is not surprising; genera are more widespread than species, having a geographical distribution that is the sum of the geographic ranges of the component species.

The nonendemics include a broad array of eurytopic taxa, such as bats and carnivores, that range widely throughout all macrohabitats. These are core taxa that will be preserved regardless of which macrohabitats are conserved. Endemics, however, are more common in the more extensive macrohabitats, the Amazon lowlands and, especially, the drylands. These data make it clear that, as far as mammal species richness is concerneds, the tropical rain forest enjoys no special advantage. Its diversity comes from the same processes that prevail in other places (*19*).

On the basis of these findings, if one could choose only a single macrohabitat in which to preserve the greatest amount of mammalian biodiversity in South America, one would work in the largely continuous deserts, scrublands, and grasslands. This is exactly the converse of the funding, research, and conservation strategies that have been employed to date. The emphasis on developing additional lowland rain forest parks and reserves may be misguided as far as mammals are concerned; a greater amount of mammalian diversity would be preserved by increasing the number of protected areas in the drylands. Unfortunately, scientists and the ubiquitous popular media have paid scant attention to the need to preserve deserts, grasslands, or scrublands. These dry areas are very likely far more highly threatened than the largely inaccessible rain forests of the lowland tropics (*7, 8, 20*).

REFERENCES AND NOTES

1. P. R. Ehrlich and E. O. Wilson, *Science* **253**, 758 (1991).
2. E. O. Wilson, Ed., *Biodiversity* (National Academy Press, Washington, DC, 1988).
3. N. Myers, *Conversion of Tropical Moist Forests* (National Academy of Sciences, Washington, DC, 1980).
4. M. A. Mares, *Science* **233**, 734 (1986); J. A. McNeely, *Economics and Biological Diversity: Developing and Using Economic Incentives to Conserve Biological Diversity* (IUCN, Gland, Switzerland, 1988); M. K. Tolba, *Environ. Conserv.* **17**, 105 (1990).
5. M. E. Soulé and K. A. Kohm, Eds., *Research Priorities for Conservation Biology* (Island Press, Washington, DC, 1989).
6. P. M. Fearnside, *Environ. Conserv.* **17**, 213 (1990).
7. W. J. Boecklen and N. J. Gotelli, *Biol. Conserv.* **29**, 63 (1984); W. J. Boecklen, in *Latin American Mammalogy: History, Biodiversity, and Conservation*, M. A. Mares and D. J. Schmidly, Eds. (Univ. of Oklahoma Press, Norman, OK, 1991), pp. 150–166.
8. K. H. Redford, A. Taber, J. A. Simonetti, *Conserv. Biol.* **4**, 328 (1990).
9. G. T. Prance, Ed., *Biological Diversification in the Tropics* (Columbia Univ. Press, New York, 1982).
10. K. E. Campbell, Jr., and D. Frailey, *Quat. Res.* **21**, 369 (1984); P. A. Colinvaux *et al.*, *Nature* **313**, 42 (1985).
11. M. Eigen *et al.*, *Science* **244**, 673 (1989).
12. M. L. Oldfield, *The Value of Conserving Genetic Resources* (Sinauer, Sunderland, MA, 1989).
13. R. I. Vane-Wright, C. J. Humphries, P. H. Williams, *Biol. Conserv.* **55**, 235 (1991).
14. M. V. Ashley, D. J. Melnick, D. Western, *Conserv. Biol.* **4**, 71 (1990).
15. C. G. Sibley and J. E. Ahlquist, *J. Mol. Evol.* **20**, 2 (1984); R. J. Britten, *Science* **231**, 1393 (1986).
16. J. C. Avise *et al.*, *Annu. Rev. Ecol. Syst.* **18**, 489 (1987); D. Goodman, P. R. Giri, S. J. O'Brien, *Evolution* **43**, 282 (1989); D. P. Mindell and R. L. Honeycutt, *Annu. Rev. Ecol. Syst.* **21**, 541 (1990).
17. E. C. Pielou, *Ecological Diversity* (Wiley, New York, 1975).
18. D. S. Simberloff, *Annu. Rev. Ecol. Syst.* **19**, 473 (1988).
19. The rain forest may prove to be unusually rich for other taxa, especially plants, insects, and fish, but quantitative data comparing the diversity of other groups across macrohabitats are lacking, as are data comparing higher taxonomic level diversity and endemism. The lowland forest may also play a novel role in maintaining gaseous balance in the global atmosphere [G. M. Woodwell *et al.*, *Science* **222**, 1081 (1983); R. P. Detwiler and C. A. S. Hall, *ibid.* **239**, 42 (1988)], but this also remains to be clarified.
20. E. Medina, *Interciencia* **10**, 224 (1985); P. M. Fearnside, *Environ. Conserv.* **17**, 213 (1990); V. Roig, in *Latin American Mammalogy: History, Biodiversity, and Conservation*, M. A. Mares and D. J. Schmidly, Eds. (Univ. of Oklahoma Press, Norman, OK, 1991), pp. 239–279.
21. Map modified from P. Hershkovitz [in *Evolution, Mammals, and Southern Continents*, A. Keast, F. C. Erk, B. Glass, Eds. (State University of New York Albany, 1972), pp. 311–431] and H. Walter [*Die Vegetation der Erde* (Fischer Verlag, Stuttgart, 1968), vol. 2].

