The LEF-1 High-Mobility Group Domain Undergoes a Disorder-to-Order Transition upon Formation of a Complex with Cognate DNA[†]

John J. Love,[‡] Xiang Li,[§] John Chung, H. Jane Dyson, and Peter E. Wright*

Department of Molecular Biology and Skaggs Institute of Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037

Received February 27, 2004; Revised Manuscript Received April 28, 2004

ABSTRACT: Lymphoid enhancer-binding factor-1 (LEF-1), a member of the high-mobility group (HMG) family of proteins, functions as an architectural transcription factor. In complex with its cognate DNA, the LEF-1 domain is highly ordered, and its NMR spectra are characteristic of a folded globular protein. In contrast, the uncomplexed protein exhibits NMR evidence of substantial conformational heterogeneity, although circular dichroism spectra indicate that much of the α-helical secondary structure of the DNA-bound state is retained in the free protein. Heteronuclear NMR experiments performed on the free LEF-1 domain reveal that helix II and helix III of the HMG domain are folded, although helix III is truncated at its C-terminal end relative to the DNA-bound protein. The major hydrophobic core between helices II and III appears to be formed, but the minor core near the C-terminus of helix III is unstructured in the free protein. Backbone resonances of helix I are undetectable, probably as a result of exchange broadening due to fluctuations between two or more conformations on a microsecond-to-millisecond time scale. On the basis of the circular dichroism spectrum, this region of the polypeptide appears to adopt helical structure but the helix is not fully stabilized in the absence of DNA. These findings argue that, prior to binding, bending, and distorting DNA, the HMG domain of LEF-1 exists in a segmentally disordered or partially folded state. Upon complex formation, the protein domain undergoes a cooperative folding transition with DNA to a highly ordered and well-folded state.

Lymphoid enhancer-binding factor-1 (LEF-1)[1] is a sequence-specific and cell type-specific transcription factor that plays an important role in regulation of the T cell receptor (TCR)-α gene enhancer (*1−3*). By bending DNA through an angle of approximately 100°, LEF-1 facilitates interactions between transcription factors (e.g., Ets-1, PEBP2α, and ATF-CREB) bound at sites flanking the LEF-1 binding site (*1*, *4*, *5*). LEF-1 appears to regulate the enhancer through its architectural role in assembling and stabilizing this higher-order nucleoprotein complex (*5*, *6*). LEF-1, and the closely related T cell factors (TCF), have also been shown to interact with β-catenin and play important roles as nuclear mediators of Wnt signaling (*7*, *8*). Sequence homology between the DNA binding domain of LEF-1 and the high-mobility group HMG1 and -2 proteins places LEF-1 in the HMG family of proteins and classifies its DNA binding domain as an "HMGB" domain (*6*, *9*).

The family of proteins containing HMGB domains can be divided into two subclasses. One subclass consists of a ubiquitous set of proteins (typified by HMG1 and HMG2) that usually have multiple HMG domains, bind DNA with little or no specificity, and are found in most cell types. Proteins in this subclass recognize DNA structure rather than sequence and therefore preferentially bind prebent DNA, e.g., Holliday junctions or DNA damaged by chemotherapeutic agents such as cisplatin (*10*, *11*). Members of the second subclass, which includes LEF-1, usually contain a single HMG domain and are cell type-specific transcription factors that recognize and bend specific sequences of DNA. Proteins of both subclasses have been shown to bind DNA primarily in the minor groove (*6*, *12*, *13*). High-resolution structures have been determined for HMG domains from both subclasses, including examples from the sequence-specific subclass in complex with their cognate DNA sequences (*14*, *15*) and from the non-sequence-specific subclass with various DNA sequences and DNA adducts modified by cisplatin (*16−21*).

LEF-1 is one of the best-characterized examples of how an HMG domain functions to regulate transcription through the direct architectural modification of a specific sequence of DNA. Biochemical analysis (*22*) and subsequent solution structure elucidation of the LEF-1 HMG domain−DNA complex (*14*) revealed that the LEF-1 domain binds predominantly in the minor groove of DNA and results in distortion of the DNA from normal B-form geometry. The DNA bending and distortion are achieved by the combination of relatively extensive hydrophobic interactions in the widened minor groove and favorable electrostatic interactions in the proximity of the compressed major groove
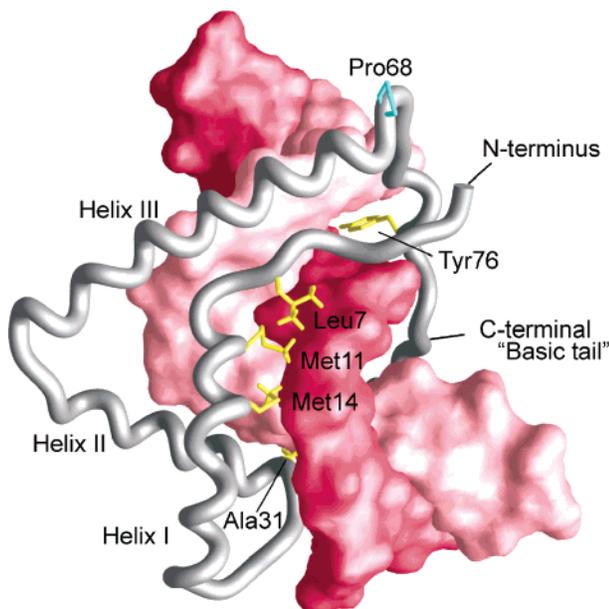
FIGURE 1: GRASP image illustrating the hydrophobic contacts between the concave side of the L-shaped LEF-1 domain (backbone worm) within the minor groove of DNA. Each DNA strand is depicted as a surface image with the Cyt 1−Cyt 15 strand colored pink and the Gua 16−Gua 30 strand colored red. Side chains that form the closest hydrophobic contacts with the DNA are colored yellow. The side chains of Leu 7, Met 11, and Met 14 form a hydrophobic patch with the methyl group of Met 11 penetrating deeply between Ade 23 and Ade 24 at the primary site of intercalation. The breaking of helix III by Pro 68 allows the protein backbone to cross back over the compressed minor groove where Tyr 76 is inserted and packs against Ade 11 and Gua 12. The secondary site of partial intercalation is formed by the penetration of the methyl group of Ala 31 between the bases of Cyt 5 and Thy 6. This view also illustrates the close fit of the LEF-1 domain within the vastly widened minor groove (adapted from ref *14*).

(Figure 1). The minor groove interactions involve the partial insertion of hydrophobic side chains into cavities where optimal DNA base stacking has been disrupted. On the side opposite the hydrophobic interaction, residues from a highly basic region, located immediately C-terminal to the canonical HMG domain (termed here the "basic tail"), bind across the compressed major groove and function to neutralize the negative charges of the phosphate oxygen atoms as they come into unfavorable proximity upon DNA bending. The contact between the LEF-1 domain and DNA is relatively extensive and quite intimate and results in the burial of approximately 1600 Å^2 of molecular surface area upon binding (*14*).

Structural and dynamic information for free HMG domains from the non-sequence-specific subclass have been gathered, including HMG1-A (*16, 23*), HMG1-B (*24, 25*), HMG-D (*26*), and NHP6A (*27*). In addition, the structures of two of the six tandem domains of hUBF were determined by NMR (*28, 29*). The hUBF protein is unique in that it partly bridges the two subclasses because it binds specifically to promoters of RNA polymerase I-transcribed genes but with "relaxed" specificity and without any discernible DNA recognition sequence (*30, 31*). To date, only two structures of free HMG domains from the sequence-specific subclass have been reported: Sox-4 (*32*) and Sox-5 (*33−35*).

All of the currently available high-resolution structures of the HMG domains, whether in the presence or absence of

DNA, have the same fundamental architecture. The proteins are L-shaped, with the short arm formed by two antiparallel helices (helices I and II) and the long arm formed by helix III and by an extended region at the N-terminus (Figure 1). In all structures, helices I and II and the N-terminal region of helix III enclose a central hydrophobic core, which will be termed here the major hydrophobic core. Domains from the non-sequence-specific subclass [e.g., the HMG1 and HMG-D domains (*24−26*)] form a secondary and less extensive hydrophobic core or strip between helix III and the N-terminal strand. In the DNA-bound states of sequence-specific domains LEF-1 and SRY, this minor hydrophobic core is largely restricted to a small cluster of residues in a turn that terminates helix III and a single hydrophobic side chain near the beginning of the N-terminal strand (*14, 15*).

