Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 8452–8456, November 1990
Biochemistry

Role of tryptophan repeats and flanking amino acids in Myb–DNA interactions

(nuclear oncogenes/DNA binding/mutagenesis)

POTHANA SAIKUMAR, RAMACHANDRAN MURALI, AND E. PREMKUMAR REDDY*

The Wistar Institute of Anatomy and Biology, 3601 Spruce Street, Philadelphia, PA 19104

Communicated by Herbert Weissbach, August 6, 1990 (received for review June 7, 1990)

ABSTRACT **The c-*myb* protooncogene codes for a sequence-specific DNA-binding protein that appears to act as a transcriptional regulator and is highly conserved through evolution. The DNA-binding domain of Myb has been shown to contain three imperfectly conserved repeats of 52 amino acids that constitute the amino-terminal end. Within each repeat, there are three tryptophans that are separated by 18 or 19 amino acids and are flanked by basic amino acids. To determine the role of tryptophans and the flanking basic amino acids in the DNA-binding activity of Myb proteins, we have selectively mutagenized individual tryptophans as well as some of the amino acid residues that flank these tryptophans. Replacement of these tryptophans with glycine, proline, or arginine abolished the DNA-binding activity whereas replacement with other aromatic amino acids or leucine or alanine did not appreciably affect this activity. On the other hand the replacement of two amino acids, asparagine and lysine, that flank the last tryptophan with acidic amino acids completely abolished their DNA-binding activity. These results are consistent with a model we present in which the tryptophans form a hydrophobic scaffold that plays a crucial role in maintaining the helix-turn-helix structure of the DNA binding domain. Basic and polar amino acids adjacent to these tryptophans seem to participate directly in DNA binding.**

The transforming gene of avian myeloblastosis virus, v-*myb*, codes for a nuclear protein that binds DNA in a sequence-specific manner and seems to function as a transcriptional transactivator (1, 2). The normal cellular counterpart of this oncogene c-*myb* is expressed predominantly in immature hematopoietic cells and codes for at least two translational products of 75 kDa and 89 kDa (3). The c-*myb* proteins, like that of v-*myb*, are localized in the nucleus and bind DNA in a sequence-specific manner and appear to function as nuclear transcription factors (2, 4). c-*myb* is highly conserved through evolution, and homologs of this gene have been detected in *Drosophila*, *Zea mays* (corn), and yeast in addition to avian and mammalian species (5–10). When the *myb* sequences from various species are compared (Fig. 1), it becomes apparent that the highly conserved region is the amino-terminal region that contains three imperfectly conserved tandem repeats of 52 amino acids that have been found to mediate the DNA-binding activity of the protein. These repeats have several interesting features, including the presence of nine tryptophans that are highly conserved and are spaced 18 or 19 amino acids apart (ref. 11 and Fig. 1) and appear to be part of a helix-turn-helix motif (12). In addition, each tryptophan residue has been found to be flanked by basic and polar amino acids that are also conserved through evolution (Fig. 1). To examine the role of tryptophans and the flanking amino acids in the DNA-binding activity of Myb proteins, we have selectively mutagenized individual tryptophans, as well as some of the amino acid residues that flank these tryptophans, and examined the ability of the mutant proteins to bind to DNA in a sequence-specific manner.

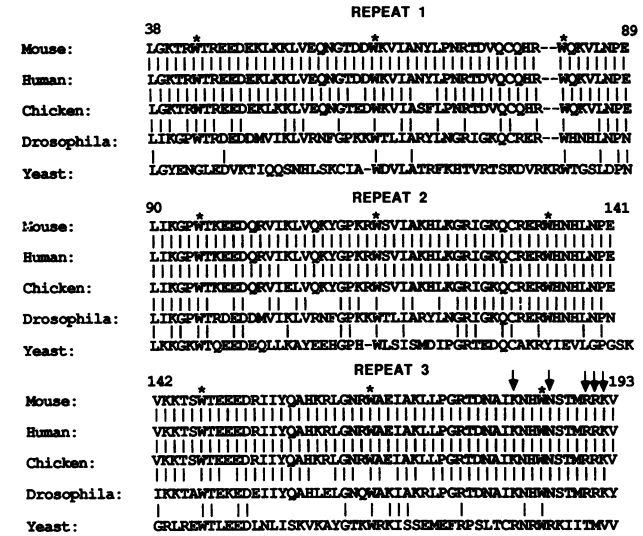

FIG. 1. Comparison of the amino acid sequence of the DNA-binding domains of *myb* gene homologs of mice, humans, chickens, *Drosophila*, and yeast (5–10). The numbers on top indicate the position of amino acids from the amino-terminal end of mouse c-Myb protein (9). The vertical bars denote the homology among various species. Asterisks indicate the position of tryptophans. The arrows indicate the position of amino acids that have been mutagenized in this study.