22. I thank J. K. Braun for technical and editorial assistance; T. E. Lacher, Jr., M. R. Willig, L. Vitt, J. Caldwell, and L. B. Mares for their comments on the manuscript; M. R. Willig for statistical assistance; and C. Kacmarcik for graphic and computational assistance. Supported by National Science Foundation grant BSR-8906665.

29 August 1991; accepted 19 December 1991

Molecular Characterization of Helix-Loop-Helix Peptides

SPENCER J. ANTHONY-CAHILL, PAMELA A. BENFIELD,* ROBERT FAIRMAN, ZELDA R. WASSERMAN, STEPHEN L. BRENNER, WALTER F. STAFFORD III, CHRISTIAN ALTENBACH, WAYNE L. HUBBELL, WILLIAM F. DEGRADO*

A class of regulators of eukaryotic gene expression contains a conserved amino acid sequence responsible for protein oligomerization and binding to DNA. This structure consists of an arginine- and lysine-rich basic region followed by a helix-loop-helix motif, which together mediate specific binding to DNA. Peptides were prepared that span this motif in the MyoD protein; in solution, they formed α-helical dimers and tetramers. They bound to DNA as dimers and their α-helical content increased on binding. Parallel and antiparallel four-helix models of the DNA-bound dimer were constructed. Peptides containing disulfide bonds were engineered to test the correctness of the two models. A disulfide that is compatible with the parallel model promotes specific interaction with DNA, whereas a disulfide compatible with the antiparallel model abolishes specific binding. Electron paramagnetic resonance (EPR) measurements of nitroxide-labeled peptides provided intersubunit distance measurements that also supported the parallel model.

INTERACTIONS BETWEEN DIFFERENT members of the "helix-loop-helix" class of transcription factors play a key role in cell cycle progression and developmental gene regulation (*1*). This motif (*2, 3*) contains a dimerization domain consisting of a conserved amino acid sequence that is predicted to form two amphiphilic α helices connected by a more variable loop. Immediately NH_2-terminal to this sequence is an approximately 15-residue basic region; together these two units form the b-HLH motif (basic region, helix-loop-helix), capable of binding specifically to DNA. Although the name "helix-loop-helix" implies a known three-dimensional structure, the structure of this motif has not yet been determined. Therefore, we studied the conformational properties of $MyoD_{rec}$, a bacterially expressed peptide spanning residues 102 to 166 of MyoD (*4*).

The ultraviolet (UV) circular dichroism (CD) spectrum of $MyoD_{rec}$ depends on concentration; the protein undergoes a transition from random coil (or aperiodic structure) to primarily α-helix in the micromolar range. The concentration dependence of the ellipticity at 222 nm [$(\theta)_{222}$, a measure of

S. J. Anthony-Cahill, P. A. Benfield, R. Fairman, Z. R. Wasserman, S. L. Brenner, W. F. DeGrado, Biotechnology Department, DuPont Merck Pharmaceutical Co., P.O. Box 80328, Wilmington, DE 19880–0328.
W. F. Stafford III, Boston Biomedical Research Institute, 20 Staniford Street, Boston, MA 02114.
C. Altenbach and W. L. Hubbell, Jules Stein Eye Institute and Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, CA 90024–7008.

*To whom correspondence should be addressed.

the helical content, (*5*)] of $MyoD_{rec}$ is well described by a monomer-dimer equilibrium with K_{obs} approximately 5 μM, and values of −4,000 deg cm^2/dmol and −19,000 deg cm^2/dmol for $(\theta)_{222}$ of the monomer and dimer, respectively (*6*). The value for the monomer indicates that it is largely unfolded; the dimer is predicted to be approximately 54 percent helical, which would be expected if the two putative helical regions in the HLH motif were indeed α-helical. Sedimentation equilibrium ultracentrifugation of the peptide at concentrations considerably greater than K_{obs} show that it was predominantly tetrameric. Thus, in the absence of DNA, peptide dimers can further aggregate. Starovasnik and Klevit have also observed that this peptide at high concentrations forms tetramers in solution (*7*).