Although the HMG domain of LEF-1 is well-folded in the presence of DNA (*14*), NMR spectra of the free protein reveal that it is not well-structured. LEF-1 thus joins a growing number of DNA-binding proteins that are incompletely folded in the absence of DNA and undergo folding transitions upon binding to their specific DNA targets (*36−39*). This paper describes the combined use of heteronuclear multi-dimensional NMR and circular dichroism (CD) to characterize the structure of the free HMG domain of LEF-1 and to elucidate the conformational changes and increase in structural order that occur upon binding cognate DNA.

MATERIALS AND METHODS

Protein and DNA Preparation and Purification. The HMG domain of LEF-1 was prepared by overexpression in *Escherichia coli*. The gene corresponding to the minimal HMG domain was subcloned from a 105-amino acid GST fusion construct supplied by R. Grosschedl (*22*). The gene corresponding to the 86-amino acid minimal DNA-binding domain was subcloned into a T7 vector (pET21a, Novagen) and transformed into *E. coli* strain BL21(DE3) for protein expression. Cells were grown in M9 minimal medium to an OD$_{600}$ of 1.2 before induction with IPTG (∼1 mM). The protein was expressed as inclusion bodies (i.e., it pelleted with the insoluble portion of disrupted *E. coli* cells) that were separated from the soluble cell contents by centrifugation at 6000 rpm after cellular disruption by sonication. The inclusion bodies were solubilized with 6 M guanidine hydrochloride (GuHCl) and separated from insoluble aggregates by centrifugation. The soluble fraction was dialyzed to remove GuHCl and centrifuged. The supernatant containing the LEF-1 domain was purified to homogeneity by reverse phase HPLC and lyophilized. Protein expression levels were enhanced by the cotransformation of pUBS520, containing the gene *argU* (dnaY) which encodes the minor arginine tRNA$_{AGA/AGG}$ (*40, 41*). The yield of the purified protein from expression in minimal medium was ∼20 mg/L of medium. Labeled protein for NMR studies was prepared by standard methods, using [^{15}N]ammonium sulfate (2 g/L) and [^{13}C]-glucose (2 g/L). The protein purity was verified by standard SDS−PAGE and reverse phase HPLC, and the correct molecular weight was confirmed by ion spray mass spectrometry. Residue numbering for the LEF-1 HMG domain in this paper is slightly different from that in the paper with the published structure (*14*). To match the formatting found in the Protein Data Bank file for the complex structures (2LEF), the N-terminal methionine residue is now numbered
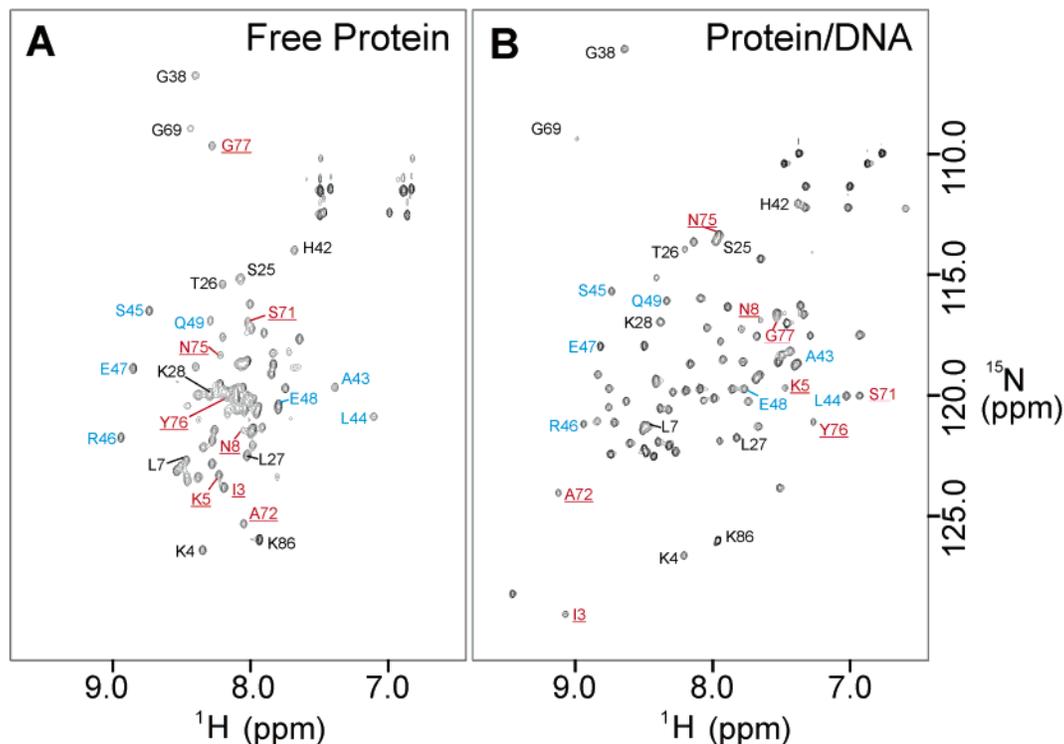
FIGURE 2: $^1H-^{15}N$ HSQC spectra of the (A) free and (B) DNA-bound LEF-1 HMG domain. Residues that exhibit minimal (blue), moderate (black), and significant (red) differences in chemical shifts between states are denoted.

residue 1, resulting in an increase of 1 for the numbering of each residue.

The 15 bp DNA duplex, containing the consensus recognition sequence for the LEF-1 HMG domain (5'-CACCCTTTGAAGCTC), was synthesized and purified as previously described (*14*).

NMR Spectroscopy. NMR spectra were acquired on Bruker spectrometers operating at 500, 600, and 750 MHz. The triple-resonance experiments used to obtain backbone assignments were carried out at 295 K on the AMX-500 MHz spectrometer equipped with a triple-resonance, triple-axis gradient probe. Proton chemical shifts were referenced to ^1H_2O or residual $^1H^2HO$ at 4.79 ppm relative to 2,2-dimethyl-2-silapentane-5-sulfonate (DSS). ^{13}C and ^{15}N chemical shifts were referenced indirectly to DSS using ratios of 0.251 449 530 for ^{13}C and 0.101 329 188 for ^{15}N (*42*).

Triple-resonance data were collected on 2 mM doubly ^{13}C- and ^{15}N-labeled LEF-1 HMG domain at pH 6.0 in a 90% H_2O/10% ^2H_2O mixture containing 10 mM KCl, 50 μM NaN$_3$, and 10 μM EDTA. The following triple-resonance experiments were performed: CBCA(CO)NH (*43*) with a spectral width of 4032 Hz and 512 complex points in ω_3 (^1H), a spectral width of 1050 Hz and 31 complex points in ω_2 (^{15}N), and a spectral width of 8333 Hz and 57 complex points in ω_1 (^{13}C); HNCACB (*44*) with a spectral width of 4032 Hz and 512 complex points in ω_3 (^1H), a spectral width of 1050 Hz and 30 complex points in ω_2 (^{15}N), and a spectral width of 8333 Hz and 60 complex points in ω_1 (^{13}C); C(CO)NH-TOCSY (*45*) with a spectral width of 4032 Hz and 512 complex points in ω_3 (^1H), a spectral width of 1050 Hz and 31 complex points in ω_2 (^{15}N), and a spectral width of 8333 Hz and 60 complex points in ω_1 (^{13}C), with a mixing time of 15.2 ms.

Exploration of Optimal Solution Conditions for NMR. $^1H-^{15}N$ HSQC spectra of the free LEF-1 domain were collected under a variety of conditions on the AMX-500, AMX-600, and DRX-750 MHz spectrometers. The various solution conditions that were explored included the following: temperature (5, 10, 12, 22, and 30 °C), pH (5.0, 6.0, 6.75, and 7.0), salt concentration (50, 100, and 150 mM NaCl, 150 mM Na$_2$SO$_4$, and 150 mM sodium phosphate), and protein concentration (0.3, 1.0, 1.5, and 2.5 mM). Regardless of the conditions, all $^1H-^{15}N$ HSQC spectra of the protein free in solution exhibit heterogeneous line broadening and are consistently poorly dispersed.

Circular Dichroism. CD spectra were collected on an AVIV model 61 DS spectropolarimeter. The spectra were recorded at 22 °C. To eliminate experimental error and avoid subtraction of the free DNA signal from that of the protein–DNA complex, all spectra were recorded in a 10 mm quartz tandem mix cell. Both the free LEF-1 domain and the duplex DNA were in 10 mM potassium phosphate buffer (pH 7) containing 50 mM KCl. Before mixing was carried out, one cell of the cuvette contained 5 μM protein while the other contained 5 μM duplex DNA (15 bp). Spectra were scanned from 320 to 195 nm in 0.5 nm steps with a bandwidth of 1.50 nm and an averaging time of 4 s/point.

