Gene Expression, Vol. 4, 311-317, 1995
Printed in the USA.

1052-2166/95 $5.00 + .00

Structural Constraints for DNA Recognition by Myc and Other b-HLH-ZIP Proteins: Design of Oncoprotein Analogues

CLIFFORD TAKEMOTO AND DAVID E. FISHER[1]

Division of Pediatric Oncology, Dana-Farber Cancer Institute, 44 Binney Street, Harvard Medical School, Boston, MA 02115

DNA recognition is a critical property of many transcription factors, some of which play important roles in human disease. Disruption of this recognition may profoundly influence the biology of these factors. One such factor, the Myc oncoprotein, utilizes a basic/helix-loop-helix/leucine zipper motif to recognize the DNA target CACGTG. As discussed here, this recognition appears to occur through recognition by one face of a basic region α helix utilizing amino acid side chains highly conserved among CACGTG binding proteins. This basic domain α helix, however, requires DNA binding for stabilization. To circumvent this energetic requirement, analogues were produced that introduce multiple alanines, displaying substantially increased spontaneous α helicity and significantly enhanced DNA affinity. These studies simplify our understanding of the structural constraints for DNA recognition by this family and may serve as a template for the design of small molecule transcription-targeted therapeutics.

Transcription DNA binding b-HLH-ZIP Myc Oncoprotein

CERTAIN human oncoproteins carry out their biological functions by modulating gene transcription in such a fashion as to produce disregulated cell growth. A prominent example of such a factor is the Myc oncoprotein, one of the most commonly amplified or translocated dominant oncogenes in human cancer. Myc, like a number of the other nuclear oncoproteins, binds to DNA with sequence specificity (Blackwell et al., 1990) and appears to function as a transcriptional regulator (Kretzner et al., 1992). Although it has not yet been established exactly which genes are the targets of Myc-mediated transcriptional regulation, it is clear from deletional and mutational studies that Myc DNA binding is critical to the oncogenic function of the protein (Stone et al., 1987).

Upstream transcription factors have in many cases been found to be organized in a modular fashion with discrete DNA binding and transcriptional regulation domains. In the case of the Myc oncoprotein, the DNA binding motif is the basic/helix-loop-helix/leucine zipper (b-HLH-ZIP) motif. The recognition of this motif within Myc placed this oncoprotein into the context of a broadly studied transcription factor family, which includes proteins involved in mammalian development, such as the helix-loop-helix protein MyoD (Olson, 1990; Weintraub et al., 1991), and other related myogenic factors. In numerous experimental systems, it has been possible to demonstrate the biological requirements for both DNA binding regions and transcriptional activation domains within a particular transcription factor. Truncated proteins that lack, for example, the activation domain (Roman et al., 1991) while retaining the DNA binding domain, may function in a dominant negative fashion, repressing the gene target rather than activating it. The generality of this observation for numerous transcription factors has suggested that interference with transcription factor function at the level of the protein-DNA interaction could be a potent means of disrupting transcription factor function. In the case of oncoproteins such as Myc, such interference may dis-

Accepted December 6, 1994.
[1]Address correspondence to David E. Fisher, M.D., Ph.D.

rupt the oncogenic function of Myc protein. The approach of interfering with wild-type Myc protein is additionally attractive for human tumors, because DNA binding function of the Myc protein appears to be wild-type in the human malignancies associated with Myc overexpression. Myc's pathologic expression occurs primarily through gene amplification or chromosome translocation and does not involve mutation of the DNA binding b-HLH-ZIP region. Therefore, studies of DNA binding constraints for the Myc family of proteins may provide clues to the design of molecules capable of specifically inhibiting Myc function at the level of DNA recognition.

