THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 275, No. 35, Issue of September 1, pp. 27266–27273, 2000

Printed in U.S.A.

Sequence-specific High Mobility Group Box Factors Recognize 10–12-Base Pair Minor Groove Motifs*

Received for publication, May 15, 2000, and in revised form, June 20, 2000
Published, JBC Papers in Press, June 23, 2000, DOI 10.1074/jbc.M004102200

Moniek van Beest‡, Dennis Dooijes‡, Marc van de Wetering‡, Søren Kjaerulff§, Alexandre Bonvin¶, Olaf Nielsen§, and Hans Clevers‡‖

From the ‡Department of Immunology, University Medical Center Utrecht, Heidelberglaan 100 Rm F03.821, 3584 CX Utrecht, The Netherlands, §Department of Genetics, University of Copenhagen, Oster Farimagsgade 2A, DK-1353 Copenhagen, Denmark, and ¶Bijvoet Institute for Biomolecular Chemistry, Section NMR Spectroscopy, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands

Sequence-specific high mobility group (HMG) box factors bind and bend DNA via interactions in the minor groove. Three-dimensional NMR analyses have provided the structural basis for this interaction. The cognate HMG domain DNA motif is generally believed to span 6–8 bases. However, alignment of promoter elements controlled by the yeast genes *ste11* and *Rox1* has indicated strict conservation of a larger DNA motif. By site selection, we identify a highly specific 12-base pair motif for Ste11, AGAACAAAGAAA. Similarly, we show that Tcf1, MatMc, and Sox4 bind unique, highly specific DNA motifs of 12, 12, and 10 base pairs, respectively. Footprinting with a deletion mutant of Ste11 reveals a novel interaction between the 3′ base pairs of the extended DNA motif and amino acids C-terminal to the HMG domain. The sequence-specific interaction of Ste11 with these 3′ base pairs contributes significantly to binding and bending of the DNA motif.

Tjian and co-workers (1) originally recognized the HMG[1] box in the RNA polymerase I transcription factor UBF as a novel type of DNA-binding domain. UBF carries several regions of homology to high mobility group-1 proteins. One of these so-called HMG box regions was shown to mediate binding to a DNA-affinity column. Numerous HMG box proteins have since been identified. An evolutionary study of the HMG box family has led to the notion that two types of HMG box proteins can be distinguished (2). The first type usually contains multiple HMG boxes that bind DNA in a structure-dependent but sequence-independent fashion. Prototype factors are UBF and HMG-1. The second type of protein contains a single HMG box that can bind to DNA in a sequence-specific fashion. Examples of the latter are encoded by the Tcf/Lef family members, the mammalian sex-determining gene *Sry* and related *Sox* genes, and by several genes involved in fungal mating-type determination such as the *Schizosaccharomyces pombe* genes *matmc* and *ste11* (2, 3).

The HMG box binds DNA as a monomer (2, 3). Various biochemical experiments involving footprinting techniques and (A/T)-to-(I/C) substitutions have demonstrated that sequence-specific HMG boxes bind predominantly in the minor groove of the DNA helix (4, 5). Circular permutation assays indicated the concomitant induction of a strong bend of approximately 70–130° in the DNA helix (6–10). Based on methylation interference footprinting performed for the T-cell-specific transcription factor Tcf1, we originally proposed a heptamer-binding site for sequence-specific HMG boxes, (A/T)(A/T)CAAAG (11). Subsequent studies on Lef-1, Sry, Sox4 and -5, and MatMc were in agreement with this notion (*e.g.* Refs. 4, 6, 7, and 12–14).

A number of NMR studies have provided a structural basis for the unusual mode of DNA binding by HMG boxes. The non-sequence-specific HMG boxes of HMG-1 (15, 16), *Saccharomyces cerevisiae* NHP6A (17), and *Drosophila* HMG (18) were found to adopt highly similar structures consisting of three α-helices, arranged in an unusual L-shape or arrowhead. Based on the presence of very similar secondary structural elements, an analogous structure was suggested for the sequence-specific HMG box of Sox-5 (12). The structure of the non-complexed Sox4 HMG box (19) confirmed the similarity in overall structure to the three α-helix/L-shapes of HMG-1. A major difference is a shortened third helix in Sox4, caused by a helix-breaking proline residue conserved between all sequence-specific HMG boxes (2), followed by an irregularly structured C terminus.