RESULTS

Solution Conditions for the NMR Study of the LEF-1 HMG Domain. Two-dimensional $^1H-^{15}N$ HSQC spectra of the uniformly ^{15}N-labeled LEF-1 domain were recorded for both the free protein (Figure 2A) and the protein in complex with a 15 bp DNA duplex containing its consensus binding sequence (Figure 2B). The spectrum of the protein bound to the DNA is clearly better than that of the free form. The spectrum of the complex contains the expected 83 backbone
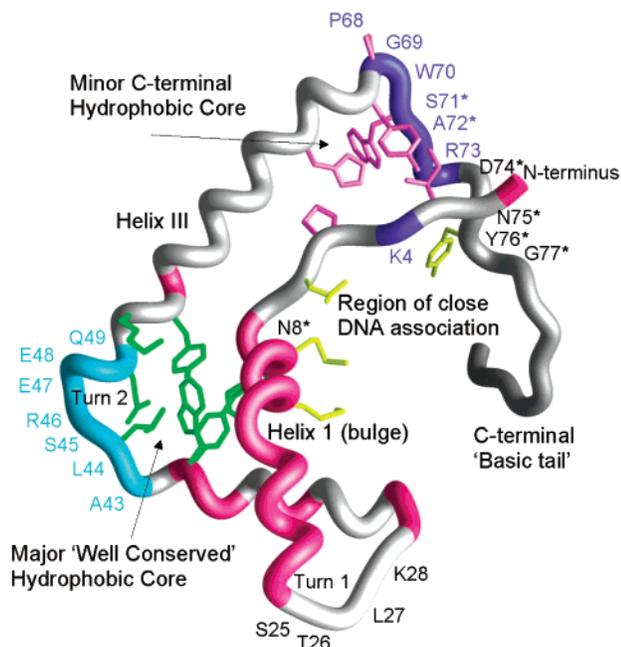
FIGURE 3: Assignments and solution characteristics of the free LEF-1 HMG domain mapped onto the structure of the DNA-bound form (2LEF). The coloring scheme for the backbone worm is as follows. Regions colored pink correspond to residues for which no assignments were made due to the absence of signals. Regions of the worm that are colored light blue exhibit a minimal change in the $^1H-^{15}N$ chemical shift assignments when the free state is compared to the DNA-bound state (turn 2). Regions of the worm colored dark blue correspond to residues that exhibit two sets of NMR signals in the free state. The coloring scheme for the alphanumeric labels is as follows. Residues highlighted with an asterisk (*) are those which exhibit significant changes in the $^1H-^{15}N$ chemical shift assignments upon DNA binding; residues labeled in black exhibit moderate changes, and those labeled in light blue exhibit little to no change in the $^1H-^{15}N$ chemical shift assignments when the free state is compared to the DNA-bound state. The side chains of residues within the central well-conserved, major hydrophobic core are colored green, while those in the smaller C-terminal, minor hydrophobic core are colored violet. The side chains of residues that interact very closely with DNA are colored yellow.

amide peaks, exhibits significant dispersion, and is homogeneous in line width and peak intensity. In contrast, the $^1H-^{15}N$ HSQC spectrum of the free protein is poorly dispersed, contains only ~63 of a potential 83 backbone amide peaks, and exhibits heterogeneous line width and intensity, all characteristics of a flexible and dynamic molecule. A wide variety of solution conditions were explored in an effort to improve the spectra (see Materials and Methods), but the line widths and signal dispersion remained consistently poor and very similar to those of the spectrum shown in Figure 2A.

The 86-amino acid fragment of LEF-1 used in the NMR studies contains a nine-residue basic tail located at the C-terminus (Figures 1 and 3). Although this basic tail has been shown to enhance the affinity of the protein for DNA and modulate the DNA bend angle (46), it is not part of the canonical HMGB domain as originally classified (9). The substantial amount of structural and functional data subsequently published on HMGB domains has revealed the important role that N- and C-terminal basic regions play in binding and bending DNA. To ascertain whether the basic tail contributes to the poor quality of the NMR spectra of

the free protein, a truncated version of the LEF-1 domain was expressed lacking the nine-residue basic tail. The spectrum of this 77-amino acid fragment was even poorer in quality than that of the longer 86-amino acid construct.

To facilitate determination of the three-dimensional structure of the LEF-1−DNA complex, the sole cysteine residue in the HMG domain, Cys 25, was mutated to serine to stabilize the protein against oxidation during NMR data acquisition (14). It was experimentally ascertained that the C25S mutation did not affect the DNA binding function (R. Grosschedl, personal communication). To determine whether the disorder exhibited by the free domain was due to this functionally silent mutation, the wild-type protein (i.e., containing Cys at residue 25) was expressed. The $^1H-^{15}N$ HSQC spectrum of this species is as poorly dispersed as that of the C25S mutant domain. All subsequent experiments were therefore performed on the original C25S mutant.

Self-Association of the Free LEF-1 HMG Domain. At NMR concentrations (e.g., ~1 mM), the LEF-1 HMG domain exhibits variable states of association as a function of temperature. In a temperature range from approximately 25 to 35 °C, the protein undergoes a relatively slow aggregation to a gel-like state. For example, a 2.5 mM sample of the LEF-1 domain in NMR buffer [i.e., 10% ^2H_2O, 10 mM KCl, 50 μM NaN_3, and 10 μM EDTA (pH 6.0)] formed a viscous, gel-like solution within 24 h at 28 °C. Once formed, the gel could not be solubilized by dilution with buffer and was only resolubilized with 6 M guanidine hydrochloride. Following denaturation and repurification by HPLC, the protein has the same molecular weight and refolds to its pre-gel-like state as determined by mass spectrometry and NMR, respectively.

Gel formation was monitored using one-dimensional ^1H NMR spectra and two-dimensional $^1H-^{15}N$ HSQC experiments; the signals do not broaden appreciably, nor are any specific peaks rendered nonobservable. Instead, the overall signal intensity decreases as if the sample concentration were being reduced over time. This implies that for the portion of the protein that enters the gel state the NMR signals are broadened to the point where they are no longer detectable; the observable signals correspond to molecules that remain in a monomeric state in solution. Gel formation was not observed below 25 °C, and therefore, all NMR and CD experiments performed on the free LEF-1 domain were carried out at 22 °C. Above 35 °C, and at NMR concentrations between 1.0 and 2.5 mM, the LEF-1 domain forms insoluble aggregates that cannot be resolubilized even with 6 M guanidine hydrochloride. The melting temperature of the LEF-1 domain, as determined by thermal melts monitored by CD, is ~37 °C (data not shown). Interestingly, a truncated 77-residue form of the HMG domain, lacking the C-terminal basic tail, gels more rapidly, more completely, and at a lower temperature than the 86-amino acid construct. This may indicate that intermolecular electrostatic repulsion caused by the presence of the basic tail partially inhibits the protein−protein interactions that result in the formation of the gel.

Resonance Assignments. Backbone resonance assignments were made for the free LEF-1 HMG domain using HNCACB, CBCA(CO)NH, and C(CO)NH TOCSY experiments. Only 68 amide proton resonances are observable, five of which arise from splitting of peaks due to conformational heterogeneity. No assignments could be made for residues in the

contiguous region from Ala 9 through Glu 24, which encompasses the entire first helix, or for the following residues: Glu 29, Ser 30, Gln 35, Ile 36, Arg 39, Arg 40, Trp 41, and Tyr 53.

Backbone ^{15}N and ^{1}H assignments were made for 56 residues, and ^{13}Cα assignments were obtained for 61 residues. These residues are located predominantly within turns 1 and 2, helix III, and the N- and C-termini. Resonances of amino acids located within the N- and C-terminal regions (Met 1−Ala 9 and Pro 68−Lys 86, respectively) are sharp and intense, indicative of rapid backbone motions of a disordered random coil conformation, while signals from residues within turns 1 and 2 and helix III are of average intensity. Connectivities can be traced in the triple-resonance spectra from turn 2 back into the C-terminal region of helix II (i.e., traced from Ser 45 back to His 42), but the signals thereafter (i.e., from residue 41 to residue 29) weaken and lose intensity to the point where they are no longer observable. Peaks in the HNCACB and CBCA(CO)NH spectra that correspond to the Cα and Cβ atoms of the other five residues within helix II for which assignments were obtained are also quite weak and were assigned on the basis of the unique chemical shift patterns of their sequences. The location of the residues for which assignments could be made is shown in Figure 3, where they are mapped onto the structure of the DNA-bound form (*14*).