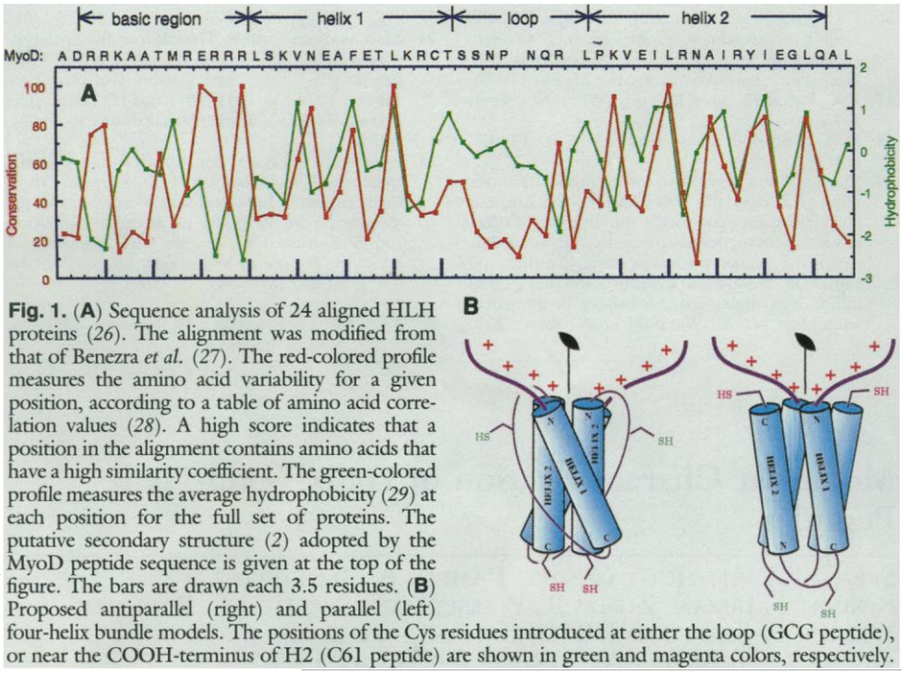

Fig. 1. (**A**) Sequence analysis of 24 aligned HLH proteins (*26*). The alignment was modified from that of Benezra *et al.* (*27*). The red-colored profile measures the amino acid variability for a given position, according to a table of amino acid correlation values (*28*). A high score indicates that a position in the alignment contains amino acids that have a high similarity coefficient. The green-colored profile measures the average hydrophobicity (*29*) at each position for the full set of proteins. The putative secondary structure (*2*) adopted by the MyoD peptide sequence is given at the top of the figure. The bars are drawn each 3.5 residues. (**B**) Proposed antiparallel (right) and parallel (left) four-helix bundle models. The positions of the Cys residues introduced at either the loop (GCG peptide), or near the COOH-terminus of H2 (C61 peptide) are shown in green and magenta colors, respectively.

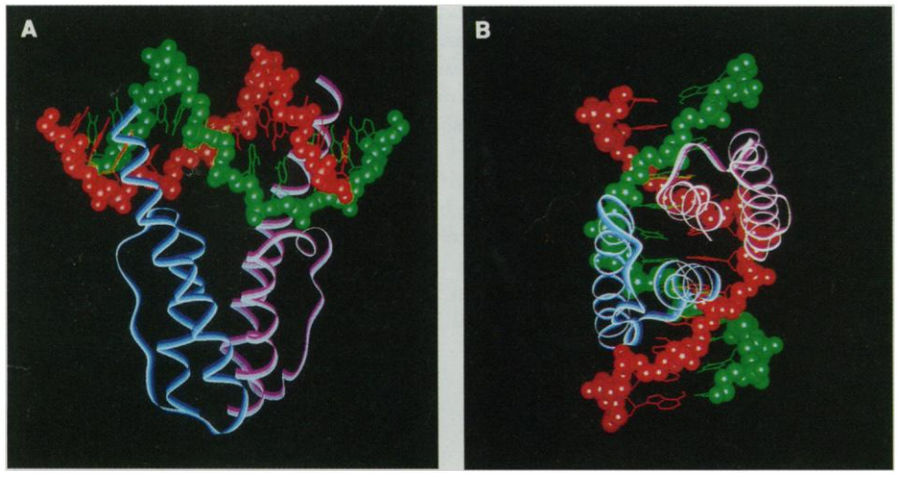

Fig. 2. Computer-generated model of a b-HLH dimer fit to a segment of DNA containing its recognition sequence NCANNTGN (here AA*CATATGT*T T). The model was constructed as follows. The H2's form a coiled-coil (*30*). H1 from each monomer was positioned interactively for good packing of hydrophobic surfaces. This was followed by an iterative series of energy minimizations and interactive graphic adjustments of interhelical angles and distances or rotations (or both) of interior side chains. The procedure was repeated until no improvements in inter-helical packing could be achieved. Each H1 was then extended in the NH_2-terminal direction to include the sequence of the basic region. The separation between the basic region helices was appropriate for insertion of the helices into the major groove of the DNA at the CA and TG sites. A second iteration of energy minimizations and manual adjustments of side chain conformations resulted in a good interaction between the phosphate backbone and the Arg and Lys side chains. The last step was the insertion of loops, which were modeled on, but are not identical to, those of triose phosphate isomerase residues A226 to A234, connecting helices G and H. A third set of minimizations and graphic adjustments was necessary to alleviate steric interference between the side chains of the helices and the backbone of the loop region. The calculations were carried out with the program AMBER (*31*) with the united atom force field, a dielectric of 5R, and a 9.5 Å cutoff for the nonbonded interactions. (**A**) View perpendicular to the DNA helical axis. (**B**) View looking down the axis of the coiled coil of the H2 segments. The phosphodiester backbone is shown as a solid surface, the bases are given as a vector representation, and the peptide backbone is shown as a ribbon.

Because peptide tetramers were observed in solution in the absence of DNA, we confirmed earlier reports that MyoD protein binds to DNA as a dimer (*2, 3*), by conducting electrophoretic mobility shift assays (EMSA) with MyoD fragments of different sizes. This technique has been used to identify dimeric, trimeric, and tetrameric DNA-protein complexes (*8*). When a recombinant protein fragment spanning residues 101 to 317 of MyoD (*9*) was mixed with $MyoD_{rec}$ (5 μM each) and run on a native gel, only three species were observed, even after prior incubation of the protein and peptide in the absence of DNA at 80°C. Thus, MyoD binds DNA as a dimer.