The reported NMR structures of Sry·DNA (20) and Lef-1-DNA (21) complexes have elucidated the nature of the sequence-specific HMG box-DNA interaction. One arm of the L-shape is formed by helix 1 and helix 2 and the second by helix 3 and the adjacent extended C-terminal segment. The concave surface of the twisted L-shape is docked into a widened minor groove of a strongly bent DNA molecule. Most base contacts occur with the helix 1/2 region in the minor groove. For Lef-1, the studied DNA motif was TTCAAAGG. Sry was studied in complex with its CAAAC core DNA motif. The bending of the DNA helix away from the HMG box is, in part, mediated through the intercalation of a hydrophobic residue (Met in Lef-1; Ile in Sry) between the second and third base pairs of the CAAA(G/C)AAAC core. Additional base contacts are mediated by a tyrosine residue located C-terminal to helix 3 (Tyr-74 in Sry; Tyr-75 in Lef-1) and occur with AT base pairs directly 5′ of the core cognate DNA motif. A major difference between the proposed structures for Sry and Lef-1 lies in their C termini. A unique feature in the Lef-1 structure is a contact of Arg-81 with the backbone phosphate directly 3′ of the core. Thus, the irregularly structured C terminus of Lef-1 makes a sequence-specific contact 5′ of the cognate core through Tyr-75 but also mediates a backbone contact 3′ of the core through Arg-81. This is possible only because the two ends of the DNA motif are brought together by bending.

* This work was supported by the Netherlands Organization for Scientific Research-Chemical Research The Netherlands (NWO-SON). 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 1734 solely to indicate this fact.

‖ To whom correspondence should be addressed. Tel.: 31-30-2507674; Fax: 31-30-2517107; E-mail: h.clevers@lab.azu.nl.

[1] The abbreviations used are: HMG, high mobility group; PCR, polymerase chain reaction; DMS, dimethyl sulfate; bp, base pair.

Many HMG box factors have been demonstrated to bind DNA in a sequence-specific fashion and to transactivate transcription in transient co-transfection assays (*e.g.* Refs. 14 and 22–25). The *S. pombe* Ste11 transcription factor has provided a unique opportunity to study the *in vivo* effects of HMG box genes. *ste11* is indispensable for nitrogen starvation response and mating type determination of the fission yeast. Multiple target genes for Ste11 have been found, based on the dependence of Ste11 protein and/or of intact Ste11-binding sites in the respective promoters. Kjaerulff *et al.* (26) compared the Ste11-binding sites from these biological target genes and thereby defined the consensus Ste11 "response element" (or TR box) as AACAAAGAAA. This consensus TR box compared well with the original consensus DNA motif YAACAAAGAA (27), which was based on the alignment of 10 conserved elements in the promoters of genes induced upon sexual development of *S. pombe*. Thus, the TR box was considerably longer than the 6–8-bp motifs reported for other HMG box proteins (2). Based on similar considerations, a 12-bp motif has been proposed for the *S. cerevisiae* protein Rox1, GAGAACAATYYY (28). The available biochemical and structural data for HMG boxes do not predict these extended cognate DNA motifs. In order to provide a biochemical basis for physical recognition of this biologically defined DNA motif, we have performed binding site selections with Ste11, Tcf1, MatMc, and Sox4. Furthermore, binding characteristics of the Ste11 HMG box to the selected 12-base pair motif were analyzed by methylation interference footprinting, by circular permutation, and by competition experiments in a gel retardation assay.

