The EMBO Journal vol.11 no.8 pp.3039-3044, 1992

Sequence-specific interaction of the HMG box proteins TCF-1 and SRY occurs within the minor groove of a Watson-Crick double helix

Marc van de Wetering and Hans Clevers[1]

Department of Immunology, University Hospital, PO Box 85500, 3508 GA, Utrecht, The Netherlands

[1]Corresponding author

Communicated by P.Borst

The high mobility group I (HMG) box is proposed to mediate DNA binding in a novel group of transcription-regulating proteins. Two of these, the proteins encoded by the T cell-specific TCF-1 and the mammalian sex-determining gene SRY, carry a single HMG box with specificity for the heptamer motif A/T A/T C A A A G. We have now analysed the mode of interaction of the HMG boxes of TCF-1 and SRY with this motif. Methylation interference footprinting revealed that both HMG boxes contacted adenines on both strands in the minor groove, whereas no major groove guanine contacts were discerned. Diethylpyrocarbonate (DEPC) carbethoxylation interference footprinting of TCF-1 indicated the absence of major groove contacts on positions 5, 6 and 7 of the motif. Carbethoxylation interference was observed, however, on positions 2, 3 and 4 and to a lesser extent on position 1 in the major groove. Combined T → C and A → I substitution, which changes the surface of the major groove but leaves the minor groove intact, did not interfere with sequence-specific binding by TCF-1 and SRY. These observations indicate that recognition of the heptamer motif by the HMG boxes of the distantly related TCF-1 and SRY proteins predominantly occurs through nucleotide contacts in the minor groove.

Key words: DEPC/DMS/DNA binding/HMG box/transcription factor

Introduction

Many of the eukaryotic DNA-binding proteins cloned to date fall into a small number of gene families based on the presence of conserved structural motifs such as the zinc finger (Evans and Hollenberg, 1988), the basic leucine zipper (Landschulz *et al.*, 1988), the homeodomain (Levine and Hoey, 1988) and the helix-loop-helix (Murre *et al.*, 1989). Tjian and co-workers recently recognized a novel type of DNA-binding domain upon analysis of the RNA polymerase I transcription factor UBF. This nucleolar factor carries four regions of homology to high mobility group (HMG) I proteins. One of these so-called HMG box regions was shown to mediate binding to a DNA affinity column (Jantzen *et al.*, 1990). Several HMG box proteins have since been identified, including the fungal mating type genes *Mat-Mc* of *Schizosaccharomyces pombe* (Kelly *et al.*, 1988) and *Mt al* of *Neurospora crassa* (Staben and Yanofsky, 1990), the mammalian sex-determining gene *sry* (Gubbay *et al.*, 1990; Sinclair *et al.*, 1990), and the mitochondrial transcription factor mtTF1 (Parisi and Clayton, 1991). The consensus HMG box comprises ~80 amino acid residues; the average sequence identity between individual HMG boxes is in the order of 25%. The HMG box is believed to interact with DNA as a monomer (Giese *et al.*, 1991; M.van de Wetering and H.Clevers, unpublished).

The T lymphocyte-specific transcription factor TCF-1 contains a single HMG box with a well defined DNA sequence specificity. We originally cloned TCF-1 based on its affinity for the AACAAAG motif in the CD3-ϵ gene enhancer (van de Wetering *et al.*, 1991). We have since shown that high affinity binding of recombinant TCF-1 involves the recognition of the moderately degenerate heptamer motif $^{A}/_{T}{}^{A}/_{T}$CAAAG (Oosterwegel *et al.*, 1991a,b). Two groups reported the cloning of the human and murine homologues of a highly related transcription factor, which was termed TCF-1α (Waterman *et al.*, 1991) and LEF-1 (Travis *et al.*, 1991), respectively. The binding specificity of TCF-1α/LEF-1 appears very similar to that of TCF-1 (Giese *et al.*, 1991; Oosterwegel *et al.*, 1991a; van de Wetering *et al.*, 1991). In line with this, a particularly high sequence similarity (93% identity at amino acid level) is apparent in the region of the DNA-binding HMG boxes of TCF-1 and TCF-1α/LEF-1. We have since cloned two additional human genes with similar levels of sequence identity, termed *TFC-3* and *TFC-4* (Castrop *et al.*, 1992). Sequence identity of the consensus TCF HMG box compared with the other known HMG box genes is in the order of 25%.

We have originally shown that minor groove N3 methylation of each adenosine in AACAAAG abrogates binding of the TCF-1 HMG box, whereas major groove N7 methylation of guanosines in or flanking the motif had no effect (van de Wetering *et al.*, 1991). This observation was later confirmed for the virtually identical LEF-1 HMG box on the TTCAAAG motif (Giese *et al.*, 1991). As argued by the latter authors, these data suggested that the TCF-1 and LEF-1 HMG boxes contacted DNA in the minor groove, but did not prove such highly unusual types of complex formation. For example, the Antennapedia homeodomain reportedly displayed clear N3 adenine methylation interference (Affolter *et al.*, 1990), whereas X-ray crystallography revealed a predominance of major groove contacts (Kissinger *et al.*, 1990).

