Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 11779–11783, December 1992
Biochemistry

Myc/Max and other helix–loop–helix/leucine zipper proteins bend DNA toward the minor groove

(DNA bending/transcription factors/immunoglobulin enhancer)

DAVID E. FISHER†‡, LANA A. PARENT†, AND PHILLIP A. SHARP†

†Center for Cancer Research, Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139; and ‡Department of Medical Oncology, Dana–Farber Cancer Institute and Boston Children's Hospital Departments of Pediatric Hematology/Oncology, Harvard Medical School, Boston, MA 02115

Contributed by Phillip A. Sharp, September 4, 1992

ABSTRACT **A distinct family of DNA-binding proteins is characterized by the presence of adjacent "basic," helix–loop–helix, and leucine zipper domains. Members of this family include the Myc oncoproteins, their binding partner Max, and the mammalian transcription factors USF, TFE3, and TFEB. Consistent with their homologous domains, these proteins bind to DNA containing the same core hexanucleotide sequence CACGTG. Analysis of the conformation of DNA in protein–DNA complexes has been undertaken with a circular permutation assay. Large mobility anomalies were detected for all basic/helix–loop–helix/leucine zipper proteins tested, suggesting that each protein induced a similar degree of bending. Phasing analysis revealed that basic/helix–loop–helix/leucine zipper proteins orient the DNA bend toward the minor groove. The presence of in-phase spacing between adjacent binding sites for this family of proteins in the immunoglobulin heavy-chain enhancer suggests the possible formation of an unusual triple-bended structure and may have implications for the activities of Myc.**

Sequence-specific binding of DNA by regulatory proteins is a critical step in the activation of gene transcription. These regulatory proteins have been assigned into families based upon the structure of domains responsible for specific recognition of DNA. One such family, characterized by adjacent basic, helix–loop–helix, and leucine zipper domains (b-HLH-ZIP), contains ≈10 proteins identified to date. These proteins include the Myc oncogene family (1); Max (2); the transcription factors USF (3), TFE3 (4), and TFEB (5); as well as several yeast DNA-binding proteins (6–8).

In addition to containing similar protein domains, these proteins have also been shown to recognize an identical core DNA target sequence of CACGTG. This DNA sequence was originally analyzed as a gene-specific promoter/enhancer element in the major late promoter of adenovirus (MLP). Although DNA binding has been shown to be critical for the expression of the oncogenic potential of Myc, the specific genes regulated by Myc that produce this phenotype remain to be determined. In addition to binding the symmetrical CACGTG site, proteins in this family appear capable of recognizing the related core sequence CATGTG (4, 9). This asymetric site is found in the immunoglobulin heavy-chain enhancer (μE3 site), where two additional core sequences related by the consensus CANNTG are also found (10). At least one member of the b-HLH-ZIP family, transcription factor TFE3, has been demonstrated to activate transcription of genes containing multiple copies of the μE3 sites (4).

Although specific DNA binding is probably a function of the basic domain of the Myc-related factors, little is known about the structure of the protein–DNA complex. Members of this family appear to bind DNA as homo- or heterodimers dependent, at least in part, on specificities contained within the leucine zipper (2–4, 9). Based on methylation-interference studies, contacts are made within the major groove of DNA (4, 5, 9). It is likely that the protein dimer binds symmetrically at the center of the palindromic site with the basic domains extending symmetrically into the major grooves of each half site. The configuration of the DNA in the complex is unknown and might change when complexed with protein.

Bending of DNA occurs upon binding of the Jun and Fos proteins that contain the related structural motifs of adjacent basic region and leucine zipper (11). A similar study involving circular permutation and phasing analyses (11–15) of GCN4, a yeast protein of the same family, did not reveal bending of DNA (16). Numerous other proteins of diverse function have also been shown to produce oriented DNA bends, including the recent report (17) that truncated c-Myc bends DNA differently from heterodimer c-Myc/Max, suggesting that alterations of DNA structure may not be rare. To further assess whether such oriented DNA bending occurs upon binding by b-HLH-ZIP proteins, several members of this family were tested and found to produce striking minor groove-oriented bends of identical direction and nearly identical magnitude.