Addition of specific DNA to $MyoD_{rec}$ leads to a large increase in helical structure, reminiscent of that observed when fragments of leucine zipper proteins bind to DNA (*10*). For instance, $(\theta)_{222}$ of 5 μM MyoD is −14,000 deg cm^2/dmol, which increases to −30,000 deg cm^2/dmol when a 1.2-fold excess of a 25–base pair oligonucleotide bearing the MyoD binding site is present (*11*). The magnitude of the latter value is greater than the maximum value observed at high peptide concentrations suggesting that additional portions of the peptide become helical upon binding to DNA. These data, in conjunction with reports that the basic region determines DNA binding (*2*), and that helix-destabilizing mutations in this region disrupt DNA binding (*3*), suggest that the basic region adopts a helical structure when bound to its target sequence.

To obtain more information concerning the location of helices in the b-HLH motif, we examined 24 b-HLH proteins for amino acid variability (*12*) and hydrophobicity (Fig. 1A); all of these proteins bind CANNTG (where N is any nucleotide) with high affinity. Conserved, functionally important residues often segregate onto one side of an α helix, leading to a periodic arrangement of conserved and variable residues. The variability profile for the b-HLH sequences shows a strong 3.6-residue periodicity from the beginning of the basic region through the end of the first predicted helix (H1) of the HLH motif. In the aligned sequences, there are no insertions, deletions,

