The Sequence-specific High Mobility Group 1 Box of TCF-1 Adopts a Predominantly \( \alpha \)-Helical Conformation in Solution*

(Received for publication, January 19, 1993, and in revised form, April 26, 1993)

Leo van Houthe†‡, Annemiek van Oers‡, Marc van de Wetering‡, Dennis Dooijes‡, Rob Kaptein‡, and Hans Clevers‡†

From the †Department of Immunology, University Hospital, 3508 GA Utrecht, The Netherlands and the ‡Department of NMR Spectroscopy, Bijvoet Center for Biomolecular Research, University of Utrecht, 3584 CH, Utrecht, The Netherlands

The High Mobility Group (HMG) 1 box is a protein motif that mediates DNA binding in a novel family of transcription-regulating proteins. Several members of this family, including the lymphoid-specific proteins TCF-1 and LEF-1 and the mammalian sex-determining factor SRY, carry a single HMG box with affinity for the minor groove of the heptamer motif ACAAGG or variations thereof. To initiate studies on the structural characteristics of the TCF-1 HMG box, we have expressed the 87-amino acid HMG box in milligram quantities in Escherichia coli and purified the soluble peptide to >95% homogeneity. The peptide bound DNA with the same specificity as the complete protein and was capable of inducing DNA bending. Circular dichroism (CD) analysis revealed the TCF-1 HMG box to adopt an approximately 60% \( \alpha \)-helix/40% random coil conformation in solution. In the presence of an equimolar amount of double-stranded DNA containing the cognate motif, the CD spectrum changed significantly, implying the induction of a structural modification upon DNA/protein association.

A number of structural protein motifs are involved in the recognition of DNA. These include the helix-turn-helix (1), the helix-loop-helix (2), the zinc finger (3), the leucine zipper/zipper (4), and the homeodomain (5). Recently, a novel type of DNA-binding domain, the so-called HMG box, has been identified upon analysis of the RNA polymerase I transcription factor UBF (6). The HMG1 box was named after its homology with high mobility group (HMG) 1 proteins and is defined by a loose consensus sequence of about 80 amino acids (7). To date, more than 60 proteins with one or more HMG boxes have been recognized, including the products of the fungal mating type genes Mat-Mc of Schizosaccharomyces pombe (8) and Mt a1 of Neurospora crassa (9), the products of the mammalian sex-determining SRY gene and of SRY-like ("Sox") genes (10, 11), and the mitochondrial transcription factor mtTF1 (12).

The T lymphocyte-specific transcription factor TCF-1 contains a single HMG box (13) with a well-defined DNA-binding specificity for the heptamer oligonucleotide motif A/T A/T C A A A G (14, 15). Very similar DNA-binding specificities were found for the highly related HMG box of the lymphoid transcrip-tion factor LEF-1 (16) as well as for the less-related HMG boxes of the SRY gene product (17–19), of the S. pombe transcription factor Ste11+ (20), and of Sox-5 (21). The sequence-specific HMG box proteins are believed to bind to DNA as monomers (22, 23). They recognize the minor groove, as indicated by methylation and diethylpyrocarbonate carboxylation interference footprinting, and I/G nucleotide substitution (13, 19, 22).

Computer-aided construction of an evolutionary tree of the HMG box superfamily places all sequence-specific HMG box genes (i.e. the TCF/LEF genes, the SRY/Sox genes, and the fungal mating type genes) on the same branch. Over 1,000 million years ago, this branch separated from the relatively non-sequence-specific HMG box supergene family members such as HMG-1 and -2, UBF, and mtTF1. Typically, all members of the sequence-specific branch contain a single HMG box, whereas all other proteins carry at least two of these domains.

An intriguing property of the sequence-specific HMG box proteins LEF-1 and SRY is the ability to bend or supercoil the DNA helix (22, 23). The LEF-1 HMG box reportedly bends DNA by an angle of 130°. Moreover, a bend-swap experiment demonstrated that LEF-1 and its specific DNA-binding motif can functionally replace bending induced by the Integration Host Factor at the attP locus in the phage \( \lambda \) integrase reaction (22). The non-sequence-specific HMG boxes of HMG-1 appear to recognize DNA structure, i.e. they bind with high affinity to cruciform DNA (24, 25). A recent study reports the puzzling finding that the HMG box of SRY, which displays high affinity toward the linear motif AA-CACAAG, binds with even higher affinity to four-way junction DNA irrespective of sequence, thus resembling the structure-specific HMG-1 protein (23).

