Folding Topology of the Disulfide-Bonded Dimeric DNA-Binding Domain of the Myogenic Determination Factor MyoD†

Melissa A. Starovasnik,‡ T. Keith Blackwell,§ Thomas M. Laue,‖ Harold Weintraub,§ and Rachel E. Klevit*,‡

Department of Biochemistry, University of Washington, Seattle, Washington 98195, Department of Genetics, Fred Hutchinson Cancer Research Center, 124 Columbia Street, Seattle, Washington 98104, and Department of Biochemistry, University of New Hampshire, Durham, New Hampshire 03824

Received March 11, 1992; Revised Manuscript Received June 30, 1992

ABSTRACT: The myogenic determination factor MyoD is a member of the basic–helix–loop–helix (bHLH) protein family. A 68-residue fragment of MyoD encompassing the entire bHLH region (MyoD-bHLH) is sufficient for protein dimerization, sequence-specific DNA binding in vitro, and conversion of fibroblasts into muscle cells. The circular dichroism spectrum of MyoD-bHLH indicates the presence of significant α-helical secondary structure; however, the NMR spectrum lacks features of a well-defined tertiary structure. There is a naturally occurring cysteine at residue 135 in mouse MyoD that when oxidized to a disulfide induces MyoD-bHLH to form a symmetric homodimer with a defined tertiary structure as judged by sedimentation equilibrium ultracentrifugation and NMR spectroscopy. Oxidized MyoD-bHLH retains sequence-specific DNA-binding activity, albeit with an apparent 100–1000-fold decrease in affinity. Here, we report the structural characterization of the oxidized MyoD-bHLH homodimer by NMR spectroscopy. Our findings indicate that the basic region is unstructured and flexible, while the HLH region consists of two α-helices of unequal length connected by an as yet undetermined loop structure. Qualitative examination of interhelical NOEs suggests several potential arrangements for the two helix 1/helix 2 pairs in the symmetric oxidized dimer. These arrangements were evaluated for whether they could incorporate the disulfide bond, satisfy loop length constraints, and juxtapose the two basic regions. Only a model that aligns helix 1 parallel to helix 1′ and antiparallel to helix 2 was consistent with all constraints. Thus, an antiparallel four-helix bundle topology is proposed for the symmetric dimer. This topology is hypothesized to serve as a general model for other bHLH protein domains.

MyoD is a sequence-specific DNA-binding protein that has been shown to activate transcription of muscle-specific genes (Lassar et al., 1989; Weintraub et al., 1989). When MyoD is transfected into a large number of cell types, they are converted into muscle. A 68-residue MyoD fragment that is necessary and sufficient for DNA binding and promoting myogenesis has been identified by deletion analysis (Tapscott et al., 1988; Lassar et al., 1989). Sequence analysis of residues comprising this region reveals that MyoD does not contain any of the DNA-binding motifs that have been described structurally, such as helix–turn–helix, zinc finger, or leucine zipper. Instead, MyoD is a member of a recently identified family of DNA-binding proteins known as the "basic–helix–loop–helix" (or bHLH)[1] protein family. Members of this family include many proteins involved in myogenesis, in cell growth and transformation, and in cell-type determination and differentiation (Sun & Baltimore, 1991, and references therein).

The bHLH family was originally defined and named on the basis of a conserved sequence found in a number of proteins that included a region rich in basic residues N-terminal and immediately adjacent to a stretch of residues that could potentially form two amphiphilic helices connected by a variable loop (Murre et al., 1989a). CD spectroscopy confirmed the presence of α-helical secondary structure in MyoD and E47 bHLH peptides (Anthony-Cahill et al., 1992); no other direct structural studies have been reported as yet about this important motif. Members of the bHLH protein family have been shown to bind DNA as both homo- and/or heterodimers with other bHLH proteins (Murre et al., 1989b). Deletion and mutational analyses indicate that dimerization is mediated through the HLH region, while the specificity of DNA binding is mediated through the basic region (Davis et al., 1990; Voronova & Baltimore, 1990). Since the 68-residue bHLH domain of MyoD (MyoD-bHLH) retains the ability to form dimers and to bind DNA, it may serve as a structural model for the bHLH family. This domain has been overexpressed in *Escherichia coli*, and its properties have been characterized, as reported here.

^{1}H homonuclear 2D NMR can yield detailed information concerning both the secondary and tertiary structure of small soluble proteins [see, for example, Wüthrich (1989) and references therein]. Regular secondary structural elements,

† This research was supported by a grant from the Muscular Dystrophy Association (R.E.K.), by an American Heart Association Established Investigator Award (R.E.K.), and by grants from the National Science Foundation (T.M.L., DIR-8914571 and DIR-9002027). M.A.S. was supported by an NSF Predoctoral Fellowship and an NIH Molecular Biophysics Training Grant. T.K.B. is a Burroughs Wellcome Fund Fellow of the Life Sciences Research Foundation.

* Author to whom correspondence should be addressed.
‡ University of Washington.
§ Fred Hutchinson Cancer Research Center.
‖ University of New Hampshire.

[1] Abbreviations: bHLH, basic–helix–loop–helix; NMR, nuclear magnetic resonance; 1D, one dimensional; 2D, two dimensional; CD, circular dichroism; EMSA, electrophoretic mobility shift assay; MCK, muscle creatine kinase; MDA, MyoD binding site A; COSY, correlation spectroscopy; RELAY, relayed coherence transfer spectroscopy; TOCSY, total correlation spectroscopy; HMQC, heteronuclear multiple-quantum coherence; NOE, nuclear Overhauser effect; NOESY, NOE spectroscopy; IPTG, isopropyl β-D-thiogalactopyranoside; EDTA, ethylenediaminetetraacetic acid; PMSF, phenylmethanesulfonyl fluoride; HPLC, high-performance liquid chromatography; TFA, trifluoroacetic acid; DTT, dithiothreitol; SDS–PAGE, sodium dodecyl sulfate–polyacrylamide gel electrophoresis; TSP, sodium 3-(trimethylsilyl)propionate-*2,2,3,3-d*$_4$.

0006-2960/92/0431-9891$03.00/0

such as α-helices, can be identified by the distinctive spectral patterns they produce. These techniques have been applied to a number of different DNA-binding proteins and domains, including helix–turn–helix proteins, homeodomains, zinc-finger domains, and leucine-zipper domains [for a review see Párraga and Klevit (1991)]. In most cases where the proteins or domains had fewer than 100 residues, the ^{1}H spectra could be assigned and the structural information could be extracted. Therefore, we reasoned that a dimer composed of two identical 68-residue MyoD-bHLH polypeptides would be amenable to this approach. As will be shown here, however, reduced MyoD-bHLH is not well suited for structural analysis by NMR. But, when Cys-135 is oxidized to a disulfide, the resulting covalent homodimer exhibits properties of a stable tertiary structure. The NMR-based structural analysis of oxidized MyoD-bHLH is described.

The incorporation of NMR-active isotopes of heavy atoms (^{15}N, ^{13}C) offers significant improvements in the resolution of protein NMR spectra [for review see Griffey and Redfield (1987) and Clore and Gronenborn (1991)]. This approach has been applied to many proteins including the DNA-binding domain of the yeast α_2 repressor (Phillips et al., 1991) which gives ^{1}H spectra that are too ill-resolved to assign and analyze. In spite of its small size, MyoD-bHLH also gives poorly resolved ^{1}H spectra. To circumvent this problem, ^{15}N was incorporated both uniformly and selectively into MyoD-bHLH. Analyses of spectra obtained for these molecules have yielded spectral assignments for a majority of the residues in the helix–loop–helix sequence. Our models for the secondary structure, tertiary fold, and quaternary topology are presented here.

MATERIALS AND METHODS

MyoD-bHLH Expression. The 68 amino acid coding region of the MyoD-bHLH expression plasmid is derived from that of the MyoD deletion mutant DM:4-101; TM:167 (Tapscott et al., 1988), which was mutated so that its initiator ATG formed part of an *Nde*I site. The resulting *Nde*I–*Hin*dIII fragment which included the MyoD-bHLH coding region was subcloned into *Nde*I/*Hin*dIII-digested pRK171a (McLeod et al., 1987), a T7-driven expression vector (Studier & Moffatt, 1986). This vector was transformed into BL21(DE3) cells containing the pLysS plasmid (Studier, 1991) to produce MyoD-bHLH. The resulting sequence of MyoD-bHLH contains the entire bHLH region starting at residue 102 plus three amino acids at the N-terminus corresponding to the native N-terminus of mouse MyoD. The amino acid sequence using the one-letter code is

($^+$H$_3$N–)^{1}MEL^{102}KRKTTNADRRKAATMRERRRLSKVNEAFE-
TLKRCTSSNPNQRLPKVEILRNAIRYIEGLQALLRD^{166}(–CO$_2^-$)

Cells containing the MyoD-bHLH plasmid were grown at 37 °C on minimal medium until the A_{650} reached 0.4. Then, IPTG was added to 1 mM to induce expression of a chromosomal copy of T7 RNA polymerase, which in turn begins transcription of the bHLH gene. Cells were harvested 3 h after induction.

Isotopic Labeling. Uniformly ^{15}N-labeled samples were produced from cells grown in minimal medium containing 99% [^{15}N]ammonium chloride as the sole nitrogen source. Selectively ^{15}N-labeled samples were produced from cells grown in minimal medium that was supplemented with all amino acids (at natural isotope abundance), except the amino acid that is to be labeled. The amount of unlabeled amino acids present in the medium is double the suggested amount for nutritional supplement (Davis et al., 1980). Fifteen minutes prior to induction the ^{15}N-labeled amino acid is added along with a small addition (12% of the initial concentration) of all other unlabeled amino acids to minimize the extent of cross-labeling. The amount of ^{15}N-labeled amino acid added to the medium was 250 mg (Ala), 85 mg (Glu), 70 mg (Leu), and 100 mg (Lys) of L-amino acid per liter of medium (Torchia et al., 1989).

Protein Purification. Cells were lysed by freeze–thawing in the presence of lysis buffer containing 50 mM Hepes, pH 7.6, 0.1 M KCl, 5 mM DTT, 1 mM PMSF, 1 mM $MgCl_2$, and 0.1% Triton X-100. DNase (Boehringer Mannheim) was used to break up the DNA until the suspension was nonviscous and homogeneous. The lysed cell extract was then centrifuged at 13 000 rpm for 20 min and the supernatant saved for column purification. Pellet resuspension and centrifugation were repeated two more times to achieve maximum protein extraction.

Crude cell extract was loaded onto an S-Sepharose column (Pharmacia) preequilibrated in 50 mM Hepes, pH 7.6, 0.4 M NaCl, 1 mM DTT, 1 mM EDTA, and 1 mM PMSF. The column was then washed extensively in the same buffer until the A_{280} stabilized near zero. Most of the cell protein flowed through. MyoD-bHLH eluted when the NaCl concentration reached ~0.6 M by a linear 0.4 → 2.0 M NaCl gradient. Column progress was monitored by UV absorption at 280 nm, conductivity measurements, and 15% SDS–PAGE [19:1 acrylamide/bis(acrylamide)]. MyoD-bHLH was >90% pure after this procedure.