^{15}N and ^{1}H Chemical Shift Comparison of the Free and DNA-Bound States. Although there are a number of resonances missing from the spectra of the free protein, comparison of ^{15}N and ^{1}H chemical shifts for the observable signals of the free state to those of the corresponding residues in the DNA-bound state proved to be quite informative in discerning structural similarities and differences between the two states. The strongest similarity observed is in the ^{15}N and ^{1}H chemical shift values of residues that make up the turn 2 region located between helices II and III, i.e., the major hydrophobic core. These residues are labeled in light blue in the two spectra (Figure 2A,B) and consist of Ala 43, Leu 44, Ser 45, Arg 46, Glu 47, Glu 48, and Gln 49. The chemical shifts of these residues vary little between states, with a mean proton chemical shift difference of 0.04 ± 0.03 ppm compared to a value of 0.27 ± 0.28 ppm for all observable signals. Similarly, the mean nitrogen chemical shift difference between states over this span of residues (i.e., Ala 43−Gln 49) is much lower than the difference between states for all observable signals (0.79 ± 0.30 and 1.59 ± 1.46 ppm, respectively). The four residues for which signals can be observed from the turn 1 region exhibit moderate chemical shift differences when compared to those of the domain in complex with DNA. These residues are labeled in black in Figure 2 and consist of Ser 25, Thr 26, Leu 27, and Lys 28.

Examples of residues that exhibit large differences in chemical shift values between the free and DNA-bound states are underlined and labeled in red in Figure 2 and denoted with asterisks in Figure 3. These include Ser 71, Ala 72, and Gly 77. The proton chemical shift value for Ala 72 moves downfield by 1.05 ppm upon binding DNA, and the ^{15}N and ^{1}H chemical shifts for Ser 71 move 1.01 ppm upfield and 2.83 ppm downfield, respectively. The ^{15}N resonance of Gly 77 moves significantly, shifting 7.42 ppm downfield, while its ^{1}H resonance moves 0.77 ppm upfield.

Proline Isomerization and Conformational Heterogeneity. Heterogeneity is observed for the resonances of a number of residues in the vicinity of turn 3, at the beginning of the C-terminal basic tail. Doubling of backbone resonances is observed for several residues, including Pro 68, Gly 69, Ser 71, Ala 72, and Arg 73, and may be indirectly attributed to the *cis−trans* isomerism of Pro 68. Proline isomerization is known to occur on a time scale of seconds (*47*), i.e., at a rate that is relatively slow on the NMR chemical shift time scale. Proline 68 is the residue that in the context of the protein−DNA complex breaks the third helix and is located within the turn that helps form the minor, C-terminal hydrophobic core in the DNA-bound form (*14*) (Figure 3). It is likely that the multiplicity of resonances observed for residues in the vicinity of Pro 68 is due to isomerization of this residue. The residue immediately preceding Pro 68 is a tyrosine; aromatic residues preceding prolines are well-known to favor the formation of a minor population of the *cis* isomer in unfolded and partly folded peptides and proteins (*48−50*). Evidence for Pro 68 isomerism comes directly from the ^{13}Cβ and ^{13}Cγ chemical shifts. On the basis of random coil chemical shifts for the *trans* (^{13}Cβ at 31.9−32.3 ppm and ^{13}Cγ at 27.0−27.6 ppm) and *cis* (^{13}Cβ at 34.5−34.8 ppm and ^{13}Cγ at 24.6−24.9 ppm) proline isomers relative to DSS (*51*), the dominant conformer (∼88% population) of Pro 68 appears to be in the *trans* configuration (^{13}Cβ and ^{13}Cγ resonances at 31.76 and 27.54 ppm, respectively), with a minor population (∼12%) in the *cis* configuration (^{13}Cβ at 33.8 ppm and ^{13}Cγ at 24.54 ppm). Only a single set of resonances was observed for Pro 68 in the DNA complex, with ^{13}Cβ and ^{13}Cγ chemical shifts (31.12 and 27.33 ppm, respectively) that are similar to those observed for the *trans* isomer of the free LEF-1 domain. Thus, it appears that Pro 68 is fully in the *trans* form in the DNA complex. Interestingly, the backbone and Cβ resonances of Lys 4 are also split in the spectra of free LEF-1. It is possible that this splitting arises from relatively weak interactions with the Pro 68 region, since the backbone of Lys 4 is in the proximity of the side chain of Tyr 67 in the structures of the LEF-1−DNA complex. For the sequence-specific HMG domain Sox-5, it was reported that, in the free state, NOEs were observed between the side chains of residues from the third helix (i.e., Leu 59, His 63, and Leu 64) and backbone atoms from residues in the N-terminal strand (i.e., Met 7/Asn 8, Arg 5, and Lys 4). This indicates that, for the free Sox-5 domain, some degree of interaction occurs between the N-terminal strand and the region encompassing the minor hydrophobic core. This interaction is most likely similar in the free LEF-1 domain, but possibly to a lesser extent.

Circular Dichroism. Circular dichroism (CD) spectra were obtained for the free LEF-1 HMG domain, free 15 bp cognate DNA, and the protein−DNA complex (Figure 4). The CD spectrum of the free protein in the absence of DNA (Figure 4A) contains minima at ∼222 and ∼207 nm indicative of the presence of α-helical structure. The fact that the minimum at ∼207 nm is more negative indicates contributions from random coil states. The overall helicity of the free protein is estimated to be ∼50%, based on the ellipticity at 222 nm (i.e., −18174 deg cm^2 dmol^{-1}) and assuming that θ_{222} for 100% helix equals −36000 deg cm^2 dmol^{-1} (*52, 53*). The spectrum of the free 15 bp DNA duplex (Figure 4B) (minima at ∼210 and ∼245 nm, a sharp maximum at ∼200 nm, and
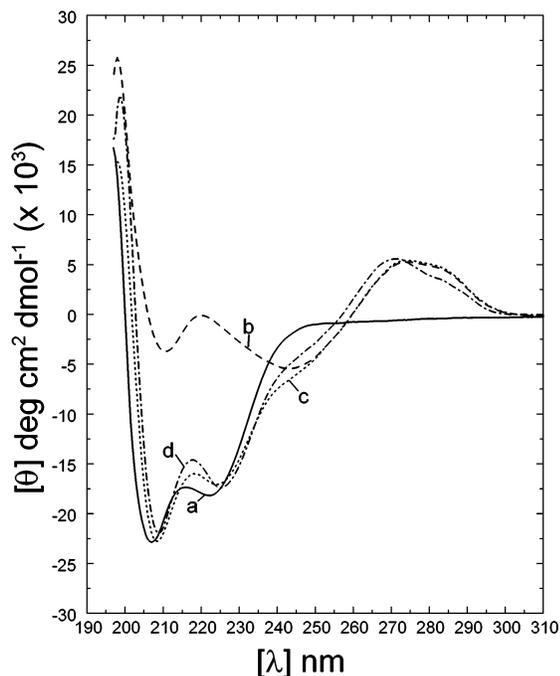
FIGURE 4: CD spectra of the free and DNA-bound states of the LEF-1 HMG domain, obtained in a tandem mixing cell: (a) free LEF-1 HMG domain alone (buffer only, no DNA in the second cell), (b) free 15 bp DNA alone (buffer only, no protein in the second cell), (c) free LEF-1 HMG domain in cell 1 and free 15 bp DNA in cell 2 before mixing, and (d) free LEF-1 HMG domain in cell 1 and free 15 bp DNA in cell 2 after mixing of the two cells for trace c, resulting in the formation of the protein−DNA complex.

a broad maximum between 265 and 285 nm) is indicative of normal B-form structure (*54, 55*).

The spectrum obtained from the tandem mix cell prior to mixing is the summation of the spectrum of the free HMG domain protein and free DNA (Figure 4C). The protein region of the spectrum (200−230 nm) remains very similar after the components are mixed (Figure 4D), but the CD spectrum of the DNA changes significantly upon formation of the complex, reflecting conformational changes in the DNA and DNA bending. The changes in the DNA ellipticity at ∼217 nm and, especially, above ∼250 nm indicate that the DNA structure takes on a more A-form-like geometry (*54, 55*) which is consistent with the NMR-determined structures of the protein−DNA complex (*14*). A qualitative interpretation of the CD spectra can be used to assess changes in secondary structure on binding to DNA. The ellipticity of the protein changes little, indicating that there is probably little change in the protein secondary structure upon complex formation.