EXPERIMENTAL PROCEDURES

Recombinant HMG Box Proteins—Production of glutathione *S*-transferase fusion protein Ste113 (Ste11 amino acids 1–113) (26), MalE-MatMc-HMG fusion protein (7), and the His-tagged HMG boxes of Sox4 (14) and Tcf1 (29) have been described elsewhere. The deletion mutant Ste91, encoding amino acids 1–91 of the Ste11 clone, was generated by PCR using the primers 5′ GCACCCGGGTCTGCTTCTTTAACAGCC 3′ and 5′ GCAGAATTCTCTACGAACAGTAGACCG 3′. The PCR product was subsequently cloned into the *Sma*I and *Eco*RI restriction sites of pGEX-2T (Amersham Pharmacia Biotech). The GST-Ste91 plasmid was introduced into *Escherichia coli* strain DH5, and the protein was induced and purified according to the manufacturer's instructions.

PCR-mediated Site Selection—A random probe was generated by labeling primer A (5′ GTTACCGCGGATCCGAATTCCC 3′) with [^{32}P]ATP using T4 polynucleotide kinase and subsequent annealing of this primer A to primer BSS (5′ CTCGGTACCTCGAGTGAAGCTTGANNNNNNNNNNNNNNNNNNGGGAATTCGGATCCGCGGTAAC 3′, where N is A/C/G/T). The independent Tcf1 site selections were performed using the modified primer BSS-TCAA (5′ CTCGGTACCTCGAGTGAAGCTTGANNN**TCAA**NNNNNNNNNNNNNNNNNGGGAATTCGGATCCGCGGTAAC 3′). Bold letters indicate the fixed part of primer BSS-TCAA. MatMc, Sox4, and Tcf1 HMG domain proteins were subjected to a gel retardation assay, as described previously (11). In a binding reaction, the Ste11, MatMc, Sox4, and Tcf1 proteins (50 ng) were incubated in a volume of 15 μl containing 10 mM HEPES, 60 mM KCl, 1 mM EDTA, 1 mM dithiothreitol, and 15% glycerol. After addition of 10,000 cpm (equaling 0.5 ng) of the random probe, the reaction was left at room temperature for 20 min. Samples were electrophoresed through a 5% non-denaturing polyacrylamide gel in 0.25× TBE at room temperature. The wet gel was scanned using a Molecular Dynamics PhosphorImager; retarded protein-DNA complexes were excised from the gel, and the probe was isolated by electroelution. After a phenol/chloroform step, and subsequent precipitation with NaAc and ethanol, the eluted probe was amplified in a PCR containing the labeled primer A and unlabeled primer B (5′ CTCGGTACCTCGAGTGAAGCTTGA 3′) according to the following protocol: 5 min at 94 °C, 25 times (30 s at 94 °C; 30 s at 55 °C; 30 s at 72 °C), 5 min at 72 °C. The probe was purified on a 5% polyacrylamide gel and subsequently electroeluted. New gel retardation reactions were repeated using recombinant protein and "enriched" random probe. The final enriched probe was amplified using unlabeled primers A and B, and the PCR product was subsequently digested with *Eco*RI and *Xho*I and cloned into pBluescriptSK (Stratagene). The cloned products were sequenced using the T7 Sequenase dGTP reagent kit (Amersham Pharmacia Biotech) according to the manufacturer's protocol.