Like TCF-1, SRY carries a single HMG box with affinity for the motifs AACAAAG and TTCAAAG (Harley *et al.*, 1991; Nasrin *et al.*, 1991). The low level of sequence identity between the SRY and TCF-1 HMG boxes (26% at the amino acid level) in combination with an apparent conservation of their sequence specificity provided the opportunity to analyse in detail the mode of interaction of

two distantly related HMG box proteins with their cognate motifs as a paradigm for HMG box−DNA complex formation.

Results

Many chemical footprinting techniques have been developed to study mechanisms of DNA−protein interaction. As we were particularly interested in determining the face of the DNA helix contacted directly by the HMG box, demonstration of backbone contacts, e.g. by ethylation interference footprinting, appeared less useful. We therefore sought to apply techniques that involve direct base modification and that discriminate between the major and the minor groove.

Methylation interference footprinting

The highest affinity site so far determined for TCF-1 is TTCAAAG, as defined in the TCR-α enhancer (Oosterwegel *et al.*, 1991a). To confirm our initial observations made on the AACAAAG motif of the CD3-ϵ enhancer, we first performed methylation interference footprinting on the TTCAAAG motif. In this assay, dimethyl sulphate (DMS) is used to methylate a double-stranded end-labelled probe, which is then subjected to gel retardation analysis. As a source of recombinant TCF-1, we used a previously described staphylococcal protein A−TCF-1A fusion protein (van de Wetering *et al.*, 1991).

As shown in Figure 1A, methylation interference footprinting performed on the TTCAAAG motif with TCF-1 extended our original observations on the AACAAAG motif: N3 methylation of A4, A5 and A6 on the positive strand, as well as of the two A residues on the negative strand opposite T1 and T2, interfered completely with binding. No interference was observed by N7 methylation of G residues in or near the motif on either of the two strands. The TCF-1 methylation interference footprint on the TTCAAAG motif is very similar to that reported for the related HMG box factor LEF-1 on this motif (Giese *et al.*, 1991).

The SRY HMG box was cloned from male human genomic DNA by the polymerase chain reaction (PCR), and expressed as a maltose binding protein (MBP) fusion product. DMS methylation footprinting with purified recombinant SRY on the AACAAAG motif and the TTCAAAG motif yielded results that were strikingly similar to those of TCF-1 and LEF-1, as reported here and elsewhere (Giese *et al.*, 1991; van de Wetering *et al.*, 1991) (see Figure 1B). These observations strengthened the suggestion that HMG boxes interact with DNA within the minor groove.

