Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 6428–6432, July 1992
Biochemistry

Solution structure of a DNA-binding unit of Myb: A helix–turn–helix-related motif with conserved tryptophans forming a hydrophobic core

(nuclear oncogenes/transcriptional activator/NMR/distance geometry)

KAZUHIRO OGATA*†, HIRONOBU HOJO‡, SABURO AIMOTO‡, TAKAHISA NAKAI§, HARUKI NAKAMURA§, AKINORI SARAI¶, SHUNSUKE ISHII¶, AND YOSHIFUMI NISHIMURA*‖

*Graduate School of Integrated Science, Yokohama City University, Seto, Kanazawa-ku, Yokohama 236, Japan; †Department of Biochemistry, School of Medicine, Yokohama City University, Fukuura, Kanazawa-ku, Yokohama 236, Japan; ‡Institute for Protein Research, Osaka University, Yamadaoka, Suita, Osaka 565, Japan; §Protein Engineering Research Institute, Furuedai, Suita, Osaka 565, Japan; and ¶Tsukuba Life Science Center, Institute of Physical and Chemical Research, Tsukuba, Ibaraki 305, Japan

Communicated by Donald M. Crothers, March 31, 1992

ABSTRACT **The DNA-binding domain of the c-*myb* protooncogene product consists of three imperfect tandem repeats of 51 or 52 amino acids, each of which contains three conserved tryptophans, spaced 18 or 19 amino acids apart. The structure of the third repeat, which is essential for sequence-specific DNA binding, has been determined by NMR with distance geometry calculation. It includes three well-defined helices (residues 149–162, 166–172, and 178–187) maintained by a hydrophobic core that includes the three conserved tryptophans, together with two histidines. Helices 2 and 3 form a structure related to but distinct from a canonical helix–turn–helix motif. In particular, the turn between these helices is one amino acid longer than the corresponding turn in bacterial repressors and homeodomains and contains a proline residue. In addition, the architecture of the three helices is different from those of homeodomains and DNA-binding domains of bacterial repressors. Based on the present structure, the binding mode of Myb repeat 3 with a specific DNA is also discussed.**

The c-*myb* protooncogene product (c-Myb) binds to DNA in a sequence-specific manner and acts as a transcriptional regulator (1–5). The DNA-binding domain of c-Myb consists of three imperfect tandem repeats of 51 or 52 amino acids (6–8), which are reported to have a similarity with the homeodomain (9). It has been suggested that the third repeat recognizes the specific base sequence, whereas the first repeat cooperates to bind to DNA (8, 10–12). The role of the second repeat is unclear. It was suggested that the second repeat could also participate in recognition of a specific base sequence by assuming that the second and third repeats take a similar structure related to the helix–turn–helix (HTH) motif (12). However, no three-dimensional structures of these repeats have been established. In addition, each repeat contains three conserved tryptophans, spaced 18 or 19 amino acids apart, which play a critical role in sequence-specific DNA binding (10, 11, 13). A conserved triplet of tryptophans with spacing similar to that of c-Myb is also found in the DNA-binding domains of the products of *ets* gene family. This indicates that these conserved tryptophans may represent a characteristic property of a group of DNA-binding proteins. To clarify the role of three tryptophans as well as to find the exact role of three repeats, we have determined the three-dimensional structure of the third repeat in aqueous solution by NMR.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

MATERIALS AND METHODS

Sample Preparation. A peptide corresponding to the third repeat of the DNA-binding domain of mouse c-Myb, amino acids 142–193, was chemically synthesized and purified by reversed-phase high-performance liquid chromatography as described (14). The purity of the synthesized c-Myb-(142–193)-NH_2 was examined by ion-exchange chromatography as well as by amino acid composition analysis; no contamination by side products was detected (14).