s1	tGA	GAACAAAGAAAatactcg	ggg
s2	tGA	GAACAAAGgAAtatcagg	ggg
s3	tga	AACAAAGAAActcgtggg	ggg
s4	tga	gcAGAACAAAGAAAcact	ggg
s5	tGA	agACAAAGAAAtagccaa	ggg
s6	tGA	GAACAAAGAAAaatcccc	ggg
s7	tGA	tAACAAAGAAAttgttca	ggg
s8	tGA	GAACAAAGAAActatatt	ggg
s9	tGA	GAACAAAGAAAgctgttt	ggg
s10	tGA	GgACAAAGAAAgtgtgct	ggg
s11	tga	AACAtAGAAAcgaacaaa	ggg
s12	tga	gcGcaAACAAAGAAAatc	ggg
s13	tGA	GAACAAAGAActagatga	ggg
s14	tGA	GAACAAAGAAActacaaa	ggg
s15	tGA	GAACAAAGAAAatgtatg	ggg
s16	tga	gttAGAACAAAGAAAttc	ggg
s17	tGA	GAACAAAGAAActatatt	ggg
s18	tga	gACAAAGAAAcctggcga	ggg
s19	tGA	GAACAAAGAAAtgtactt	ggg
s20	tGA	agACAAAGAAAattcact	ggg
s21	tGA	GAACAAAGAAttgttttg	ggg
s22	tGA	tAACAAAGAAAccacatg	ggg
s23	tGA	GAACAAAGAAAttaatgt	ggg
s24	tga	gGAGAACAAAGAAttaaa	ggg
s25	tga	ctAGAACAAAGAAAccca	ggg
s26	tGA	GAACAAAGAActctcgtt	ggg
s27	tGA	GAACAAAGAAAgtgatta	ggg
s28	tGA	GAACAAAGAAAcgaagag	ggg
s29	tga	gACAAAGAAAtagttgtt	ggg
s30	tGA	GAACAAAGAAAttcattt	ggg
s31	tGA	GAACAAAGAAtcagcaaa	ggg
s32	tGA	cAACAAAGAAAtcgtttg	ggg
s33	tga	tctagACAAAGAAAgttt	ggg
s34	tGA	GAACAAAGAAAcattcag	ggg
s35	tga	taGAGAACAAAGAAActc	ggg
s36	tGA	GAACAAAGAAttagcgtt	ggg
s37	tGA	GAACAAAGAAAgttttt	ggg
s38	tGA	GAACAAAGAAActctaaa	ggg
s39	tga	gcAGAACAAAGAAActat	ggg
s40	tga	ctAGAACAAAGAAAccca	ggg

FIG. 1. **Compilation of sites selected with the recombinant Ste11 HMG domain.** A site selection assay was performed for Ste11 as described under "Experimental Procedures." Forty independent sequences were compiled, as judged from differences in the sequences directly adjacent to the selected DNA motif. A 12-bp consensus site can be predicted from this compilation, indicated in *capital letters*. The fixed parts of the BSS primer are indicated on the *right* and *left margins* of each sequence. Those bases of the fixed primers that appeared to belong to the selected DNA motif are depicted in *capital letters*.

Gel Retardation Assay—Experiments were carried out as described previously (11). T4 polynucleotide kinase was used to label annealed oligonucleotides with [γ-^{32}P]ATP. Oligonucleotides were purified on a 10% non-denaturing polyacrylamide gel and electroeluted. In a binding reaction, the recombinant Ste11, MatMc, Sox4, and Tcf1 proteins (50 ng) together with 1 μg of poly[d(I-C)] were incubated and electrophoresed as described above. The oligonucleotides used are as follows: Ste1, 5′ GGGGAGAACAAAGAAAGGG 3′, and Ste2, 5′ CCCTTTCTTTGTTCTCCCC 3′; Mat1, 5′ GGGAAGAACAATGGGGGGG 3′, and Mat2, 5′ CCCCCCCATTGTTCTTCCC 3′; Tcf1, 5′ GGGAAGATCAAAGGGGGG 3′, and Tcf2, 5′ CCCCCCCTTTGATCTTCCC 3′; Sox1, 5′ GGGCAGAACAAAGGCCGGG 3′, and Sox2, 5′ CCCGGCCTTTGTTCTGCCC 3′.

Methylation Interference Footprinting—Probes were labeled either at the positive or negative strand using [γ-^{32}P]ATP and T4 polynucleotide kinase and purified as described above. The labeled probes were partially methylated at purine residues using dimethyl sulfate (30). 100,000 cpm of methylated probe was used in a 5-fold scale up of the gel retardation binding reaction. After separation by gel retardation, the wet gel was subjected to autoradiography. The bound and free probes were excised and recovered by electroelution. After cleavage by NaOH at the G and A residues, the reaction products were analyzed on a 12.5% polyacrylamide, 8 M urea sequencing gel. The probe used, 5′ CCTTCCAAGGTAGAACAAAGAAAGGAATTAAGG 3′, annealed with the complementary strand. The underline indicates the area within the primer corresponding to the Ste11-binding site referred to in Fig. 4.