DISCUSSION

Structural Differences between Free and Bound States. Contiguous regions of the free LEF-1 HMG domain exhibit conformational heterogeneity that renders NMR signals from these regions nonobservable. Although this precludes calculation of high-resolution structures, the combined use of heteronuclear NMR and circular dichroism provided significant insight into the structure and dynamic behavior of the free protein.

The $^{13}C\alpha$ chemical shift is highly sensitive to local ϕ and ψ dihedral angles and secondary structure (*56, 57*). The $^{13}C\alpha$

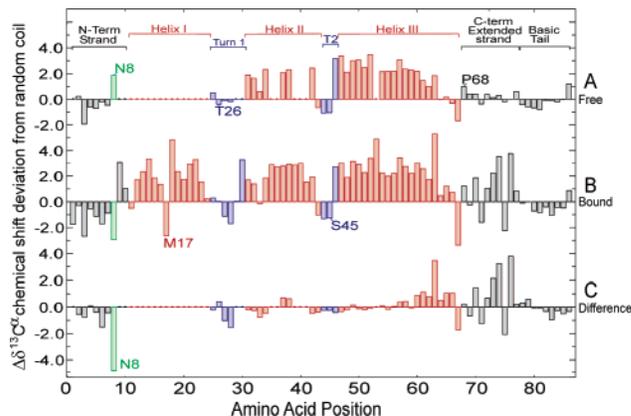

FIGURE 5: $^{13}C\alpha$ chemical shift differences ($\Delta\delta^{13}C\alpha$) (A) between the free LEF-1 and random coil values, (B) between the values for LEF-1 bound to DNA and the corresponding random coil values, and (C) between the DNA-bound values and the free state values. The secondary structure elements shown at the top of the figure correspond to those seen in the solution structures of the DNA-bound form (*14*). Orange bars represent $\Delta\delta^{13}C\alpha$ values for residues present in helices in the DNA-bound form of the protein. The bars for the intervening turn regions are colored blue. Regions in which NMR signals are not observable are marked with a dot (·). The green bar indicates Asn 8, which shows the greatest chemical shift difference between the free and DNA-bound states.

chemical shift differences of the DNA-bound state from random coil, and of the free state from random coil, together with the $^{13}C\alpha$ chemical shift difference between the two states, are plotted in Figure 5. The secondary structure predicted by the $^{13}C\alpha$ chemical shift comparison of the bound state and random coil (Figure 5B) is in good agreement with that obtained from calculated structures (*14*). Helices in the calculated structures (residues 9−24, 30−41, and 46−66) are depicted in orange and interhelical turns in blue. Downfield $^{13}C\alpha$ shifts are observed for most residues in the helices. Met 17, located in a kink near the center of helix 1, is an exception in that its $^{13}C\alpha$ resonance is shifted 1.5 ppm upfield, diagnostic of nonhelical ϕ and ψ angles. This bend is a characteristic of HMG domains from both subclasses and occurs at the site where helix I crosses helix II. The nine-residue basic tail located at the C-terminus exhibits $^{13}C\alpha$ chemical shift values that are close to those of a random coil. This is consistent with the ill-defined secondary structure in this region of the DNA-bound protein, and the large rms deviations between calculated solution structures for this region (*14*). This feature is also observed for the basic tail region located C-terminal to the HMG-D domain (*58*). Results from both circular dichroism and differential scanning calorimetry indicated that the 26-residue basic tail of the HMG-D100 domain is unfolded even at 5 °C (*58*). Residues in the N-terminal extended region (His 2−Asn 8) and the region between Tyr 67 and Tyr 76, which forms the minor hydrophobic core at the C-terminus of helix III and makes minor groove contacts with the DNA (*14*), display relatively large deviations from random coil $^{13}C\alpha$ chemical shifts, indicative of ordered backbone structure.

Deviations of the $^{13}C\alpha$ chemical shifts from random coil values are shown for the free LEF-1 HMG domain in Figure 5A. As noted above, backbone resonances of all residues in helix I and of some residues in helix II were broadened beyond detection and could not be assigned. However, the secondary shifts observed for those residues that can be

assigned in helix II are comparable in sign and magnitude to the values found for the same residues in the LEF-1−DNA complex. This provides evidence that helix II is folded in the free protein, but it appears that local regions are subject to conformational fluctuations that cause resonance broadening. Comparison of the secondary $^{13}C\alpha$ shifts in the helix III region shows that this helix is also folded in the free protein over approximately 60% of its length, from Arg 46 to Glu 59. Between residues Arg 60 and Leu 66, however, the secondary shifts are significantly smaller for the free protein, suggesting destabilization of the helix with probable fraying at its C-terminal end. It is likely that in the free state the helical structure is stabilized in the middle and N-terminal regions of helix III through packing involving the central, major hydrophobic core. Variable stability in the helix III region was also observed for the free Sox-4 domain (*32*). In particular, backbone amide proton exchange was rapid in the region encompassing residues 55−59 of helix III, consistent with a less rigid and more exposed character for helix III in the free state (*32*).

The change in $^{13}C\alpha$ chemical shift (bound − free) is shown in Figure 5C. Chemical shift differences between the free and DNA-bound states for the $^{13}C\alpha$ resonances of residues 25−28, which form the turn between helices I and II, suggest conformational changes in this region upon DNA binding. In addition, the ^1HN and ^{15}N shifts (not shown) are closer to random coil values for the free protein, suggesting that this region may be partially disordered in the absence of DNA. We note that ^{15}N relaxation experiments with the non-sequence-specific HMG1-A and HMG-D domains in the absence of DNA indicate increased local motion in this interhelical region (*23, 26*). Additionally, solution structure analysis of the free Sox-4 sequence-specific domain revealed that the assignments of residues found in this loop (i.e., Asn 28 and Ala 29) could not be established, also implying the possible existence of conformational heterogeneity in this region (*32*).

The largest difference in Cα chemical shift between the two states is displayed by Asn 8 (bar colored green in Figure 5) which has a positive $\Delta\delta^{13}C\alpha$ in the free state and a negative value in the DNA-bound state. Although Asn 8 functions to cap helix I in the bound state, it is not actually part of the helix (helix I starts at Ala 9). In complex with DNA, Asn 8 forms important bipartite hydrogen bonds with specific bases (*14*). In the absence of these interactions, the N-terminus of helix I may propagate further and include Asn 8, giving rise to a Cα chemical shift that is more characteristic of a helix in the free state of the protein.

The minor hydrophobic core formed in the DNA-bound state by Ile 3 in the N-terminal extended region and by residues Tyr 67 and Trp 70 near the C-terminal end of helix III (*14*) is largely unfolded in the free protein. Formation of this smaller hydrophobic core and stabilization of the C-terminal end of helix III are clearly dependent upon interactions with the DNA. Furthermore, residues 71−76 are unstructured in the free LEF-1 HMG domain, with nearly random coil chemical shifts. As noted above in the Results, the LEF-1 domain exhibits variable states of association as a function of temperature. In a range of temperatures from approximately 25 to 35 °C, the protein undergoes a relatively slow aggregation to a (chemically reversible) gel-like state, whereas above 35 °C (and at NMR concentrations), the LEF-1 domain forms insoluble aggregates. The melting temperature of the LEF-1 domain, as determined by thermal melts monitored by CD, is ∼37 °C (data not shown). A comprehensive thermodynamic description of the sequence-specific Sox-5 domain revealed that the protein unfolds in two separate stages (*34*). The temperature midpoint of the first stage (∼34 °C) was attributed to the minor wing (i.e., the N-terminal strand packed against helix III), while the midpoint of the second unfolding stage (∼46 °C) was attributed to the major wing (i.e., helices I and II and the major hydrophobic core). We suspect, in a manner similar to that described for the Sox-5 domain, that it is the unfolding of the LEF-1 major wing above 35 °C that results in insoluble aggregation, and by analogy, it may be the observed lack of stability in the minor wing that results in the reversible, gel-like aggregation that occurs between 25 and 35 °C for the free LEF-1 domain. Upon binding to DNA, the residues that make up the C-terminal minor hydrophobic core adopt a defined structure and Tyr 76 forms important interactions in the minor groove.