MATERIALS AND METHODS

Construction of c-myb, t-myb, and Δ-myb Expression Vectors. The c-myb cDNA was cloned into Bluescript vector (Stratagene) such that it can be transcribed using the T7 RNA polymerase. A truncated version of this protein, which lacks the carboxyl-terminal 240 amino acids, is present in NFS-60 cells (13, 14). A clone that expresses this protein was generated by site-directed mutagenesis from the total c-myb cDNA and was designated t-myb. A third mutant that lacks 311 amino acids from the carboxyl-terminal end was generated by digestion with *Sma* I and was designated as Δ-myb-1. Δ-myb-2 was derived from a cDNA clone isolated from ABPL-2 tumor cell line that contains a rearranged locus and produces mRNA and protein that lack the first repeat at the amino-terminal end of c-*myb* (14, 15). Δ-myb-3 was generated by site-directed mutagenesis of Δ-myb-1 construct.

Site-Directed Mutagenesis. Site-directed mutagenesis for the creation of deletions and point mutations was carried out

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.

*To whom reprint requests should be addressed.

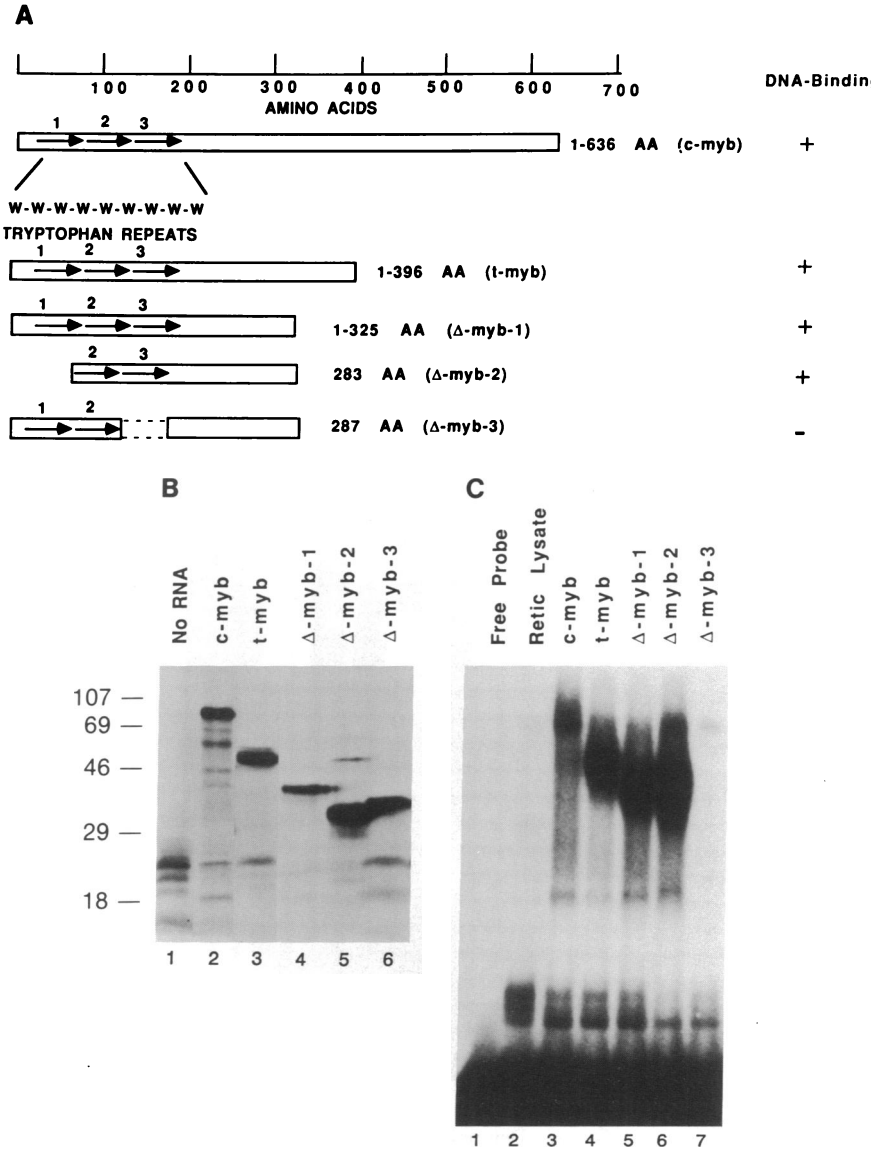

FIG. 2. Analysis of *myb* deletion mutants. (*A*) Structure of various deletion mutants used for DNA-binding activity. Deletions in the c-myb cDNA were made either by restriction digestion or by site-directed mutagenesis. t-myb lacks the region that codes for the carboxyl-terminal 240 amino acids. Δ-myb-1 was created by digestion of c-myb cDNA with *Sma* I. Δ-myb-2 was derived from a cDNA clone isolated from ABPL-2 tumor cell line, which lacks the first repeat at the amino-terminal end (14, 15). This cDNA was digested with *Sma* I to generate a truncated version of this cDNA. Δ-myb-3, which lacks the third repeat of c-*myb*, was generated by site-directed mutagenesis of Δ-myb-1 construct. AA, amino acids. (*B*) SDS/PAGE analysis of the reticulocyte translational products of c-*myb* and its deletion mutants. c-myb cDNA and its deletion mutants cloned into the Bluescript vector were used for *in vitro* transcription and translation in the presence of [^{35}S]methionine and the labeled translational products were resolved on a 12% polyacrylamide gel. Reticulocyte lysate incubated with no RNA (lane 1), c-myb RNA (lane 2), t-myb RNA (lane 3), Δ-myb-1 RNA (lane 4), Δ-myb-2 RNA (lane 5), and Δ-myb-3 RNA (lane 6) are shown. (*C*) Mobility-shift analysis of c-Myb and its deletion mutants with a 26-base-pair oligonucleotide derived from the *mim*-1 gene promoter (19). The sequence of the oligonucleotide used was 