NMR Measurements. The lyophilized sample was dissolved in a mixture of either 90% H_2O/10% ^2H_2O (1.8 mM peptide/150 mM potassium phosphate buffer/0.4 mM NaN_3, pH 6.76) or 99.6% ^2H_2O (1.9 mM peptide/150 mM potassium phosphate buffer/0.4 mM NaN_3, p^{2}H 6.86), and NMR spectra were recorded at 500 MHz on a Bruker AMX-500 spectrometer. In ^2H_2O solution at 27°C the spectra of double quantum filtered correlated spectroscopy (DQF-COSY) (15), total correlated spectroscopy (TOCSY) (16) with mixing times of 30, 50, and 89 ms, and nuclear Overhauser effect spectroscopy (NOESY) (17) with mixing times of 50, 150, and 200 ms were measured. In H_2O solution DQF-COSY, NOESY with mixing times of 50 and 150 ms, and TOCSY with a 101-ms mixing time were measured. Each measurement was performed with 2048 data points along the t_2 dimension over 8-kHz spectral width and 512 points along the t_1 dimension. After data acquisition, zero-filling along t_1 was performed to produce 1024 data points. The number of scans per t_1 value was 64 or 128. A window function (phase-shifted squared sine bell) was applied before two-dimensional Fourier transformation. The water signal was suppressed by presaturation. Using the jump-and-return method with a homo spoil pulse, a NOESY spectrum in H_2O solution (150-ms mixing time) was also measured. Baseline correction of all columns was performed in each case. The appropriate baseline was determined by using a statistical method and then matched to the polynomial of degree 3. In addition, a highly resolved DQF-COSY spectrum in H_2O solution with 8192 data points along the t_2 dimension over 6 kHz was also measured.

Calculation Procedures. Embedded structures with 673 atoms with a pseudoatom representation were obtained by the vectorized embedding program EMBOSS (data not shown). Of these, structures whose distance violations were small were subjected to 1000 steps of unrestrained energy minimi-

Abbreviations: HTH, helix–turn–helix; MD, molecular dynamics; RMSD, root-mean-square-deviation; DQF-COSY, double quantum filtered correlated spectroscopy; TOCSY, total correlated spectroscopy; NOESY, nuclear Overhauser effect spectroscopy.

‖To whom reprint requests should be addressed.

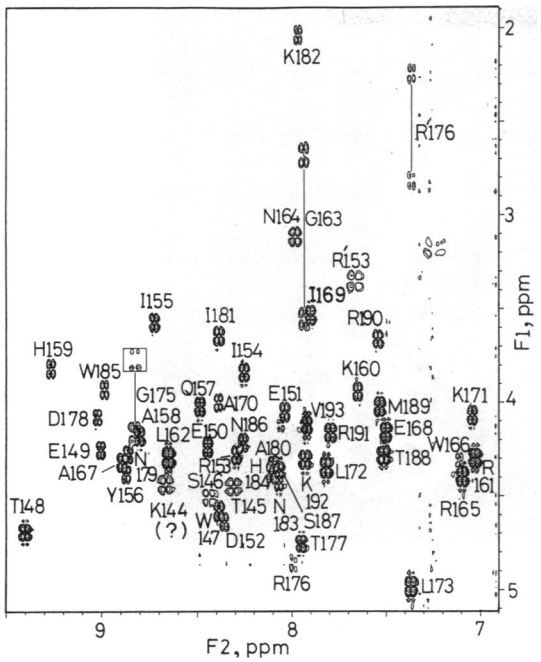

FIG. 1. Part of a DQF-COSY spectrum of Myb repeat 3 in H_2O at 300 K showing the through-bond connectivity between amide and CαH protons of each residue. At each cross-peak position, the amino acid number corresponding to mouse c-Myb is indicated. In addition, the cross-peaks between NεH and CδH_2 protons for arginines are shown as R'.

zation with low repulsive forces (k_{rep} = 0.15 kcal/mol per Å^4; 1 cal = 4.18 J) instead of the 6–12 Lennard–Jones nonbonded interaction to remove contacts that are too close. Next, 3-ps temperature-regulated molecular dynamics (MD) calculations with a time step of 0.5 fs were carried out to raise the temperature from 0 K to 3000 K. During MD, k_{rep}, and the weights of the distance restraint, k_{dist}, and the torsion angle restraints, k_{tors}, were increased from 0.15 to 5.4 kcal/mol per Å^4, from 0.15 to 6.0 kcal/mol per Å^4, and from 60 to 270 kcal/mol per rad^2, respectively. Another 2-ps MD slowly cooled the temperature to 300 K, using the values of the force constants and the restraint weights at 3000 K. After such simulated annealing, the structures were improved by 2000 steps of energy minimization with distance and torsion angle restraints, using the normal AMBER force field (18) with 6–12 Lennard–Jones potential instead of the repulsive forces. Of these refined structures, those whose total distance violations were <6.4 Å were retained. All energy minimizations and restrained MD calculations were performed using the program PRESTO (data not shown).