Column fractions were pooled, dialyzed against 3 × 3.5 L of 0.1% acetic acid, and lyophilized. Final purification was achieved by reverse-phase HPLC using a Waters Delta Pak C4 column with a water/acetonitrile gradient in the presence of 0.1% TFA. Protein was loaded onto a column that was preequilibrated in 30% acetonitrile. MyoD-bHLH eluted at about 45% acetonitrile by a linear 30% → 48% acetonitrile gradient at a rate of 0.5 %/min; the flow rate was 3.0 mL/min. The MyoD-bHLH peak was pooled and lyophilized. Purity is >99% as determined by HPLC and SDS–PAGE. Protein was quantified by UV absorption at 275 nm using an estimated extinction coefficient of 1420 M^{-1} cm^{-1} (extinction coefficient of tyrosine; Creighton, 1984). Correct amino acid composition and sample homogeneity was verified by electrospray mass spectrometry.

Preparation of Oxidized MyoD-bHLH. To obtain oxidized MyoD-bHLH, purified protein was dissolved in 50 mM Tris and 25 mM acetate, pH 7.5, and warmed in a 50 °C water bath for 30 min. The bath was then turned off, and the sample was allowed to remain in the bath and cool slowly overnight. Production of oxidized protein was verified by analytical reverse-phase HPLC using the same system described above for protein purification. The retention times of oxidized and reduced MyoD-bHLH differ by 3.3 min. Also the NMR spectra of reduced and oxidized MyoD-bHLH differ dramatically (Figure 1) such that the relative population of the two forms can be monitored by 1D NMR. The oxidized MyoD-bHLH sample used for DNA-binding experiments had undergone the above oxidation protocol 8 months prior to use. High-affinity DNA binding could be recovered from this sample by addition of DTT.

CD Spectroscopy. Spectra were obtained on a Jasco-720 spectropolarimeter fitted with a temperature-controlled water bath. Spectra were obtained on samples containing 150 μM MyoD-bHLH, 10 mM Tris, and 150 mM NaCl, pH 7.3, using a 0.1-mm path length. The spectrum of freshly reduced MyoD-bHLH was acquired at 24 °C; then, while remaining

in the cuvette, the sample was heated to 50 °C for 30 min and allowed to cool slowly for 8 h prior to acquisition of the oxidized MyoD-bHLH spectrum at 24 °C. This heating procedure generates >90% oxidized MyoD-bHLH as determined by reverse-phase HPLC. Eight transients were acquired for each spectrum. The spectrum obtained from buffer alone was subtracted from each protein spectrum prior to calculation of molar ellipticity; no smoothing of the data was performed.

Ultracentrifugation. Short-column (15-μL sample volume, 700-μm column height; Yphantis, 1960) equilibrium sedimentation was carried out at a constant temperature of 23.3 °C in a model E analytical ultracentrifuge equipped with an on-line Rayleigh interference optical system using either a 633-nm He–Ne laser or a 670-nm laser diode light source (Laue, 1992). Samples were dialyzed as described previously (Laue et al., 1984) into 50 mM Tris, 25 mM acetate, and 100 mM NaCl, pH 7.2, and the final dialysate was used for the reference. Reduced MyoD-bHLH was examined at cell-loading concentrations of 660, 330, 170, 83, 42, and 21 μM (monomer concentration) and at rotor speeds of 20 000, 28 000, and 36 000 rpm. Oxidized MyoD-bHLH was examined at 600, 300, 150, and 75 μM at rotor speeds of 20 000, 24 000, and 28 000 rpm. Equilibrium was assessed by subtraction of consecutive images taken 10 min apart. The fringe displacement data were analyzed using nonlinear least squares (Johnson et al., 1981). Molecular weights were calculated from the reduced molecular weight (σ) using a measured buffer density of 1.004 g/mL (23.3 °C). Calculated partial specific volumes ($\bar{v}$) of 0.73_5 mL/g without ^{15}N enrichment and 0.72_5 mL/g with 99% ^{15}N were estimated from the amino acid composition (McMeekin & Marshall, 1952; Laue et al., 1992). However, these values of $\bar{v}$ consistently resulted in overestimates in the apparent molecular weight, most likely due to both the large number of charged groups (+18 e, –8 e) and the high net charge (up to +10 e per monomer) anticipated for this relatively small (M_r = 8057) peptide. Accordingly, $\bar{v}$ was decreased by –0.028 mL/g, which corresponds to electrostriction of –25 mL/mol and 9 mol of charge/mol of monomer (McMeekin & Marshall, 1952; Kharakoz, 1989). An uncertainty of ±0.015 in $\bar{v}$ was included in calculating the confidence interval for the molecular weight. Association models were chosen, and the conversion of units was accomplished as described previously (Laue et al., 1984) using refractive increments of 3.52 fringes·mL/mg (633 nm) and 3.33 fringes·mL/mg (670 nm) for the two lasers (Perlman & Longsworth, 1948).

DNA-Binding Assay. Single-stranded oligonucleotides were end-labeled with ^{32}P by T4 polynucleotide kinase as previously described (Davis et al., 1990). They were then denatured and annealed to a 3-fold molar excess of their respective unlabeled complementary strands, which were designed so that a four-base overhang was present on both ends of the annealed DNA. Unincorporated label was removed by passage over G-50 columns (Boehringer Mannheim). DNA-binding reactions were performed for 20 min at room temperature and then analyzed by EMSA, as previously described (Lassar et al., 1989). The reactions shown in Figure 2 each contained 100 ng of poly[d(IC)], but essentially the same results were obtained without poly[d(IC)] present (data not shown). EMSA data that were obtained using the phosphorimager (Molecular Dynamics) were analyzed quantitatively with the program Excel (Microsoft).

NMR Spectroscopy. Most samples used to acquire 2D NMR spectra for assignment purposes contained 1 mM (monomer concentration) oxidized MyoD-bHLH, 50 mM Tris-d_{11}, and 25 mM acetate-d_3, pH 5.4, in 84% H_2O/8% D_2O/8% acetonitrile-d_3. pH values given are uncorrected for isotope composition. In the absence of acetonitrile, at concentrations of oxidized MyoD-bHLH greater than 0.7 mM a higher order oligomeric form (probably tetramer as determined by FPLC gel filtration chromatography; data not shown) is observed in the NMR spectrum. The addition of 8% (v/v) acetonitrile favors the dimer form of the protein at 1 mM MyoD-bHLH without disrupting its structure as determined by a comparison of NOESY spectra obtained with and without acetonitrile. Although the protein is highly soluble, the maximum concentration that elicited a single homogeneous population was 1 mM MyoD-bHLH, thus limiting the sensitivity of the NMR experiments.

All NMR spectra were acquired on a Bruker AM-500 spectrometer at 25 °C. Two-dimensional spectra were obtained in the pure-phase absorption mode using time-proportional phase incrementation (Marion & Wüthrich, 1983). Data were processed using the software FTNMR (Hare Research). Two-dimensional data sets were usually 600 × 2K points, processed with sine-bell filters skewed toward $t = 0$, shifted by $\pi/3$ in both dimensions, and zero-filled to produce 1K × 1K matrices. ^1H chemical shifts were referenced to external TSP. ^{15}N chemical shifts were referenced to external $^{15}NH_4Cl$ using the value 24.93 ppm relative to NH_3 by ^{15}N detection of a sample that was 2.9 M $^{15}NH_4Cl$ in 1.0 M HCl (Levy & Lichter, 1979).

Heteronuclear 2D $^{15}N/^1H$ spectroscopy was carried out with proton detection, using a 5-mm inverse probe. Heteronuclear decoupling was achieved during acquisition with the GARP-1 composite pulse (Shaka et al., 1985). HMQC spectra (Bax et al., 1983; Bendall et al., 1983) were acquired using presaturation for water suppression or with a modified jump–return pulse sequence (Roy et al., 1984; Griffey et al., 1985). 2D HMQC-COSY (Clore et al., 1988) and 2D HMQC-NOESY (Gronenborn et al., 1989) spectra were recorded on uniformly and selectively labeled MyoD-bHLH samples.

Three NOESY spectra were obtained on samples at pH 5.4, pH 6.2, and pH 7.4, and two HMQC-NOESY spectra were obtained on uniformly ^{15}N-labeled samples at pH 5.4 and pH 7.4, all using 150-ms NOESY mixing time. In these experiments, the water signal was suppressed by saturation between acquisitions and during the NOESY mixing time. The chemical shifts of amide protons are often sensitive to pH, resolving some of the assignment ambiguities arising from resonance overlap. Also, signals from relatively unstructured regions of the molecule (comprising about one-third of the peptide; see text) were missing in spectra acquired with presaturation at pH 7.4, thereby isolating many of the resonances involved in helical regions of secondary structure and allowing their unambiguous assignment. In addition, a 150-ms NOESY in 92% D_2O/8% acetonitrile-d_3 at pH 6.3 was acquired.

Detection of amide protons resistant to exchange with solvent was achieved by rapid measurement of 2D HMQC spectra, after uniformly ^{15}N-labeled protein was lyophilized out of H_2O and dissolved in D_2O (Marion et al., 1989). The spectrometer was modified with a timing device (Tschudin Associates, Kensington, MD) which allowed for buffered acquisition. Five minutes after D_2O was added to the protein, the first experiment was begun, and subsequent experiments were performed over a 21-h period. Spectra were collected 15, 35, 55, 122, 239, 351, 460, 567, and 1242 min after mixing. The first three time points were collected with 4 scans per increment, and the later time points were collected with 16

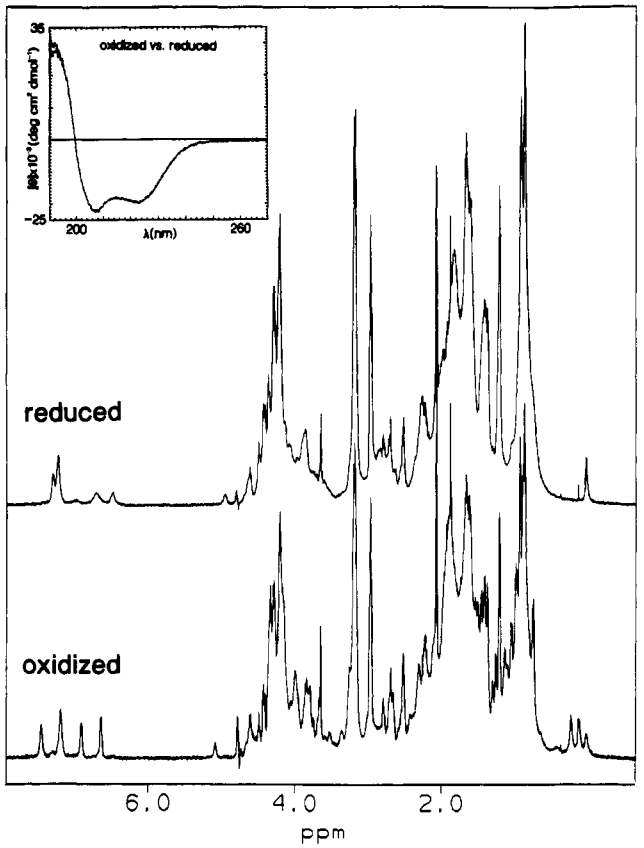

FIGURE 1: 1D NMR and CD spectra of reduced and oxidized MyoD-bHLH. The NMR samples contained 0.5 mM MyoD-bHLH, 50 mM Tris-d_{11}, and 25 mM acetate-d_3, pH 7.6, in 99.96% D_2O. Fully oxidized protein for the NMR was prepared as described under Materials and Methods. The reduced NMR sample was obtained from the oxidized sample by addition of 5 mM DTT-d_{11} and storage at room temperature overnight prior to NMR acquisition. The CD spectra of reduced (—) and oxidized (- - -) MyoD-bHLH are shown superimposed and are indistinguishable except at very low wavelengths. CD samples were prepared as described under Materials and Methods.

scans per increment to provide adequate signal-to-noise ratios. Data sets of 256 × 2K points were collected.