There are no marked differences in the ^1H, ^{15}N, and $^{13}C\alpha$ chemical shifts of residues located in the turn between helices II and III between the free and DNA-bound states (Figures 2 and 5). This region encompasses the major hydrophobic core of the protein, and the similarity of the NMR spectra of the two states implies a similar spatial geometry of the core residues and the presence of tertiary structure in this area in the free state. This region appears to be relatively rigid, and therefore, the angle between helices II and III, a defining characteristic of many HMG domains, does not vary to a significant degree between states.

Structure of Helix I in the Free State. The origin of the conformational averaging observed for the free LEF-1 HMG domain in the helix 1 region is unclear. However, the structure of the LEF-1−DNA complex (*14*) reveals a pronounced kink near the middle of helix 1, where it crosses helix II (Figure 3). This kink disrupts backbone hydrogen bonding interactions, which might lead to instability of the helix in the absence of stabilizing interactions with the DNA. Conformational heterogeneity in this region in the absence of DNA is not surprising, since some of the most intimate DNA contacts, including intercalated residues Met 11 and Met 14, are made by residues in helix I.

Although NMR cannot provide direct information about the regions of the LEF-1 domain that cannot be assigned, the use of circular dichroism, combined with the observable NMR signals, has provided substantial insight into the similarities and differences between the free and bound states. None of the NMR signals for the helix I region can be observed, since the backbone resonances are too broad to be detected, but the CD spectra suggest this helix is at least partly formed. The calculated structures of the protein−DNA complex indicate that the bound protein is ∼57% helical. On the basis of the measured ellipticity at 222 nm in the CD spectrum, the overall helicity of the free protein is estimated to be ∼50%. If helix I was not formed to any degree, then the overall helicity would be only ∼38%. The CD-based estimate of ∼50% helicity for the free HMG domain and the fact that the ellipticities at 208 and 222 nm do not change much upon DNA binding (Figure 4) argue strongly that helix I is partly formed in the absence of DNA. The most probable explanation for the absence of NMR

signals for helix I residues is that the signals are broadened by slow (i.e., microsecond to millisecond) time scale conformational fluctuations; a single conformer becomes stabilized upon DNA binding such that a single set of resonances with widths comparable to those for other regions of the protein is observed for the complex. Such behavior is not without precedent: resonances of the F helix of apomyoglobin are broadened beyond detection by conformational fluctuations but become observable after stabilization of helical structure when heme is bound to form the holoprotein (*59*).

Members of the subgroup of sequence-specific HMG domains that are designated Sox, in relation to the Sry box (the mammalian testis-determining factor encoded by the Y chromosome), also have been shown to display lower stability than non-sequence-specific HMG domains (*60*). The calculated NMR structure of the Sox-4 domain indicates an α-helical content of 53% which is consistent with the reported CD-based estimate of secondary structure that revealed an α-helical content of 54% (*32*). Our results argue that the LEF-1 domain may be even less stable than the Sox-4 domain and that it is probably the observed increase in mobility and fraying of the LEF-1 helices that preclude high-resolution structure elucidation.

Is the Disorder Observed for the Free State of the LEF-1 HMG Domain Functional? In comparison to the partially disordered state of the free LEF-1 domain, members of the structure-specific subclass (e.g., HMG1-A, HMG1-B, and HMG-D) exhibit quite different conformational stability in the absence of DNA. The NMR spectra of these free domains are high in quality and thus indicative of relative structural stability. In fact, rigorous analysis of ^{15}N relaxation experiments performed on HMG1-A (*23*) and HMG-D (*26*) indicates that, although certain regions within these domains exhibit mobility on a picosecond time scale, and there is some evidence of minor structural heterogeneity, the molecules are well-ordered overall and, due to their L shape, tumble anisotropically as rigid ellipsoids.

On the other hand, experimental results indicate that members of the sequence-specific subclass (i.e., Sox-4, Sox-5, and LEF-1) are less ordered in comparison to members of the structure-specific subclass. Structural heterogeneity may have contributed to the lack of a complete NMR data set for Sox-4. Additionally, the rmsd values for the Sox-4 solution structures increase significantly at the N- and C-termini, in particular near the C-terminus of the third helix. This is the case even though the third helix of Sox-4 superimposes on itself well (rmsd = 0.76 ± 0.26 Å) which is also in good agreement with CD data (*32*). These findings imply that the secondary structure elements are likely formed but are possibly more mobile and dynamic at helix termini and not locked into one unique structure.

The presence of increased structural dynamics in the N-terminal strand and at the C-terminus of helix III has also been reported for the free Sox-5 domain (*35*). For helix III residues, the $^3J_{HN\alpha}$ values steadily increase from 3.0 Hz (characteristic of backbone dihedrals in helices) to 7.0 Hz (indicative of extended structures) from the midpoint of the helix to the C-terminal end, implying a loosening or fraying of the helix at the end furthest from the major hydrophobic core. This conclusion is bolstered by low S^2 and $^1H-^{15}N$ values reported for residues at the C-terminal end of helix

III. These findings indicate that, in a manner similar to that observed for the free LEF-1 domain, regions of Sox-5 exhibit some degree of dynamic disorder when studied free in solution. Interestingly, $^3J_{HN\alpha}$ coupling constants for residues at the C-terminal end of helix III and for N-terminal residues (i.e., Lys 4, Arg 5, and Met 7) are lowered when the temperature is lowered from 25 to 12 °C, indicating an increase in structural order at lower temperatures (*34, 35*). This is a marked difference from the behavior of the free LEF-1 domain; i.e., there is no improvement in the quality of the NMR spectra of the free LEF-1 domain regardless of the temperature that was tested (5, 10, 12, 22, and 30 °C). This difference, in combination with other findings described herein, argues that the LEF-1 domain may possess even less structural order than the Sox-4 or Sox-5 domain.

It has been reported that the yeast NHP6A HMG domain, a member of the non-sequence-specific subclass, appears to be largely unfolded at 37 °C (*27*). The yeast NHP6A domain is somewhat unusual in that, although it belongs to the non-sequence-specific subclass, it only contains one HMG domain. The only other structurally characterized member of this subclass with one HMG domain is the HMG-D domain from *Drosophila melanogaster* (*26, 61, 62*). $^1H-^{15}N$ TROSY experiments performed on the free NHP6A domain revealed a dramatic increase in NMR signal dispersion and quality when the temperature was reduced from 37 to 20 °C (*27*). This improvement allowed the structure of the free NHP6A domain to be determined at 20 °C. This finding is similar to that observed for the sequence-specific Sox-5 domain, but, as described above, not for the LEF-1 domain. The high-resolution structure of the NHP6A domain in complex with a 15 bp DNA duplex comprising the SRY recognition sequence has also been determined (*20*). As in the LEF-1−DNA complex structure, there are significant structural changes in the DNA, but unlike the LEF-1 domain, there are only minor changes in the protein structure when the DNA-bound NHP6A domain (determined at 37 °C) is compared to the free NHP6A structure determined at 20 °C (mean backbone rmsd = 2.58 Å). Thus, upon binding DNA at 37 °C, the NHP6A domain appears to undergo a transition to higher structural order similar to that observed for the free NHP6A domain at lower temperatures. By contrast, the LEF-1 domain becomes more ordered only upon binding DNA (at 30 °C). Deuterium exchange experiments performed on the free and DNA-bound states of the NHP6A domain indicate that the major hydrophobic core is stable in the free state (there are 12 slowly exchanging amide protons in this region) and that 18 additional amides are stabilized against deuterium exchange upon DNA binding (*20*). The additional amides protected upon complex formation extend away from the primary hydrophobic core along the helices and N-terminal extended strand. The fact that there is little change in chemical shift upon DNA binding for residues that make up the primary hydrophobic core region of the LEF-1 domain implies that a similar increase in structural order (i.e., toward the helical termini and the N-terminal strand) occurs within the LEF-1 domain upon formation of a complex with DNA. Thus far, there is no obvious reason the LEF-1 domain should exhibit the highest degree of conformational heterogeneity and lack of structural order when free in solution.

A recent alanine scanning mutagenesis study performed on the Sox-9 HMG domain from zebra fish revealed the

importance of certain residues and their influence on protein stability and DNA binding (*63*). In addition, the authors calculated a theoretical "instability index" and reported the average instability indices to have contrasting values of 67 and 30 for the sequence-specific and non-sequence-specific subclasses, respectively (the higher the value, the greater the instability). The authors hypothesize that the observed differences in structural stability between the two subclasses are correlated to their biological function and *in vivo* half-life; i.e., the non-sequence-specific HMG proteins are ubiquitous and expressed throughout an organism's lifetime and therefore necessarily have longer half-lives, whereas the sequence-specific HMG transcription factors are only expressed during specific times during development and therefore have functionally shorter half-lives. This supposition, which quite possibly may reflect part of the reason for the stability differences between subclasses, has yet to be verified experimentally.