RESULTS AND DISCUSSION

NMR Analysis of Myb Repeat 3. Sequence-specific ^1H NMR assignments were obtained following the conventional strategy (19, 20). Spin systems of the individual amino acids were identified in a ^2H_2O solution at 27°C using DQF-COSY, TOCSY, and NOESY spectra. Corresponding spectra in a H_2O solution were used to detect exchangeable protons. All observable protons in individual amino acids (except for NH_2 and NεH protons of arginine and NH_2 protons of lysine) were identified for 50 of 52 residues. Using the NOESY spectrum with a 150-ms mixing time in a H_2O solution, sequence-specific assignments were completed.

Fig. 1 shows the fingerprint region of the DQF-COSY spectrum in H_2O. The amino acid number indicated here corresponds to mouse c-Myb. Fig. 2 shows sequential and short-range NOE connectivities in the peptide. The intense sequential NOEs between amide protons as well as the connectivities of dαN(i, i + 3) and dαN(i, i + 4) indicate that three helical structures are present in the peptide.

Distance Geometry Calculations of Myb Repeat 3. Intensities of NOEs between backbone protons were classified as strong, medium, or weak, assumed to represent distance ranges of 2–3 Å, 2–4 Å, and 2–5 Å, respectively. Intensities between side chain protons or backbone and side chain protons were classified as strong or weak, assumed to represent distances of 2–4 Å and 2–5 Å, respectively. A set of 496 distance constraints was derived from NOEs (170 sequential, 172 medium-range, and 154 long-range NOEs). In addition, $^3J_{NH\alpha}$ coupling constants were measured, and some Cβ methylene protons were assigned stereospecifically; these provided 35 ϕ and 17 χ^1 angle constraints (Fig. 2).

At the first stage of the distance geometry calculation, geometrically consistent structures were generated by the conventional metric matrix algorithm (21). From 150 embedded structures, 45 structures were selected and then subjected to the simulated annealing procedure (22). Of the 45 refined structures, 29 structures were selected whose total distance violations were <6.4 Å. The conformation of the

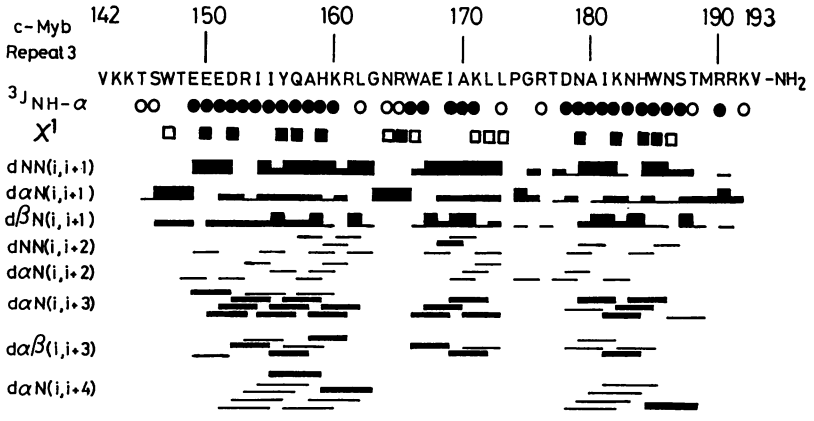

FIG. 2. Sequence of Myb repeat 3 together with a summary of short-range ($|i - j| \leq 4$) NOEs, $^3J_{NH\text{-}\alpha}$ coupling constants, and χ^1 angle ranges. Myb repeat 3 corresponds to amino acid sequence 142–193 of mouse c-Myb as indicated by the number above the sequence. The presence of a certain NOE is indicated by a line connecting the two residues. The thickness of the lines corresponds to the intensity of the NOE. $^3J_{NH\text{-}\alpha}$ coupling constants <7 Hz and >8.5 Hz are indicated by closed and open circles, respectively. For χ^1 angle ranges, closed and open squares correspond to 180° ± 60° and −60° ± 60°, respectively. As NMR signals of the N-terminal two lysines at positions 143 and 144 were not well identified in the spectra examined so far (Fig. 1), the N-terminal three residues could not be connected sequentially. In the case of Pro-174 the sequential connectivity with CαH of Leu-173 is to δCH_2, which shows that this peptide bond is the trans form.