5′-TCGACACATTATAACGGTTTTTTA-GC-3′, where the Myb recognition sequence is underlined. The mobility-shift assays were performed as described (18). Lanes: 1, free probe; 2, probe incubated with control (incubated with no RNA) reticulocyte lysate; 3, probe incubated with c-Myb; 4, probe incubated with t-Myb; 5, probe incubated with Δ-Myb-1; 6, probe incubated with Δ-Myb-2; 7, probe with Δ-Myb-3.

using the polymerase chain reaction technique described by R. Higuchi *et al.* (16).

Transcription and Translation Reactions and Mobility-Shift Assays. The cDNA clones were transcribed using T7 RNA polymerase (17) and *in vitro* translaltion reactions were carried out using a rabbit reticulocyte translation kit supplied by Promega. The gel-shift assays were carried out as described by Singh *et al.* (18).

RESULTS AND DISCUSSION

DNA-Binding Properties of c-Myb and Truncated Myb Proteins. An *in vitro* DNA binding assay for c-Myb and truncated forms of Myb was developed using proteins that were expressed in an *in vitro* reticulocyte translation system. The structures of these proteins are shown in Fig. 2*A*. When a complete c-myb cDNA was used in this transcription–translation system, a 75-kDa form of the protein was synthesized, while the truncated forms of this cDNA (t-myb and Δ-myb-1, Δ-myb-2, and Δ-myb-3) generated proteins of 50 kDa, 40 kDa, 31 kDa, and 35 kDa, respectively (Fig. 2*B*). After the translation, we tested their ability to bind to synthetic oligonucleotides containing the Myb-binding motif. Ness *et al.* (19) identified a gene termed *mim*-1 that appears to be transcriptionally regulated by Myb and contains a Myb-binding motif, YAACG/TG (where Y is a pyrimidine) (1), in its promoter/enhancer region. An oligonucleotide derived from the *mim*-1 sequence, which was 26 base pairs long containing the Myb-binding motif, was used in these studies. With this oligonucleotide, we observed binding by the normal c-Myb as well as the truncated forms of this protein in a gel-shift assay, which is shown in Fig. 2*C*. In addition to the Myb-specific DNA complexes, several batches of reticulocyte lysates were found to contain various amounts of a factor that nonspecifically bound to the oligonucleotide used in this study and migrated at a position close to the c-Myb–DNA complexes (marked with a small arrow). To avoid confusion resulting from comigration of c-Myb with a nonspecific band, all additional experiments were carried out with truncated forms of Myb, whose DNA complexes migrated faster than the nonspecific band and so could be visualized readily.