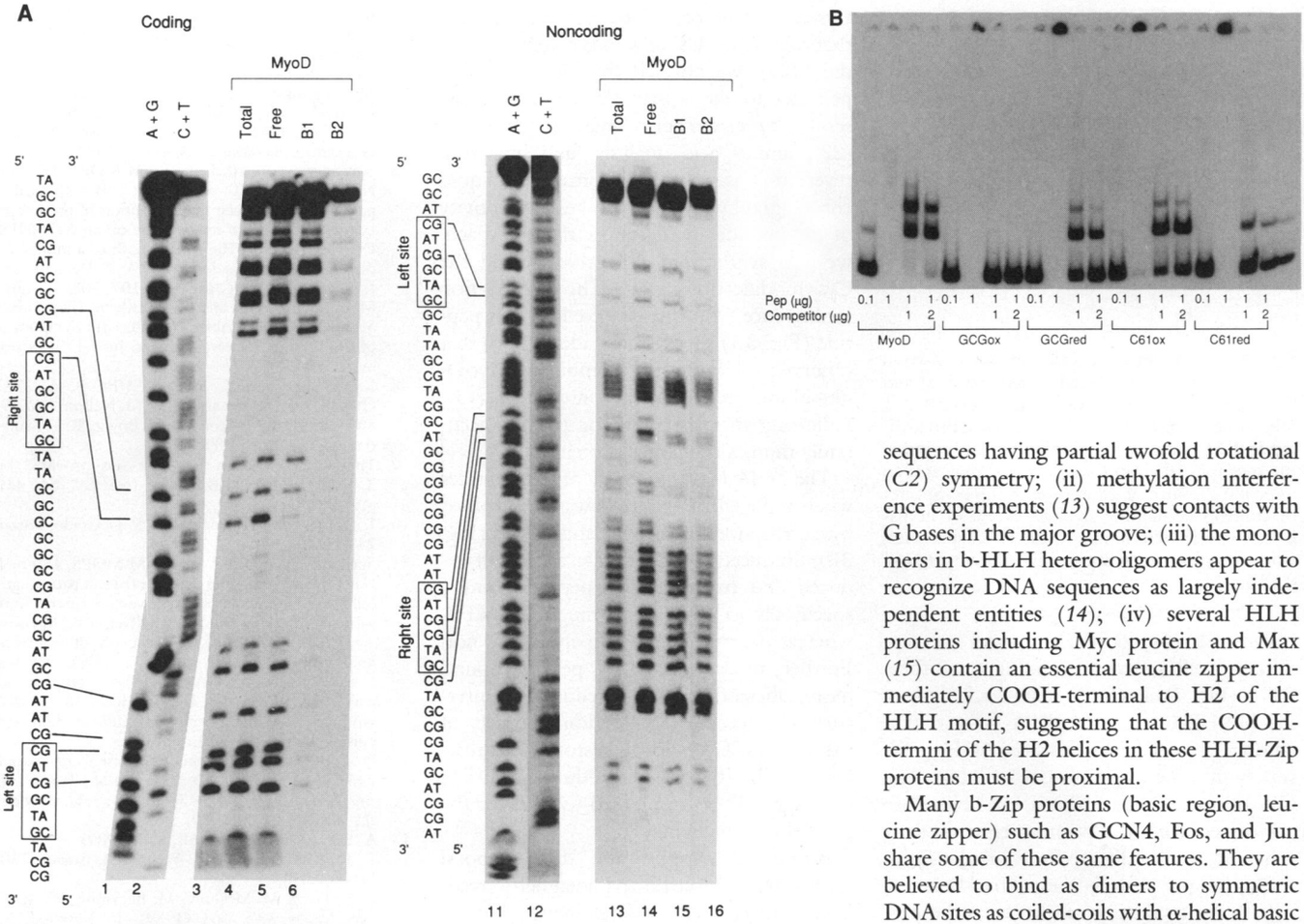

Fig. 3. (**A**) Methyl interference footprinting on the MCK enhancer obtained with the synthetic MyoD peptide. B1 and B2 refer, respectively, to singly and doubly bound species. The relevant portions of the enhancer fragment with the E box–binding sequences are enclosed in rectangles. The sequence of the right site E-box is shown enclosed in a rectangle (bottom), and G residues where methylation interferes with binding are indicated by arrows. Fragments of the rat MCK enhancer were generated from a synthetic modular enhancer construct PM255 (*32*). Plasmid PM255 was first digested with either Aat II or Hind III and then treated with alkaline phosphatase. A portion of the digested DNA (10 μg) was then partially methylated on G residues by treatment with dimethyl sulfate. Methylated DNA was labeled with [γ-^{32}P]ATP (adenosine triphosphate) and cut with Hind III or Aat II. This generated an enhancer fragment running from the Aat II (−1179) to the Hind III site (−1063) and labeled at either the Aat II site (coding strand) or the Hind III site (noncoding strand). Gel-purified fragments were used in binding reactions (*13*), and bound and unbound material were separated by electrophoresis on 6% nondenaturing gels. (**B**) DNA binding of reduced and oxidized MyoD peptides. Binding reactions were performed as above on 1 ng of DNA (residues −1026 to −1180 of rat MCK enhancer), with peptide and competitor [poly(dI · dC)] added as indicated. The symbols GCGox and C61ox refer to oxidized peptides, and GCGred and C61red refer to peptides in the reduced state; MyoD here refers to the synthetic peptide lacking any cysteine.

or strong helix-breaking residues such as Pro or Gly between the basic region and H1, an indication that the basic region might be an α-helical extension of H1. A second region with strong 3.6-residue periodicity comprises the second helix (H2) of the HLH motif, while the loop region is less conserved. The hydrophobicity profile (Fig. 1A) shows two regions with strong 3.6-residue periodicity, corresponding to H1 and H2. In these regions, the variability and hydrophobicity profiles are in phase, indicating that the most conserved residues tend to be the most hydrophobic, as is frequently observed in globular proteins. In contrast, the loop and basic regions are both hydrophilic, suggesting that they project from the protein surface.

On the basis of the following experimental observations, we next attempted to determine how these secondary structural units might be arranged in the DNA-bound dimer. (i) HLH proteins recognize DNA sequences having partial twofold rotational (*C2*) symmetry; (ii) methylation interference experiments (*13*) suggest contacts with G bases in the major groove; (iii) the monomers in b-HLH hetero-oligomers appear to recognize DNA sequences as largely independent entities (*14*); (iv) several HLH proteins including Myc protein and Max (*15*) contain an essential leucine zipper immediately COOH-terminal to H2 of the HLH motif, suggesting that the COOH-termini of the H2 helices in these HLH-Zip proteins must be proximal.

Many b-Zip proteins (basic region, leucine zipper) such as GCN4, Fos, and Jun share some of these same features. They are believed to bind as dimers to symmetric DNA sites as coiled-coils with α-helical basic regions projecting into the major groove of DNA (*10, 16*). We therefore then devised molecular models for the b-HLH motif with broadly similar attributes. The antiparallel four-helix bundle shown in Fig. 1B appeared to be a good candidate because it is a frequently observed folding motif (*17*). This fold places the ends of the basic regions on the same side of the bundle within about 12 Å of one another (the interhelical distance of the H1 helices), as in leucine zippers, which position their basic regions within about 10 Å. However, certain aspects of this structure seem inconsistent with experiment. For instance, the minimum loop length among the known HLH proteins is approximately seven residues, whereas shorter loops could bridge H1 and H2 in the antiparallel four-helix bundle model. Also, the leucine zipper regions of proteins such as Myc or Max would protrude from the bundle on the same end as the basic regions, and potentially interfere with binding to DNA.

These findings led us to build other, less well precedented helical models, such as the parallel four-helix model shown in Fig. 1B. Here the dimer is modeled as two H2 regions associated as coiled-coils and two extended helices (each comprised of the basic and H1 regions) connected by an

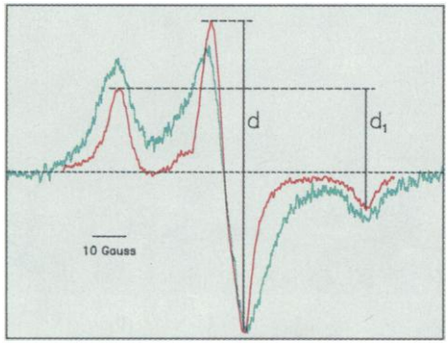

Fig. 4. Comparison of first derivative X-band EPR spectra of spin-labeled GCG (red line) and C61 (green line) in the presence of excess MCK-1 DNA in frozen solution. The additional broadening in C61 is due to dipolar interaction between the two spin labels in the dimer and indicates an interspin distance of 18 Å (d_1/d = 0.54). Spectra of controls containing fivefold excess of nonlabeled peptide were similar to the GCG spectrum, where dipolar interaction is insignificant (d_1/d = 0.38).

"overhand" (*18*) loop. This model is exceptionally well suited to binding to B-form DNA (Fig. 2). Each H1 helix docks against the H2-H2 interface, and the basic region (shown as a helical extension of H1) fits snugly into the major groove of DNA. The NH_2-terminal ends of both H2 helices are positioned with the positive ends of their α-helical dipoles directed toward the negatively charged phosphodiester backbone.