MATERIALS AND METHODS

DNA Vectors and Probes. For circular permutation analyses an oligonucleotide containing the MLP binding site (CTAGAACCCGGTCACGTGGCCTACTGCAG) was cloned into the *Xba* I/*Sal* I sites of the vector pBend2 (provided by S. Adhya, National Institutes of Health, Bethesda, MD, ref. 18) (pBend–MLP). Probes numbered 1–4 were generated by digestion with *Bgl* II, *Nhe* I, *Spe* I, and *Bam*HI, respectively, and labeled by using Klenow subunit of DNA polymerase and [α-^{32}P]dATP. pBend–μE3 probes were made analogously by inserting an oligonucleotide containing the μE3 site (CTAGACAGGTCATGTGGCAAGGCTGCAG) into the *Xba* I/*Sal* I sites of pBend2. Probes labeled A–C were generated by restriction endonuclease cleavage with *Mlu* I, *Xho* I, and *Bam*HI, respectively, followed by end-filling with Klenow subunit of DNA polymerase in the presence of radionucleotides. Phasing probes were constructed by using double-stranded oligonucleotide fragments containing three adjacent, in-phase $(dA)_5$ tracts and the MLP binding site spaced 21, 23, 26, 28, and 30 base pairs (bp) from the central deoxyadenosine in the middle adenosine tract. Synthetic oligonucleotides were cloned into the *Xba* I/*Sal* I sites of pBS-SK (Stratagene). The insert sequences are (respectively): CTAGACCGGTCACGTGGCCGCAAAAACGGG-

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.

Abbreviations: b-HLH-ZIP, adjacent basic, helix–loop–helix, and leucine zipper domains; MLP, adenovirus major late promoter; EMSA, electrophoretic mobility-shift assay; GST, glutathione-*S*-transferase.

CAAAAACGGGCAAAAACGGCTGCAG; CTAGACCGGTCACGTGGCCTAGCAAAAACGGGCAAAAACGGGCAAAAACGGCTGCAG; CTAGACCGGTCACGTGGCCTACTGGCAAAAACGGGCAAAAACGGGCAAAAACGGCTGCAG; CTAGACCGGTCACGTGGCCTACTGTAGCAAAAACGGGCAAAAACGGGCAAAAACGGCTGCAG; CTAGACCGGTCACGTGGCCTACTGTACTGCAAAAACGGGCAAAAACGGGCAAAAACGGCTGCAG. The variable spacer lengths contain differences of one DNA helical turn and are underlined. Phasing probes were made from these vectors by the PCR using T3 and T7 primers and standard reaction conditions, except 1 μM dATP and 100 μCi of [α-^{32}P]dATP (1 Ci = 37 GBq) were used. The probes used in the analysis (Fig. 1*C*) were generated by digestion of pBend–MLP with *Bgl* I, *Nhe* I, *Spe* I, *Eco*RV, *Sma* I, *Stu* I, *Ssp*, I, and *Bam*HI. Probes were labeled by using Klenow subunit of DNA polymerase, phosphatase treatment followed by polynucleotide kinase, or the PCR using ^{32}P-labeled dNTPs. These probes, when unbound, all migrated indistinguishably (data not shown).