The v-Myb protein as well as the aberrant Myb proteins produced by ABPL (Abelson virus-induced plasmacytoid lymphosarcomas) tumors contain only the second and third repeats at the amino-terminal end (14, 15). Since these proteins are known to bind DNA, it could be argued that only two of the three repeats are essential for the DNA-binding activity of this protein. To further examine the role of these repeats in DNA binding activity, we created mutants that

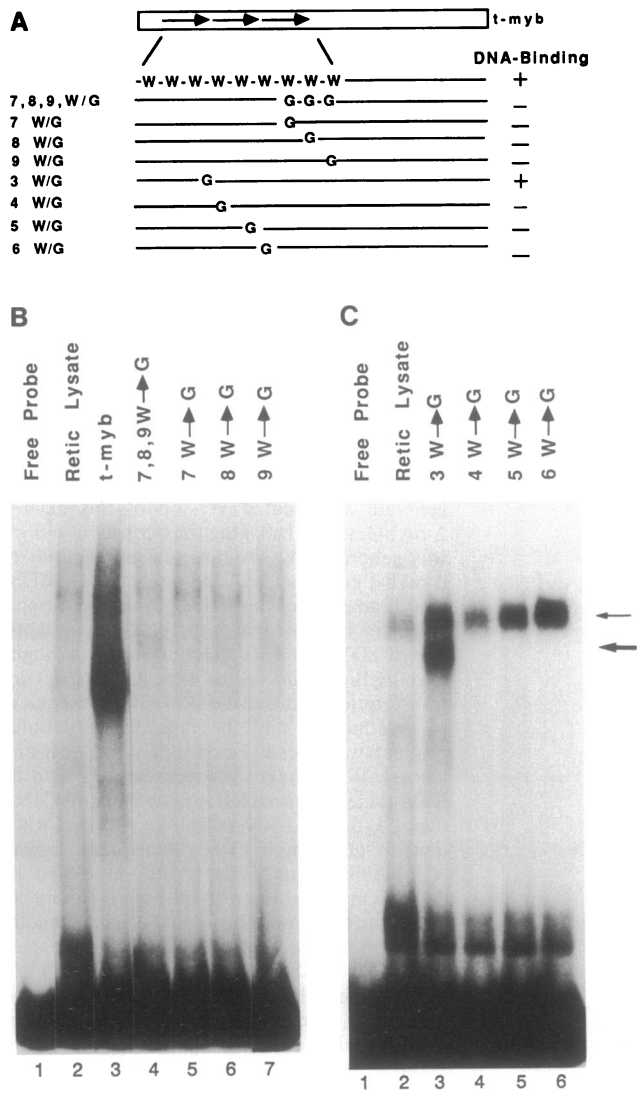

FIG. 3. Effect of amino acid substitutions in the tryptophan repeat of the DNA-binding domain of Myb. By using the t-myb subcloned into Bluescript vector, mutations were created as described by Higuchi *et al.* (16), and the structure of each mutant was confirmed by sequence analysis. (*A*) Schematic representation of *myb* mutants. (*B*) Mobility-shift analysis of *myb* mutants with Trp→Gly substitutions in the third repeat. The analysis was performed using the 26-base-pair oligonucleotide described in Fig. 2. Lanes: 1, free probe; 2, probe with control reticulocyte lysate; 3, probe with t-Myb; 4, probe with mutant 7,8,9 W/G; 5, probe with mutant 7 W/G; 6, probe with mutant 8 W/G; 7, probe with mutant 9 W/G. (*C*) Mobility-shift analysis of Myb mutants with Trp→Gly substitutions in the first and second repeats. Lanes: 1, free probe; 2, probe with control reticulocyte lysates; 3, probe with mutant 3 W/G; 4, probe with mutant 4 W/G; 5, probe with mutant 5 W/G; 6, probe with mutant 6 W/G. The Myb-specific band is shown by a large arrow and the nonspecific band is shown by a small arrow.