As a first step toward a biophysical characterization of the sequence-specific HMG box protein motif, we have overexpressed a peptide corresponding to the HMG box region of the T-cell specific transcription factor TCF-1 in Escherichia coli in milligram quantities. Here, we describe the overexpression and purification of the TCF-1 HMG box domain, its binding properties, and its spectroscopic characterization by circular dichroism (CD).

MATERIALS AND METHODS

**Plasmid Construction**—The region encoding amino acids Ala-149 to Ser-234 was excised by polymerase chain reaction amplification from pTF-1 (19). Primers were: 5'-ATACATTAGGCTAGAAAGGCAACCATC-3' and 5'-TATAGATCTACAGCGCCTCTTTTCTTCTT-3'. The polymerase chain reaction primers added a methionine translation start codon coinciding with a NdeI site at the 5' end, and a translation stop codon and a BglII cloning site at the 3' end of the

*The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1794 solely to indicate this fact.

†To whom correspondence and reprint requests should be addressed.

‡The abbreviations used are: HMG, high mobility group; PAGE, polyacrylamide gel electrophoresis; bp, base pair(s).

\( ^{2} \)V. Lauaret, D. Stehelin, and H. Clevers, submitted for publication.
amplified fragment. The fragment was subcloned into the NdeI/BamHI-digested pET-3c vector (26). The resulting plasmid pET-TCF/HMG was transformed into E. coli strain BL21(DE3). The identity of the inserted fragment was confirmed by sequencing.

**Production and Purification of the Recombinant TCF-1 Peptide**

Cells were grown at 37°C to midlog phase in LB containing 100 µg/ml ampicillin and induced with 0.3 mM isopropyl-1-thio-β-D-galactopyranoside for 3–4 h. The bacteria were harvested by centrifugation (20 min, 4000 × g, 4°C), resuspended in ice-cold lysis buffer (50 mM Tris-HCl, 1 mM EDTA, 10% glycerol, 50 mM NaCl, 1 mM NaF, and 5 mM dithiothreitol, 4 mM CaCl$_2$, 40 mM MgCl$_2$, 0.5 mM phenylmethylsulfonyl fluoride, pH 8.0), and stored at −20°C until further processing. Next, Triton X-100 (0.1%) and lysozyme (0.5 mg/ml) were added to the cell suspension, which was then incubated for 1 h at room temperature. Lysis of cells was achieved by additional sonication (3 × 3′). To reduce the viscosity of the cell lysate, the DNA was broken down with DNase I (10 µg/ml) for 15 min at room temperature. The cellular debris was removed by centrifugation (15 min, 15,000 × g, 4°C). DNA in the supernatant was precipitated with 0.2% polyethyleneimine. The HMG box protein was collected in a two-step (NH)$_4$SO$_4$ precipitation (30% and 60%). The 30% precipitate was discarded. The 60% precipitate was resuspended in 50 mM Tris-HCl, 1 mM EDTA, 10% glycerol, 50 mM NaCl, 1 mM NaF, and 5 mM dithiothreitol, pH 7.5, and dialyzed against the same buffer at 4°C. To purify the HMG box peptide, the protein solution was applied to a 30 × 1-cm Accell Plus CM cation exchange column (Waters), which was eluted with a linear salt gradient (0.1–1 M NaCl). The HMG fractions were pooled, concentrated, and taken up in the desired buffer by Amicon ultrafiltration.