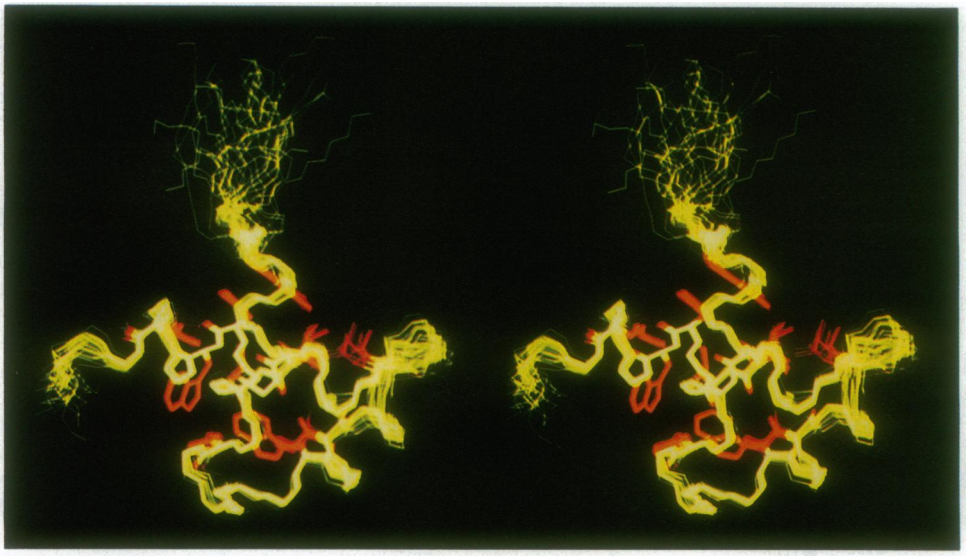

FIG. 3. Stereo view of backbones (yellow) with amino acid side chains in a hydrophobic core (red) of the superimposed 29 solutions from distance geometry calculations. Of the 45 refined structures, 29 structures were selected according to the low violations of the distance and dihedral angle restraints and the low conformation energies. Their total violations of the distance restraints were all below 6.4 Å, and any distance violations more than 0.5 Å did not exist. Their total violations of the dihedral restraints were below 8.2°, without any violations of more than 4°.

backbone of residues 147–189 was well defined by the NMR data. The average root-mean-square deviations (RMSDs) between 29 calculated structures are 0.45 Å for the backbone and 0.93 Å for all heavy atoms (residues 147–189). The average structure was further optimized by 2000 steps of energy minimization with distance and torsion angle restraints.

Three-Dimensional Structure of Myb Repeat 3. Fig. 3 shows the ensemble of 29 calculated structures. It consists of three well-defined helices (residues 149–162, 166–172, and 178–187) and two less-well-defined turns (residues 163–165 and 173–177). Fig. 4 shows the final calculated structure. The three helices define a hydrophobic core formed by Trp-147, Ile-154, Ile-155, His-159, Trp-166, Ile-169, Leu-173, Ile-181, His-184, Trp-185, and Met-189. Of these amino acids, the three tryptophans and two histidines seem to play a crucial role; His-159, Trp-166, and Trp-185 are in contact with each other in the core, and Trp-147 (in the N-terminal region) participates through a stacking interaction with His-184 in helix 3.

HTH-Related Structure. Helices 2 and 3 appear to form a HTH structure. Superimposing this part on the canonical HTH motifs of 434 cro (23) and the engrailed homeodomain (24) reveals that these structures are related, as shown in Fig. 5. RMSDs between the Cα coordinates of 21-residue segments of Myb repeat 3 and of several classical HTH proteins range from 1.4 to 1.7 Å. However, within this group of HTH proteins, RMSD values were calculated to be within 1.0 Å (25, 26). As a control, RMSDs between the HTH motif and other protein fragments of the same length taken at random from the protein databank were mostly larger than 2 Å and rarely around 1.6 Å (25). Thus, we conclude that the structure of Myb repeat 3 is closely related to, but distinct from, the canonical HTH motif. This difference is caused by (*i*) the turn, which contains a proline residue and is one residue longer, and (*ii*) a difference in the relative orientation of helices 2 and 3 (Fig. 5). The overall structure of Myb repeat 3 is significantly different from those of the homeodomain and the DNA-binding domain of 434 cro. In Myb repeat 3, helices 1 and 2 are aligned obliquely with a short turn and helix 3 is arranged almost perpendicular to helix 1.