Distance Geometry. Relative alignment of secondary structural elements was evaluated using a variable target function algorithm, as implemented in the distance geometry program DIANA (Güntert et al., 1991), which was licensed from Professor Kurt Wüthrich at ETH-Zurich. Structures obtained from DIANA were displayed using QUANTA (Polygen Corp., Waltham, MA) on a Silicon Graphics 4D/35 workstation. Input constraints were derived from a total of 185 unambiguously assigned, nonredundant NOE peaks including 60 sequential ($i,i+1$), 20 medium-range ($i,i+2$; $i,i+3$; $i,i+4$), and 12 long-range ($i,>i+4$) NOEs identified in NOESY and/or HMQC-NOESY spectra. All unambiguously assigned NOEs observed were used as distance constraints with lower and upper bounds of 2.0 and 5.0 Å, respectively. Interhelical constraints consisted of six backbone–side-chain NOEs and six side-chain–side-chain NOEs involving four unique interhelical residue pairs (Figure 6). In addition, the absence of an NOE between Phe-129 and Tyr-156 ring resonances was input as a lower bound of 4.0 Å between ring protons. Pseudoatom corrections were used for all methylene groups, γ-methyls of Val, δ-methyls of Leu, and aromatic δ and ε protons of Phe and Tyr (Güntert et al., 1991). Helical hydrogen-bonding interactions were identified for amide protons that were detectable in an HMQC spectrum acquired 15 min after dissolution in D_2O and displayed NOE patterns consistent with α-helical secondary structure. Helical H-bonds (6 in helix 1; 16 in helix 2) were input as distance constraints between NH_i and O_{i-4} of 1.7–2.1 Å. Calculations were performed using only the 45-residue sequence from MyoD-bHLH starting at Leu-122 through Asp-166 (i.e., the 23 N-terminal amino acids were excluded from the calculation). Minimization parameters were as suggested by Güntert and co-workers (1991) with slight modification. A total of 300 structures were calculated, of which 15 had no distance-bound violations greater than 0.5 Å. A copy of the file of upper bounds is available as supplementary material.

RESULTS

MyoD-bHLH has been expressed in *E. coli* and purified to homogeneity as described under Materials and Methods. The 1D NMR spectrum of MyoD-bHLH under reducing conditions is composed of very broad peaks with chemical shifts near those of a random-coil polypeptide (Figure 1). The spectrum lacks all features suggestive of a single well-defined tertiary structure. However, the circular dichroism (CD) spectrum of reduced MyoD-bHLH indicates the presence of α-helical secondary structure (Figure 1, inset). All attempts to obtain a satisfactory NMR spectrum of reduced MyoD-bHLH by manipulating sample conditions (e.g., pH, temperature, buffer composition, protein concentration, ionic strength) were unsuccessful (data not shown). A mutant bHLH domain in which Cys-135 is replaced with tyrosine is similarly refractive to NMR analysis, as is a form of MyoD-bHLH in which Cys-135 is alkylated by reaction with iodoacetic acid (data not shown). In contrast, when the single cysteine in the molecule (Cys-135) is oxidized to an intermolecular disulfide, the NMR spectrum is markedly improved (i.e., peaks are sharper and there is more chemical shift dispersion; Figure 1), indicating that oxidized MyoD-bHLH exists in a definable tertiary structure. Oxidized MyoD-bHLH maintains the same amount of α-helical secondary structure as reduced MyoD-bHLH, when measured by CD spectroscopy (Figure 1, inset). Reduced MyoD-bHLH may be recovered from an oxidized sample by treatment with reducing agent (DTT), indicating the only chemical difference between the two forms is the oxidation state of the single cysteine residue in the sequence, Cys-135.

Sedimentation Equilibrium Ultracentrifugation. Experiments were performed with both oxidized and reduced MyoD-bHLH, and the results are summarized in Table I. There was no evidence of sample heterogeneity for either form. When the data obtained for reduced MyoD-bHLH were analyzed on the assumption of the existence of a single component, the resulting "monomer" M_r is 25 800. Since the actual M_r of a single polypeptide chain of MyoD-bHLH is 8057, this indicates that the system involves self-associating species. When the data were fit as a nonideal reversible association, the best fit was obtained from a model involving a single self-association reaction, with an apparent "monomeric" M_r of 16 610. Thus the association observed is best described as dimer–tetramer, with a K_D of 17.3 μM. There was no evidence of higher oligomers, nor was there any evidence for detectable levels of monomeric MyoD-bHLH, even at the highest rotor speed and lowest concentration (21 μM). Only models that included nonideality fit the data adequately, presumably due to the relatively high net charge on MyoD-bHLH. The second virial coefficient is consistent with the dimer having a net apparent charge (i.e., including all counterion shielding) between +3 and +4 e (Tanford, 1961), consistent with the peptide possessing its full net charge.

Table I: Molecular Weights As Determined by Sedimentation Equilibrium Ultracentrifugation

form of MyoD-bHLH model	monomer M_r[a]	BM,[b] mL/g	K_D, μM	rms[c] fringes
reduced				
single species[d]	25 800 (24 000–27 150)	5[f]		0.022
associating species[e]	16 610 (15 570–16 700)	12[f]	17.3 (15.9–17.4)	0.020
oxidized				
single species	16 400 (15 300–17 700)	7 (6–8)		0.021
associating species	15 900 (14 800–17 200)	8 (7–9)	>3000[g]	0.022

[a] All values are based on the simultaneous fit of data acquired at the concentrations and rotor speeds given under Materials and Methods. The confidence interval (in parentheses) includes an uncertainty of ±0.015 mL/g in $\bar{v}$ as well as the uncertainty in σ. Monomer refers to the smallest oligomeric form present; the calculated M_r for the MyoD-bHLH monomer is 8057. [b] Product of the molecular weight and the second virial coefficient returned by NONLIN (Johnson et al., 1981). Units were converted from reciprocal fringes to inverse weight concentration using the appropriately scaled refractive increment. These units were chosen to be consistent with those of Tanford (1961). [c] Root mean square of the variance of the fit. [d] Single species = data fit as a single, nonideal thermodynamic component. [e] Associating species = data fit as a nonideal, reversible self-association. [f] Value held fixed during the final fit. [g] Only the lower limit could be established. See text for details.

In contrast, tetramer formation is effectively abolished when MyoD-bHLH is oxidized. The best fit was obtained when the data were analyzed on the assumption of a single component. This results in a M_r of 16 400, which is consistent with the oxidized form existing as a single dimeric species under the concentration range examined. When a nonideal reversible self-association model is used, the fit is not as good, and the value of the best-fit dissociation constant would result in less than 10% of the tetramer being present at the highest concentration examined. Therefore, the K_D given in Table I should be considered a lower limit (Johnson et al., 1981).

DNA-Binding Activity. Figure 2 shows an electrophoretic mobility shift assay (EMSA) of DNA binding by oxidized and reduced MyoD-bHLH. MyoD binds with high affinity (Lassar et al., 1989) to a sequence that is required for activity of the muscle creatine kinase (MCK) enhancer (Buskin & Hauschka, 1989) and that contains the CANNTG motif that is characteristic of bHLH protein binding sites (Blackwell & Weintraub, 1990). In (A), lanes 1 and 2 compare the binding affinities of oxidized and reduced MyoD-bHLH to a labeled oligonucleotide (MCK) that contains this sequence. At this concentration of oxidized MyoD-bHLH, only a very small fraction of the MCK DNA in the reaction is bound (Figure 2A, lane 1). However, reduced MyoD-bHLH, at the same concentration, binds nearly all of the added MCK DNA (Figure 2A, lane 2). The results of this and of numerous other experiments (Figure 2A; not shown) suggest that oxidized and reduced MyoD-bHLH differ in their affinities for the MCK site by a factor of at least 10^2. At this concentration, which is 100-fold lower than the lowest assayed by ultracentrifugation (Table I), reduced MyoD-bHLH is presumably present in a monomer–dimer equilibrium (Sun & Baltimore, 1991), with only the dimeric form capable of binding DNA (Davis et al., 1990; Voronova & Baltimore, 1990). Therefore, DNA binding by reduced MyoD-bHLH requires two steps, dimerization and DNA binding, whereas the oxidized form, as a covalent dimer, only requires the single DNA-binding step. Thus, the actual difference in the DNA-binding affinities of oxidized and reduced MyoD-bHLH may be greater than that suggested by EMSA (Sauer et al., 1986).

The relative binding affinities of oxidized and reduced MyoD-bHLH for another high-affinity MyoD binding site, the MDA oligonucleotide (Blackwell & Weintraub, 1990), are similar to those observed for the MCK probe (Figure 2A, lanes 1 and 2, 3 and 9). Significantly, binding of either form of MyoD-bHLH to a labeled MDA probe can be competed by addition of excess unlabeled MDA DNA (lanes 4–6, 10–12) but not by addition of an excess of an altered MCK site ("mutant", Figure 2A, lanes 7 and 8, 13 and 14) that differs from MCK at three positions and is bound only very weakly by reduced MyoD (Lassar et al., 1989; not shown). These findings demonstrate that although oxidized MyoD-bHLH binds to the MCK and MDA sites with relatively low affinity, it nevertheless does so through nativelike sequence-specific interactions.

Whether reduced MyoD-bHLH is present in a DNA-binding reaction at the concentration shown in Figure 2A or at 1:100 of this concentration, the EMSA mobilities of the resulting protein–DNA complexes are identical (not shown). In contrast, the mobility of the complex of oxidized MyoD-bHLH bound to DNA differs reproducibly from that of the reduced protein–DNA complex (Figure 2B, lanes 2 and 3). Significantly, when both oxidized MyoD-bHLH and a low concentration of reduced MyoD-bHLH are present in the same reaction, the presence of both protein–DNA complexes in the reaction is indicated by a doublet (Figure 2B, lane 1). The difference in mobilities of these complexes and the presence of both in the doublet in Figure 2B, lane 1, are clearly apparent when a phosphorimage of the gel is subjected to quantitative graphic analysis (Figure 2C). These findings indicate that the sequence-specific DNA binding observed for oxidized MyoD (Figure 2A) is indeed derived from this species and not from the unlikely possibility that a residual amount of reduced MyoD-bHLH might be present in the DNA-binding reaction.