The b-HLH-ZIP motif is present in a number of mammalian transcription factors. Most of these factors are ubiquitously expressed and lack clear biological roles, based on our current understanding. Two major exceptions are Myc and a recently cloned melanocyte survival gene known as *Microphthalmia* (Hodgkinson et al., 1993; Hemesath et al., 1994). The basic region consists of approximately 15 amino acids and is located at the amino-terminus of this DNA binding motif. The basic region makes specific DNA contacts and thereby mediates base pair-specific recognition of DNA sequences. The adjacent helix–loop–helix and leucine zipper motifs are each well-known dimerization interfaces. The helix–loop–helix is found in a number of mammalian transcription factors without an adjacent leucine zipper and functions in these factors for dimerization. It consists of two amphipathic helices separated by a flexible loop region. The C-terminal leucine zipper appears to function as a continuation of the amphipathic helix, comprising helix 2 of the HLH domain. The specific amphipathic helix in this case contains the heptad repeats of leucine, which characterize the b/ZIP family of transcription factors. Protein–protein interactions for the b-HLH-ZIP family are largely mediated by the HLH-ZIP domains (Ferre-D'Amare et al., 1993). For most of these factors, removal of the leucine zipper has been shown to abrogate DNA binding and dimerization despite the presence of an intact HLH (Dang et al., 1989; Gregor et al., 1990; Beckmann and Kadesch, 1991; Blackwood and Eisenman, 1991; Fisher et al., 1991; Prendergast et al., 1991; Blanar and Rutter, 1992; Roman et al., 1992). Furthermore, specificity of dimerization partners appears to be encoded within the HLH zipper region. For example, Myc is unable to bind DNA as a homodimer but binds avidly as a heterodimer with its partner Max. Another example, TFEB, is a protein capable of either homodimeric or heterodimeric DNA binding with several partners, constituting a discrete subfamily of b-HLH-ZIP factors (Fisher et al., 1991).

The design of transcription factor inhibitors could potentially occur at a number of levels. These include inhibitors of transactivation, inhibitors of DNA binding, or inhibitors of protein oligomerization, because oligomerization is a feature common to the vast majority of DNA binding transcription factors. For the b-HLH-ZIP family, oligomerization is apparently more complex than simple homo- or heterodimeric interactions. Several of these factors, Myc and TFEB, have been shown to form stable tetramers in the absence of DNA. The helix–loop–helix, possibly with some contribution from the leucine zipper, appears to mediate this tetramerization (Dang et al., 1989; Fisher et al., 1991; Anthony-Cahill et al., 1992; Farmer et al., 1992; Fairman et al., 1993). Tetrameric forms have also been observed for HLH proteins such as MyoD and myogenin. Although it remains unclear what the biological function of these tetrameric forms is, the fact that they are unable to bind DNA suggests that they may carry out a regulatory role in limiting the accessibility of the protein in a form capable of binding DNA.

This manuscript presents experiments aimed at systematically determining certain of the structural constraints for DNA recognition by b-HLH-ZIP proteins and Myc in particular. By examining the amino acid requirements within the basic domain and secondary structural features of this domain, it has been possible to determine the minimally required side chain and secondary structures needed for DNA recognition. In this way, a simplified understanding of the ground rules for DNA binding by this family has been achieved. Furthermore, using this information it has been possible to design peptide analogues capable of recognizing the identical DNA targets with substantially enhanced affinity. These analogues provide potential molecular blueprints for the design of small molecule transcriptionally targeted inhibitors.

MATERIALS AND METHODS

Myc and TFEB plasmid constructs were as described previously (Fisher et al., 1993). Site-directed mutagenesis was performed as recommended by the manufacturer (Amersham) and verified by DNA sequencing. Recombinant-purified TFEB protein was made using the pET-15b vector (Novagen) and purified from BL21 *E.*

coli extracts using nickel chelate chromatography according to the manufacturer's recommendations (Qiagen Corp.). Circular dichroism (CD) spectroscopy was performed using an Aviv 60DS spectrapolarimeter over the wavelength range shown in Fig. 3, at 25°C with signal averaging of 10 min for each data point. Purified protein was used directly for CD analysis (Fisher et al., 1993) in either the absence or presence of a double-stranded oligonucleotide containing the CACGTG consensus binding site plus 5 ng per ml poly(dI-dC), after dialysis into 10 mM sodium phosphate buffer, pH 7.0, with 100 mM sodium chloride. Trifluoroethanol (TFE) titrations were carried out by substituting the indicated proportions of trifluoroethanol in the samples undegoing CD measurements. Electrophoretic mobility shift assays (EMSA) were performed using proteins generated through in vitro transcription and translation (Fisher et al., 1991, 1993). ^{32}P-labeled DNA probes were generated by Klenow end-filling of restriction-digested inserts from plasmids containing the CACGTG consensus binding site (Carr and Sharp, 1990) as it occurs in the adenovirus major late promoter. Polyacrylamide gels were run in Tris-glycine-EDTA buffer system. Quantitation of binding was determined by counting protein-bound DNA probe using a Phosphorimager (Molecular Dynamics). The linear range for DNA binding was found to be approximately 5- to 10-fold above or below the protein levels used in these experiments.