One of the characteristics that distinguishes Myb from canonical HTH proteins is the presence of three tandemly repeated domains. Comparison of the amino acid sequences of repeat 3 with those of repeat 1 or 2 shows that the two histidines in the hydrophobic core in repeat 3 (see above) are not conserved in repeats 1 and 2; the first histidine is replaced by valine and second is replaced by arginine. To examine the

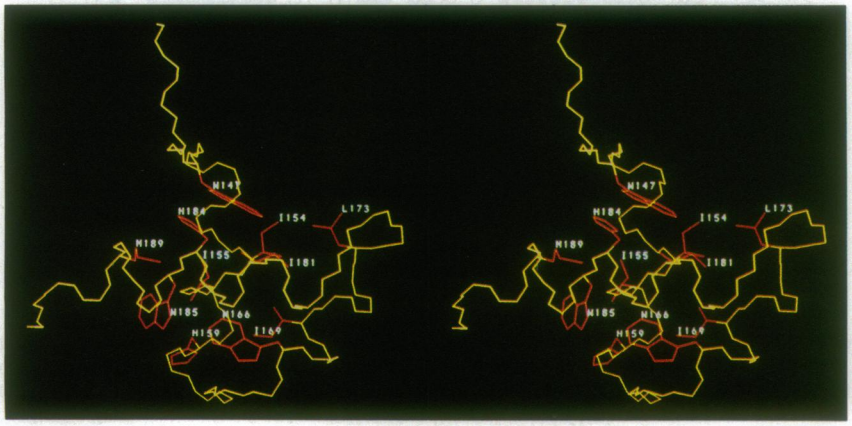

FIG. 4. Stereo view of the final structure of Myb repeat 3 of the backbone atoms (yellow) and amino acid side chains that make a hydrophobic core (red).

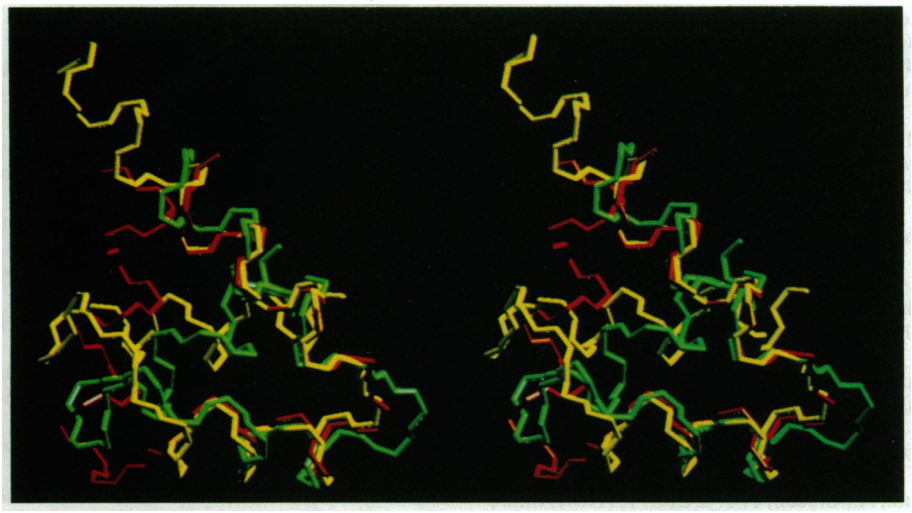

FIG. 5. Stereo view of the structure of Myb (green), 434 cro (red), and homeodomain (yellow). Only backbone structures are shown. The Myb backbone structure was optimized after averaging the 29 calculated structures in Fig. 3. N- and C-terminal residues outside helices are omitted. These structures were fitted with each other only inside the 21-residue HTH regions. The RMSDs for the Cα coordinates of the 21-residue segment between Myb and 434 cro and repressor are both 1.4 Å.

effects of such substitutions, we built models of repeats 1 and 2 based on the NMR structure of repeat 3 using computer graphic modeling. Comparison of the structures of the predicted hydrophobic cores of the three repeats suggests that repeats 1 and 2 would be less stable than repeat 3. Preliminary NMR analysis and calorimetric measurements of a fragment containing all three repeats have indicated that in the fragment repeat 3 is most stable and appears to take the same conformation as an isolated peptide.