Assignment Strategy. Even though MyoD-bHLH is small, containing only 68 residues per monomer, resonance assignment by conventional ^1H homonuclear 2D NMR strategies was impossible due to the problem of severe spectral overlap: 69% of the amide protons (45 peaks) resonate within a region of 0.6 ppm and 78% of the α protons (53 peaks) resonate within 0.5 ppm. Complete spin-system assignment by ^1H/^1H-correlated spectroscopy (COSY, RELAY, TOCSY) was possible for only 6 of the 68 residues in oxidized MyoD-bHLH. Therefore, we relied on ^1H/^{15}N heteronuclear NMR using samples labeled with ^{15}N either uniformly or selectively to obtain amino acid type assignments. An HMQC spectrum of a uniformly ^{15}N-labeled protein contains a peak for every backbone amide in the protein plus peaks from side-chain NH_2 groups; so for MyoD-bHLH we expect 65 amide peaks ($68_{total} - 2_{proline} - 1_{N\text{-}terminus} = 65_{observable}$) plus 7 pairs of peaks representing the side-chain NH_2 groups of the five asparagine and two glutamine side chains. Figure 3 (top) shows the ^1H/^{15}N HMQC spectrum for a sample uniformly labeled with ^{15}N (in red). The spectrum obtained from uniform labeling contains about 60 distinguishable amide peaks. Resolution is improved through the use of a ^{15}N frequency dimension; however, spectral dispersion is still poor (80% of amide ^{15}N peaks resonate within 12 ppm).

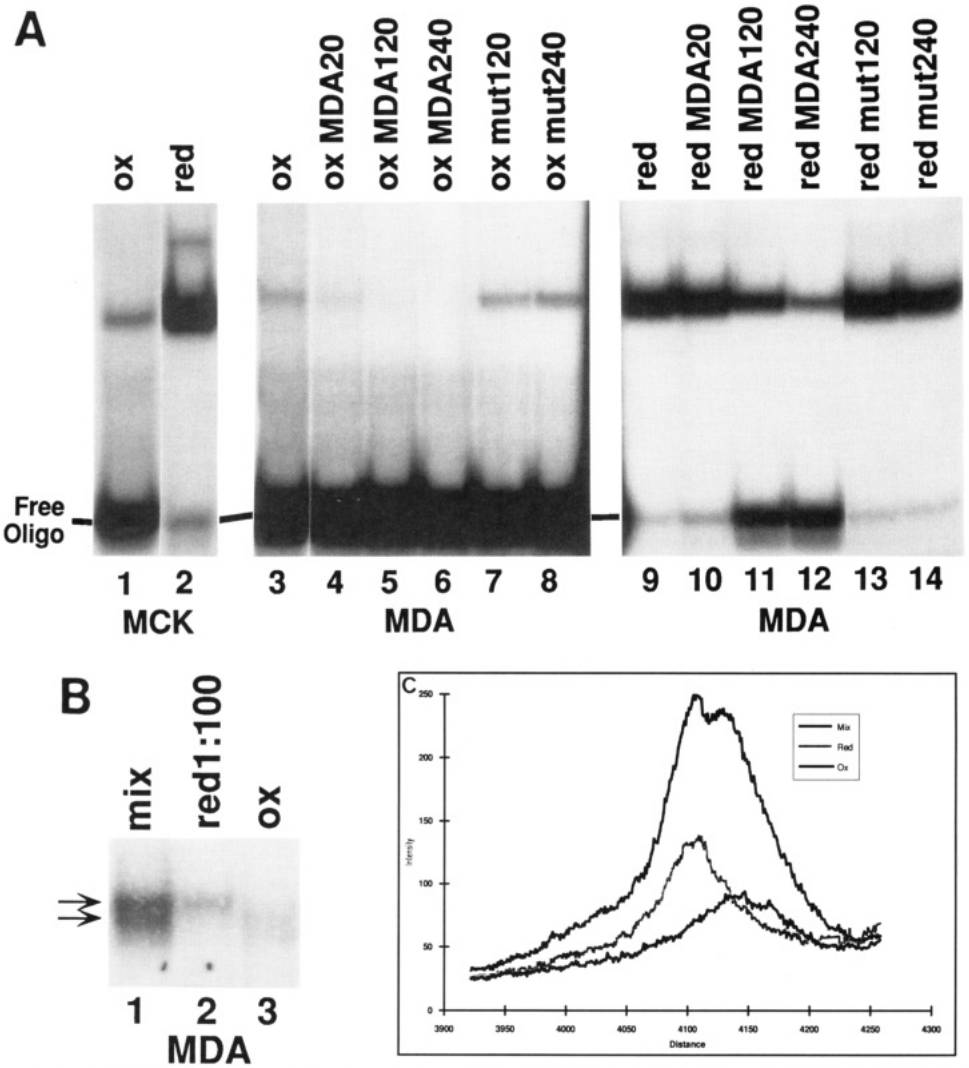

FIGURE 2: EMSA analysis of DNA binding by oxidized and reduced MyoD-bHLH. (A) Relative affinity and sequence specificity of DNA binding by oxidized and reduced MyoD. Lanes 1 and 2, respectively, show binding of oxidized (ox) and reduced (red) MyoD-bHLH, both at 1.6×10^{-7} M (monomer), to the double-stranded ^{32}P-labeled MCK oligonucleotide at 6.5×10^{-10} M. In the binding reactions assayed in lanes 3–14, the indicated proteins were present at 8.6×10^{-8} M, and the double-stranded ^{32}P-labeled MDA oligonucleotide was present at 1.6×10^{-9} M. Reactions that also contained unlabeled double-stranded MDA or mutant (mut) competitor DNA are thus designated above the autoradiogram, with the amount of competitor added indicated as a multiple of the concentration of labeled probe. For example, the reaction assayed in lane 4 contained unlabeled MDA competitor at a concentration that was 20 times that of the labeled MDA probe. EMSA was performed on a 5% polyacrylamide gel. The MCK oligonucleotide sequence is GATCCCCCCAACACCTGCTGCCTGA, the MDA sequence is GATCCCCCCAACAGCTGTTGCCTGA, and the mutant sequence is GATCCCCCCAACACGGTCTGCCTGA, all read from 5′ to 3′, with the CANNTG bHLH protein binding consensus (Blackwell & Weintraub, 1990) underlined. (B) EMSA mobilities of oxidized and reduced MyoD-bHLH bound to DNA. These DNA-binding reactions contained the double-stranded ^{32}P-labeled MDA oligonucleotide at 1.6×10^{-9} M, oxidized MyoD-bHLH at 1.3×10^{-7} M, and reduced MyoD at 1.3×10^{-9} M (lane 1) or each protein separately at the same respective concentration (lanes 2 and 3). They were analyzed by EMSA as in (A) but on a 6% polyacrylamide gel which was run until the bound complexes reached the bottom of the gel. Arrows indicate the mobilities of the different complexes. (C) Quantitative analysis of the EMSA shown in (B). The gel was analyzed with a phosphorimager; then the summed intensities in each entire lane, as scanned from top to bottom, were displayed from left to right in graph form. The lanes from which these data were derived are indicated as in (B).

Four different specifically labeled samples were prepared from bacteria grown on media containing either [^{15}N]Ala, [^{15}N]Leu, [^{15}N]Lys, or [^{15}N]Glu. The HMQC spectrum of a sample specifically labeled with [^{15}N]Ala contains only six peaks (Figure 3, top, black). The alanine labeling allowed immediate assignment of the six peaks to the amide ^{15}N^{1}H's of the six Ala residues in MyoD-bHLH. Likewise, peaks were classified as to whether they are present in the [^{15}N]-Leu, [^{15}N]Lys, and [^{15}N]Glu samples or whether they do not appear in the spectra of any of the four specifically labeled samples. Some cross-labeling was observed in the [^{15}N]Leu sample to Ile, Val, and Ala and in the [^{15}N]Glu sample to Gln, Asp, Asn, and Ala, as has been observed by others (Torchia et al., 1989; Dekker et al., 1991). The fact that only a single set of peaks is observed in oxidized MyoD-bHLH (6 Ala peaks rather than 12) indicates that the dimer is symmetric (i.e., each subunit of the dimer is indistinguishable from the other).

Sequence-specific assignments were obtained by conventional methods relying on sequential NOE connectivities (Wüthrich, 1986) combined with specific amino acid type classification obtained from ^{15}N labeling. For example, residues 150–162 were assigned by first identifying a 13-residue stretch of d_{NN} connectivities (i.e., an NOE between the NH of residue i and the NH of residue i+1; see Figure 3, bottom). Of the 13 peaks connected in this way, peaks at positions 4 and 13 were present in the [^{15}N]Ala spectrum, and peaks at positions 1, 5, 8, and 11 were present in the [^{15}N]Leu spectrum. The only sequence in MyoD-bHLH that matches this pattern is between residues 150–162 (i.e., L-R-N-A-I-