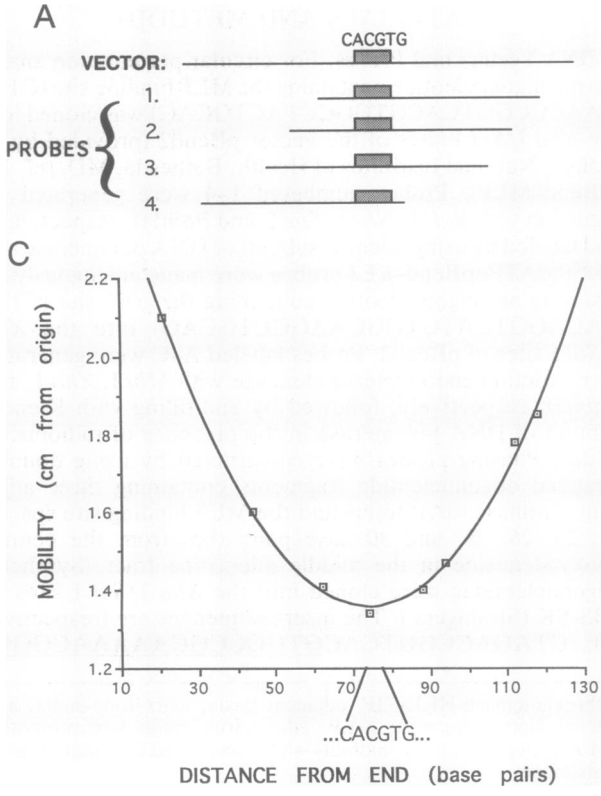
Proteins and Electrophoretic Mobility-Shift Assays (EMSAs). Binding conditions were as described (9), except that 10 μg of bovine serum albumin was added to binding reactions for TFEB, Max, USF, and c-Myc/Max. Gels (8% acrylamide/0.27% bisacrylamide/25 mM Tris/190 mM glycine (pH 8.9)/1 mM ethylenediaminetetraacetic acid) were run at 4°C for various times. TFEB (amino acids 265–514) was cloned into the pET-15b vector (TFEB–His), overexpressed, and purified as a histidine fusion according to the manufacturer's protocol (Novagen, Madison, WI). Max (provided by C. Dang, Johns Hopkins University School of Medicine, Baltimore) was overexpressed and purified as described (17). USF was purified from HeLa cells (19). The DNA-binding domain of c-Myc was produced by cloning the *Tth*3I–*Eco*RI fragment of pHSR-1 (American Type Culture Collection 41010) into the *Eco*RI site of pGEX-3X (Pharmacia). This glutathione-*S*-transferase (GST)–Myc fusion and pGEX vector alone (for GST) were overexpressed, and the proteins were purified, as described (20). TFE3 (provided by T. Kadesch, University of Pennsylvania, Philadelphia) was transcribed and translated *in vitro* and used directly as described (9). Purified proteins were included in EMSAs at concentrations of 15–75 ng/μl, except for c-Myc/Max reactions, in which Max at 1 ng/μl was mixed with either c-Myc–GST or GST.

RESULTS

DNA bending induced by the b-HLH-ZIP proteins was assessed by circular permutation analysis (12). DNA probes were constructed that contained a naturally occurring Myc-binding site from the MLP at various distances from the ends of the fragments (Fig. 1*A*). Fig. 1*B* shows that the human proteins TFEB, TFE3, USF, Max, and c-Myc/Max all produced mobility anomalies that were proportional to the distance of the binding site from the nearest end of the probe. The mobility anomalies are measured from the origin and compared as the ratio of slowest to fastest migrating species (see below). This ratio was very similar for each of the proteins tested. The DNA probes, which were of identical length, migrated with the same mobility when not bound by proteins (Fig. 1*B*). TFEB, USF, TFE3, and Max proteins bind DNA as homodimers (2–4, 9). Heterodimer binding of c-Myc and Max was tested by mixing a fusion protein containing the b-HLH-ZIP domains of c-Myc and GST with purified Max protein. The heterodimer complexes (Fig. 1*B*, **) migrated more slowly than the Max homodimer complexes (*) but migrated with a similar degree of anomaly. Mobilities of all protein-bound complexes correlated with the sizes of the respective proteins (data not shown).

The center of flexure was mapped by plotting mobility as a function of distance from the binding site to the end of the probe (Fig. 1*C*). In this way, the site of flexure was determined to coincide with the position of the MLP binding site centered on the CACGTG palindromic sequence.