On the basis of the collective data published for HMG domains, and on the basis of the results reported herein, we suggest that members of the sequence-specific subclass tend to be stable in complex with DNA yet, under physiological conditions, are partially mobile and disordered when free in solution. Conversely, members of the more ubiquitous, non-sequence-specific (structure-specific) subclass are usually more stable and ordered free in solution. This hypothesis was recently strengthened with the experimentally determined values for the enthalpy and entropy of the second unfolding transitions of the sequence-specific Sox-5 domain and non-sequence-specific HMG-D domain (*58*). Extrapolation of the enthalpy and entropy of the second transition of the Sox-5 domain to 41.7 °C gives 154 kJ/mol for the enthalpy and 486 J K^{-1} mol^{-1} for the entropy. The averaged values for the enthalpy and entropy of the second transition for HMG-D are 193 kJ/mol and 611 J K^{-1} mol^{-1}, respectively. The authors conclude that since the unfolded states of these HMG domains are not expected to differ significantly, the globular part of the HMG-D protein is more rigid and less flexible than that of the Sox-5 domain (*58*). The difference in stability between the free state of LEF-1 (and other sequence-specific HMG domains) and structure-specific domains may reflect the various modes of binding of these subclasses. The well-formed structures of HMG1-A, HMG1-B, and HMG-D recognize and bind prebent DNA, while the flexible and partially disordered LEF-1 domain binds to, distorts, and bends a specific sequence of linear DNA. We therefore propose that the experimentally observed flexibility of the LEF-1 HMG-1 domain may be functionally significant and is yet another example of induced folding upon binding (*36*–*38*). The disorder and conformational heterogeneity displayed by the free LEF-1 domain may be necessary to allow both it and the DNA to bind one another and simultaneously collapse to the relatively compact, lower-energy conformation attained upon final complex formation.

REFERENCES

1. Travis, A., Amsterdam, A., Belanger, C., and Grosschedl, R. (1991) LEF-1, a gene encoding a lymphoid-specific protein with an HMG domain, regulates T-cell receptor α enhancer function, *Genes Dev. 5*, 1–2.
2. Waterman, M. L., Fischer, W. H., and Jones, K. A. (1991) A thymus-specific member of the HMG protein family regulates the human T cell receptor C α enhancer, *Genes Dev. 5*, 656–669.
3. Held, W., Clevers, H., and Grosschedl, R. (2003) Redundant functions of TCF-1 and LEF-1 during T and NK cell development, but unique role of TCF-1 for Ly49 NK cell receptor acquisition, *Eur. J. Immunol. 33*, 1393–1398.
4. Giese, K., Cox, J., and Grosschedl, R. (1992) The HMG domain of lymphoid enhancer factor 1 bends DNA and facilitates assembly of functional nucleoprotein structures, *Cell 69*, 185–195.
5. Giese, K., Kingsley, C., Kirshner, J. R., and Grosschedl, R. (1995) Assembly and function of a TCRα enhancer complex is dependent on LEF-1-induced DNA bending and multiple protein–protein interactions, *Genes Dev. 9*, 995–1008.
6. Grosschedl, R., Giese, K., and Pagel, J. (1994) HMG domain proteins: architectural elements in the assembly of nucleoprotein structures, *Trends Genet. 10*, 94–100.
7. Eastman, Q., and Grosschedl, R. (1999) Regulation of LEF-1/TCF transcription factors by Wnt and other signals, *Curr. Opin. Cell Biol. 11*, 233–240.
8. Roel, G., Hamilton, F. S., Gent, Y., Bain, A. A., Destree, O., and Hoppler, S. (2002) Lef-1 and Tcf-3 transcription factors mediate tissue-specific Wnt signaling during *Xenopus* development, *Curr. Biol. 12*, 1941–1945.
9. Baxevanis, A. D., and Landsman, D. (1995) The HMG-1 box protein family: classification and functional relationships, *Nucleic Acids Res. 23*, 1604–1613.
10. Ferrari, S., Harley, V. R., Pontiggia, A., Goodfellow, P. N., Lovell-Badge, R., and Bianchi, M. E. (1992) SRY, like HMG1, recognizes sharp angles in DNA, *EMBO J. 11*, 4497–4506.
11. Pil, P. M., and Lippard, S. J. (1992) Specific binding of chromosomal protein HMG1 to DNA damaged by the anticancer drug cisplatin, *Science 256*, 234.
12. Webb, M., Payet, D., Lee, K. B., Travers, A. A., and Thomas, J. O. (2001) Structural requirements for cooperative binding of HMG1 to DNA minicircles, *J. Mol. Biol. 309*, 79–88.
13. Klass, J., Murphy, F. V., Fouts, S., Serenil, M., Changela, A., Siple, J., and Churchill, M. E. (2003) The role of intercalating residues in chromosomal high-mobility-group protein DNA binding, bending and specificity, *Nucleic Acids Res. 31*, 2852–2864.
14. Love, J. J., Li, X. A., Case, D. A., Giese, K., Grosschedl, R., and Wright, P. E. (1995) Structural basis for DNA bending by the architectural transcription factor LEF-1, *Nature 376*, 791–795.
15. Werner, M. H., Huth, J. R., Gronenborn, A. M., and Clore, G. M. (1995) Molecular basis of human 46X,Y sex reversal revealed from the three-dimensional solution structure of the human SRY-DNA complex, *Cell 81*, 705–714.
16. Hardman, C. H., Broadhurst, R. W., Raine, A. R. C., Grasser, K. D., Thomas, J. O., and Laue, E. D. (1995) Structure of the A-domain of HMG1 and its interaction with DNA as studied by heteronuclear three- and four-dimensional NMR spectroscopy, *Biochemistry 34*, 16596–16607.
17. Ohndorf, U. M., Rould, M. A., He, Q., Pabo, C. O., and Lippard, S. J. (1999) Basis for recognition of cisplatin-modified DNA by high-mobility-group proteins, *Nature 399*, 708–712.
18. Payet, D., Hillisch, A., Lowe, N., Diekmann, S., and Travers, A. (1999) The recognition of distorted DNA structures by HMG-D: A footprinting and molecular modelling study, *J. Mol. Biol. 294*, 79–91.
19. Cerdan, R., Payet, D., Yang, J. C., Travers, A. A., and Neuhaus, D. (2001) HMG-D complexed to a bulge DNA: An NMR model, *Protein Sci. 10*, 504–518.
20. Masse, J. E., Wong, B., Yen, Y. M., Allain, F. H., Johnson, R. C., and Feigon, J. (2002) The *S. cerevisiae* architectural HMGB protein NHP6A complexed with DNA: DNA and protein conformational changes upon binding, *J. Mol. Biol. 323*, 263–284.
21. Wong, B., Masse, J. E., Yen, Y. M., Giannikopoulos, P., Feigon, J., Johnson, R. C., and Giannikoupolous, P. (2002) Binding to cisplatin-modified DNA by the *Saccharomyces cerevisiae* HMGB protein Nhp6A, *Biochemistry 41*, 5404–5414.
22. Giese, K., Amsterdam, A., and Grosschedl, R. (1991) DNA-binding properties of the HMG domain of the lymphoid-specific transcriptional regulator LEF-1, *Genes Dev. 5*, 2567–2578.
23. Broadhurst, R. W., Hardman, C. H., Thomas, J. O., and Laue, E. D. (1995) Backbone dynamics of the A-domain of HMG1 as studied by ¹⁵N NMR spectroscopy, *Biochemistry 34*, 16608–16617.
24. Read, C. M., Cary, P. D., Crane-Robinson, C., Driscoll, P. C., and Norman, D. G. (1993) Solution structure of a DNA-binding domain from HMG1, *Nucleic Acids Res. 21*, 3427–3436.

25. Weir, H. M., Kraulis, P. J., Hill, C. S., Raine, A. R. C., Laue, E. D., and Thomas, J. O. (1993) Structure of the HMG box motif in the B-domain of HMG1, *EMBO J. 12*, 1311−1319.

26. Jones, D. N. M., Searles, M. A., Shaw, G. L., Churchill, M. E. A., Ner, S. S., Keeler, J., Travers, A. A., and Neuhaus, D. (1994) The solution structure and dynamics of the DNA-binding domain of HMG-D from *Drosophila melanogaster*, *Structure 2*, 609−627.