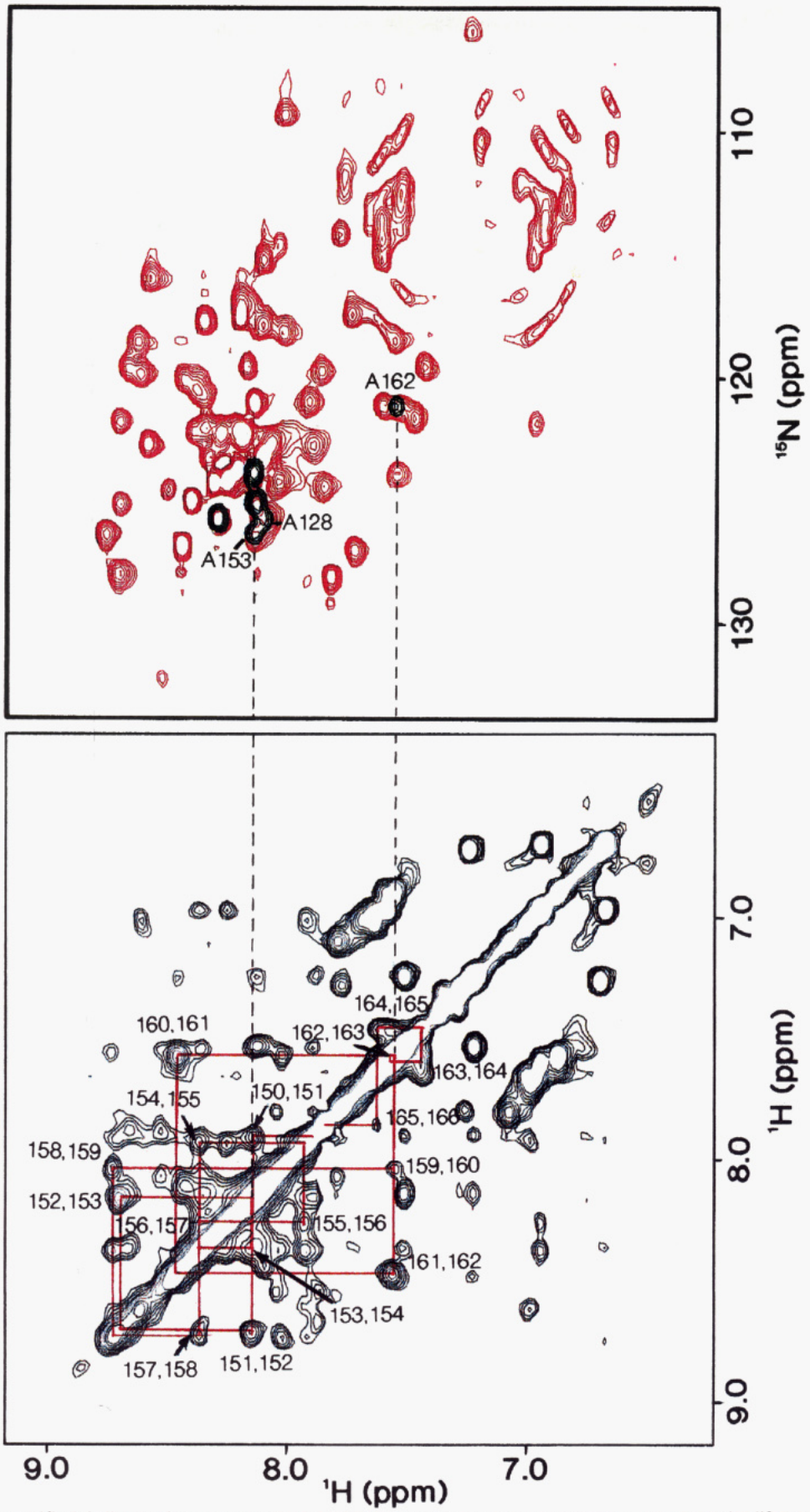

FIGURE 3: Use of specific labeling and consecutive NH–NH NOEs to assign helical segments. (Top) 2D ^{1}H/^{15}N HMQC spectrum of MyoD-bHLH labeled with ^{15}N uniformly (red) or selectively with [^{15}N]Ala (black). HMQC spectra were collected using a jump–return pulse sequence for water suppression. (Bottom) Amide region of the 2D NOESY spectrum of MyoD-bHLH. The NOESY mixing time is 150 ms. Consecutive d_{NN} connections observed for residues 150–166 are indicated. Sample conditions are 1 mM MyoD-bHLH, 50 mM Tris-d_{11}, 25 mM acetate-d_3, pH 5.4 (top) and pH 6.2 (bottom), 8% acetonitrile-d_3, 8% D_2O, 84% H_2O, and 25 °C.

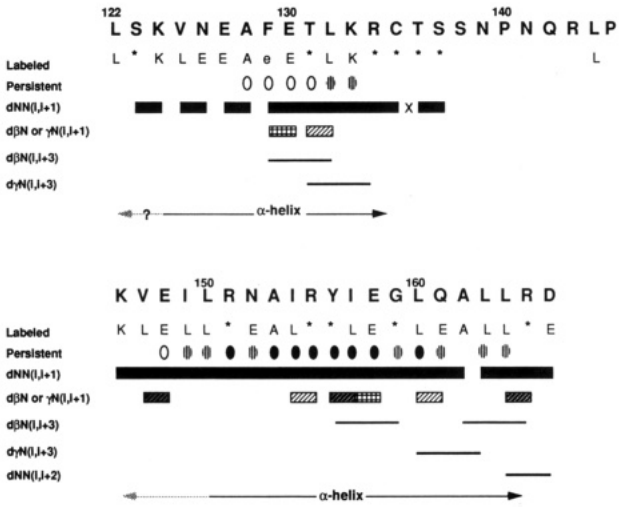

FIGURE 4: Sequential connectivities observed in NOESY spectra of oxidized MyoD-bHLH. Sequential NOEs assigned in NOESY and/or HMQC-NOESY spectra are included. This table is not inclusive of all NOEs observed due to severe spectral overlap. A blank space indicates that the presence of a given connection could not be confirmed because it would occur in a region that is overlapped with other peaks so it could not be unambiguously assigned. An x indicates the clear lack of a cross peak in a resolved region of the spectrum. $d_{\beta N}$ and $d_{\gamma N}$ connections are differentiated by grids and hatches, respectively; dense hatches indicate that both $d_{\beta N}$ and $d_{\gamma N}$ are assigned. *Labeled* indicates whether the amide peak for a given residue is present in [^{15}N]Ala (A), [^{15}N]Leu (L), [^{15}N]Glu (E), [^{15}N]Lys (K), or none of the four (*) HMQC spectra of specifically labeled MyoD-bHLH. An e indicates the amide peak for Phe-129 is overlapped with an amide peak that is labeled in the [^{15}N]Glu spectrum. Persistent NH's are those that persist after a sample that had been in H_2O is lyophilized and dissolved in D_2O. Those amides that persisted at 25 °C, pH 5.4, are indicated by open circles (>15 min), hatched circles (>2 h), and filled circles (>21 h).

R-Y-I-E-G-L-Q-A). In addition, peaks at positions 3, 9, and 12 are present in the [^{15}N]Glu spectrum, while peaks at positions 2, 6, 7, and 10 are not present in any of the four-labeled spectra, confirming these assignments (data not shown). The same logic was used to assign residues 123–135 and 146–149 and to extend assignments out to the C-terminus of the peptide. Sequential connectivities were identified using both 2D homonuclear NOESY and heteronuclear HMQC-NOESY spectra. A summary of the assigned sequential connectivities is given in Figure 4. The NOEs included represent a conservative set of those observed; chemical shift degeneracy limits the ability to unambiguously assign many sequential connectivities. Note that there are no sequential assignments for the residues from Pro-136 through Pro-145. Since our assignment protocol relies on amide protons, prolines are stopping points in the assignment process. Leu-144 was assigned as the remaining unassigned peak in the [^{15}N]Leu HMQC spectrum. It also exhibits NOEs consistent with a leucine side chain in the [^{15}N]Leu HMQC-NOESY. ^{1}H and ^{15}N chemical shifts are available as supplementary material.

Amino acid residues from the basic region of MyoD-bHLH have been omitted in Figure 4. Residues in this portion of the molecule have little chemical shift dispersion and do not generate many NOEs. By specific labeling and a process of elimination, peaks corresponding to the three Ala, one Leu, and 3 Lys residues in the basic region were identified. In HMQC spectra acquired without H_2O presaturation, these peaks are more intense and have narrower line widths than the peaks assigned to residues within the HLH domain. Also, they are much more labile to solvent exchange and can be further differentiated from HLH residues as those peaks (~20) that are not detectable in HMQC spectra acquired using H_2O presaturation at pH 7.2 (not shown). These properties indicate that the basic region in MyoD-bHLH (in the absence of DNA) is unstructured.

Secondary Structure. Sequential d_{NN} and $d_{\beta N_{i,i+3}}$ connectivities, combined with persistent NH's (see below and Figure 4), are highly indicative of α-helical structure (Wüthrich et al., 1984). By these criteria, two α-helices are defined in MyoD-bHLH: one includes approximately residues 124–135 (helix 1) and the other includes approximately residues 144–165 (helix 2). The beginning of helix 1 is presently uncertain due to chemical shift degeneracy for peaks belonging to residues 122–126. The amide ^{15}N^{1}H peaks from these five residues resonate in a region of the HMQC spectrum that contains a group of approximately 18 peaks centered around 8.2 ppm (^{1}H) and 123 ppm (^{15}N) (see Figure 3, top). Thus, for these residues α-helical structure is not confirmed by the expected sequential connectivities since they are overlapped with other peaks. However, amide protons protected from solvent exchange are observed for residues 128–133, suggesting helical H-bonds begin at residue 124. Helix 1 clearly terminates by Cys-135 where there is a resolved break in the sequential d_{NN} connections. The beginning of helix 2 is also difficult to define exactly. The uninterrupted sequential d_{NN} connectivities begin at residue 146 and persistent amides begin at residue 148, suggesting that helical secondary structure begins as early as Leu-144. Furthermore, the residue at position 145 is a proline. Prolines are often found at the N-cap +1 position in helices and have been argued to act as "helix initiators" (Richardson & Richardson, 1988), consistent with helix 2 beginning at Leu-144. Chemical shift degeneracy in the 2D spectra prevents assignment of confirmatory (e.g., $d_{\beta N_{i,i+3}}$) NOEs. The structure of residues 136–143 cannot be defined by our current analysis due to the absence of sequential assignments for these residues.

The relative degree of protection of individual amide protons from solvent exchange was measured by recording at various time points 2D HMQC spectra on uniformly ^{15}N-labeled MyoD-bHLH that had been lyophilized out of H_2O and redissolved in D_2O (Marion et al., 1989). Amide proton-exchange rates are dependent on H-bonding interactions and/or solvent accessibility (Englander & Kallenbach, 1984; Goodman & Kim, 1991). There are two segments of contiguous residues in oxidized MyoD-bHLH that are protected from hydrogen exchange. These regions are coincident with helical patterns of NOE connectivities and are defined as helix 1 and helix 2. As summarized in Figure 4, the degree of protection from solvent exchange differs significantly between the two helices. There are 6 amide resonances from helix 1 and 16 from helix 2 that persist for at least 15 min after being dissolved in D_2O. As well, 2 resonances from helix 1 and 15 from helix 2 are detectable 2 h after dissolution, and 8 resonances from helix 2 continue to persist for more than 21 h. These results demonstrate that helix 1 is more labile to solvent exchange than helix 2. Such differences can be accounted for in one of two ways: either there is more solvent access to helix 1 than helix 2 in the 3D structure or helix 1 is less stable than helix 2 due to increased flexibility or strain in the helix. It is possible that disulfide oxidation of Cys-135 at the C-terminus of helix 1 is responsible for inducing helical strain or simply that the differential chain length of helices 1 and 2 accounts for the different amide proton-exchange behaviors (Rohl et al., 1992).

Interhelical Interactions and Dimer Model. The relative positioning of secondary structural elements was evaluated by examining the NOEs observed between residues that are

far apart in the amino acid sequence. For example, the ring protons of Phe-129 exhibit NOEs to the two upfield-shifted methyl peaks of Leu-160 (Figure 5, top). These methyl groups also exhibit NOEs to the ring protons of Tyr-156; however, there is clearly no NOE between the Phe and Tyr ring protons (Figure 5, bottom). This single set of interactions through Leu-160 indicates that helix 1 and helix 2 are in close contact with one another in the 3D structure and are tethered at least through Phe-129 and Leu-160. NOEs are also observed between peaks from the following residue pairs: Phe-129/Ile-157, Phe-129/Leu-164, and Ala-128/Leu-160 (Figure 6). Many more NOEs are observed; however, due to overlap in the ^1H frequency dimension, they cannot be unambiguously assigned.