Because the mobility anomaly observed by circular permutation analysis does not provide information on whether the bending was oriented toward the major or minor groove, phasing analysis (11, 13–15) was done. DNA probes were



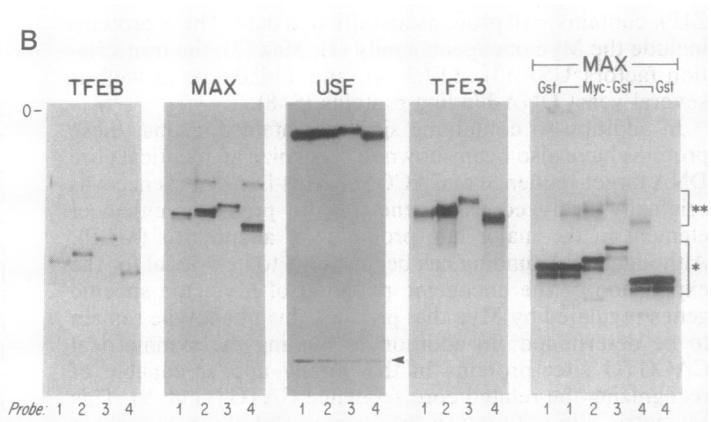


FIG. 1. Circular permutation of b-HLH-ZIP proteins on CACGTG binding site. (*A*) Construction of probes. Four radiolabeled DNA probes of identical length were made that placed the MLP binding site at various positions along the DNA fragment. (*B*) EMSA. Gel origin is indicated (O). Unbound DNA probe is indicated by the arrowhead for the lanes containing transcription factor USF–DNA complexes but is not shown for other protein–DNA complexes. Max–DNA complexes are indicated by brackets. **, c-Myc/Max heterodimer bands; *, Max homodimer–DNA bands; Gst, GST. (*C*) Circular-permutation analysis. TFEB–His was mixed with probes containing the MLP site at various positions. Migration of shifted probe bands was plotted against distance in base pairs from end of probe to MLP site. Probes were 154 bp long, and position of core CACGTG is indicated. The curve was drawn by using polynomial curve fitting (Cricketgraph Software, Malvern, PA).