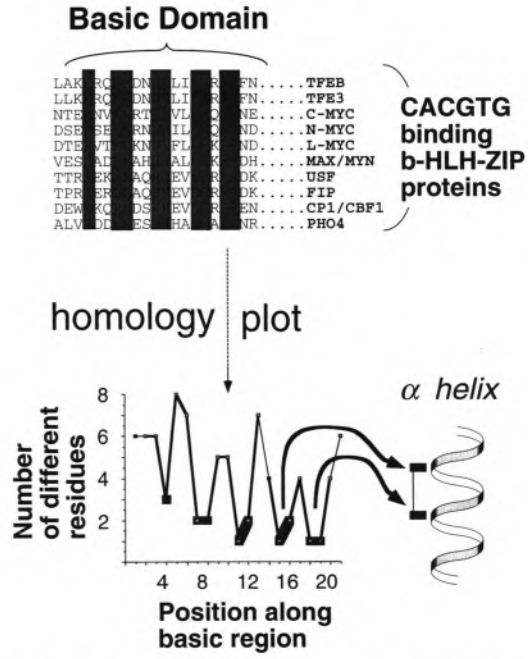

FIG. 1. Basic domain amino acid sequences of 10 CACGTG binding b-HLH-ZIP proteins were aligned and plotted as number of different amino acids at each position (see Fisher et al., 1991). Shaded residues are conserved. The spacing of conserved positions (every 3–4 amino acids) predicts that these amino acids could constitute a single face of peptide α helix.

RESULTS

Amino Acid Sequence Homologies in the Basic Domain

As a first means of determining which residues within the basic domain are specifically responsible for DNA recognition, the amino acid sequences of all known b-HLH-ZIP proteins were aligned and analyzed for sequence homologies as a function of position. As shown in Fig. 1, several positions were highly conserved whereas numerous other positions were not conserved within the basic domain. Because all of these proteins share the property of recognizing the CACGTG core sequence, it appears likely that the conserved amino acids provide nucleotide contacts responsible for that recognition. Interestingly, the spacing between conserved positions along the basic domain is every three to four amino acids. This spacing is reminiscent of the spacing of consecutive turns of a peptide α helix along one face. On this basis, it was initially proposed (Fisher et al., 1991) that DNA recognition mediated by the basic domain for this family of proteins occurs through interactions along one face of a peptide α helix lying within the major groove of DNA.

To further examine the amino acid side chain requirements for this DNA recognition, alanine scanning mutagenesis was carried out across the basic domain. For these experiments, the protein TFEB was utilized rather than Myc, because TFEB is capable of homodimeric DNA binding, simplifying the experimental design. Alanine scanning was utilized to avoid disruption of α helical character (because alanine is the amino acid most commonly represented in α helices and least likely to disrupt α helices). As shown in Fig. 2, mutagenic substitutions of alanine revealed that amino acid positions critical for DNA recognition correlated very closely with the conserved positions predicted from the sequence homology plot shown in Fig. 1. Amino acid positions that were not conserved tended to tolerate alanine substitutions, whereas those that were conserved did not tolerate alanine substitutions when measured for DNA binding. In addition, conservative amino acid substitutions were analyzed to examine more stringently the amino acid side chain requirements for

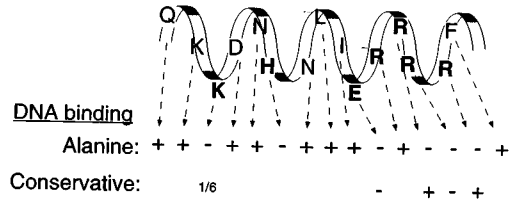

FIG. 2. Summary of DNA binding results for exhaustive mutation within the basic domain of TFEB. Either alanine substitutions or conservative (Arg:Lys or Glu:Asp) changes were made at indicated positions. + Represents DNA binding affinity at least relative to twofold of wild-type, − represents unmeasurable binding; 1/6 represents affinity sixfold lower than wild-type.