In a symmetric dimer, an NOE observed between a residue, X, on helix 1 and a residue, Y, on helix 2 could arise because (a) X is close to Y on the same polypeptide chain (intrasubunit NOE), (b) X is close to Y′ on different polypeptide chains (intersubunit NOE), or (c) X is close to both Y and Y′ (combination intra- and intersubunit NOE). Without more information we cannot distinguish among these possibilities, and hence the 3D structure of the dimer cannot be derived directly from the NMR analysis. However, several salient features of the oxidized dimer have been revealed:

(1) The dimer is symmetric.

(2) Each subunit in the dimer contains two amphipathic α-helices, helix 1 and helix 2, connected by an approximately eight-residue "loop".

(3) Helix 2 is longer than helix 1.

(4) The two subunits are covalently linked through a disulfide at the C-terminus of helix 1.

(5) The alignment of helix 1 with respect to helix 2 is such that Phe-129 is close to Ile-157, Leu-160, and Leu-164 and Ala-128 is close to Leu-160.

The distance geometry algorithm DIANA (Güntert et al., 1991) was used to examine the types of models that would satisfy the assigned NOE distance constraints. Model structures were calculated for a single polypeptide chain that included the 45 residues from Leu-122 to Asp-166 as described under Materials and Methods. If all NOEs are input as distance constraints between residues on a single polypeptide chain, then only structures that aligned helix 1 antiparallel to helix 2 satisfied all distance constraints. The acute interhelical angle of the 15 models with the lowest target functions averaged 18° ± 44°. In order to treat the distance constraints as if they were between helix 1 and helix 2′ on different polypeptide chains, residues 138–143 were replaced by 16 glycine residues and DIANA calculations were performed as before. The polyglycine loop is sufficiently long and flexible such that it does not restrict the alignment of helices 1 and 2. In this case, both parallel and antiparallel models were found that satisfied the limited interhelical distance constraints. These results indicate that the length of the loop is critical in preventing a single HLH sequence from aligning helix 1 parallel to helix 2 with Phe-129 close to Ile-157, Leu-160, and Leu-164.

Using only the secondary structural information and the interhelical NOEs derived from the NMR analysis, symmetric models were constructed with two helix 1/helix 2 units (Figure 7). "Antiparallel" models were generated from two units that have helix 1 aligned antiparallel to helix 2, and "parallel" models were generated from units that have helix 1 aligned parallel to helix 2. It is important to note that no assumption was made about which helix 1 and which helix 2 are on the same polypeptide chain. Two helix 1/helix 2 units can be packed together in a symmetric manner by aligning the units either in the same orientation (Figure 7, classes A.1 and P.1) or in the opposite orientation (Figure 7, classes A.2, A.3, P.2, and P.3) to one another. Classes A.3 and P.3 differ from classes A.2 and P.2 by a translation of the two units with

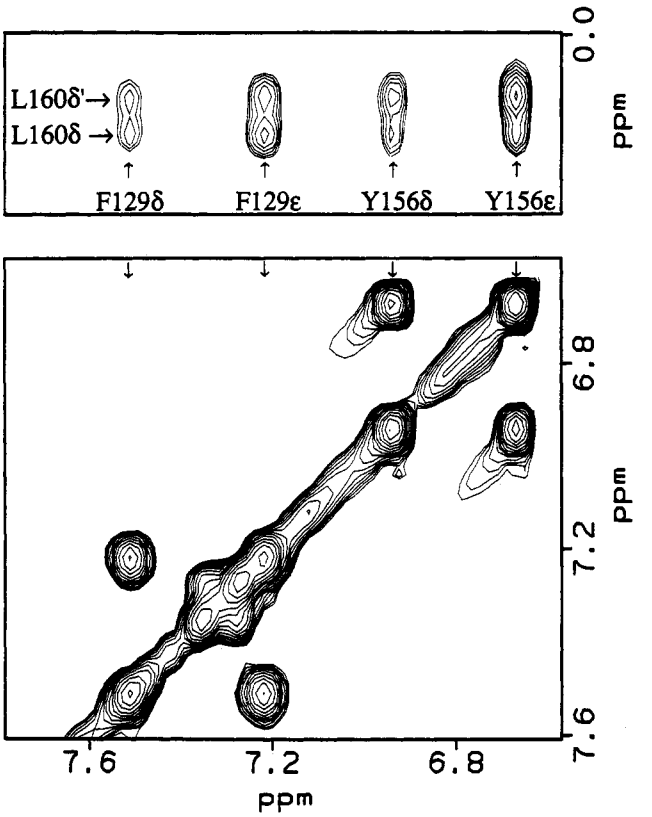

FIGURE 5: NOEs observed between aromatic resonances and upfield-shifted methyl peaks. Subsections of the D_2O NOESY spectrum of oxidized MyoD-bHLH are shown. NOEs are observed between Leu-160 and both Phe-129 and Tyr-156 (top). No NOE is observed between Phe-129 and Tyr-156 (bottom).

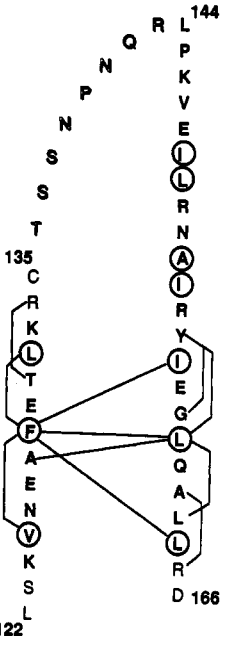

FIGURE 6: Summary of the medium- and long-range NOEs observed in oxidized MyoD-bHLH. A total of 12 long-range connectivities have been assigned between helix 1 and helix 2: six backbone–side-chain NOEs and six side-chain–side-chain NOEs. Proposed helical residues are shown boldfaced; residues from the loop are shown outlined. Hydrophobic residues conserved in bHLH protein sequences are circled (Benezra et al., 1990).

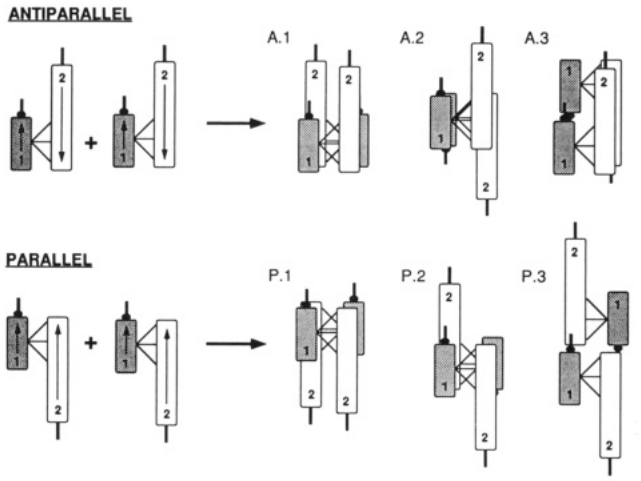

FIGURE 7: Classes of models evaluated for the oxidized MyoD-bHLH dimer. Antiparallel and parallel are classifications that refer to the rough orientation of helix 1 with respect to helix 2, with no specification of exact interhelical angle. No assumption is made about which helix 1 is connected to which helix 2. Helices 1 and 2 are labeled at their N-terminal ends with 1 or 2, respectively. The NOEs observed between Phe-129 on helix 1 and Ile-157, Leu-160, and Leu-164 are indicated by lines connecting the two helices. Cys-135 is represented by the filled circle at the C-terminus of helix 1; a disulfide connects helices 1 and 1′ through Cys-135. A covalent linker (the loop) exists between the C-terminus of helix 1 and the N-terminus of helix 2; the ends of helices 1 and 2 where the loop is attached are indicated by the thick line extensions.

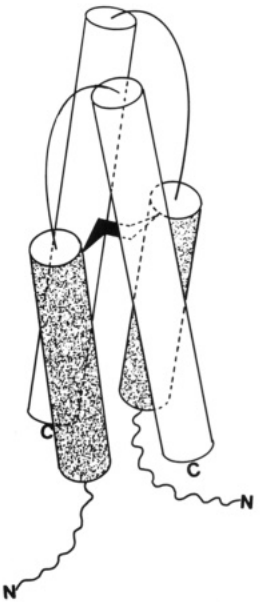

FIGURE 8: Model of the Cys-135 disulfide-bonded dimer of MyoD-bHLH. The overall topology is that of an antiparallel four-helix bundle with helix 1 diagonally opposed, roughly parallel, and covalently linked to helix 1′ by a disulfide bond. The N-termini of helices 2 and 2′ extend approximately two turns beyond the top of the bundle. The bundle is arbitrarily drawn as right-handed; however, a left-handed bundle would be equally justified. The structure of the loop is not defined, and the basic region is unstructured and flexible in the absence of DNA.

respect to each other. The models were first evaluated for whether the natural MyoD loop sequence is long enough to connect the C-terminal end of a given helix 1 to the N-terminal end of either helix 2. Models from class P.1 were ruled out by this criterion, because the loop sequence is too short to satisfy the interhelical alignment directed by the NOEs with helix 1 parallel to helix 2 on a single chain, as described above. Models were next evaluated for their ability to incorporate the Cys-135/Cys-135′ disulfide bond between the C-terminal ends of helices 1 and 1′. Models from classes A.2 and P.2 were ruled out by this criterion. This leaves only model classes A.1, A.3, and P.3 that are not restricted by the physical constraints defined thus far for the oxidized MyoD-bHLH dimer.

We turned next to results from mutagenesis and DNA-binding studies to evaluate the remaining models. The MyoD dimer requires two functional basic regions for high-affinity DNA binding (Davis et al., 1990) and preferentially binds to palindromic DNA sites (Blackwell & Weintraub, 1990). Moreover, residues at the junction between the basic region and helix 1 are critical for DNA binding (Davis et al., 1990), and mutation of basic-region residues near the junction can alter the sequence specificity of DNA binding (T. K. Blackwell and H. Weintraub, unpublished results). Thus these residues from each subunit are predicted to be juxtaposed in the DNA-bound dimer. In models A.3 and P.3, the N-termini of helices 1 and 1′ are far away from one another and thus so are their attached basic regions. Also, P.3 models have large hydrophobic surfaces exposed, particularly on helix 2. This leaves only models belonging to class A.1 that satisfy the NOE, disulfide, and loop contraints; juxtapose the two basic regions; and form a relatively compact structural domain.

Models from class A.1 represent a common tertiary structural motif known as an antiparallel four-helix bundle (Richardson, 1981). The features of this topology proposed for the oxidized MyoD-bHLH dimer are illustrated in Figure 8. The constraints described above favor a bundle arrangement with all neighboring helices roughly antiparallel to one another; however, the specific angle between helices cannot be specified by the present analysis. Furthermore, the handedness of the bundle (i.e., which helix 1 is connected to which helix 2) cannot be defined at this stage. In this model, Val-125, Phe-129, and Leu-132 from helix 1 align approximately across from Leu-164, Leu-160, and Ile-157 from helix 2, respectively (see Figure 6). As mentioned earlier, we have no detailed information about the structure of the loop region; however, the specific alignment of helix 1 relative to helix 2 is such that approximately two turns at the N-terminus of helix 2 extend beyond the C-terminal end of helix 1, thereby placing the loops on the outside of the helix 2/helix 2′ extension. Moreover, the ends of the two helices connected by the loop are not adjacent, thus placing a constraint on the minimum length required for a functional loop even in an antiparallel model.