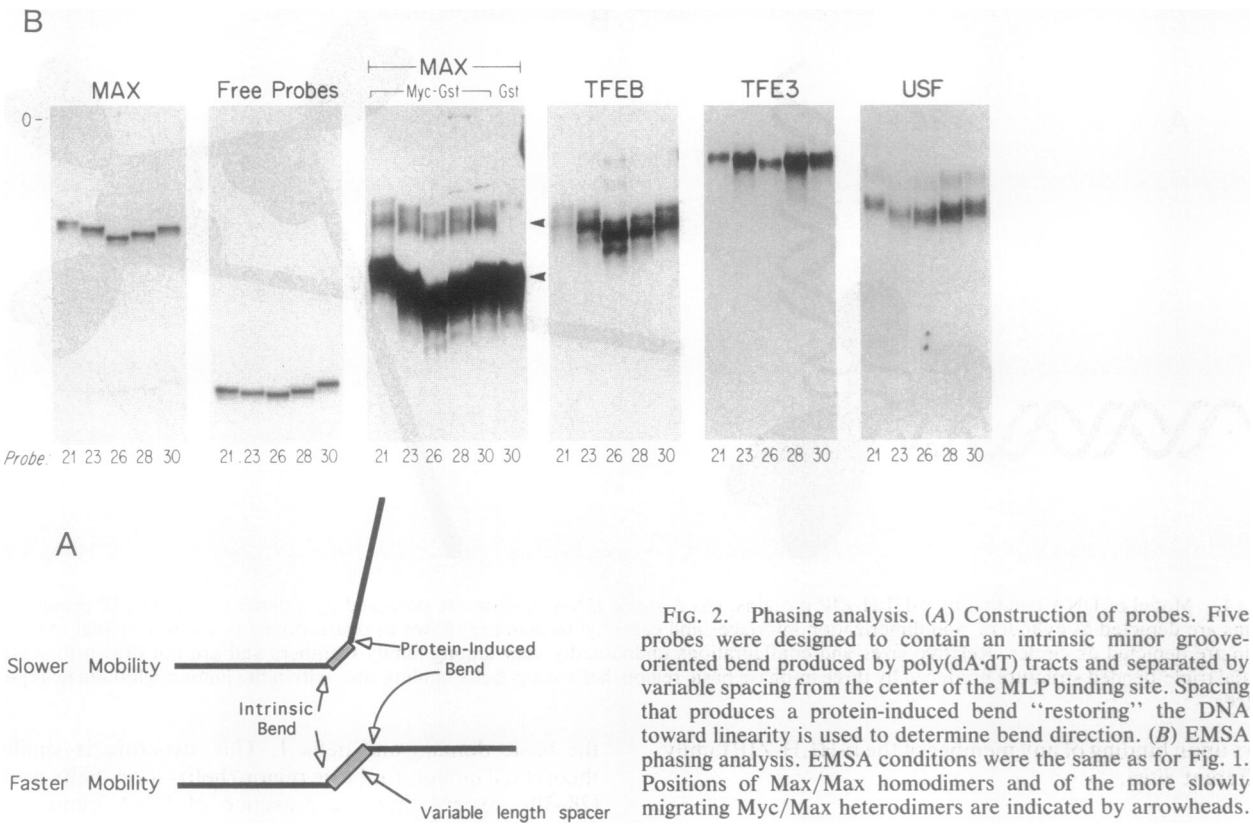


FIG. 2. Phasing analysis. (*A*) Construction of probes. Five probes were designed to contain an intrinsic minor groove-oriented bend produced by poly(dA·dT) tracts and separated by variable spacing from the center of the MLP binding site. Spacing that produces a protein-induced bend "restoring" the DNA toward linearity is used to determine bend direction. (*B*) EMSA phasing analysis. EMSA conditions were the same as for Fig. 1. Positions of Max/Max homodimers and of the more slowly migrating Myc/Max heterodimers are indicated by arrowheads.

devised (Fig. 2*A*) that contain an intrinsic DNA bend in which the central nucleotide was spaced 21, 23, 26, 28, or 30 bp from the center of the CACGTG binding site. These probes, thus, represent incremental steps spacing the two bends between two and three complete turns of double-helical DNA (assuming 10.5 bp per complete turn). The intrinsic bend was produced by insertion of three adjacent tracts of $(dA{\cdot}dT)_5$ bp, which generate minor-groove-directed bends of 18°–22° per tract (13, 14, 21–24). The phasing analysis determines the spacing—i.e., rotation—between intrinsic and protein-induced bends that restore linear-type (faster) gel mobility by cancellation of the two adjacent bends. For each b-HLH-ZIP protein, the fastest mobility of protein-bound DNA was observed at a spacing of 26 bp between the center of the intrinsic bend and the center of the CACGTG palindrome (Fig. 2*B*). This spacing suggests that the orientation of bending induced by the b-HLH-ZIP proteins is toward the minor groove because fastest (less bent) mobility is seen when the binding site is rotated to the opposite face relative to an intrinsic minor-groove-oriented bend. The same results were obtained for all of the b-HLH-ZIP proteins tested. In this analysis, the free DNA probes had anticipated slight variation in their mobilities due to differences in spacer length. The magnitude of variation in mobility $[1-(\mu_M/\mu_E)]$ of the free probes was ≈8-fold smaller than that seen for the protein-induced mobility anomalies. Similar small variations have been observed previously with probes containing the Jun/Fos (11) and GCN4 (16)-binding sites.

To estimate the bend angle, ratios of slowest to fastest migrating species from the circular permutation analysis (μ_M/μ_E) were fit to the equation $\mu_M/\mu_E = \cos(\alpha/2)$, were α describes the bend angle (25). Bend angles were also interpolated from published standard curves generated with longer DNA molecules than those used in this study (25). For all proteins, μ_M/μ_E was in the range of 0.76–0.80, and the estimated angles for bending were within the range of 74°–82°, whether calculated from the cosine relationship or interpolated from the standard curve. An alternative method to estimate bend angle uses data from phasing analysis to generate computer-derived phasing amplitudes and likely would produce smaller angle estimates for these proteins, as it has for others when compared to circular-permutation amplitudes (26). It is noteworthy that all of these b-HLH-ZIP proteins appear to bend DNA to nearly the same angle and with the same orientation.