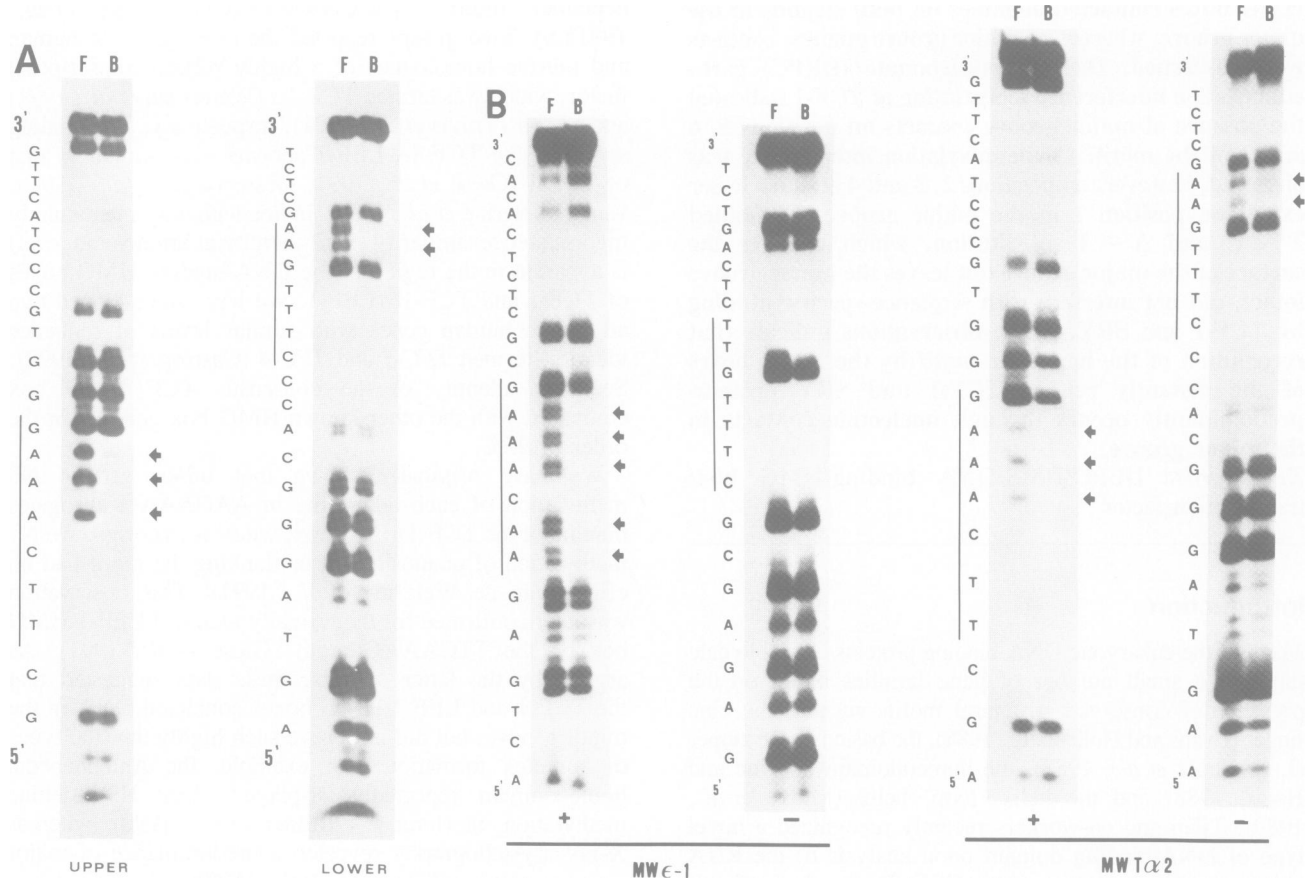

Fig. 1. Methylation interference footprinting. (**A**) Methylation interference footprinting of TCF-1 on the TTCAAAG motif. The left panel represents interference analysis of the positive strand; the right panel that of the negative strand. F, cleavage products of the free probe eluted after gel retardation analysis; B, cleavage products of the bound probe eluted after gel retardation analysis. Arrows point to methylated bases that interfere with binding. A vertical bar indicates the binding motif of TCF-1. Interference is only seen with N-3 methyl groups on adenosine residues within the heptamer motif. (**B**) Methylation interference analysis of binding of SRY. The left panel represents interference analysis performed on the AACAAAG motif; the right panel that on the TTCAAAG motif. +, positive strand; −, negative strand. F, cleavage products of the free probe eluted after gel retardation analysis; B, cleavage products of the bound probe eluted after gel retardation analysis. Arrows point to methylated bases that interfere with binding. A vertical bar indicates the binding motif of SRY. Interference is only seen by N3 methyl groups on adenosine residues within the heptamer motif.

DEPC carbethoxylation interference footprinting

DEPC carbethoxylation interference footprinting allows modification of purines in the major groove. The proposed reaction mechanism for DEPC modification of adenine is opening of the imidazole ring and carbethoxylation of N7, and to a lesser extent of N6 (Leonard *et al.*, 1971), both situated in the major groove. The modification of G residues has been studied in less detail, but reportedly proceeds along the same lines (Herr, 1985). Thus, the resulting purine modifications are much more extensive than those obtained with DMS, and remain confined to the major groove. After modification of the end-labelled probe by DEPC, the assay is essentially identical to the methylation interference footprinting described above.

As demonstrated in Figure 2, DEPC footprinting on the TTCAAAG motif indicated that no major groove contacts occurred on A5, A6 and G7. Complete interference was induced by the modification of A4 and of the A and G opposite T2 and C3. Partial interference was observed on the A opposite T1 and on the G directly preceding the motif.

These observations definitively prove that the second half of the motif (A5−G7) was not contacted in the major groove. In contrast, DEPC footprinting would indicate that recognition of the first half of the motif also involved the major groove. The measured interference on G3 in the negative strand was at variance with the lack of interference observed upon methylation of the same residue (see above). We propose that this apparent contradiction resulted from the different types of modification, as subtle N7 methylation did not interfere with TCF-1 binding, whereas ring opening and carbethoxylation completely blocked the same event. The adenines at positions 1 (partially) and 2 and 4 (completely) appeared to be contacted in the major groove, as well as in the minor groove (see Figure 1). Interference on the G residue at position −1 might again result from steric hindrance by the 'bulky' DEPC-induced modification; we have no indication that the identity of the base at position −1 contributes to TCF-1 binding, as G, T and C are tolerated at this position (Oosterwegel *et al.*, 1991b).

Thus, the HMG box−DNA interaction in the region A5−G7 occurred exclusively in the minor groove, whereas DEPC interference suggested that the HMG box might protrude into the major groove at the first four positions of the TTCAAAG motif. The data obtained for TCF-1 in the methylation and DEPC carbethoxylation interference footprinting are presented in Figure 3.