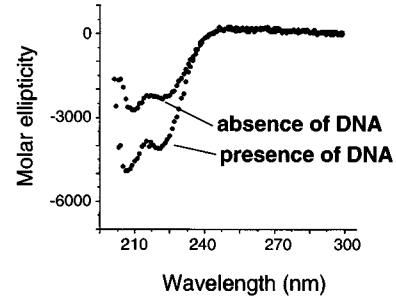

FIG. 3. Circular dichroism spectroscopy reveals DNA induced α helical folding by the basic domain of TFEB. Protein spectra in the absence or presence of DNA are shown. The difference spectrum of protein in the presence of DNA was derived by subtracting the DNA spectrum from protein plus DNA and appears to be valid because at higher wavelengths the DNA spectrum is unaffected by presence of protein (Fisher et al., 1993). Increased 222 nm amplitude represents increased α helical content.

DNA recognition at positions that were sensitive to alanine substitution. In this way it was observed that several positions such as the glutamic acid or an arginine near the C-terminus of the basic domain were intolerant not only to alanine substitutions, but even substitutions that are chemically conservative (aspartate or lysine, respectively). Three amino acids shown in Fig. 2 were sensitive to both conservative mutations and alanine substitutions, and on this basis were predicted together with the histidine to make critical base pair or phosphate contacts, stabilizing DNA recognition for the basic region. The spacing of these correlated with the spacing predicted from sequence homology blots.

α-Helicity of the Basic Domain

To more directly assess the secondary structure of the basic domain when bound to DNA, circular dichroism was employed using a purified protein encompassing the b-HLH-ZIP domain of TFEB. These studies were carried out both in the absence of DNA and in a slight molar excess of DNA containing the core sequence CACGTG. As shown in Fig. 3, the CD spectra derived from these protein and DNA combinations revealed the typical double minima at 208 and 222 nm, which typify protein α helices. Importantly, however, in the presence of DNA, the amplitude of those minima was substantially increased, suggesting that in the presence of DNA the protein segment analyzed in these studies has obtained a new α helical domain. Because the basic domain appears to be accountable for DNA recognition (based on deletional and mutagenic studies as in Fig. 2), this CD result suggests that the basic domain has become α helical upon binding to DNA. This α helical signal, derived from the basic domain–DNA interaction, supports the prediction that recognition occurs through an α helical structure. In addition, it also suggests that the basic domain in the absence of DNA is intrinsically a very poor α helix and is predominantly disordered in structure. Therefore, one of the energy requirements for DNA recognition appears to be α-helical folding, an event that requires the presence of DNA for stable protein secondary structure.

High-Affinity Analogues

Utilizing this information, it was next of interest to determine whether the intrinsic α-helicity of the basic domain could be artificially enhanced. The strategy employed was to systematically place alanine (the amino acid with highest intrinsic α-helicity) at as many positions as possible in the basic domain without disrupting DNA recognition. From the alanine scanning experiments shown in Fig. 2, it was possible to predict which amino acids would tolerate these substitutions. By systematically adding alanine to multiple positions within the basic domain of TFEB (Fig. 4), it was possible ultimately to place 12 alanines within the basic domain of TFEB, while retaining just six of the basic domain amino acids. The amino acids that were not substituted in these studies were the positions shown to be critical for DNA recognition in the same alanine scanning and other mutagenesis studies shown in Fig. 2. When DNA binding affinity was compared for these analogues, it was discovered that not only could DNA recognition occur in the presence of this alanine basic domain analogue, but the affinity was substantially enhanced in a manner that reflected the increased alanine content of the basic domain. Fur-