DISCUSSION

The analysis described here represents the first structural study of a bHLH DNA-binding domain. Because the reduced form of MyoD-bHLH is not amenable to a detailed structural study by NMR, we have presented an NMR analysis of oxidized MyoD-bHLH. This form binds DNA in a sequence-specific manner but with significantly decreased affinity compared to reduced MyoD-bHLH. The results of sedimentation equilibrium ultracentrifugation indicate that the oxidized species is only a dimer, while the reduced species is primarily tetrameric at NMR concentrations. While it would be preferable to study the structure of the native bHLH domain, its behavior precludes such analysis. MyoD is recovered from myocyte nuclear extracts as a heterocomplex with E12 or E47 (Murre et al., 1991). It may be that the structural properties of reduced MyoD-bHLH are ill-behaved by design to allow preferential heterodimer formation with E12 or E47 (Sun & Baltimore, 1991). We do not expect Cys-135 to be oxidized in the cell due to the strongly reducing

intracellular environment; however, specific examples of regulatory mechanisms mediated through controlled disulfide bond formation have been reported [see, for example, Abate et al. (1990) and references therein]. The fact that oxidized MyoD-bHLH retains its ability to recognize specific-sequence DNA indicates that the formation of a disulfide bond between the two helices 1 and 1′ in the dimer leads to a structure that is capable of, but not optimal for, sequence-specific DNA binding. Therefore, we believe that the oxidized dimer of MyoD-bHLH offers insights into the HLH domain structure.

The regular secondary structural elements of oxidized MyoD-bHLH include two α-helices, helix 1 (approximately residues 124–135) and helix 2 (approximately residues 144–165). The N-terminal basic region is unstructured and flexible. The structure of the loop has not been determined due to lack of sequential assignments. Originally, bHLH domains were predicted to contain two amphipathic α-helices; putative helix 1 was comprised of 13 residues (123–135; numbering of full-length MyoD), and putative helix 2 was comprised of 12 residues (149–160) (Murre et al., 1989a). The location of residues involved in helix 1 as determined by the NMR analysis reported here is in good agreement with the prediction. The location of helix 2 is similar to what was predicted; however, helix 2 as defined by NMR includes an additional 5 residues at the C-terminus and 3–5 residues at the N-terminus. By NMR it is clear that Leu-164 is an integral part of helix 2 (Figures 4 and 6), yet has previously been overlooked as an important conserved hydrophobic residue. In fact, the position occupied by Leu-164 is conserved as a hydrophobic residue in all bHLH proteins of the MyoD subfamily [e.g., MyoD, E47, E12, *daughterless*, *twist*, and the *achaete-scute* proteins; see Figure 3 in Murre et al. (1989a)]; however, it is *not* hydrophobic, and usually charged, in the Myc subfamily of bHLH proteins that contain a leucine repeat sequence immediately following helix 2 [e.g., c-Myc, N-Myc, L-Myc, Max, AP-4, and USF; see Figure 3 in Blackwood and Eisenman (1991)]. Reduced and oxidized MyoD-bHLH have indistinguishable CD spectra, indicating that both forms contain the same amount of α-helical secondary structure. It is expected that the same residues as found for oxidized MyoD-bHLH would participate in forming helix 1 and helix 2 in reduced MyoD-bHLH.

In the dimer model proposed here, the specific alignment of helix 1 with respect to helix 2 places the C-terminus of helix 2 across from the basic region/helix 1 junction (Figure 8). Since helix 2 is about two helical turns longer than helix 1, the N-terminus of helix 2 extends beyond the C-terminus of helix 1. Residues involved in this N-terminal extension include hydrophobic residues that are highly conserved among HLH proteins (Val-147, Ile-149, Leu-150), suggesting that these residues are structurally important to the domain. It is clear that these residues are not interacting with helix 1 and so, instead, are likely to be involved in helix 2/helix 2′ interactions (Figure 8). Because helix 2 is parallel to helix 2′, any interactions between them would not be distinguishable from intrachain interactions in the symmetric dimer.

Recently, two models for HLH dimer structure were proposed on the basis of sequence analysis and computer modeling (Anthony-Cahill et al., 1992). One model has the four helices arranged in an antiparallel bundle, while the other aligns all helices roughly parallel to one another. These models were evaluated by cysteine mutagenesis. The authors favor the parallel structure due to results that appeared inconsistent with their antiparallel model. However, the antiparallel bundle that had been evaluated differs significantly from the antiparallel bundle presented here [Figure 8; compare with Figure 1B in Anthony-Cahill et al. (1992)]. The essential differences are the length of helix 2 and the specific alignment of helix 1 with respect to helix 2. Helix 2 was proposed to begin at Leu-144, consistent with our results on oxidized MyoD-bHLH. However, the results presented here indicate that helix 2 extends to Arg-165, whereas in the proposed model helix 2 terminates at Gln-161. This results in the two antiparallel models differing in the specific alignment of helix 2 with respect to helix 1 by a translation of approximately one helical turn. We believe that the failure of the oxidized cysteine loop mutant [GCGox; Figure 3B in Anthony-Cahill et al. (1992)] to bind DNA is consistent with the antiparallel bundle structure presented here because, in this model, the centers of the loops are not proximal.

Models for bHLH domains that contain all four helices in the dimer aligned roughly parallel to one another have been proposed by several groups (Halazonetis & Kandil, 1992; Anthony-Cahill et al., 1992; Vinson & Garcia, 1992). In all cases the models do not align helices 1 and 2 such that the observed NOEs in oxidized MyoD-bHLH would be satisfied. The key feature of the MyoD-bHLH sequence that rules out an entirely parallel bundle arrangement is the number of residues (approximately 8) that must stretch from the C-terminal end of helix 1 to the N-terminal end of helix 2 (Figure 7, class P.1). The loop lengths of many bHLH proteins are identical to that of MyoD [see Benezra et al. (1990)]. Furthermore, two bHLH domains from the maize proteins R-Lc and B-Peru have loop lengths that are two residues shorter than that of MyoD (Ludwig et al., 1989; Radicella et al., 1991). None of these proteins would be expected to be able to fold into parallel bundles with the alignment of the conserved hydrophobic residues as determined here for oxidized MyoD-bHLH. To examine in more detail how the DNA-binding activity of MyoD depends on the length of the loop, MyoD proteins containing a deletion of 1, 2, or 4 residues in the loop [corresponding to deletions of residue(s) S138, S138–N139, and/or S138–N141, respectively] were prepared and their DNA-binding activities were examined by EMSA. High-affinity DNA-binding activity was maintained when either one or two residues were deleted (H. Weintraub, unpublished results). The observation that proteins with a sequence even shorter than that of native MyoD can still bind DNA with high affinity further refutes a parallel alignment of helices 1 and 2 in the DNA-bound structure. When four residues were deleted from the loop, high-affinity DNA-binding activity was lost (H. Weintraub, unpublished result). This is in agreement with the antiparallel alignment proposed here that requires a sufficiently long loop to traverse the portion of helix 2 that extends beyond the end of helix 1 (Figure 8).

The quaternary topology of the oxidized MyoD-bHLH dimer suggested by the studies presented here is an antiparallel four-helix bundle, that is, a nearly antiparallel arrangement of four α-helices to form a left-twisted bundle (Argos et al., 1977; Weber & Salemme, 1980). Among proteins whose secondary structure is primarily α-helical, the antiparallel four-helix bundle is the most common tertiary structural motif and is found in at least 20 crystal structures of proteins, many of which are unrelated by sequence homology or function (Richardson, 1981; Presnell & Cohen, 1989). In most cases, the bundle is composed of helices from a single polypeptide chain; however, the Rop protein from plasmid ColE1 builds its four-helix bundle structure from two identical subunits (Banner et al., 1987). Rop forms a symmetric dimer from two pairs of antiparallel helices such that all adjacent helices

are antiparallel. The way that the two subunits come together in Rop differs from that proposed for MyoD-bHLH. Helices 1 and 1′ (and helices 2 and 2′) of Rop are antiparallel neighbors in the bundle (similar to class A.2, Figure 7), whereas they are parallel and at opposite corners in the MyoD-bHLH bundle. This topology results in the two basic regions of MyoD-bHLH extending from the same end of the bundle structure, allowing each basic region access to interactions with DNA.

If the overall topology of oxidized MyoD-bHLH were generally the same as reduced MyoD-bHLH, what might account for the observed difference in DNA-binding affinities? It has been demonstrated repeatedly that residues at the basic region/helix 1 junction of each subunit are essential for DNA binding and are responsible for determining DNA-binding specificity at the central base pairs of the bHLH consensus sequence (Davis et al., 1990; Dang et al., 1992; Halazonetis & Kandil, 1992; T. K. Blackwell and H. Weintraub, unpublished results). These results suggest that the N-terminal ends of helix 1 and helix 1′ not only must be close in space but must have a specific orientation with respect to one another to allow the adjacent basic regions to simultaneously fit in the major groove and make specific contacts with DNA (Lassar et al., 1989). The oxidized dimer model presented here places the two basic regions close to one another; however, the covalent link between helix 1 and helix 1′ fixes the relative orientation and distance between the two helices and, therefore, the two basic regions as well. The decrease in DNA-binding affinity exhibited by the oxidized relative to the reduced form of MyoD-bHLH suggests that the orientation dictated by the disulfide bond is not optimal for DNA binding. Decreased flexibility of the structure imposed by the disulfide cross-linking may also contribute to the loss in DNA-binding affinity. In several studies of λ and cro repressor proteins, the introduction of nonnative disulfides that did not appear to disrupt structure still resulted in a decrease in DNA-binding affinity (Sauer et al., 1986; Hubbard et al., 1990; Shirakawa et al., 1991). In these cases, the observed loss in binding affinity was attributed to a loss in flexibility.

The bHLH domain has been identified in many protein sequences. A subset of these proteins also contains a sequence immediately adjacent to the C-terminus of the bHLH domain that consists of conserved leucines every seventh residue with sequence similarity to leucine-zipper proteins. Members of this subfamily are identified as bHLH-Zip proteins and include the Myc protein family, AP-4 (Hu et al., 1990), Max (Blackwood & Eisenman, 1991), USF (Gregor et al., 1990), TFE3 (Beckmann et al., 1990), and TFEB (Fisher et al., 1991). MyoD does *not* contain an adjacent leucine repeat sequence and neither do the proteins with which it forms heterodimers (e.g., E47 and E12; Murre et al., 1989a,b). Does the bHLH region form the same structure in all bHLH-containing proteins regardless of whether a leucine repeat sequence is attached? There are two notable differences in HLH sequence conservation between the MyoD subfamily and the Myc subfamily. The positions occupied by Leu-122 and Leu-164 (MyoD numbering) are conserved as hydrophobic only in proteins lacking the leucine repeat. In bHLH-Zip proteins, an arginine replaces the corresponding Leu-122, and this arginine is involved in determining DNA-binding specificity (Dang et al., 1992; Halazonetis & Kandil, 1992; T. K. Blackwell and H. Weintraub, unpublished results). These differences in hydrophobic amino acid sequence conservation suggest that at least slight differences in structure may exist between the two subfamilies.