In order to distinguish between the parallel and antiparallel models, we used a disulfide cross-linking approach similar to that described (*19*). Peptides were prepared that contained a single Cys residue either in the middle of the putative loop region (GCG), or at the COOH-terminus of H2 (C61) (Fig. 1B). If the parallel four-helix model is correct, then the C61 peptide should be able to form an intermolecular disulfide bond with only minor structural adjustments. Its Cys was placed at the "a" position of the coiled-coil formed by the H2 helices in the parallel four-helix bundle model. At this position disulfides are easily formed and can be slightly stabilizing (*20*). However, if the peptides adopted the antiparallel dimer, disulfide formation would require unwinding of the ends of the H2 helices. By contrast, an intramolecular disulfide could readily be formed by the GCG peptide in the antiparallel model, in which the loops are proximal (Fig. 1B). The Cys in this peptide was inserted near the center of loops and is flanked by two Gly residues to provide extra flexibility for disulfide formation. However, disulfide bond formation between loops in the parallel model would greatly disrupt the proposed structure.

To anticipate the introduction of spectroscopic labels and unnatural amino acids, the Cys-containing peptides were prepared synthetically (*21*). Using a two-column procedure (*21*), we purified the Cys-containing peptides to more than 95% purity as assessed by electrospray mass spectroscopy (*22*), amino acid analysis, and analytical reversed-phase high-performance liquid chromagraphy (HPLC). To test the activity of the peptides prepared by this approach, we also synthesized a peptide lacking the Cys modifications (*21*). The methylation interference footprint observed for this peptide (Fig. 3A) was virtually identical to that observed for a longer recombinant MyoD-glutathione reductase fusion protein (*13*), indicating that the synthetic peptide accurately mimics the larger protein.

The DNA binding ability of the peptides when reduced or oxidized was determined by electrophoretic mobility shift assays (Fig. 3B). Reduced GCG peptide and both reduced and oxidized C61 peptides bound specifically to DNA containing CANNTG, whereas the oxidized GCG peptide did not. Further, the oxidized C61 peptide bound more efficiently than its reduced counterpart. The specific DNA binding activity of oxidized GCG could be restored by treatment with 10 mM dithiothreitol (DTT) overnight. These results led us to favor the parallel four-helix model.

Additional support for this proposal comes from electron paramagnetic resonance (EPR) data. A cysteine-specific nitroxide spin label was attached to the Cys residues in C61 or GCG (*23, 24*). The derivatized peptides were mixed with the oligonucleotide used in the CD experiments (*11*), and the EPR spectra were recorded at 177°K (Fig. 4). Dipolar interactions between spin labels are reflected by the intensity ratio d_1/d (Fig. 4). Significant dipolar interaction was observed with the spin labels attached to the COOH-termini, suggesting a distance of less than 20 Å between the unpaired electrons (*25*). No dipolar interaction was observed for the derivatized GCG peptide, indicating a separation greater than 35 Å when the spin label was attached to the loops. Measurements at 298 K show qualitatively the same results; however, a quantitative distance determination is difficult at this temperature. Again, the foregoing data are most consistent with the parallel four-helix model.

REFERENCES AND NOTES

1. N. Jones, *Cell* **61**, 9 (1990).
2. C. Murre, P. S. McCaw, D. Baltimore, *ibid.* **56**, 777 (1989); C. Murre *et al.*, *ibid.* **58**, 537 (1989).
3. R. L. Davis, P.-F. Cheng, A. B. Lassar, H. Weintraub, *ibid.* **60**, 733 (1990).
4. The peptide was expressed from a plasmid provided by H. Weintraub, and purified with S-Sepharose (Pharmacia) ion exchange chromatography and C-18 reversed-phase HPLC. It was homogeneous by criteria of analytical reversed-phase HPLC, amino acid analysis, and electrospray mass spectroscopy. The NH_2-terminus of the peptide includes the sequence Met-Glu-Leu in addition to the wild-type MyoD sequence spanning residues 102 to 166.
5. R. W. Woody, in *The Peptides* (Academic Press, New York, 1985), vol. 7, p. 15; N. Greenfield and G. Fasman, *Biochemistry* **8**, 4108 (1969).
6. Spectra were recorded in 10 mM MOPS, 150 mM NaCl, 1 mM DTT, pH 7.5, in cells with various path lengths (to keep the absorption of the sample below 1.0) at room temperature on an Aviv 62DS CD spectrometer. The data were fit to a monomer-dimer equilibrium as described [W. F. DeGrado and J. D. Lear, *J. Am. Chem. Soc.*, **107**, 7685 (1985)].
7. M. A. Starovasnik and R. E. Klevit, *The Protein Society Symposium*, Abstract M26, Fifth Symposium of the Protein Society, 22 to 26 June 1991, Baltimore, MD.
8. I. A. Hope and K. Struhl, *EMBO J.* **6**, 2781 (1987); P. K. Sorger and H. C. M. Nelson, *Cell* **59**, 807 (1989); B. M. Brown, J. U. Bowie, R. T. Sauer, *Biochemistry* **29**, 11189 (1990).
9. The purified protein fragment was provided by X.-H. Sun and D. Baltimore [see *Cell* **64**, 459 (1991)].
10. K. T. O'Neil, R. H. Hoess, W. F. DeGrado, *Science* **249**, 774 (1990).
11. Spectra were recorded in 10 mM MOPS, 150 mM NaCl, 1 mM DTT, pH 7.5, in a 0.1-cm path length cell at room temperature, and baseline spectra were subtracted as described [K. T. O'Neil *et al.*, *Biochemistry* **30**, 9030 (1991)]. The sequence of the oligonucleotide (MCK-1) was GATCCCCCCAA-*CACCTG*CTGCCTGA (and its complement) which was prepared on an ABI model 380A DNA synthesizer by the phosphoramidite method and purified by reversed-phase HPLC.
12. D. C. Rees, L. DeAntonio, D. Eisenberg, *Science* **245**, 510 (1989); T. O. Yeates *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 6438 (1987); J. U. Bowie *et al.*, *ibid.* **86**, 2152 (1989).
13. A. B. Lassar *et al.*, *Cell* **58**, 823 (1989).
14. T. K. Blackwell and H. Weintraub, *Science* **250**, 1104 (1990).
15. C. V. Dang, M. McGuire, M. Buckmire, W. M. F. Lee, *Nature* **337**, 664 (1989); E. Kerkhoff, K. Bister, K.-H. Klempnauer, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 4323 (1991); E. M. Blackwood and R. N. Eisenman, *Science* **251**, 1211 (1991).
16. C. R. Vinson, P. B. Sigler, S. L. McKnight, *Science* **246**, 911 (1989).
17. J. S. Richardson, *Adv. Protein Chem.* **34**, 167 (1981).
18. S. R. Presnell and F. E. Cohen, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 6592 (1989).
19. J. J. Falke and D. E. Koshland, Jr., *Science* **237**, 1596 (1987); E. K. O'Shea, R. H. Rutkowski, P. S. Kim, *ibid.* **243**, 538 (1989).
20. R. C. Hodges, N. E. Zhou, C. M. Kay, P. D. Semchuk, *Peptide Res.* **3**, 123 (1990).
21. The synthetic peptides span residues 106 to 166 of MyoD protein, and contain an NH_2-terminal acetyl and a COOH-terminal carboxamide. To simplify the characterization of the peptides, the wild-type Cys^{134} was replaced by Ser. In the C61 peptide, the COOH-terminal amino acid was Cys; the GCG peptide contains a Gly-Cys-Gly inserted between Pro and Asn at positions 34 and 35 of the peptide. The peptides were synthesized and partially purified as described (*10*), providing materials that were 60 to 70% pure. Peptides of this purity gave methylation interference footprints virtually identical to recombinantly derived peptides (Fig. 3A), and further purification failed to affect their affinity for DNA as assessed by the electrophoretic mobility shift assay. Nevertheless, the Cys-containing peptides were purified to greater than 95% purity by reversed-phase HPLC with buffered (pH 2.5) triethylammonium phosphate–acetonitrile gradients [C. Hoeger *et al.*, *BioChromatography* **2**, 134 (1987)]. The fully purified Cys-containing peptides were oxidized by incubating a solution (1 mg/ml containing 100 mM tris-HCl, 200 mM KCl, 1 mM EDTA, pH 8.7) at room temperature overnight. The oxidized peptide was purified by reversed-phase HPLC.
22. J. B. Fenn, M. Mann, C. K. Meng, S. F. Wong, C.