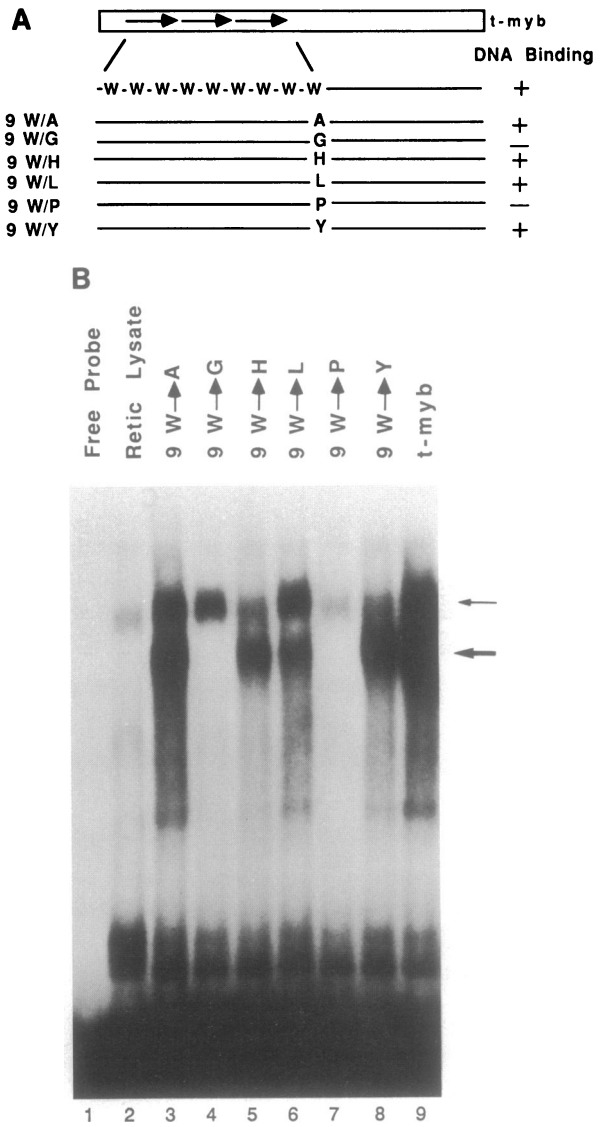

FIG. 4. Determination of the effect of replacement of tryptophans with other amino acids. The last tryptophan of the third repeat was replaced with various amino acids indicated in *A*, and the effect of these substitutions was evaluated on the DNA binding of the mutant proteins, which is shown in *B*. Lanes: 1, free probe; 2, probe with control reticulocyte lysate; 3, probe with mutant 9 W/A; 4, probe with mutant 9 W/G; 5, probe with mutant 9 W/H; 6, probe with mutant 9 W/L; 7, probe with mutant 9 W/P; 8, probe with mutant 9 W/Y; 9, probe with wild type t-Myb. The Myb-specific band is shown by a large arrow and the nonspecific band is shown by a small arrow.

lack one of the three repeats and examined their ability to bind to DNA. The results of these experiments (Fig. 2) demonstrate that while the deletion of the first repeat had little or no effect on the DNA-binding activity of the mutant protein, the deletion of the third repeat completely abolished the DNA-binding activity. These results are in agreement with the suggestion that the structural integrity of repeats 2 and 3 is an absolute requirement for the biochemical activity of this protein (20, 21).

Role of Tryptophan Repeats in the DNA-Binding Property of Myb. To examine the role of tryptophan residues within each repeat in DNA-binding activity of Myb, site-directed mutagenesis (16) was used to specifically mutate individual tryptophans. The nature of mutations introduced is shown in Fig. 3*A*, and their DNA-binding properties are summarized. Initially, the tryptophan residues were converted to glycines. The eight mutant *myb* gene products were expressed *in vitro* in the presence of [^{35}S]methionine and analyzed by SDS/PAGE. All the mutant forms used in this study expressed appropriately sized proteins (data not shown) and so were then utilized in a DNA-binding (gel shift) assay. For this assay, the oligonucleotide was end-labeled and the Myb proteins were unlabeled. When all three tryptophans of the third repeat were converted to glycines, the DNA-binding activity of the mutant protein was completely lost (Fig. 3*B*; lane 4). In addition, mutation of any one tryptophan in the second or third repeat resulted in a loss of binding activity (Fig. 3 *B* and *C*). The only exception was mutant 3W/G, which had a mutation of a tryptophan in repeat 1 and retained normal DNA-binding activity (Fig. 3*C*, lane 3). These results