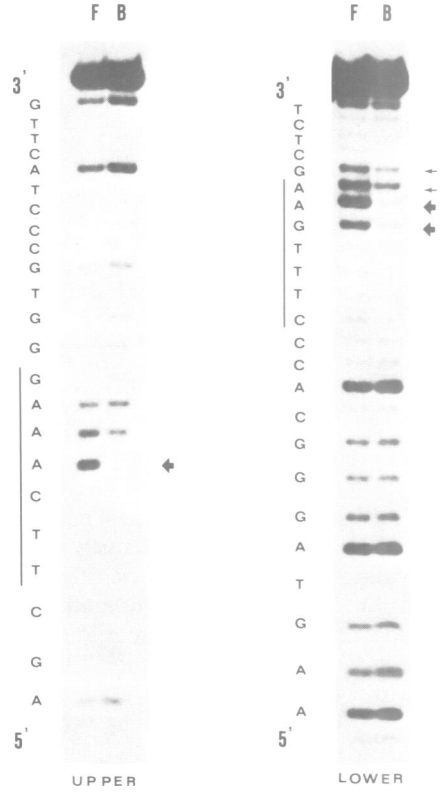

Fig. 2. DEPC carbethoxylation footprinting of TCF-1 on the TTCAAAG motif. The left panel represents interference analysis of the positive strand; the right panel that of the negative strand. F, cleavage products of the free probe eluted after gel retardation analysis; B, cleavage products of the bound probe eluted after gel retardation analysis. A bar indicates the TCF-1 binding motif. Carbethoxylated purine residues that abolish binding are indicated by large arrows. Smaller arrows indicate the carbethoxylated residues that only partially interfere with TCF-1 binding.

Replacement of AT by IC base pairs

As a third assay to discriminate between major and minor groove recognition, we applied an elegant technique recently designed by Star and Hawley to study the interaction of TFIID within the minor groove of the TATAAA motif (Star and Hawley, 1991). These authors reasoned that the exchange of thymines and adenines for cytosines and inosines, respectively, leaves the hydrogen-bonding surface of the minor groove intact, whereas it mimics substitution of G for A and C for T in the major groove. TFIID was still found to bind to the substituted motif, interpreted as definitive proof of its minor groove sequence recognition.

In the gel retardation probe from the TCR-α enhancer used throughout this study, we replaced the positive strand

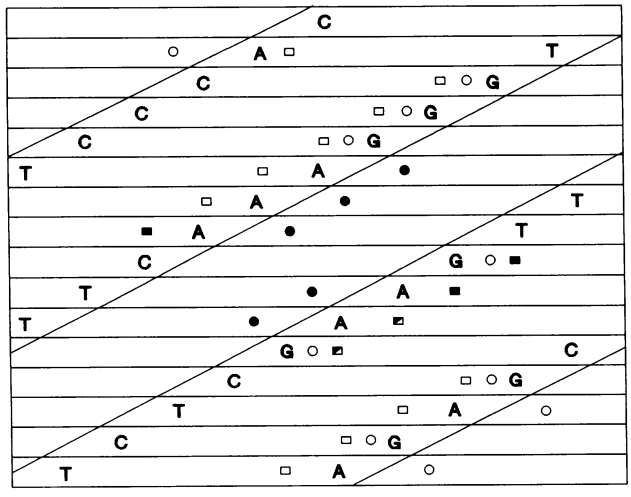

Fig. 3. Summary of methylation and DEPC carbethoxylation interference footprinting of TCF-1. The DNA double helix is displayed in a planar representation. Diagonals indicate DNA backbones and horizontal lines represent the orientation of hydrogen-bonding base pairs. The minor groove thus occupies the centre of the figure from bottom left to top right. Residues tested for interference by methylation are indicated by a circle, those tested for interference by DEPC modification are indicated by a box. Closed circles and boxes in minor and major grooves indicate the occurrence of complete inhibition at that position; opened circles and boxes indicate absence of interference; hatched boxes indicate partial interference. For interpretation of the observed interference patterns, see text.