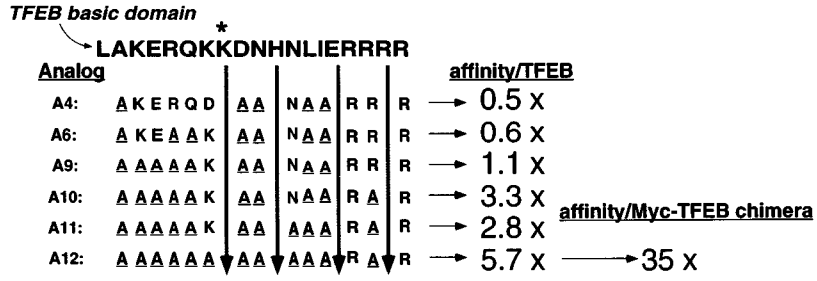

FIG. 4. Multiple alanine substitutions in the basic domain enhance DNA affinity. Alanine was systematically substituted into positions of the basic domain that do not appear to contact DNA and compared for DNA affinity with wild-type TFEB or a Myc-TFEB chimera (Myc's basic region substituted into TFEB to allow homodimeric assessment of Myc DNA affinity). The position indicated by asterisk is lysine in TFEB, but arginine in Myc. This single change (Lys to Arg) confers sixfold lower DNA affinity.

thermore, when comparing the DNA recognition properties of Myc with TFEB, it was noted that one of the critical amino acid positions for DNA recognition (a lysine shown with an asterisk in Fig. 4 for TFEB, but an arginine in Myc) substantially altered the affinity of these basic domains for DNA. Using TFEB containing either arginine or lysine at that position, the affinity was found to be sixfold enhanced for the lysine-containing protein compared to the arginine-containing protein, as occurs in Myc. Therefore, a lysine substitution at that position, as well as the multiple alanine substitutions produced a DNA binding analogue of Myc that recognized the same target sequence with an affinity that was approximately 35-fold higher than the basic domain of Myc studied in a chimera of TFEB so that the dimerization of the HLH-ZIP would not confound the analysis. Footprint and competition analyses revealed the recognition to have retained specificity (Fisher et al., 1993).

To quantitatively assess whether these alanine substitutions had in fact altered the α-helical character of the basic domain, peptides were examined that corresponded to either the wild-type basic domain or the high-affinity analogue containing multiple alanines. Because the peptides alone lack dimerization domains and correspond only to the basic regions, DNA binding could not be assessed. However, the α-helical character of these factors could be analyzed by looking at the ability to induce α-helical secondary structure using the solvent trifluoroethanol (TFE). This solvent induces α-helical folding of peptides that may otherwise be less prone to α-helical folding and in this way can be used in titration studies to determine the α-helical propensity of a given peptide. As shown in Fig. 5, using titrations of TFE on these two peptides and by determining the α-helical content as measured by the molar ellipticity at 222 nm using circular dichroism spectroscopy, the alanine-substituted basic domain had substantially enhanced intrinsic α-helicity and was much more easily induced into an α-helical form than the wild-type protein. Importantly, even in the absence of TFE this peptide was approximately 50% α-helical whereas the wild-type basic domain peptide was less than 10% α-helical (the exact degree of α-helicity for the wild-type peptide was difficult to determine because it was so near the baseline and therefore may represent close to zero α-helix). Therefore, the substitution of multiple alanines in the basic domain does appear to have induced greater intrinsic α-helical stability in the basic region and the affinity studies above indicate that this enhanced α-helicity correlates with higher affinity for DNA.

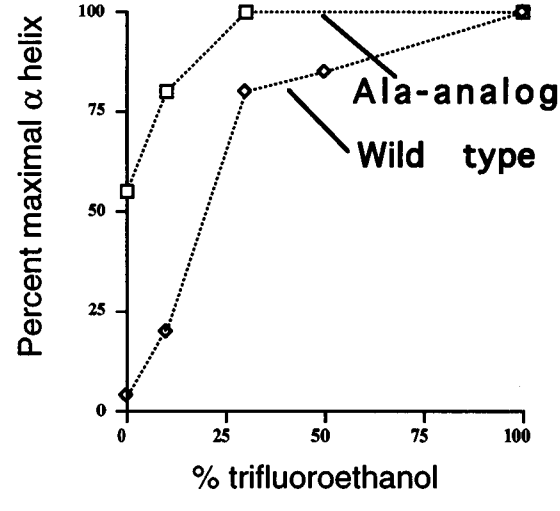

FIG. 5. High-affinity alanine basic domain peptide analogue displays enhanced spontaneous α helicity. Titration of trifluoroethanol induces α helical folding more readily in alanine analogue, which is otherwise >50% α helical compared to wild-type basic region peptide (<5%).