Predicted Binding Mode of Repeat 3 with DNA. The present structure of repeat 3 supports the idea that the helix 3 in repeat 3 recognizes the "core" AAC nucleotide sequence of the consensus Myb-binding site (10, 11). Fig. 6 shows a model

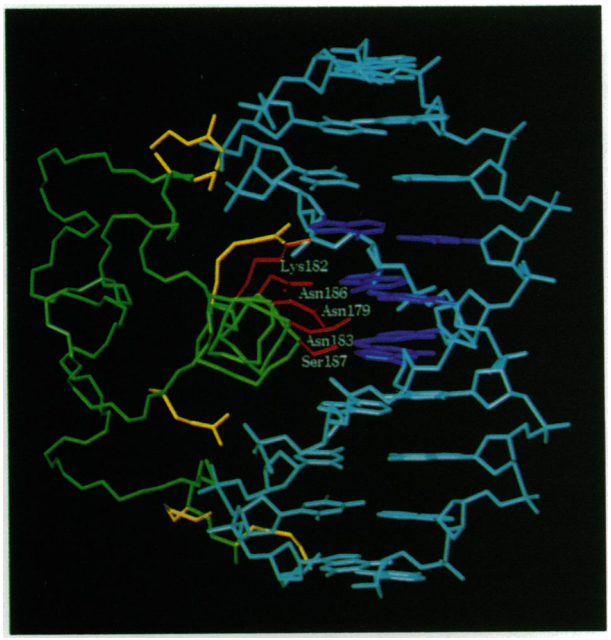

FIG. 6. Model of the DNA-binding mode for Myb repeat 3. DNA is colored in cyan except for the core AAC sequence, which is shown in blue. The Myb backbone is shown in green. Yellow-colored residues, Arg-165, Asn-164, Arg-190, Arg-176, Lys-144, and Lys-143 (from top to bottom), are likely to interact with phosphate backbones of DNA. On the other hand, red-colored residues, Asn-179, Lys-182, Asn-183, Asn-186, and Ser-187 from helix 3, are presumed to interact with the core AAC sequence of the consensus Myb-binding site.

of interaction between Myb repeat 3 and a DNA fragment containing the AAC sequence. The model structure was built by superimposing the Myb HTH unit and DNA onto the cocrystal structure of 434 repressor (27) and modifying the Myb position to obtain better interactions between amino acids and base pairs. This model structure suggests several possible interactions between the DNA backbone and specific amino acids. These amino acids can form a firm grip on the DNA, putting helix 3 in the right place into the major groove. In helix 3, Asn-179, Lys-182, Asn-183, Asn-186, and Ser-187 are facing DNA and can make contact with base pairs. Asparagines can form double hydrogen bonds with N6 and N7 of adenosine residues, as observed in the cocrystal structures of DNA fragments and λ repressor (28) or the homeodomain (24), and, thus, are good candidates for forming specific interactions with either A of the AAC sequence. This is also supported by an earlier experiment in which the replacement of Asn-186 with aspartic acid or Lys-182 with glutamic acid resulted in a loss of specific DNA-binding activity (11).

The role of repeat 1 in specific binding to DNA should be insignificant, since repeat 1 can be deleted without serious effects on the overall binding (7, 8). The role of repeat 2 is unclear. It was shown that the replacement of Glu-129 or Asn-136 in repeat 2 with alanine residue eliminated specific DNA-binding activity as well as the replacement of Asn-183 or Asn-186 in repeat 3 (12); these authors suggested that a HTH-related motif exists in both repeats of 2 and 3. At least for repeat 3, this is established in the present structural study. However, repeat 2 is likely to take a different conformation from a HTH-related motif as mentioned above. Thus, the replacement of Glu-129 or Asn-136 with an alanine residue disrupts a characteristic conformation of repeat 2 or forms aggregates, causing DNA-binding loss.

In this study, NMR analysis revealed that the conserved tryptophans in the Myb repeat 3 contribute to a hydrophobic core in the DNA-binding structure. As with other Myb-related proteins, Ets-related proteins also contain three conserved tryptophans with a spacing similar to that of c-Myb. Therefore, the three-dimensional structure obtained here may also help to elucidate how this group of proteins binds to DNA and regulates transcription.