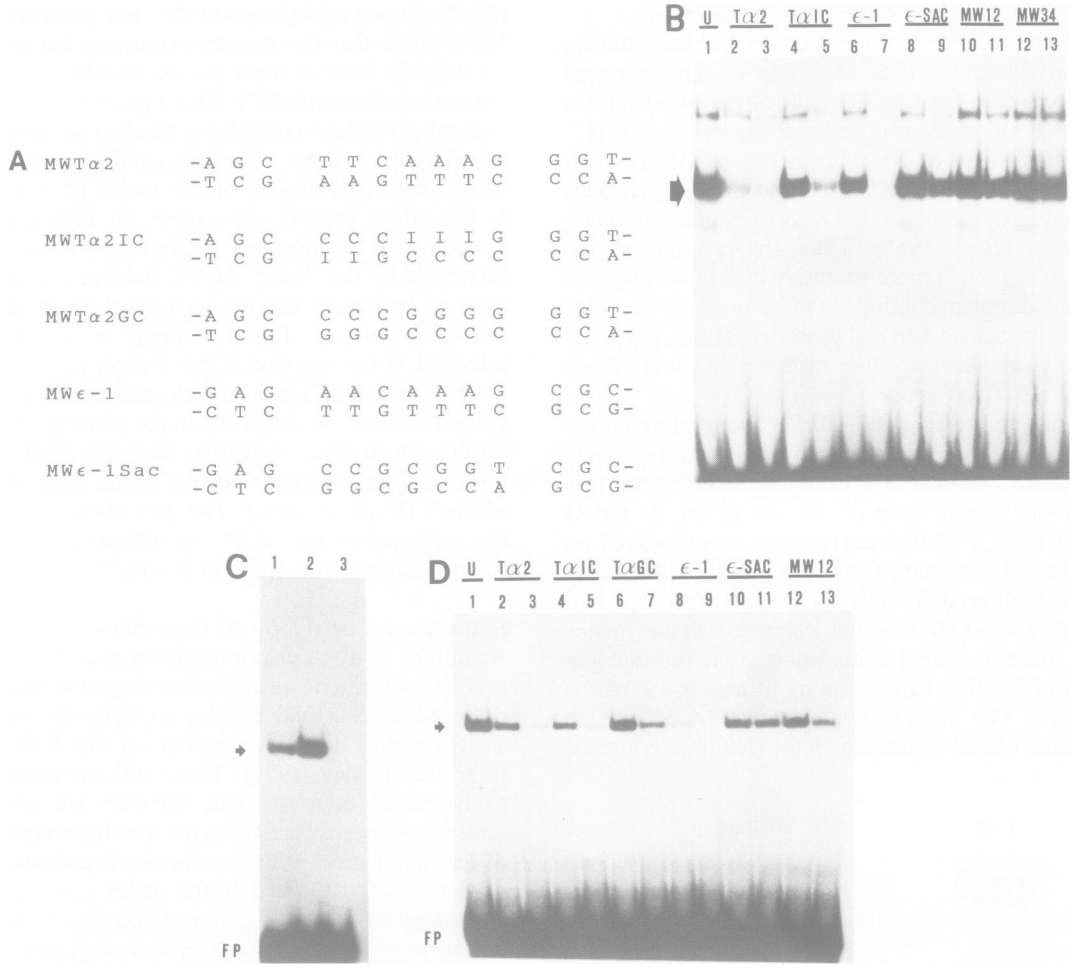

Fig. 4. TCF-1 and SRY bind in a sequence-specific fashion to an A-to-I and T-to-C substituted gel retardation probe. (**A**) The sequences of the heptamer binding site in the positive and negative strand MWTα2 and MWε-1 oligonucleotides, and variants thereof, are given with three flanking base pairs on each side. (**B**) Gel retardation analysis performed with recombinant TCF-1 resulted in a retarded band with the radiolabelled MWTα2IC oligonucleotide. Sequence specificity was confirmed by competition with: lane 1, no competition; lanes 2 and 3, 10- and 100-fold excess of MWTα2; lanes 4 and 5, 10- and 100-fold excess of MWTα2IC; lanes 6 and 7, 10- and 100-fold excess of MWε-1; lanes 8 and 9, 10- and 100-fold excess of MWε-1Sac; lanes 10 and 11, 10- and 100-fold excess of MW12; lanes 12 and 13, 10- and 100-fold excess of MW34. The arrow indicates the position of the TCF-1 retarded band. (**C**) Gel retardation analysis performed with recombinant SRY: lane 1, MWTα2; lane 2, MWTα2IC; lane 3, MWTαGC. SRY binds with comparable affinity to normal and inosine-substituted probe. (**D**) Sequence specificity of SRY binding to MWTα2IC: lane 1, no competition; lanes 2 and 3, 20- and 200-fold excess of MWTα2; lanes 4 and 5, 20- and 200-fold excess of MTα2IC; lanes 6 and 7, 20- and 200-fold excess of MWTα2GC; lanes 8 and 9, 20- and 200-fold excess of MWε-1; lanes 10 and 11, 20- and 200-fold excess of MWε-1Sac; lanes 12 and 13, 20- and 200-fold excess of MW12. The arrow indicates the position of the SRY retarded band.

TTCAAAG by CCCIIIG and the negative strand CTTTGAA by CCCCGII (see Figure 4A). From the minor groove, the substituted motif would still look like the TTCAAAG motif in terms of hydrogen bonding. From the major groove, however, the motif should resemble CCCGGGG. Gel retardation analysis demonstrated that the substituted probe was specifically recognized by recombinant TCF-1 (Figure 4B). Binding could be competed away by excess of unlabelled double-stranded oligonucleotides containing the cognate motifs TTCAAAG, CCCIIIG and AACAAAG. No competition was observed with the CD3-ε enhancer-derived oligonucleotide probe in which the AACAAAG motif was replaced by CCGCGGT, or with two non-specific double-stranded oligonucleotides (van de Wetering *et al.*, 1991). As expected, gel retardation analysis revealed that a control probe in which TTCAAAG had been replaced by CCCGGGG (thus mimicking the major groove aspect of the CCCIIIG probe, Figure 4A) did not form a complex with recombinant TCF-1 (not shown). It should be noted that the TTCAAAG motif competed more effectively for binding to CCCIIIG than did the CCCIIIG motif itself (Figure 4B), the data indicating an estimated 10-fold difference in affinity.

Essentially identical observations were made for SRY. Gel retardation analysis demonstrated that the TTCAAAG and the CCCIIIG probes were bound to a similar extent by recombinant SRY HMG box; no binding was observed to the control CCCGGGG probe (Figure 4C). This observation is particularly striking, as a previous report has demonstrated that single base substitutions at positions 2, 4, 5 and 6 of the cognate motif each abrogate SRY binding. The major groove aspects of three of these substitutions (C at position 2, and G at positions 5 and 6) are equivalent to those of the substituted bases in CCCIIIG.