A mobility anomaly has also been observed for binding of transcription factor TFEB to the immunoglobulin enhancer μE3 site, which contains the core sequence CATGTG (Fig. 3). The amplitude of the mobility anomaly was nearly identical for the same protein when bound to the μE3 site as compared with the MLP site (Fig. 3). TFE3 has also previously been reported to exhibit a mobility anomaly when bound to the μE3 site (27). Thus, DNA bending probably

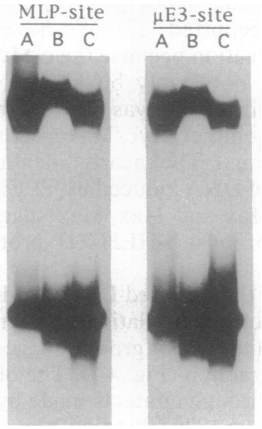


FIG. 3. Similar mobility anomaly for TFEB binding to MLP or μE3 sites. Probes of identical length containing the MLP (CACGTG) or μE3 (CATGTG) core sequences at corresponding positions within the DNA fragment (lanes A–C) were tested by EMSA. Virtually identical mobility anomalies were observed for both sites.

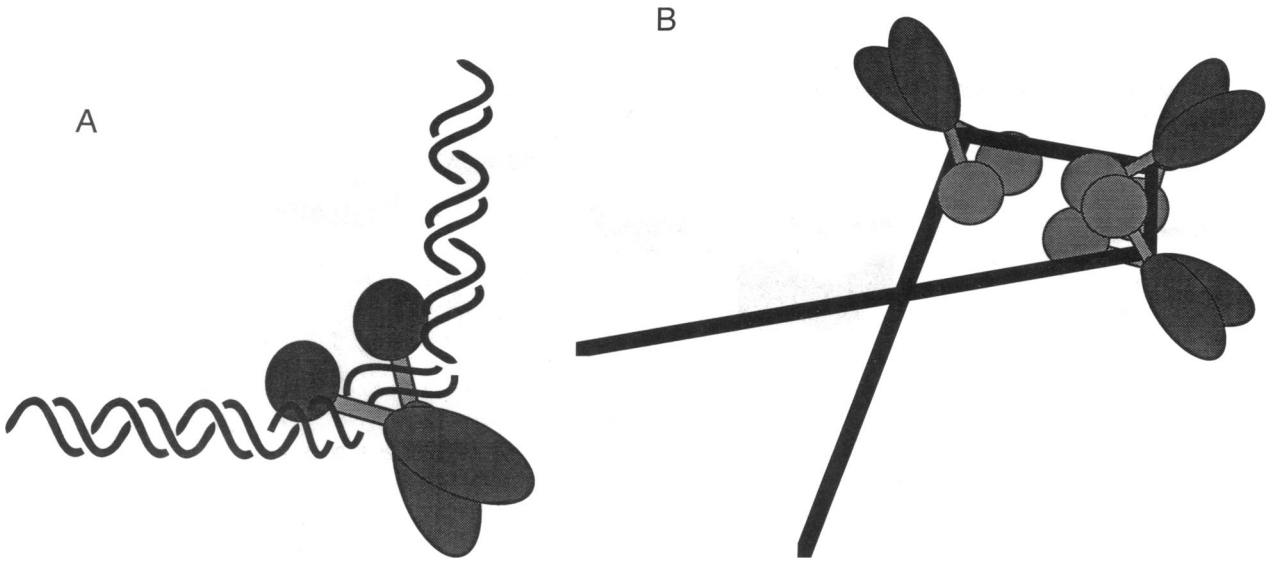


FIG. 4. Model of DNA bending by b-HLH-ZIP proteins. (*A*) A single DNA-binding site occupied by a dimeric b-HLH-ZIP protein. Basic domains are depicted as cylinders, and dimerization interfaces and carboxyl termini are shown as ovals; domains amino-terminal to the basic domain are depicted as circles. Relative sizes and configurations undoubtedly vary among family members and are not drawn to scale. (*B*) Potential triple-bended structure produced by three in-phase basic region/helix–loop–helix binding sites within the immunoglobulin heavy-chain enhancer.

occurs upon binding of any member of the b-HLH-ZIP family to different sites.