We thank Drs. T. J. Gonda and R. G. Ramsay for critical reading of the manuscript and Dr. C. O. Pabo for giving us the three-

dimensional coordinates of the engrailed homeodomain structure. This work was supported by Yokohama City University and special grants-in-aid (02263104 and 03259104) from the Ministry of Education, Science and Culture of Japan.

1. Biedenkapp, H., Borgmeyer, U., Sippel, A. E. & Klempnauer, K.-H. (1988) *Nature (London)* **335,** 835–837.
2. Nakagoshi, H., Nagase, T., Kanei-Ishii, C., Ueno, Y. & Ishii, S. (1990) *J. Biol. Chem.* **265,** 3479–3483.
3. Nishina, Y., Nakagoshi, H., Imamoto, F., Gonda, T. J. & Ishii, S. (1989) *Nucleic Acids Res.* **17,** 107–117.
4. Weston, K. & Bishop, J. M. (1989) *Cell* **58,** 85–93.
5. Ness, S. A., Marknell, A. & Graf, T. (1989) *Cell* **59,** 1115–1125.
6. Klempnauer, K.-H. & Sippel, A. E. (1987) *EMBO J.* **6,** 2719–2725.
7. Sakura, H., Kanei-Ishii, C., Nagase, T., Nakagoshi, H., Gonda, T. J. & Ishii, S. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 5758–5762.
8. Howe, K. M., Reakes, C. F. L. & Watson, R. J. (1990) *EMBO J.* **9,** 161–169.
9. Frampton, J., Leutz, A., Gibson, T. J. & Graf, T. (1989) *Nature (London)* **342,** 134 (lett.).
10. Kanei-Ishii, C., Sarai, A., Sawazaki, T., Nakagoshi, H., He, D.-N., Ogata, K., Nishimura, Y. & Ishii, S. (1990) *J. Biol. Chem.* **265,** 19990–19995.
11. Saikumar, P., Murali, R. & Reddy, E. P. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 8452–8456.
12. Gabrielsen, O. S., Sentenac, A. & Fromageot, P. (1991) *Science* **253,** 1140–1143.
13. Anton, I. A. & Frampton, J. (1988) *Nature (London)* **336,** 719 (lett.).
14. Hojo, H. & Aimoto, S. (1991) *Bull. Chem. Soc. Jpn.* **64,** 111–117.
15. Rance, M., Sorensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R. & Wuthrich, K. (1983) *Biochem. Biophys. Res. Commun.* **117,** 479–485.
16. Bax, A. & Davis, D. G. (1985) *J. Magn. Reson.* **65,** 355–366.
17. Kumar, A., Ernst, R. R. & Wuthrich, K. (1980) *Biochem. Biophys. Res. Commun.* **95,** 1–6.
18. Weiner, S. J., Kollman, P. A., Nguyen, D. T. & Case, D. A. (1986) *J. Comp. Chem.* **7,** 230–252.
19. Wuthrich, K. (1986) *NMR of Proteins and Nucleic Acids* (Wiley, New York).
20. Englander, S. W. & Wand, A. J. (1987) *Biochemistry* **26,** 5953–5958.
21. Crippen, G. M. (1981) *Distance Geometry and Conformational Calculations*, Chemometrics Research Studies Series, ed. Bawden, D. (Wiley, New York), Vol. 1.
22. Driscoll, P. C., Gronenborn, A. M., Beress, L. & Clore, G. M. (1989) *Biochemistry* **28,** 2188–2198.
23. Mondragon, A., Wolberger, C. & Harrison, S. C. (1989) *J. Mol. Biol.* **205,** 179–188.
24. Kissinger, C. R., Liu, B., Martin-Blanco, E., Kornberg, T. B. & Pabo, C. O. (1990) *Cell* **63,** 579–590.
25. Qian, Y. Q., Billeter, M., Otting, G., Muller, M., Gehring, W. J. & Wuthrich, K. (1989) *Cell* **59,** 573–580.
26. Brennan, R. G. & Matthews, B. W. (1989) *J. Biol. Chem.* **264,** 1903–1906.
27. Aggarwal, A. K., Rodgers, D. W., Drottar, M., Ptashne, M. & Harrison, S. C. (1988) *Science* **242,** 899–907.
28. Jordan, S. R. & Pabo, C. O. (1988) *Science* **242,** 893–899.