Ste11 Competition—Ste113 or Ste91 recombinant protein was bound to the optimal Ste11 oligonucleotide as described above. After a 30-min binding reaction, cold oligonucleotides were added in a 10-, 30-, 100-, or 300-fold excess. The cold oligonucleotides consisted either of the optimal Ste11 oligonucleotide described above or a non-optimal oligonucleotide

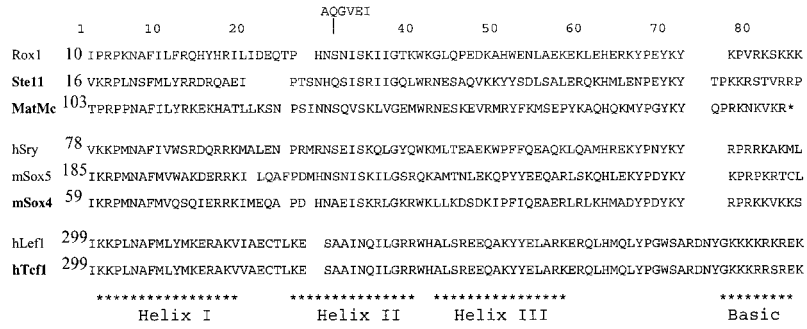

FIG. 2. **Alignment of some sequence-specific HMG domains.** The four proteins used in this study (Ste11, MatMc, Sox4, and Tcf1) are indicated in *bold*. The other proteins are closely related family members for which the structure (Lef1 and Sry) or the binding site (Rox1 and Sox5) has been elucidated. Positions of α-helices are taken from Ref. 2. For each HMG box the first amino acid is numbered according to its position in the full-length protein.

TABLE I
DNA motifs selected by the HMG boxes of Ste11, MatMc, Sox-4, and Tcf-1

The top three lines represent the binding motifs of Rox1 and Ste11 as deduced by comparison of multiple target promotor sites (26, 27, 28) and the binding motifs of Sry and Sox-5 as deduced via site selections (12, 13). The bottom four lines represent the motifs identified by site selection for the indicated HMG boxes. For each position, the frequency of the indicated bases is given relative to the total number of analyzed sequences. For Ste11, the frequency of the first two bases is relative to the number of sequences for which the random part is extended to this position. The second series of frequencies given for Tcf1 were obtained during the second site selection performed specifically with Tcf1 on a random primer containing the fixed TCAA motif.

Sry/Sox-5				A	A	C	A	A	T				
Ste11				A	A	C	A	A	A	G	A	A	A
Rox1	G	A	G	A	A	C	A	A	T	Y	Y	Y	
	0	1	2	3	4	5	6	7	8	9	10	11	12
Ste11		A 7/9	G 29/36	A 34/40	A 40/40	C 40/40	A 40/40	A 39/40	A 40/40	G 40/40	A 39/40	A 40/40	A 34/40
MatMc	A 26/33	A 26/33	G 29/33	A 32/33	A 33/33	C 33/33	A 33/33	A 33/33	T 33/33	G 33/33	G 33/33	G 31/33	
Sox-4		A 23/45	G 22/45	A 40/45	A 41/45	C 45/45	A 45/45	A 45/45	A 38/45	G 45/45	G 40/45		
Tcf-1	A 29/33 n.d.	A 25/33 42/42	G 24/33 41/42	A 25/33 42/42	T 30/33 -	C 24/33 -	A 30/33 -	A 30/33 -	A 26/33 40/42	G 33/33 40/42	G 27/33 24/42	G 26/33 24/42	

consisting of SteNon-1 (5′ GGGGAGAACAAAGACCGGG 3′) annealed with SteNon-2 (5′ CCCGGTCTTTGTTCTCCCC 3′). All retarded bands were quantified on a Molecular Dynamics PhosphorImager using ImageQuant software.