**SDS-PAGE Gel Electrophoresis—SDS-PAGE electrophoresis was performed on a Pharmacia PhastGelSystem using precast gradient 8–25% SDS-PAGE gels. Gels were developed by silver staining.**

**Gel Retardation Analysis—**Annealed oligonucleotides were labeled by T4 polynucleotide kinase with [γ-³²P]ATP. All probes were purified by nondenaturing polyacrylamide electrophoresis. For a typical binding reaction, 10 ng of purified protein was incubated in a volume of 15 µl containing 10 mM Hepes, 60 mM KCl, 1 mM EDTA, 1 mM dithiothreitol, and 12% glycerol. After a 30 min incubation at room temperature, the reaction was stopped (10,000–20,000 cpm, equaling 1 ng) was added, and the mixture was incubated for an additional 20 min. The samples were then electrophoresed through a nondenaturing 8% polyacrylamide gel run in 0.25 X TBE at room temperature.

Oligonucleotide probes were: MW-1c, GGGAGACTGAAACAAGCCGCTCTTACACC annealed to CCCGGTGAGGAGCCGCTTCTTGCAGCTC; MW-1c, ACTGAGIAACCAAGCTTCTC annealed to AGAGCGCGGGCTCTCAGCGCAGTC; MW-1c, ACTGAGIAACCAAGCTTCTC annealed to AGAGCGCGGGCTCTCAGCGCAGTC.

**DNA Bending Assay—**The oligonucleotides TCGAGAAACAAGGC and CTTGGGCTTCTTGTGGG (containing the cognate motif AACAAAG) were phosphorylated, annealed, and inserted between the XhoI and Sall sites of pBEND2 (27). Using the enzymes Mal, BglII, NheI, SpeI, XhoI, and BamHI individually, fragments of constant length were excised from the resulting plasmid pBEND-HMG, effectively yielding probes with variant positions of the TCF-1 binding site. The fragments were then phosphorylated with [γ-³²P]ATP, and purified over acrylamide. A standard gel retardation analysis was subsequently performed.

**Circular Dichroism—**Circular dichroism (CD) measurements were performed on a Jasco-6000 spectropolarimeter equipped with a temperature-controlled water bath. The CD signal was calibrated with d-10 camphor sulfoxide (28). Each spectrum represents an average of 10 scans. The CD spectra were fitted as described (29).

**RESULTS AND DISCUSSION**

An 87-amino acid peptide, consisting of a methionine residue followed by amino acids 149 to 234 of human TCF-1 (Fig. 1), was produced in a T7-based expression system (26). This peptide represents the minimal TCF-1 fragment capable of binding its cognate motif with high affinity. To express this peptide, a DNA fragment encoding the HMG box was excised by polymerase chain reaction from a human TCF-1 cDNA clone and inserted into the pET-3c vector. The resulting construct pET-TCF/HMG was transformed into E. coli BL21(DE3). Overexpression and purification of the recombinant peptide was visualized by SDS-PAGE gel electrophoresis (Fig. 2). After lysis of the cells and removal of insoluble material by centrifugation, the HMG box peptide was collected in a 60% ammonium sulfate precipitate. The HMG box in the redissolved protein fraction was subsequently purified to homogeneity by single-step cation exchange chromatography. A typical elution profile is presented in Fig. 3. The procedure reproducibly yielded 1–2 mg of purified recombinant protein per liter of bacterial culture, appearing as a single band on a silver-stained gel (Fig. 2, lane 7). Overloading of the SDS-PAGE gel revealed the presence of minor contaminants of higher molecular weight (not shown). The purity of the recombinant protein was estimated to be greater than 95%.

Specific binding of the recombinant HMG box peptide to its cognate heptamer motif was analyzed in a gel retardation experiment with the MW-1c probe (13). This probe represents the motif AACAAAG as embedded in the T-lymphocyte specific CD3-ε enhancer. As shown in lane 1 of Fig. 4A, a specific protein/DNA complex was formed under the conditions applied. No binding was observed to the control probe MWE-1c, in which the binding motif was replaced by CCCCGGTT (lane 2).

Several HMG box proteins, including TCF-1, have been demonstrated to contact the DNA helix predominantly within the minor groove (19, 22). The main line of evidence for this notion derives from the application of inosine substitutions in the binding motif. As first described by Star and Hawley (30), replacement of A/T pairs by I/C pairs changes the major groove dramatically but leaves the surface of the minor groove intact. Minor groove recognition is consequently not impaired by such base substitutions. A/T to I/C replacement at positions 1 and 2 (lane 3) or at positions 4, 5, and 6 (lane 5) of the AACAAAG motif did not affect binding. These results were in good agreement with our observations on longer versions of TCF-1 (19) and demonstrated that the isolated HMG box is capable of interacting with determinants located in the minor groove. As expected, A/T to G/C replacement at positions 1 and 2 (lane 4) or 4, 5, and 6 (lane 6) abolished binding.