M. Whitehouse, *Science* **246**, 64 (1989).
23. Peptides were derivatized with *S*-(1-oxy-2,2,5,5,-tetramethylpyrroline-3-methyl) methanthiosulfonate (*24*), then purified by reversed-phase HPLC. The EPR spectra of 2-μl degassed samples (containing 0.5 mM spin-labeled peptide and 2.5 mM MCK-1 (a 25-bp oligonucleotide) in 10 mM MOPS, 150 mM NaCl, pH 7.5) were measured with a loop gap resonator (*24*). Samples measured at 177 K contained 30% glycerol and were recorded at 10 μW to eliminate saturation effects. The distance between the spin labels was estimated from the empirical line-shape parameter d_1/d indicated in Fig. 4 (*25*).
24. C. Altenbach, S. L. Flitsch, H. G. Khorana, W. L. Hubbell, *Biochemistry* **28**, 7806 (1989).
25. G. I. Lihktenshtein, *Spin Labeling Methods in Molecular Biology* (Wiley, New York, 1976).
26. The protein sequences included in this alignment are: human c-myc, N-myc, and L-myc, mouse MyoD, rat myogenin, human Myf-5, rat MRF4, human E12 and E47, Enhancer of split proteins m8, m7, and m5, *daughterless, twist, achaete-scute* (AS-C) *achaete*, AS-C *scute*, AS-C *lethal of scute*, AS-C *asense, hairy*, human SCL (identical to *tal*), lyl-1 (*27*); TFEB, C. S. Carr and P. A. Sharp, *Mol. Cell. Biol.* **10**, 4384 (1990); TFE3, H. Beckmann, L.-K. Su, T. Kadesch, *Genes Dev.* **4**, 167 (1990); CBF-1, M. Cai and R. W. Davis, *Cell* **61**, 437 (1990).
27. R. Benezra, R. L. Davis, D. Lockshon, D. L. Turner, H. Weintraub, *Cell* **61**, 49 (1990).
28. E. Tüdös, M. Cserzö, I. Simon, *Int. J. Peptide Protein Res.* **36**, 236 (1990).
29. D. Eisenberg, R. M. Weiss, T. C. Terwilliger, W. Wilcox, *Faraday Symp. Chem. Soc.* **17**, 109 (1982).
30. R. E. Bruccoleri, J. Novotny, P. Keck, C. Cohen, *Biophys. J.* **49**, 79 (1986).
31. P. K. Weiner *et al.*, *J. Am. Chem. Soc.* **106**, 765 (1984).
32. R. A. Horlick, G. M. Hobson, J. H. Patterson, M. T. Mitchell, P. A. Benfield, *Mol. Cell. Biol.* **10**, 4826 (1990).
33. We thank R. Beran-Steed for assistance in isolating $MyoD_{rec}$; S. Jackson for assistance with peptide synthesis; B. Larsen for obtaining electrospray mass spectra; K. O'Neil for helpful discussions, J. Lear for help with mathematical modeling, J. Krywko for computer graphics, and M. Starovasnik and R. Klevit for comments on an earlier version of the manuscript. We acknowledge postdoctoral research support from NIH grants GM13731 (S.J.A.-C.) and GM14321 (R.F.).