DISCUSSION

DNA bending has been assessed for the b-HLH-ZIP family of proteins, bound to the CACGTG binding site. Circular-permutation analysis revealed the presence of a significant mobility anomaly that was remarkably similar for all members of this protein family. Quantitative estimates suggest that these proteins bend DNA in the range of 74°–82°, and phasing analysis revealed that the bend was oriented toward the minor groove. A significant aspect of these findings is the similarity in both angle and orientation of bending induced by all family members. It is also noteworthy that the mobility anomaly was indistinguishable for TFEB when bound to a related, but distinct, target DNA sequence (μE3 site).

These observations agree with the recent report that Myc/Max heterodimers and Max homodimers bend DNA toward the minor groove (17). Bending of DNA in the opposite direction (toward the major groove) by truncated c-Myc protein was also reported (17). We have not been able to confirm this observation because the c-Myc construct tested in this report did not stably bind DNA as a homodimer. Importantly, no difference was seen in the estimated bend angle for the heterodimer Myc/Max as compared with other b-HLH-ZIP proteins. The nearly identical minor-groove-oriented bending of DNA induced by TFEB, USF, and TFE3 as well as Myc/Max and Max/Max suggests that this is a common feature of most b-HLH-ZIP protein–DNA interactions.

Combination of the oriented DNA bend toward the minor groove and previous methylation-interference studies that indicate binding in the major groove suggests a model for the protein–DNA interaction (Fig. 4*A*). The dimer protein probably binds DNA through contacts made by the basic regions in the major groove. This binding places the helix–loop–helix/leucine zipper axis perpendicular to the palindromic site extending from the major-groove side. Basic domains of helix–loop–helix proteins are probably α-helical upon binding to DNA (ref. 28, D.E.F. and P.A.S., unpublished work) and, therefore, a continuous protein α-helix may form that spans the basic domain and helix I. This structure is similar to theoretical models for basic region/helix–loop–helix proteins (28–30), except that the presence of DNA bending may facilitate contacts between the central two nucleotides and carboxyl-terminal amino acids in the basic domain. Thus, the bending of DNA toward the minor groove could better accommodate the continuous α-helix spanning the basic domain and helix I and be stabilized by maximal protein–DNA contacts. In contrast to these b-HLH-ZIP proteins, the basic region/leucine zipper protein GCN4, which binds DNA in a continuous α-helix, appears not to bend DNA (16). However, the basic region/leucine zipper protein Fos also bends DNA toward the minor groove, and when compared on the basis of a single molecule, the magnitude of bending resembles that for the b-HLH-ZIP family (26). Surprisingly, the binding of Jun protein bends DNA toward the major groove (11), similarly to the observation with c-Myc. This result has been explained by proposing that the basic region may contain an elbow that clasps the DNA helix.

The bending of DNA toward the minor groove suggests the potential for a triple-bended structure in the immunoglobulin heavy-chain gene enhancer (10), where three binding sites for helix–loop–helix proteins are spaced 10 and 20 bp apart (approximately one and two DNA helical turns). This spacing would produce protein-bound sites that are "in-phase" with one another (Fig. 4*B*). It is not known whether the three sites are simultaneously occupied or whether twisting of DNA might change the phasing considerations. However, such a composite structure would be notable for the manner in which protein regions amino-terminal to the basic domains are clustered together and might easily interact (or repel). In addition, this structure may significantly alter the direction of the DNA template, affect chromatin structure that already may involve HMG (high-mobility group) protein-induced bends (31), or produce a higher-order structure recognizable by other regulatory components.

We thank Drs. C. Dang for His–Max, T. Kadesch for the TFE3 clone; S. Adhya for pBend2; Drs. T. Curran, T. Kerppola, J. Parvin, T. Kristie, S. Harper, B. Shykind, M. Moore, R. Meyers, A. Gil, and J. Pomerantz for useful discussions; and M. Siafaca for expert secretarial assistance. This work was supported by grants from the

National Institutes of Health (P.A.S.) and a Howard Hughes Medical Institute Fellowship (D.E.F.).