LEF-1 and SRY, two proteins related to TCF-1, reportedly bend the DNA helix upon binding to their cognate motifs (22, 23). To determine if the TCF-1 HMG box peptide was similarly capable of bending DNA, a circular permutation analysis was performed. The AACAAAG motif was therefore cloned into the vector pBEND2 (27), yielding pBEND2-HMG. Digestion with six individual enzymes and end-labeling of the released fragments yielded a set of probes of constant length (128 bp) but with the AACAAAG motif located at variant position. Linear DNA displays anomalous migration behavior in polyacrylamide gel electrophoresis upon bending. This effect, observed as a decrease in migration velocity, is strongest when the bend occurs in the middle of the DNA probe. As demonstrated in Fig. 4B, anomalous migration behavior was observed with the various probes excised from pBEND2-TCF. In the Mtl1 probe used in lane 1, the motif is located at the 3′ end of the DNA. From lanes 2–5, the motif is placed progressively more to the center of the probe. The BamHI probe used in lane 6 contains the motif at its 5′ end. The protein/probe complex migrated increasingly slower when the motif was shifted from the 3′ end to the center of the probe.

---

3 M. van de Wetering and H. Clevers, unpublished data.
Circular Dichroism Analysis of TCF-1 HMG Box

Chou-Fasman
Garnier

Fig. 1. Sequence of the TCF-1 HMG box (amino acids 149 to 234). The sequence is preceded by an initiator methionine. α-Helices predicted by Garnier (31) and Chou-Fasman (32) analysis are indicated above the sequence. Asterisks indicate residues strongly conserved within the HMG box family (13).

![Image of gel electrophoresis](image)

Fig. 2. Production and purification of TCF-1 HMG box peptide. Protein preparations were analyzed by SDS-PAGE chromatography and silver staining. Lane 1, molecular weight markers. Whole E. coli lysate before (lane 2) and after (lane 3) induction by isopropyl-
1-thio-β-D-galactopyranoside. Lane 4, supernatant after 30% (NH₄)₂SO₄ precipitation. Lane 5, supernatant after precipitation with 60% (NH₄)₂SO₄. Lane 6, proteins precipitated by 60% (NH₄)₂SO₄. Lane 7, TCF-1 HMG box after cation exchange chromatography.

![Image of gel electrophoresis](image)

Fig. 3. Cation exchange elution profile. The 60% (NH₄)₂SO₄ pellet was redissolved, loaded onto an Accell Plus CM cation exchange column, and eluted using a linear salt gradient of 0.1 M NaCl (10% buffer B) to 1 M NaCl (100% buffer B). The recombinant peptide elutes as a single peak indicated by HMG, well separated from endogenous E. coli proteins.

(lanes 1-5). As expected, placing the motif at either end of the probe resulted in protein/DNA complexes with identical mobilities (lanes 1 and 6). These experiments demonstrate that binding of the HMG box peptide to its cognate motif introduced a bend in the DNA helix.

From experiments very similar to those presented here, the angles of flexure have been calculated for SRY (23) and LEF-1 (22) at 85° and 130°, respectively. We have observed that the HMG boxes of TCF-1, LEF-1, Mat-Mc, SRY, and Sox-4 all yield very similar results using the set of probes from Fig.