DISCUSSION

The biochemical analysis of amino acid side chain and secondary structural requirements for DNA recognition by the basic domain of b-HLH-ZIP proteins has revealed the importance of several specific side chains located along one face of a predicted α helix. It has also demonstrated that the intrinsic α-helicity of the basic domain is very poor but can be artificially enhanced through amino acid substitutions, and these substitutions correspondingly and significantly enhance the affinity of the peptide for DNA.

A number of the predictions made by these mutational and structural studies have been borne out in the recently determined DNA protein cocrystal structure determination of Max (Ferre D'Amare et al., 1993). Specifically, the amino acids along the predicted α-helical face are indeed responsible for base pair and critical phosphate contacts, and the face containing nonconserved amino acids does not contact DNA and is oriented away from the major groove of DNA. In the protein DNA cocrystal of Max, the basic domain is configured as a continuous α-helix together with helix-1 of the HLH domain.

The instability of the basic domain α-helix, as shown here with TFEB, has more recently been confirmed by studies of the b-HLH-ZIP factor USF (Ferre D'Amare et al., 1994). Using similar purified protein and CD analysis in the presence and absence of DNA, the same structural transition from disorder to α-helical folding was noted and therefore suggests that our observations represent a general feature of DNA recognition by the b-HLH-ZIP family. Interestingly, a number of b-ZIP and b-HLH proteins have been previously shown to undergo the same α-helical folding transition upon DNA binding (O'Neil et al., 1990; Talanian et al., 1990; Weiss et al., 1990; Anthony-Cahill et al., 1992). Therefore, it appears that the disordered to α-helical folding transition is not a coincidence, but rather may be important in the regulation of either DNA binding or some other biochemical feature of these factors within cells. Our observation that the DNA affinity could be enhanced by increasing the α-helicity using alanine substitutions demonstrates that it was possible within nature to produce a basic domain peptide capable of both DNA recognition and fairly stable α-helical secondary structure. This also supports the notion that the disordered state of all these basic domain proteins in the absence of DNA is providing a biologically relevant function within cells, perhaps by diminishing DNA affinity and enhancing regulatory potential.

The use of alanine analogues achieves high-affinity DNA binding and also simplifies our understanding of the critical constraints for DNA recognition by the basic domain. The placement of multiple alanines in the basic domain suggests that several critical amino acid side chains positioned along one face of an α-helical backbone represent the minimal requirements for sequence-specific and high-affinity DNA recognition. This simplified structure may represent a blueprint usable for the design of small molecule inhibitors of Myc DNA binding. The basic domain analogues utilized in these studies required dimerization through the helix–loop–helix and leucine zipper motifs, which collectively add an additional 70–80 amino acids. This size likely precludes any therapeutic usefulness of such protein/peptide analogues. However, several strategies are being investigated to artificially dimerize basic domain peptides such that their overall size is substantially decreased. For example, the strategy of Talanian et al. (1990) utilized cystine-mediated disulfide dimerization of basic domains derived from the yeast b-ZIP protein GCN4. Other laboratories have artificially dimerized b-ZIP-derived basic domain peptides utilizing iron chelate interfaces (Cuenoud and Schepartz, 1993). These strategies may be applicable to the basic domain from Myc and other b-HLH (ZIP) factors as well. Finally, peptidomimetic strategies may be applied to the minimalist structures studied here. For example, non-α-helical backbones, which place the critical amino acid side chains in trajectories that imitate a peptide α-helix, might be more highly constrained to produce even higher DNA affinity. Such small organic molecules might also be more easily amenable to dimerization through a variety of cross-linking agents. Such mimics of DNA binding oncoproteins may function as potent transdominant suppressors of DNA recognition and thereby transcription. Their successful design could underlie a major advance in transcription targeted therapeutics.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge Phillip Sharp in whose laboratory many of the described studies were performed. This work was supported in part by a fellowship from the Howard Hughes Medical Institute and the Fundacion Internacional Jose Ferreras to D.E.F.