1. DePinho, R., Hatton, K. S., Tesfaye, A., Yancopoulos, G. D. & Alt, F. W. (1987) *Genes Dev.* **1,** 1311–1326.
2. Blackwood, E. M. & Eisenman, R. N. (1991) *Science* **251,** 1211–1217.
3. Gregor, P. D., Sawadogo, M. & Roeder, R. G. (1990) *Genes Dev.* **4,** 1730–1740.
4. Beckmann, H., Su, L. & Kadesch, T. (1990) *Genes Dev.* **4,** 167–179.
5. Carr, C. S. & Sharp, P. A. (1990) *Mol. Cell. Biol.* **10,** 4384–4388.
6. Baker, R. E. & Masison, D. C. (1990) *Mol. Cell. Biol.* **10,** 1458–1467.
7. Cai, M. & Davis, R. W. (1990) *Cell* **61,** 437–446.
8. Ogawa, N. & Oshima, Y. (1990) *Mol. Cell. Biol.* **10,** 2224–2236.
9. Fisher, D. E., Carr, C. S., Parent, L. A. & Sharp, P. A. (1991) *Genes Dev.* **5,** 2342–2352.
10. Ephrussi, A., Church, G. M., Tonegawa, S. & Gilbert, W. (1985) *Science* **227,** 134–140.
11. Kerppola, T. K. & Curran, T. (1991) *Cell* **66,** 317–326.
12. Wu, H.-M. & Crothers, D. M. (1984) *Nature (London)* **308,** 509–513.
13. Zinkel, S. S. & Crothers, D. M. (1987) *Nature (London)* **328,** 178–181.
14. Salvo, J. J. & Grindley, N. D. F. (1987) *Nucleic Acids Res.* **15,** 9771–9779.
15. Snyder, U. K., Thompson, J. F. & Landy, A. (1989) *Nature (London)* **341,** 255–257.
16. Gartenberg, M. R., Ampe, C., Steitz, T. A. & Crothers, D. M. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 6034–6038.
17. Wechsler, D. S. & Dang, C. V. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 7635–7639.
18. Kim, J., Zwieb, C., Wu, C. & Adhya, S. (1989) *Gene* **85,** 15–23.
19. Chodosh, L. A., Carthew, R. W. & Sharp, P. A. (1986) *Mol. Cell. Biol.* **6,** 4723–4733.
20. Smith, D. B. & Corcoran, L. M. (1991) in *Current Protocols in Molecular Biology*, eds. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. & Struhl, K. (Greene/Wiley-Interscience, New York), pp. 16.7.1–16.7.8.
21. Levene, S. D., Wu, H.-M. & Crothers, D. M. (1986) *Biochemistry* **25,** 3988–3995.
22. Zahn, K. & Blattner, F. R. (1987) *Science* **236,** 416–422.
23. Nelson, H. C., Finch, J. T., Luisi, B. F. & Klug, A. (1987) *Nature (London)* **330,** 221–226.
24. Koo, H.-S. & Crothers, D. M. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 1763–1767.
25. Thompson, J. F. & Landy, A. (1988) *Nucleic Acids Res.* **16,** 9687–9705.
26. Kerppola, T. K. & Curran, T. (1991) *Science* **254,** 1210–1214.
27. Roman, C., Matera, G., Cooper, C., Artandi, S., Blain, S., Ward, D. C. & Calame, K. (1992) *Mol. Cell. Biol.* **12,** 817–827.
28. Anthony-Cahill, S. J., Benfield, P. A., Fairman, R., Wasserman, Z. R., Brenner, S. L. & DeGrado, W. F. (1992) *Science* **255,** 979–983.
29. Halezonetis, T. D. & Kandil, A. N. (1992) *Science* **255,** 464–466.
30. Vinson, C. R. & Garcia, K. C. (1992) *New Biol.* **4,** 396–403.
31. Giese, K., Cox, J. & Grosschedl, R. (1992) *Cell* **69,** 185–195.