Fig. 4. DNA binding and bending by the recombinant TCF-1 peptide. A, gel retardation analysis. All probes contain variants of the AACAAAG motif of the CD3-ε enhancer. Lane 1, AACAAAG (MW=1 probe). Lane 2, CCACGGT (MW=1 sac probe). Lane 3, ICAAAG (MW=112 probe). Lane 4, GGCAAAG (MW=1456 probe). Lane 5, AACIIIIG (MW=1456 probe). Lane 6, AAGGGGG (MW=1456 probe). B, TCF-1 induced DNA bending analyzed by a circular permutation assay. A standard gel retardation assay was performed using a set of probes excised from pBEND/HMG with the enzymes indicated below. The position of the last base of the AACAAAG motif relative to the 5' end of the 128-bp probe was as follows. MluI probe, bp 120 (lane 1). BglII probe, bp 114 (lane 2). NheI, bp 108 (lane 3). Spel, bp 90 (lane 4). XhoI, bp 82 (lane 5). BamHI, bp 20 (lane 6). The position of the motif in the various probes is also displayed graphically above the lanes.

4B. As the nature of the DNA flexure and the structure of the protein/DNA complex remain to be resolved, we are reluctant to calculate bending angles from these experiments. Nevertheless, a tendency to strongly bend DNA appears to be a general characteristic of the sequence-specific HMG boxes including that of TCF-1 and does not depend upon protein sequences outside the HMG box proper.

As the recombinant peptide retained the characteristics of a sequence-specific HMG box, we next sought to determine the presence of elements of secondary protein structure. To that end, the CD spectrum of the HMG box peptide in solution was measured at 10 °C (Fig. 5). Mathematical deconvolution of the CD spectrum predicted an α-helical content of 60% and a 40% content of random coil. No β-sheet elements were predicted from the spectrum. Heating the sample resulted in a gradual decrease in α-helicity and an increase of random coil (Fig. 5A). The CD spectroscopy results are in striking agreement with secondary structure predictions according to the algorithms of Garnier (31) and Chou and Fasman (32), which yielded α-helical contents of 60% and 52%, respectively.

* D. Dooijes and H. Clevers, unpublished results.
Both theoretical methods predict two $\alpha$-helices at roughly equivalent positions in the HMG box peptide (Fig. 1). Computer-aided predictions of secondary structure from primary sequence have also been reported for the HMG boxes of LEF-1 (33) and SRY (23). In both cases, extended $\alpha$-helices were predicted at positions similar to those of TCF-1. It appears unlikely that the $\alpha$-helices can mediate base recognition, as a minor groove in its normal configuration will not accommodate a peptide in $\alpha$-helical conformation. However, the DNA flexibility induced by HMG box binding might open the minor groove or otherwise alter local DNA structure to allow for direct base recognition.

We also determined the CD spectrum of an equimolar mixture of the recombinant peptide and a short double-stranded oligonucleotide carrying the TTCAAG motif. Given a $K_d$ of $10^{-9}$ M for sequence-specific HMG box binding (22), the majority of the peptide was expected to be complexed with DNA under the experimental conditions. The CD spectrum of the peptide in the absence of DNA as well as the CD difference spectrum (spectrum of mixture minus spectrum of free DNA) are given in Fig. 5B. The CD difference spectrum is the result of the spectrum of the protein plus all spectral changes induced by complexation. It differed significantly from the spectrum of the uncomplexed protein. From these data we concluded that: 1) the peptide was indeed complexed to DNA, and 2) structural changes coincided with protein/DNA complex formation. The contribution of secondary structure elements could not be directly calculated from the CD difference spectrum. However, from the shape of the spectrum as well as from the conservation of the crossover point at about 200 nM, it appeared that $\alpha$-helicity remained prominent.

In a comparable study, CD spectroscopy has reliably predicted a change of structure for the bZIP motif of the prototypic bZIP protein GCN4 (34) upon binding to DNA. As in the present study, the CD difference spectrum deviated dramatically from the spectrum of the free bZIP peptide. The GCN4-bZIP motif was found to exist predominantly as a random coil in solution but adopted a helical structure upon binding to DNA. By analogy, it has been postulated that the HMG box changes shape upon complexation to DNA (23). The change in the CD spectrum reported here supports such a change. Unfortunately, our CD spectroscopy data do not describe the nature of this putative structural change.

In conclusion, the present study demonstrates the feasibility of the production and biophysical analysis of functional TCF-1 HMG box peptide. A detailed understanding of the three-dimensional structure of the sequence-specific HMG boxes in complex with DNA will have to await NMR and/or x-ray crystallographic analyses. The experiments presented here serve as a starting point for such studies.

REFERENCES