Binding to CCCIIIG could be competed by 20-fold and 200-fold molar excess of unlabelled double-stranded oligonucleotides containing the cognate motifs TTCAAAG,

CCCIIIG and AACAAAG (Figure 4D). Of note, the TTCAAAG and CCCIIIG oligonucleotides were equally effective competitors of SRY binding, implying that they interact with similar affinities with the SRY HMG box. No competition was observed with various control oligonucleotides at 20-fold excess, including that carrying the CCCGGGG motif. However, some competition was observed at 200-fold excess of the latter motif as well as one of the control oligonucleotides (Figure 4D, lanes 7 and 13). The significance of this observation is unclear.

We concluded that the affinity of the SRY HMG box for TTCAAAG was preserved upon I-for-A and C-for-T substitution in this motif, establishing that the predominant determinants for its sequence-specific binding are located within the minor groove of a DNA double helix.

Discussion

The T lymphocyte-specific transcription factor TCF-1 and the product of the mammalian sex-determining gene *sry* contain a novel type of DNA-binding domain, the HMG box. The consensus HMG box appears structurally unrelated to any of the well characterized eukaryotic DNA-binding protein domains. We propose, based on the present data, that the HMG box has an unusual characteristic in that it interacts with DNA predominantly in the minor groove. Three lines of evidence support this hypothesis. First, methylation interference footprinting on the AACAAAG and TTCAAAG motifs revealed that N3 methylation of all A residues (tested at positions 1 and 2 on both strands, and at positions 4, 5 and 6 on the positive strand) resulted in complete inhibition of TCF-1 binding. The footprint obtained for TCF-1 and SRY on the TTCAAAG motif was very similar to that obtained for the related factor LEF-1 (Giese *et al.*, 1991), supporting our conclusions. Secondly, DEPC carbethoxylation interference footprinting with TCF-1 proved that no major groove contacts occurred at positions 5−7. However, partial or complete interference was observed after major groove DEPC modification of bases at the first four positions. We do not know at present whether any of these four bases are indeed contacted in the major groove. However, it appears more likely in the light of the other experiments that the observed interference is an indirect result caused by the size and nature of the DEPC-induced modifications. The HMG box might thus occupy space in the major groove at positions 1−4 with limited (or no) actual base contacts. Thirdly, I-for-A and C-for-T substitution of the TTCAAAG motif established that the predominant determinants for sequence-specific binding of TCF-1 and SRY are located within the minor groove.

We propose that our findings describe general characteristics of HMG box−DNA interaction. The binding properties of two other HMG box factors have been described in less detail, but are likely to be similar to those of TCF-1 and SRY. The protein encoded by the rat gene IRE-ABP (67% identity to SRY) displays binding specificity for the motif TTCAAAG (Nasrin *et al.*, 1991), whereas the *S. pombe* HMG box transcription factor ste11 (42% identity to SRY) binds to the consensus motif AACAAAGAA (Sugimoto *et al.*, 1991). Not all HMG box factors display this sequence specificity; e.g. UBF binds to a GC-rich stretch in the promoter or rRNA genes (Jantzen *et al.*, 1990), whereas binding of yet other HMG box factors, such as mtTF1 (Parisi and Clayton, 1991), is relatively independent of sequence.

As pointed out by Seeman *et al.* (1976), AT and TA base pairs are not easily discriminated in the minor groove. Hydrogen bonding with the C2 carbonyl of thymine and with N3 of adenine is presumably equivalent, resulting in identical hydrogen-bonding surfaces for AT and TA pairs. In this light, it appears significant that TA to AT transversions are tolerated at the first two positions of the TCF-1/SRY cognate motif, arguing that base contacts at these positions primarily occur through hydrogen bonding in the minor groove. Comparison of optimal binding sites for TCF-1 and testing of specific mutations have indicated that the adenines at positions 4−6 cannot be replaced by thymines without a severe loss of binding affinity (Oosterwegel *et al.*, 1991b). The 'added' specificity of TCF-1 on these positions might be based on the recognition of local DNA backbone structure, or on non-hydrogen bonding interactions (e.g. van der Waals interactions).

Only two other sequence-specific DNA-binding proteins have been described to date that appear to utilize the minor groove as their primary contact surface: the *Escherichia coli* integration host factor (IHF) (Yang and Nash, 1989) and the TATA-binding protein TFIID (Lee *et al.*, 1991; Star and Hawley, 1991). IHF DNA-binding characteristics have been studied by hydroxyl radical and methylation interference footprinting techniques (Yang and Nash, 1989). No three-dimensional structure is available for IHF itself. However, the structure of the related protein HU of *Bacillus stearothermophilus*, which interacts with DNA in a non-specific manner, has been determined by X-ray crystallography. It is proposed that HU utilizes a pair of two-stranded β-ribbon arms which protrude to the protein and encircle the DNA helix (Tanaka *et al.*, 1984). Two independent recent studies propose, deduced from experimental evidence similar to our present data, that TFIID predominantly contacts the minor groove (Lee *et al.*, 1991; Star and Hawley, 1991). In an accompanying letter, a structural relation between TFIID and IHF is suggested based on sequence similarities (Nash and Granston, 1991).