27. Allain, F. H., Yen, Y. M., Masse, J. E., Schultze, P., Dieckmann, T., Johnson, R. C., and Feigon, J. (1999) Solution structure of the HMG protein NHP6A and its interaction with DNA reveals the structural determinants for non-sequence-specific binding, *EMBO J. 18*, 2563−2579.

28. Xu, Y., Yang, W., Wu, J., and Shi, Y. (2002) Solution structure of the first HMG box domain in human upstream binding factor, *Biochemistry 41*, 5415−5420.

29. Yang, W., Xu, Y., Wu, J., Zeng, W., and Shi, Y. (2003) Solution structure and DNA binding property of the fifth HMG box domain in comparison with the first HMG box domain in human upstream binding factor, *Biochemistry 42*, 1930−1938.

30. Jantzen, H.-M., Admon, A., Bell, S. P., and Tjian, R. (1990) Nucleolar transcription factor hUBF contains a DNA-binding motif with homology to HMG proteins, *Nature 344*, 830−836.

31. Jantzen, H.-M., Chow, A. M., King, D. S., and Tjian, R. (1992) Multiple domains of the RNA polymerase I activator hUBF interact with the TATA-binding protein complex hSL1 to mediate transcription, *Genes Dev. 6*, 1950−1963.

32. Van Houte, L. P. A., Chuprina, V. P., Van der Wetering, M., Boelens, R., Kaptein, R., and Clevers, H. (1995) Solution structure of the sequence-specific HMG box of the lymphocyte transcriptional activator Sox-4, *J. Biol. Chem. 270*, 30516−30524.

33. Connor, F., Cary, P. D., Read, C. M., Preston, N. S., Driscoll, P. C., Denny, P., Crane-Robinson, C., and Ashworth, A. (1994) DNA binding and bending properties of the post-meiotically expressed Sry-related protein Sox-5, *Nucleic Acids Res. 22*, 3339−3346.

34. Crane-Robinson, C., Read, C. M., Cary, P. D., Driscoll, P. C., Dragan, A. I., and Privalov, P. L. (1998) The energetics of HMG box interactions with DNA. Thermodynamic description of the box from mouse Sox-5, *J. Mol. Biol. 281*, 705−717.

35. Cary, P. D., Read, C. M., Davis, B., Driscoll, P. C., and Crane-Robinson, C. (2001) Solution structure and backbone dynamics of the DNA-binding domain of mouse Sox-5, *Protein Sci. 10*, 83−98.

36. Spolar, R. S., and Record, M. T. (1994) Coupling of local folding to site-specific binding of proteins to DNA, *Science 263*, 777−784.

37. Wright, P. E., and Dyson, H. J. (1999) Intrinsically unstructured proteins: Re-assessing the protein structure−function paradigm, *J. Mol. Biol. 293*, 321−331.

38. Dyson, H. J., and Wright, P. E. (2002) Coupling of folding and binding for unstructured proteins, *Curr. Opin. Struct. Biol. 12*, 54−60.

39. Gunasekaran, K., Tsai, C. J., Kumar, S., Zanuy, D., and Nussinov, R. (2003) Extended disordered proteins: targeting function with less scaffold, *Trends Biochem. Sci. 28*, 81−85.

40. Garcia, G. M., Mar, P. K., Mullin, D. A., Walker, J. R., and Prather, N. E. (1986) The *E. coli* dnaY gene encodes an arginine transfer RNA, *Cell 45*, 453−459.

41. Brinkmann, U., Mattes, R. E., and Buckel, P. (1989) High-level expression of recombinant genes in *Escherichia coli* is dependent on the availability of the dnaY gene product, *Gene 85*, 109−114.

42. Wishart, D. S., Bigam, C. G., Yao, J., Abildgaard, F., Dyson, H. J., Oldfield, E., Markley, J. L., and Sykes, B. D. (1995) ¹H, ¹³C and ¹⁵N chemical shift referencing in biomolecular NMR, *J. Biomol. NMR 6*, 135−140.

43. Grzesiek, S., and Bax, A. (1992) Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR, *J. Am. Chem. Soc. 114*, 6291−6293.

44. Wittekind, M., and Mueller, L. (1993) HNCACB, a high-sensitivity 3D NMR experiment to correlate amide-proton and nitrogen resonances with the α- and β-carbon resonances in proteins, *J. Magn. Reson. 101*, 201−205.

45. Grzesiek, S., Anglister, J., and Bax, A. (1993) Correlation of backbone amide and aliphatic side chain resonances in ¹³C/¹⁵N-enriched proteins by isotropic mixing of ¹³C magnetization, *J. Magn. Reson., Ser. B 101*, 114−119.

46. Lnenicek-Allen, M., Read, C. M., and Crane-Robinson, C. (1996) The DNA bend angle and binding affinity of an HMG box increased by the presence of short terminal arms, *Nucleic Acids Res. 24*, 1047−1051.

47. Brandts, J. F., Halvorson, H. R., and Brennan, M. (1975) Consideration of the possibility that the slow step in protein denaturation reactions is due to cis−trans isomerism of proline residues, *Biochemistry 14*, 4953−4963.

48. Grathwohl, C., and Wüthrich, K. (1976) The X-Pro peptide bond as an NMR probe for conformational studies of flexible linear peptides, *Biopolymers 15*, 2025−2041.

49. Dyson, H. J., Cross, K. J., Houghten, R. A., Wilson, I. A., Wright, P. E., and Lerner, R. A. (1985) The immunodominant site of a synthetic immunogen has a conformational preference in water for a type-II reverse turn, *Nature 318*, 480−483.

50. Yao, J., Feher, V. A., Espejo, B. F., Reymond, M. T., Wright, P. E., and Dyson, H. J. (1994) Stabilization of a type VI turn in a family of linear peptides in water solution, *J. Mol. Biol. 243*, 736−753.

51. Howarth, O. W., and Lilley, D. M. J. (1978) Carbon-13-NMR of peptides and proteins, *Prog. NMR Spectrosc. 12*, 1−40.

52. Chen, Y.-H., Yang, J. T., and Chau, K. H. (1974) Determination of the helix and β form of proteins in aqueous solution by circular dichroism, *Biochemistry 13*, 3350−3359.

53. Chakrabartty, A., Schellman, J. A., and Baldwin, R. L. (1991) Large differences in the helix propensities of alanine and glycine, *Nature 351*, 586−588.

54. Saenger, W. (1984) *Principles of Nucleic Acid Structure*, Springer-Verlag, New York.

55. Johnson, W. C., Jr. (1996) *Circular Dichroism and the Conformational Analysis of Proteins*, Plenum Press, New York.

56. Spera, S., and Bax, A. (1991) Empirical correlation between protein backbone conformation and Cα and Cβ ¹³C nuclear magnetic resonance chemical shifts, *J. Am. Chem. Soc. 113*, 5490−5492.

57. Wishart, D. S., and Sykes, B. D. (1994) The ¹³C chemical-shift index: A simple method for the identification of protein secondary structure using ¹³C chemical-shift data, *J. Biomol. NMR 4*, 171−180.

58. Dragan, A. I., Klass, J., Read, C., Churchill, M. E., Crane-Robinson, C., and Privalov, P. L. (2003) DNA binding of a non-sequence-specific HMG-D protein is entropy driven with a substantial non-electrostatic contribution, *J. Mol. Biol. 331*, 795−813.

59. Eliezer, D., and Wright, P. E. (1996) Is apomyoglobin a molten globule? Structural characterization by NMR, *J. Mol. Biol. 263*, 531−538.

60. Weiss, M. A. (2001) Floppy SOX: mutual induced fit in hmg (high-mobility group) box-DNA recognition, *Mol. Endocrinol. 15*, 353−362.

61. Churchill, M. E. A., Jones, D. N. M., Glaser, T., Hefner, H., Searles, M. A., and Travers, A. A. (1995) HMG-D is an architecture-specific protein that preferentially binds to DNA containing the dinucleotide TG, *EMBO J. 14*, 1264−1275.

62. Murphy, F. V., IV, Sweet, R. M., and Churchill, M. E. A. (1999) The structure of a chromosomal high mobility group protein-DNA complex reveals sequence-neutral mechanisms important for non-sequence-specific DNA recognition, *EMBO J. 18*, 6610−6618.

63. Hsiao, N. W., Samuel, D., Liu, Y. N., Chen, L. C., Yang, T. Y., Jayaraman, G., and Lyu, P. C. (2003) Mutagenesis study on the zebra fish SOX9 high-mobility group: comparison of sequence and non-sequence specific HMG domains, *Biochemistry 42*, 11183−11193.