REFERENCES

S. J. Anthony-Cahill, P. A. Benfield, R. Fairman, Z. R. Wasserman, S. L. Brenner, W. F. Stafford, C. Altenbach, W. L. Hubbell, and W. F. DeGrado (1992), Science 255, 979–983.

H. L. Beckmann and T. Kadesch (1991), Genes Dev 5, 1057–1066.

T. K. Blackwell, L. Kretzner, E. M. Blackwood, R. N. Eisenman, and H. Weintraub (1990), Science 250, 1149–1151.

E. M. Blackwood and R. N. Eisenman (1991), Science 251, 1211–1217.

M. A. Blanar and W. J. Rutter (1992), Science 256, 1014–1018.

C. S. Carr and P. A. Sharp (1990), Mol Cell Biol 10, 4384–4388.

B. Cuenoud and A. Schepartz (1993), Science 259, 510–513.

C. V. Dang, M. McGuire, M. Buckmire, and W. M. F. Lee (1989), Nature 337, 664–666.

R. Fairman, R. K. Beran-Steed, S. J. Anthony-Cahill, J. D. Lear, W. F. Stafford, W. F. DeGrado, P. A. Benfield, and S. L. Brenner (1993), Proc Natl Acad Sci USA 90, 10429–10433.

K. Farmer, F. Catala, and W. E. Wright (1992), J Biol Chem 267, 5631–5636.

A. R. Ferré-D'Amaré, P. Pognonec, R. G. Roeder, and S. K. Burley (1994), EMBO J 13, 180–189.

A. R. Ferré-D'Amaré, G. C. Prendergast, E. B. Ziff, and S. K. Burley (1993), Nature 363, 38–45.

D. E. Fisher, L. A. Parent, and P. A. Sharp (1991), Genes Dev 5, 2342–2352.

D. E. Fisher, L. A. Parent, and P. A. Sharp (1993), Cell 72, 467–476.

P. D. Gregor, M. Sawadogo, and R. G. Roeder (1990), Genes Dev 4, 1730–1740.

T. J. Hemesath, E. Steingrimsson, G. McGill, M. Hansen, J. Vaught, C. A. Hodgkinson, H. Arnheiter, N. G. Copeland, N. A. Jenkins, and D. E. Fisher, Genes Dev 8, 2770–2780.

C. A. Hodgkinson, K. J. Moore, A. Nakayama, E. Steingrimsson, N. G. Copeland, N. A. Jenkins, and H. Arnheiter (1993), Cell 74, 395–404.

L. Kretzner, E. M. Blackwood, and R. N. Eisenman (1992), Nature 359, 426–429.

K. T. O'Neil, R. H. Hoess, and W. F. DeGrado (1990), Science 249, 774–778.

E. N. Olson (1990), Genes Dev 4, 1454–1461.

C. Roman, L. Cohn, and K. Calame (1991), Science 254, 94–97.

G. C. Prendergast, D. Lawe, and E. B. Ziff (1991), Cell 65, 395–407.

J. Stone, R. DeLange, G. Ramsay, E. Jakobovits, J. M. Bishop, H. E. Varmus, and W. Lee (1987), Mol Cell Biol 7, 1697–1709.

R. V. Talanian, C. J. McKnight, and P. S. Kim (1990), Science 249, 774–778.

H. Weintraub, R. Davis, S. Tapscott, M. Thayer, M. Krause, R. Benezra, T. K. Blackwell, D. Turner, R. Rupp, S. Hollenberg, Y. Zhuang, and A. Lassar (1991), Science 251, 761–766.

M. A. Weiss, T. Ellenberger, C. R. Wobbe, J. P. Lee, S. C. Harrison, and K. Struhl (1990), Nature 347, 575–578.