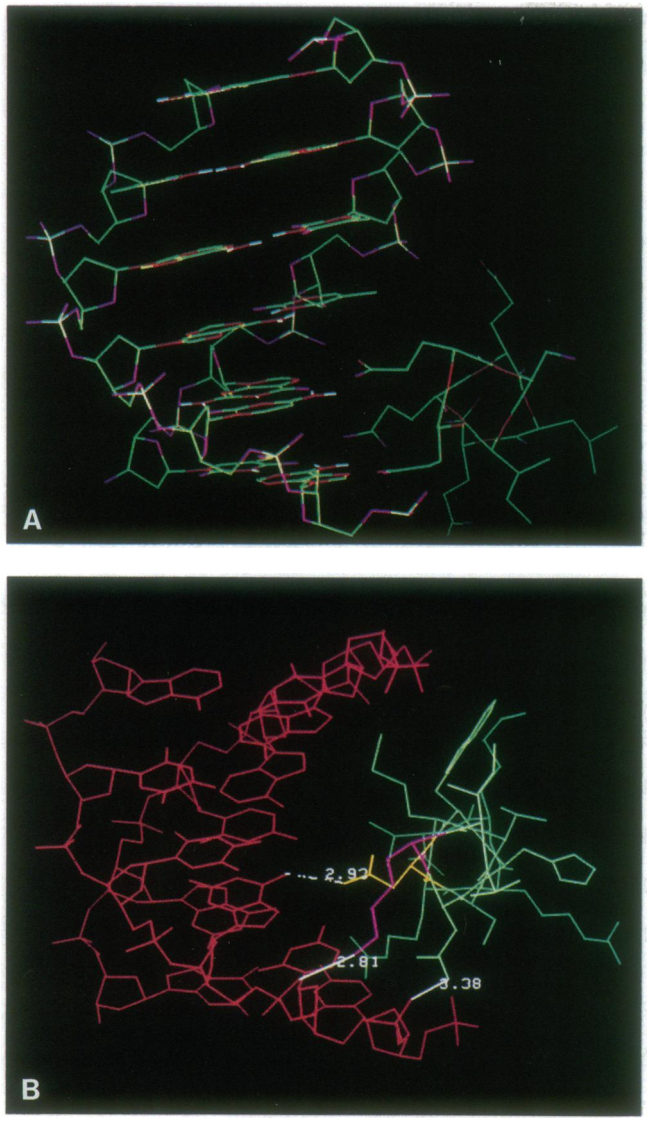

FIG. 5. Molecular modeling of Myb binding to the B-DNA. (*A*) A computer model of 434 repressor–operator–DNA complex (see ref. 22 for x-ray crystallographic structure). (*B*) A computer model of Myb–DNA complex. The Myb recognition sequence 5′-TAACGG-3′ in its double-stranded B-DNA form is shown in red. The possible contact residues in the third repeat of Myb protein are shown. Asn-186 (yellow) and Lys-182 (pink) are in contact with adenine and the phosphate backbone of DNA, respectively. The numbers in white denote the hydrogen bonding distance between the amino acid residues and the bases. Molecular modeling was performed as described in the text.

suggest that each repeat does not contribute equally to the DNA-binding activity (which is consistent with the deletion mutation data), and a single point mutation in any one of the tryptophans in the second and third repeats could abolish the DNA-binding activity of the protein.

To distinguish whether the major function of tryptophans is to maintain structural integrity of the DNA-binding region or whether these residues are involved in direct interaction with DNA, we replaced the last tryptophan of the third repeat with other amino acids (Fig. 4). Replacement of this tryptophan with glycine or proline (lanes 4 and 7) resulted in a complete loss of DNA-binding activity. However, replacement of the ninth tryptophan with an aromatic amino acid such as tyrosine (lane 8) or histidine (lane 5) did not result in the loss of DNA-binding activity. In addition, when the tryptophan residue was replaced with alanine or leucine, two

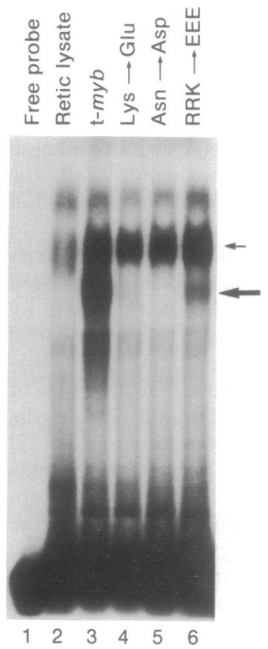

FIG. 6. Role of Lys-182 and Asn-186 in Myb–DNA interactions. Lys-182 and Asn-186 (Fig. 1) were replaced with Glu and Asp, respectively. In addition, the three amino acids Arg-Arg-Lys at positions 190–192 were replaced with Glu-Glu-Glu, and the binding activity of the three mutant proteins was assessed. Lanes: 1, free probe; 2, control reticulocyte lysate; 3, wild-type t-Myb; 4, Lys → Glu mutant; 5, Asn → Asp mutant; 6, Arg-Arg-Lys → Glu-Glu-Glu mutant. The Myb-specific band is shown by a large arrow and the nonspecific band is shown by a small arrow.