19 August 1991; accepted 26 December 1991

A Freeze-Frame View of Eukaryotic Transcription During Elongation and Capping of Nascent mRNA

JEREMIAH HAGLER AND STEWART SHUMAN

Ribonuclease footprinting of nascent messenger RNA within ternary complexes of vaccinia RNA polymerase revealed an RNA binding site that encompasses an 18-nucleotide RNA segment. The dimensions of the binding site did not change as the polymerase moved along the template. Capping of the 5′ end of the RNA was cotranscriptional and was confined to nascent chains 31 nucleotides or greater in length. Purified capping enzyme formed a binary complex with RNA polymerase in solution in the absence of nucleic acid. These findings suggest a mechanism for cotranscriptional establishment of messenger RNA identity in eukaryotes.

THE PRODUCTION OF MRNA IN EUkaryotic cells is regulated at multiple steps (*1*). A key regulatory target during transcription elongation and termination is the nascent mRNA chain, whose structure and sequence can be "sensed" by the ternary complex at distal sites on the template. Transduction of nascent RNA signals to the elongating polymerase may require the participation of accessory proteins that interact with the RNA, the polymerase, or both (*2*). Nascent chains are also the substrates for processing enzymes that cap, splice, cleave, and polyadenylate the mRNA precursor. How these proteins identify pre-mRNAs among other classes of transcripts is unknown. The mRNA identity may be established by recognition of the RNA polymerase II elongation apparatus or may be conferred upon the nascent RNA, perhaps through an RNA polymerase II–specific modification.

Program in Molecular Biology, Sloan-Kettering Institute, New York, NY 10021.

The question of how nascent mRNA interacts with the elongating transcription apparatus and with the RNA modifying enzymes can be most effectively approached in an in vitro system that mimics the process in vivo and is sufficiently pure to permit manipulation of the individual components. A model system that meets these criteria is provided by vaccinia virus. Vaccinia, which replicates in the cytoplasm of mammalian cells, encapsidates within the virion all the enzymes required for the synthesis of early mRNAs, including a virus-encoded multisubunit RNA polymerase with structural and functional similarity to cellular RNA polymerase II (*3*). Faithful synthesis of early mRNAs can be recapitulated in vitro on exogenous DNA templates with enzymes purified from virus particles (*4, 5*). Initiation and elongation of RNA chains require the vaccinia RNA polymerase and a vaccinia early transcription factor (VETF) that binds to the promoter. VETF is a heterodimer of 80-kD and 70-kD subunits, has intrinsic DNA-dependent adenosine triphosphatase (ATPase) activity (*5*), and is the functional equivalent of the several RNA polymerase II general transcription factors (*1*). Termination of early transcription requires a cis-acting sequence UUUUUNU in the nascent RNA strand (*6*) and a vaccinia-encoded termination factor (VTF) that is identical to the vaccinia mRNA capping enzyme (*4, 7*). Capping enzyme, a heterodimer of 95-kD and 31-kD subunits, is a multifunctional protein that catalyzes three separate reactions leading to the synthesis of a ^{m7}GpppN RNA terminus (*8*).

The finding that capping enzyme is involved both proximally (in capping) and distally (in 3′ end formation) raises questions about the timing of its interaction with the ternary complex. We addressed this issue by examining the structure of nascent RNA contained within ternary transcription complexes assembled in vitro. We took advantage of a series of DNA templates that allowed us to pause the elongating polymerase at discrete positions downstream of the major transcription initiation site (Table 1). These templates are named according to the position of the first G residue in the transcript (Gn). Pausing could be restricted to a single template position (Gn) by inclusion of 3′-*O*-methyl guanosine triphosphate (GTP) in the reactions (Fig. 1, pulse lanes P). The G18, G21, and G27 templates yielded a single major [^{32}P]cytidine phosphate (CMP)–labeled transcript *n* bases long; minor species of apparent length *n*+2 and *n*+3 are 3′ coterminal with the major RNA but arose via initiation at the −2U and −3U positions of the template (*9*). The major doublet RNAs transcribed from G31, G34, and G51 templates are 3′ coterminal but differed in the state of 5′ terminal modification.

The integrity of the ternary complexes containing pulse-labeled, 3′-*O*-methyl guanosine phosphate (3′OMeGMP)–paused transcripts was confirmed by the ability of these RNAs to be elongated during a "chase" in the presence of GTP (*10*). Addition of excess GTP to the 3′OMeGMP-paused ternary complexes allowed nearly quantitative elongation to the end of each linear template (Fig. 1, chase lanes C). Reversal of the elongation block was a result of pyrophosphorolytic removal of 3′OMeGMP and incorporation of GTP during the chase (*9, 11*). Some minor short species that were not elongated represented abortive transcripts. The analysis of active ternary complexes as they move away from the promoter provides a "freeze-frame" view of transcription elongation.

To study the interaction of nascent RNA with the transcriptional apparatus, we treated discretely paused ternary complexes containing [^{32}P]CMP-transcripts with increas-