An open question remains as to whether the leucine repeat sequence of bHLH-Zip proteins acts as a true leucine-zipper dimerization domain forming a parallel coiled coil (as was found for the GCN4 leucine zipper; O'Shea et al., 1991). The antiparallel bundle topology presented here places the C-termini of helices 2 and 2′ in proximity to the basic regions (see Figure 8). This topology in combination with a leucine-repeat-mediated coiled-coil extension would be incompatible with DNA binding through the basic regions due to steric crowding among the basic regions, leucine zipper, and DNA. There are no structural data obtained on proteins from the bHLH-Zip family; however, models suggesting a parallel arrangement of helices in the HLH dimer, based on the incorporation of a parallel coiled coil in the zipper region, have been proposed (Halazonetis & Kandil, 1992; Anthony-Cahill et al., 1992; Vinson & Garcia, 1992). Perhaps the HLH region of bHLH proteins belongs to a structural class different from that of bHLH-Zip proteins.

In summary, results of the analyses presented here suggest that the folding topology of oxidized MyoD-bHLH is an antiparallel four-helix bundle. Because this peptide maintains the ability to recognize DNA sequence specifically, it may serve as a general model for the bHLH protein domain structure. We expect the topologies of other bHLH homo- and heterodimers of the MyoD subfamily also to be antiparallel four-helix bundles. Future structural and mutational experiments will test this hypothesis.

ACKNOWLEDGMENT

We thank Dan Lockshon for developing the expression system, Bev Castner for protein purification, and Lars Pederson for acquiring the CD spectra. Also, we thank Ross Hoffman and Mia Schmiedeskamp for many helpful discussions and comments and Trisha Davis, Jon Herriott, David Wemmer, and Michael Wittekind for critical reading of the manuscript.

SUPPLEMENTARY MATERIAL AVAILABLE

A table of ^{1}H and ^{15}N chemical shifts and a list of the distance bounds used for DIANA calculations (6 pages). Ordering information is given on any current masthead page.

REFERENCES

Abate, C., Patel, L., Rauscher, F. J., III, & Curran, T. (1990) *Science 249*, 1157–1161.

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

Argos, P., Rossmann, M. G., & Johnson, J. E. (1977) *Biochem. Biophys. Res. Commun. 75*, 83–86.

Banner, D. W., Kokkinidis, M., & Tsernoglou, D. (1987) *J. Mol. Biol. 196*, 657–675.

Bax, A., Griffey, R. H., & Hawkins, B. L. (1983) *J. Magn. Reson. 55*, 301–315.

Beckmann, H., Su, L.-K., & Kadesch, T. (1990) *Genes Dev. 4*, 167–179.

Bendall, M. R., Pegg, D. T., & Doddrell, D. M. (1983) *J. Magn. Reson. 52*, 81–117.

Benezra, R., Davis, R. L., Lockshon, D., Turner, D. L., & Weintraub, H. (1990) *Cell 61*, 49–59.

Blackwell, T. K., & Weintraub, H. (1990) *Science 250*, 1104–1110.

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

Buskin, J. N., & Hauschka, S. D. (1989) *Mol. Cell. Biol. 9*, 2627–2640.

Clore, G. M., & Gronenborn, A. M. (1991) *Annu. Rev. Biophys. Chem. 20*, 29–63.
Clore, G. M., Bax, A., Wingfield, P. T., & Gronenborn, A. M. (1988) *FEBS Lett. 238*, 17–21.
Creighton, T. E. (1984) *Proteins: Structures and Molecular Properties*, p 17, W. H. Freeman and Co., New York.
Dang, C. V., Dolde, C., Gillison, M. L., & Kato, G. J. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 599–602.
Davis, R. L., Cheng, P.-F., Lassar, A., & Weintraub, H. (1990) *Cell 60*, 733–746.
Davis, R. W., Botstein, D., & Roth, J. R. (1980) *A Manual for Genetic Engineering: Advanced Bacterial Genetics*, p 207, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Dekker, N., Peters, A. R., Slotboom, A. J., Boelens, R., Kaptein, R., & de Haas, G. (1991) *Biochemistry 30*, 3135–3147.
Englander, S. W., & Kallenbach, N. R. (1984) *Q. Rev. Biophys. 16*, 521–655.
Fisher, D. E., Carr, C. S., Parent, L. A., & Sharp, P. A. (1991) *Genes Dev. 5*, 2342–2352.
Goodman, E. M., & Kim, P. S. (1991) *Biochemistry 30*, 11615–11620.
Gregor, P. D., Sawadogo, M., & Roeder, R. G. (1990) *Genes Dev. 4*, 1730–1740.
Griffey, R. H., & Redfield, A. G. (1987) *Q. Rev. Biophys. 19*, 51–82.
Griffey, R. H., Redfield, A. G., Loomis, R. E., & Dahlquist, F. W. (1985) *Biochemistry 24*, 817–822.
Gronenborn, A. M., Bax, A., Wingfield, P. T., & Clore, G. M. (1989) *FEBS Lett. 243*, 93–98.
Güntert, P., Braun, W., & Wüthrich, K. (1991) *J. Mol. Biol. 217*, 517–530.
Halazonetis, T. D., & Kandil, A. N. (1992) *Science 255*, 464–466.
Hu, Y.-F., Lüscher, B., Admon, A., Mermed, N., & Tjian, R. (1990) *Genes Dev. 4*, 1741–1752.
Hubbard, A. J., Bracco, L. P., Eisenbeis, S. J., Gayle, R. B., Beaton, G., & Caruthers, M. H. (1990) *Biochemistry 29*, 9241–9249.
Johnson, M. L., Correia, J. C., Yphantis, D. A., & Halvorson, H. R. (1981) *Biophys. J. 36*, 575–588.
Kharakoz, D. P. (1989) *Biophys. Chem. 34*, 115–125.
Lassar, A. B., Buskin, J. N., Lockshon, D., Davis, R. L., Apone, S., Hauschka, S. D., & Weintraub, H. (1989) *Cell 58*, 823–831.
Laue, T. M. (1992) in *Analytical Ultracentrifugation in Biochemistry and Polymer Science* (Harding, S., & Rowe, A., Eds.) Chapter 6, Royal Society of Chemistry, London (in press).
Laue, T. M., Johnson, A. E., Esmon, C. T., & Yphantis, D. A. (1984) *Biochemistry 23*, 1339–1348.
Laue, T. M., Shah, B., Ridgeway, T. M., & Pelletier, S. L. (1992) in *Analytical Ultracentrifugation in Biochemistry and Polymer Science* (Harding, S., & Rowe, A., Eds.) Chapter 7, Royal Society of Chemistry, London (in press).
Levy, G. C., & Lichter, R. L. (1979) *Nitrogen-15 Nuclear Magnetic Resonance Spectroscopy*, J. Wiley & Sons, New York.
Ludwig, S. R., Habera, L. F., Dellaporta, S. L., & Wessler, S. R. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 7092–7096.
Marion, D., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967–974.
Marion, D., Ikura, M., Tschudin, R., & Bax, A. (1989) *J. Magn. Reson. 85*, 393–399.
McLeod, M., Stein, M., & Beach, D. (1987) *EMBO J. 6*, 729–736.
McMeekin, T. L., & Marshall, K. (1952) *Science 116*, 142–143.
Murre, C., McCaw, P. S., & Baltimore, D. (1989a) *Cell 56*, 777–783.
Murre, C., McCaw, P. S., Vaessin, H., Caudy, M., Jan, L. Y., Jan, Y. N., Cabrera, C. V., Buskin, J. N., Hauschka, S. D., Lassar, A. B., Weintraub, H., & Baltimore, D. (1989b) *Cell 58*, 537–544.
Murre, C., Voronova, A., & Baltimore, D. (1991) *Mol. Cell. Biol. 11*, 1156–1160.
O'Shea, E. K., Klemm, J. D., Kim, P. S., & Alber, T. (1991) *Science 254*, 539–544.
Párraga, G., & Klevit, R. E. (1991) *Methods Enzymol. 208*, 63–82.
Perlman, G. E., & Longsworth, L. (1948) *J. Am. Chem. Soc. 70*, 2719–2724.
Phillips, C. L., Vershon, A. K., Johnson, A. D., & Dahlquist, F. W. (1991) *Genes Dev. 5*, 764–772.
Presnell, S. R., & Cohen, F. E. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 6592–6596.
Radicella, J. P., Turks, D., & Chandler, V. L. (1991) *Plant Mol. Biol. 17*, 127–130.
Richardson, J. S. (1981) *Adv. Protein Chem. 34*, 167–339.
Richardson, J. S., & Richardson, D. C. (1988) *Science 240*, 1648–1652.
Rohl, C. A., Scholtz, J. M., York, E. J., Stewart, J. M., & Baldwin, R. L. (1992) *Biochemistry 31*, 1263–1269.
Roy, S., Papastavros, M. Z., Sanchez, V., & Redfield, A. G. (1984) *Biochemistry 23*, 4395–4400.
Sauer, R. T., Hehir, K., Stearman, R. S., Weiss, M. A., Jeitler-Nilsson, A., Suchanek, E. G., & Pabo, C. O. (1986) *Biochemistry 25*, 5992–5998.
Shaka, A. J., Barker, B. P., & Freeman, R. (1985) *J. Magn. Reson. 64*, 547–552.
Shirakawa, M., Matsuo, H., & Kyogoku, Y. (1991) *Protein Eng. 4*, 545–552.
Studier, F. W. (1991) *J. Mol. Biol. 219*, 37–44.
Studier, F. W., & Moffatt, B. A. (1986) *J. Mol. Biol. 189*, 113–130.
Sun, X.-H., & Baltimore, D. (1991) *Cell 64*, 459–470.
Tanford, C. (1961) in *Physical Chemistry of Macromolecules*, p 230, J. Wiley & Sons, Inc., New York.
Tapscott, S. J., Davis, R. L., Thayer, M. J., Cheng, P.-F., Weintraub, H., & Lassar, A. B. (1988) *Science 242*, 405–411.
Torchia, D. A., Sparks, S. W., & Bax, A. (1989) *Biochemistry 28*, 5509–5524.
Vinson, C. R., & Garcia, K. C. (1992) *New Biol. 4*, 396–403.
Voronova, A., & Baltimore, D. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 4722–4726.
Weber, P. C., & Salemme, F. R. (1980) *Nature 287*, 82–84.
Weintraub, H., Tapscott, S. J., Davis, R. L., Thayer, M. J., Adam, M. A., Lassar, A. B., & Miller, A. D. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 5434–5438.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.
Wüthrich, K. (1989) *Methods Enzymol. 177*, 125–131.
Wüthrich, K., Billeter, M., & Braun, W. (1984) *J. Mol. Biol. 180*, 715–740.
Yphantis, D. A. (1960) *Ann. N.Y. Acad. Sci. 88*, 586–601.

Registry No. CK, 9001-15-4.