nonaromatic but hydrophobic amino acids, both mutant proteins, retained DNA-binding activity (lane 3 and 6). Collectively, these results suggest that (*i*) preservation of the α-helical character of the repeat region may be important for DNA-binding; (*ii*) the tryptophans probably do not interact in a stacking mode with the nucleotide bases since nonaromatic nonplanar side chains are tolerant in this position; and (*iii*) amino acids with hydrophobic character may be essential in this position. This argument is supported by our observation that replacement of the sixth or the ninth tryptophan with arginine abolished the DNA-binding activity (data not shown). The tryptophans may thus form a hydrophobic scaffold analogous to the tryptophans found in bacterial DNA-binding proteins (22, 23) that allows adjacent regions, rich in basic amino acids, to interact with DNA.

Physical Model for Myb–DNA Interactions. A common structural motif present in bacterial DNA-binding proteins appears to be a helix-turn-helix structure interacting with bases and phosphates at the major groove (22, 23). Indeed, it has been proposed that the DNA-binding domain of Myb constitutes a special case of the helix-turn-helix motif (12). To gain an understanding of the nature of physical interactions between DNA and Myb protein, we constructed a model of these interactions, which is schematically shown in Fig. 5. The region between amino acids 180 and 193, in the third repeat of Myb, was constructed as an α-helix. Then, the sequence TAACGG was built as a double-stranded B-DNA structure. The possible interactions between the helix and DNA fragment were studied using "static docking" methods by applying the program QUANTA (Polygen, Waltham, MA) on a Stellar GS 1035 graphic supercomputer. In docking the helix to the DNA fragment, the orientation of the helix is assumed to be similar to that of 434 repressor–operator complex where the helix binds to DNA (see Fig. 5*A* for a computer model of 434 repressor–operator complex and refs. 22 and 23 for x-ray crystallographic structure). To construct a similar model for Myb–DNA interaction, the orientation of the side chains in the helix of Myb were set to the statistically observed values based on several protein structures (24). By starting with the helix in the configuration of the 434 repressor–operator complex (Fig. 5*A* and ref. 22), the helix was moved and rotated slightly so that the maximal interaction was obtained (Fig. 5*B*). The model structure was then minimized using CHARMm software (Energy minimization software in Quanta, Polygen, Waltham, MA). The final mini-

mized energy was −76 kcal/mol (1 cal = 4.184 J). This study indicated that although tryptophan is important for maintaining the helical structure, no direct contact of the tryptophan residue with DNA was possible. On the other hand, the two amino acids that seem to be capable of interacting with DNA appeared to be the Lys-182 and Asn-186 (shown in pink and yellow in Fig. 5*B*). In addition to these two amino acids, the computer model indicated the possibility of somewhat weaker interaction with Asn-183 (Fig. 5*B*).

Role of Flanking Amino Acids in the DNA-Binding Activity of Myb. To assess the role of the two amino acids, Lys-182 and Asn-186, in DNA binding, predicted to be important by the computer model, we changed the lysine to glutamic acid and asparagine to aspartic acid and used them in a DNA-binding assay (Fig. 6). These results show that mutation of Asn-186 into aspartic acid or Lys-182 to glutamic acid results in a loss of DNA-binding activity. Frampton *et al.* (12) pointed to the presence of a patch of basic amino acids (residues 190–192 in Fig. 1) at the end of repeat 3 in Myb, which also seem to occur in the homeodomain. To assess the role of residues 190–192, Arg-Arg-Lys, in DNA-binding activity, we mutated the three basic amino acids to Glu-Glu-Glu. This drastic mutation did not result in a loss of DNA-binding activity even though a reduction in the intensity of DNA binding was observed. Although we cannot draw any firm conclusions regarding the role of residues 190–192, Arg-Arg-Lys, our results suggest that Lys-182 and Asn-186 could be involved in establishing a direct association with DNA. These results indicate that the physical model developed on the basis of helix-turn-helix motif is reasonably accurate and allows the prediction of the role of other amino acids whose function remains to be determined.