We have not been able to discern similarities between the AACAAAG-binding HMG boxes of TCF-1, SRY and ste11 on the one hand, and TFIID and IHF on the other. We therefore tentatively conclude that the HMG box indeed represents a novel type of DNA-binding domain. The only consistent structural prediction from comparison of Garnier analysis (Garnier, 1978) of the HMG boxes of TCF-1, SRY and ste11 is a strong tendency to adopt an extended α-helical configuration in the region of amino acids 44−75 (H.Clevers, unpublished). Definitive elucidation of the three-dimensional structure of the HMG box−DNA complex awaits NMR (currently under way) and/or X-ray crystallographic analyses.

Materials and methods

Gel retardation assay

Annealed oligonucleotides were labelled by T4 kinase with [γ-^{32}P]ATP. All probes were purified by polyacrylamide gel electrophoresis. For a typical binding reaction, recombinant protein (10 ng) and 10 ng poly(dI−dC) were incubated in a volume of 15 μl containing 10 mM HEPES, 60 mM KCl, 1 mM EDTA, 1 mM DTT and 12% glycerol. After 5 min preincubation at room temperature, probe (10 000−20 000 c.p.m., equalling 0.2 ng) was added and the mixture was incubated for an additional 20 min. The samples

were then electrophoresed through a non-denaturing 4% polyacrylamide gel run in 0.25 × TBE at room temperature. In competition experiments, non-labelled competitor DNA was added together with the poly(dI−dC). Oligonucleotides used were as follows. MW12, GTGCCTCCGCCCAGCTGCCGCT annealed to ACAGCGGCAGCTGGGCGGAGGC; MW34, AGCGCTCTCACACGGGCCTCCGCCC annealed to CTGGGCGGAGGCCCGTGTGAGAGCG; MWε-1, GGGAGACTGAGAACAAAGCGCTCTCACAC annealed to CCCGTGTGAGAGCGCTTTGTTCTCAGTCT; MWε-1Sac, GGGAGACTGAGCCGCGGTCGCTCTCACAC annealed to CCCGTGTGAGAGCGACCGCGGCTCAGTCT; MWTα2, CCCAGAGCTTCAAAGGGTGCCCTACTTG annealed to GGGCAAGTAGGGCACCCTTTGAAGCTCT; MWTα2IC, CCCAGAGCCCCIIIGGGTGCCCTACTTG annealed to GGGCAAGTAGGGCACCCCCCGIIGCTCT; and MWTα2GC, CCCAGAGCCCCGGGGGGTGCCCTACTTG annealed to GGGCAAGTAGGGCACCCCCCGGGGCTCT. All oligonucleotides were synthesized on an Applied Biosystems 381A machine.

Methylation interference footprinting

Probes were labelled either at the positive or the negative strand oligonucleotide with [γ-^{32}P]ATP using T4 polynucleotide kinase and purified on a sequencing gel. After annealing, the probes were purified on a non-denaturing acrylamide gel. The labelled probes were partially methylated at purine residues using DMS (Siebenlist and Gilbert, 1980). 100 000 c.p.m. of methylated probe was used in a 5-fold scale-up of the gel retardation binding reaction. After fractionation by gel retardation assay, the wet gel was subjected to autoradiography. The bound and free probes were cut out and recovered by electroelution. After cleavage by NaOH at the G and A residues, the reaction products were analysed on a 12.5% polyacrylamide−8 M urea sequencing gel.

DEPC carbethoxylation interference assay

Oligonucleotides were labelled and purified as described above. Since DEPC does not modify double-stranded DNA efficiently, the probes were denatured at 90°C for 5 min and then placed on ice. The single-stranded probes were partially carbethoxylated at purine residues using diethylpyrocarbonate (Herr, 1985). After precipitation and annealing, 100 000 c.p.m. were used in a 5-fold scale-up of the gel retardation binding reaction as described for the methylation interference assay. After cleavage by piperidine at the G and A residues (Maxam and Gilbert, 1980), the reaction products were analysed on a 12.5% polyacrylamide−8 M urea sequencing gel.

Production of recombinant protein in E.coli

E.coli strain N4830-1 was used as the host for expression of TCF-1A cloned in the heat-inducible protein A expression vector pRIT2-T as described previously (van de Wetering *et al.*, 1991). Induction and purification of the protein A fusion product was performed according to the manufacturer's instructions (Pharmacia) and stored at −70°C. The SRY HMG box was amplified from male human genomic DNA with PCR using the primers 5′-GGGGAATTCAAAGGCAACGTCCAGGAT-3′ and 5′-GGGAAGCTTACATCTTCGCCTTCCGACG-3′. PCR products were digested with *Eco*RI and *Hin*dIII, subcloned into pIH902 (New England Biolabs) and transformed into *E.coli* strain TB1. After induction, the MBP−SRY fusion protein was purified according to the manufacturer's instructions (New England Biolabs). The MBP−SRY fusion protein contained the HMG box from Lys53 Gly Asn . . to . . Ala Lys Met137 (Sinclair *et al.*, 1990; Harley *et al.*, 1991).

