.1 \times 10^{5}$	9.8×10^{6}	3.2×10^{6}

[a] Data were taken from Oda *et al.* (17).

[b] k_{off}, k_{on}, and $K_{a,eq}$ are the average of four experiments for the protein concentrations from 50 to 400 nM.

ing (20), then the artificial R3-G7-R3 protein would be able to recognize two AAC base sequences separated by several base pairs at the center. However, the R3-G7-R3 protein could not specifically bind to the DNAs used in this study, although each R3 was connected with the flexible linker, to accommodate unexpected conformational changes in both the protein and the DNA (14, 15, 20). These results strongly suggest that cooperativity is required for the R3 recognition, as seen in the c-Myb R2R3-DNA complex (7).

Williams and Grotewold showed that a chimeric Myb domain with a novel DNA binding specificity can be created by combining plant and animal Myb domains, the R2 of P and the R3 of v-Myb, although the opposite chimeric Myb domain, containing the R2 of v-Myb and the R3 of P, lacks DNA binding activity (9). The sequence of v-Myb is very similar to that of c-Myb, with only one residue (Val181) different between the recognition helices. They precisely investigated the roles of the different amino acid residues in the recognition helices between v-Myb and P. The substitution in v-Myb R2 of Glu132 with Leu had no significant effect on the binding to MBS-I, while the simultaneous substitutions of Ala180, Val181, and His184 in v-Myb R3 caused a complete loss of specific binding. Additionally, Solano *et al.* showed the same substitution of Glu132 in c-Myb R2 with Leu can gain the MYB.Ph3 specificity, and the reciprocal substitution can also gain the c-Myb specificity (13). Since the complex structure of the plant Myb domain with its target DNA has not been determined yet, no one knows whether the residue at position 132 in the plant Myb domains is engaged in the cooperativity and the base recognition, although the residue is fully conserved as leucine among the known sequences (9). It can be speculated that the cooperative interactions between the adjacent two repeats might not be required for the specific DNA binding by the plant Myb, and the DNA recognition mechanism between the animal and plant Myb domains might be different.

The elucidation of the principles of the DNA recognition mechanism and the design of new DNA-binding proteins are attractive challenges for the future. Structure-based design has been used to link zinc fingers with other DNA-binding domains, and has succeeded in increasing the DNA binding affinity and specificity (18–20). Even for the zinc finger proteins, the X-ray crystal structures of the variants selected using phage display have shown that several contacts between the protein and the DNA were not predicted by the previously determined structure of the wild-type protein (22). c-Myb has unique features in its DNA recognition, which are not only the DNA contacts but also the cooperativity between the two repeats. The design of new c-Myb R2R3 proteins will be difficult, but the attempts will provide useful knowledge for understanding of the DNA recognition mechanism.

ACKNOWLEDGMENTS

The original pact-c-*myb* and pAR2156NcoI plasmids were kindly provided by Dr. Shunsuke Ishii of Tsukuba Life Science Center, the Institute of Physical and Chemical Research. The authors thank Dr. Kazuhiro Ogata of Kanagawa Academy of Science and Technology (KAST), and Dr. Yoshifumi Nishimura of Yokohama City University for helpful discussions.

REFERENCES

1. Biedenkapp, H., Borgmeyer, U., Sippel, A. E, and Klempnauer, K.-H. (1988) *Nature* **335,** 835–837.
2. Weston, K. (1992) *Nucleic Acids Res.* **20,** 3043–3049.
3. Tanikawa, J., Yasukawa, T., Enari, M., Ogata, K., Nishimura, Y., Ishii, S., and Sarai, A. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 9320–9324.
4. Gonda, T. J., Gough, N. M., Dunn, A. R., and de Blaquiere, J. (1985) *EMBO J.* **4,** 2003–2008.
5. Klempnauer, K.-H., and Sippel, A. E. (1987) *EMBO J.* **6,** 2719–2725.
6. Sakura, H., Kanei-Ishii, C., Nagase, T., Nakagoshi, H., Gonda, T. J., and Ishii, S. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 5758–5762.
7. Ogata, K., Morikawa, S., Nakamura, H., Sekikawa, A., Inoue, T., Kanai, H., Sarai, A., Ishii, S., and Nishimura, Y. (1994) *Cell* **79,** 639–648.
8. 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.
9. Williams, C. E., and Grotewold, E. (1997) *J. Biol. Chem.* **272,** 563–571.
10. Thompson, M. A., and Ramsay, R. G. (1995) *BioEssays* **17,** 341–350.
11. Grotewold, E., Drummond, B. J., Bowen, B., and Peterson, T. (1994) *Cell* **76,** 543–553.
12. Solano, R., Nieto, C., Avila, J., Cañas, L., Diaz, I., and Paz-Ares, J. (1995) *EMBO J.* **14,** 1773–1784.
13. Solano, R., Fuertes, A., Sánchez-Pulido, L., Valencia, A., and Paz-Ares, J. (1997) *J. Biol. Chem.* **272,** 2889–2895.
14. Oda, M., Furukawa, K., Ogata, K., Sarai, A., Ishii, S., Nishimura, Y., and Nakamura, H. (1997) *J. Biol. Chem.* **272,** 17966–17971.
15. Oda, M., Furukawa, K., Ogata, K., Sarai, A., Ishii, S., Nishimura, Y., and Nakamura, H. (1997) *Protein Eng.* **10,** 1407–1414.
16. Oda, M., Furukawa, K., Ogata, K., Sarai, A., and Nakamura, H. (1998) *J. Mol. Biol.* **276,** 571–590.
17. Oda, M., Furukawa, K., Sarai, A., and Nakamura, H. (1999) *FEBS Lett.,* in press.
18. Pomerantz, J. L., Sharp, P. A., and Pabo, C. O. (1995) *Science* **267,** 93–96.
19. Kim, J.-S., Kim, J., Cepek, K. L., Sharp, P. A., and Pabo, C. O. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 3616–3620.
20. Kim, J.-S., and Pabo, C. O. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 2812–2817.
21. Sarai, A., Uedaira, H., Morii, H., Yasukawa, T., Ogata, K., Nishimura, Y., and Ishii, S. (1993) *Biochemistry* **32,** 7759–7764.
22. Elrod-Erickson, M., Benson, T. E., and Pabo, C. O. (1998) *Structure* **6,** 451–464.