During the past year, a number of DNA-binding proteins such as Fos, Jun, and Myc have been shown to possess a conserved heptad repeat of leucine residues that may represent a widely used motif for the generation of a hydrophobic surface through which DNA-binding proteins interact (25–27). These proteins contain a basic domain adjacent to the leucine zipper region that together constitute the DNA-binding motif. The Myb family of proteins may be the prototype of a family of DNA-binding proteins where the tryptophans seem to form a hydrophobic core that places adjacent amino acids in the appropriate spatial arrangement for interaction with DNA. This arrangement is reminiscent of the zinc finger motifs (28) where it has been shown that basic amino acids between each finger contact DNA and the zinc finger itself forms a scaffold.

We thank Drs. F. Rauscher and R. Burnett for a critical review of the manuscript. This work was supported by Grant CA44463 and CA21124 from the National Institutes of Health. Salary support of some of the investigators was provided by Grant CA10815, also from the National Institutes of Health.

1. Biedenkapp, H., Borgmeyer, U., Sippel, A. E. & Klempnauer, K. H. (1988) *Nature (London)* **335,** 835–837.
2. Weston, K. & Bishop, J. M. (1989) *Cell* **58,** 85–93.
3. Dudek, H. & Reddy, E. P. (1989) *Oncogene* **4,** 1061–1066.
4. Sakura, H., Kanei-Ishii, C., Nagase, T., Nakagoshi, H., Gonda, T. J. & Ishii, S. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 5758–5762.
5. Rosson, D. & Reddy, E. P. (1986) *Nature (London)* **319,** 604–607.
6. Tice-Baldwin, K., Fink, G. R. & Arndt, K. T. (1989) *Science* **264,** 931–935.
7. Peters, C. W. B., Sippel, A. E., Vingron, M. & Klempnauer, K. H. (1987) *EMBO J.* **6,** 3085–3090.
8. Paz-Ares, J., Ghosal, D., Wienand, U., Peterson, P. A. & Saedler, H. (1987) *EMBO J.* **6,** 3553–3558.
9. Bender, T. P. & Kuehl, W. M. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 3204–3208.
10. Majello, B., Kenyon, L. C. & Dalla-Favera, R. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 9636–9640.
11. Anton, I. A. & Frampton, J. (1988) *Nature (London)* **336,** 719.
12. Frampton, J., Leutz, A., Gibson, T. J. & Graf, T. (1989) *Nature (London)* **342,** 134.
13. Weinstein, Y., Ihle, J. N., Lavu, S. & Reddy, E. P. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 5010–5014.
14. Dudek, H. & Reddy, E. P. (1989) *Oncogene* **4,** 1489–1495.
15. Rosson, D., Dugan, D. & Reddy, E. P. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 3171–3175.
16. Higuchi, R., Krummel, B. & Saiki, R. (1988) *Nucleic Acids Res.* **16,** 7351–7367.
17. Nielson, D. & Shapiro, D. (1986) *Nucleic Acids Res.* **14,** 5536–5546.
18. Singh, H., Sen, R., Baltimore, D. & Sharp, P. (1988) *Nature (London)* **323,** 640–644.
19. Ness, S. A., Marknell, A. & Graf, T. (1989) *Cell* **59,** 1115–1125.
20. Howe, K. M., Reakes, C. F. L. & Watson, R. J. (1990) *EMBO J.* **9,** 161–166.
21. Oehler, T., Arnold, H., Bidenkapp, H. & Klempnauer, K. H. (1990) *Nucleic Acids Res.* **18,** 1703–1710.
22. Anderson, J. E., Ptashne, M. & Harrison, S. C. (1987) *Nature (London)* **326,** 846–850.
23. Pabo, C. O. & Sauer, R. T. (1984) *Annu. Rev. Biochem.* **53,** 293–321.
24. Benedetti, E., Morelli, E. G., Nemethy, G. & Scherega, H. A. (1983) *Int. J. Pept. Protein Res.* **22,** 1–6.
25. Landschulz, W. M., Johnson, P. F. & McKnight, S. L. (1988) *Science* **243,** 1695–1698.
26. Kouzarides, T. & Ziff, E. (1988) *Nature (London)* **336,** 646–651.
27. Gentz, R., Rauscher, F. J., III, Abate, C. & Curran, T. (1989) *Science* **243,** 1695–1698.
28. Berg, J. M. (1990) *J. Biol. Chem.* **265,** 6513–6516.