Acknowledgements

The authors wish to thank Drs L.van Houte, M.Oosterwegel, S.Verbeek and P.Toye for critically reading the manuscript, and W.van Driel and members of the laboratory for helpful discussions.

References

Affolter,M., Percival-Smith,A., Muller,M., Leupin,W. and Gehring,W.J. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 4093−4097.

Castrop,J., van Norren,K. and Clevers,H. (1992) *Nucleic Acids Res.*, **20**, 611.

Evans,R.M. and Hollenberg,S.M. (1988) *Cell*, **52**, 1−3.

Garnier,F. (1978) *J. Mol. Biol.*, **120**, 97−120.

Giese,K., Amsterdam,A. and Grosschedl,R. (1991) *Genes Dev.*, **5**, 2567−2578.

Gubbay,J., Collignon,J., Koopman,P., Capel,B., Economou,A., Munsterberg,A., Vivian,N., Goodfellow,P. and Lovell-Badge,R. (1990) *Nature*, **346**, 245−250.

Harley,V.R., Jackson,D.I., Hextall,P.J., Hawkins,J.R., Berkovitz,G.D., Sockanathan,S., Lovell-Badge,R. and Goodfellow,P.N. (1991) *Science*, **255**, 453−456.

Herr,W. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 8009−8013.

Jantzen,H.-M., Admon,A., Bell,S.P. and Tjian,R. (1990) *Nature*, **344**, 830−836.

Kelly,M., Burke,J., Smith,M., Klar,A. and Beach,D. (1988) *EMBO J.*, **7**, 1537−1547.

Kissinger,C.R., Beisha,L., Martin-Blanco,E., Kornbereg,T.B. and Pabo,C.O. (1990) *Cell*, **63**, 579−590.

Landschulz,W.H., Johnson,P.F. and McKnight,S.L. (1988) *Science*, **240**, 1759−1764.

Lee,D.K., Horikoshi,M. and Roeder,R.G. (1991) *Cell*, **67**, 1241−1250.

Leonard,N.J., McDonald,J.J., Henderson,E.L. and Reichman,M.E. (1971) *Biochemistry*, **10**, 3335−3342.

Levine,M. and Hoey,T. (1988) *Cell*, **55**, 537−540.

Maxam,A. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 497−559.

Murre,C., McCaw,P.S. and Baltimore,D. (1989) *Cell*, **56**, 777−783.

Nasrin,N., Buggs,C., Kong,X.F., Carnazza,J., Goebl,M. and Alexander-Bridges,M. (1981) *Nature*, **354**, 317−320.

Nash,H.A. and Granston,A.E. (1991) *Cell*, **67**, 1037−1038.

Oosterwegel,M., van de Wetering,M., Dooijes,D., Klomp,L., Winoto,A., Georgopoulos,K., Meijlink,F. and Clevers,H. (1991a) *J. Exp. Med.*, **173**, 1133−1142.

Oosterwegel,M., van de Wetering,M., Holstege,F., Prosser,H.M., Owen,M.J. and Clevers,H. (1991b) *Int. Immunol.*, **3**, 1189−1192.

Parisi,M.A. and Clayton,D.A. (1991) *Science*, **252**, 965−969.

Seeman,N., Rosenberg,J. and Rich,A. (1976) *Proc. Natl. Acad. Sci. USA*, **73**, 804−809.

Siebenlist,U. and Gilbert,W. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 122−126.

Sinclair,A.H., Berta,P., Palmer,M.S., Hawkins,J.R., Griffiths,B.L., Smith,M.J., Foster,J.W., Frischau,A.-M., Lovell-Badge,R. and Goodfellow,P.N. (1990) *Nature*, **346**, 240−244.

Staben,C. and Yanofsky,C. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 4917−4921.

Star,D.B. and Hawley,D.K. (1991) *Cell*, **67**, 1231−1240

Sugimoto,A., Iino,Y., Maeda,T., Watanabe,Y. and Yamamoto,M. (1991) *Genes Dev.*, **5**, 90−99.

Tanaka,I., Appelt,K., Dijk,J., White,S.W. and Wilson,K.S. (1984) *Nature*, **310**, 376−381.

Travis,A., Amsterdam,A., Belanger,C. and Grosschedl,R. (1991) *Genes Dev.*, **5**, 880−894.

Waterman,M.L., Fischer,W.H. and Jones,K.A. (1991) *Genes Dev.*, **5**, 656−669.

van de Wetering,M., Oosterwegel,M., Dooijes,D. and Clevers,H. (1991) *EMBO J.*, **10**, 123−132.

Yang,C.C. and Nash,H.A. (1989) *Cell*, **57**, 869−880.

Received on March 17, 1992; revised on April 23, 1992