Circular Permutation—The optimal and non-optimal Ste11-binding sites were cloned into pBend2 (31) using the following oligonucleotides: SteBend-1 (5′ CTAGGAGAACAAAGAAA 3′) annealed with SteBend-2 (5′ TGCATTTCTTTGTTCTC 3′) and SteNonBend-1 (5′ CTAGGAGAACAAA GACC 3′) annealed with SteNonBend-2 (5′ TCGAGGTCTTTGTTCTC 3′). Probes containing the binding site at different positions were generated using the restriction enzymes *Bgl*II, *Nhe*I, *Xho*I, *Eco*RV, and *Bam*HI, after which the fragments were labeled using T4 polynucleotide kinase and [γ-^{32}P]ATP. The binding of the Ste11 recombinant proteins to the different probes and the subsequent gel electrophoreses were performed as described above. Experiments comparing the different proteins or the different probes were always run on the same gel. Differences in bending angles for the different probes and proteins were determined according to the algorithm described by Thompson and Landy (32), using the center of each retarded band as indicated in Fig. 6.

Ste11 Modeling—The model of the Ste11-DNA complex was obtained via altering the existing Lef-1/DNA NMR model by Love *et al.* (21). By using the modeling program SETOR (33), the helix 1-to-2 loop in the Lef-1 NMR model was shortened by removing amino acids Val-21 to Ser-24 like the missing four amino acids in the Ste11 loop. To the C terminus, 11 amino acids were added to create a longer tail as present in the Ste113 recombinant protein. We modeled this tail to the minor groove at positions 11 and 12 where predictably interactions would occur. The resulting structure was transformed to the schematic model of Fig. 7 using the VMD modeling program (34). The resulting model is only a schematic representation of the binding site selections and footprint analyses and is not based on any NMR data of an Ste11-DNA complex.

RESULTS

Binding Site Selection—Kjaerulff *et al.* (26) compared the Ste11-binding sites from genetically defined target genes and

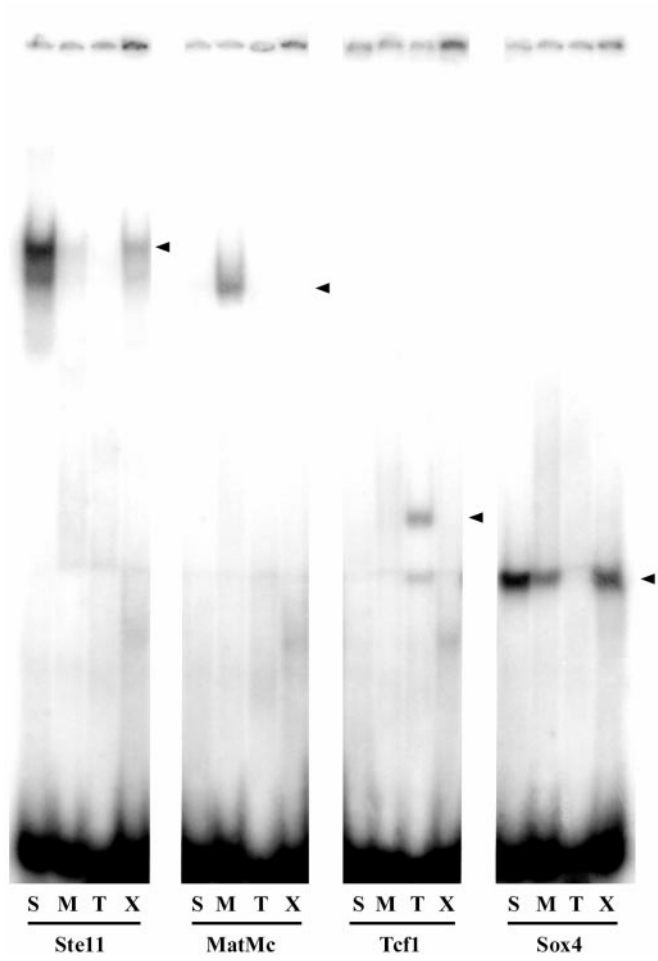

FIG. 3. **Gel retardation analysis of the HMG boxes of Ste11, MatMc, Tcf1, and Sox4.** The four recombinant HMG domains used for the site selection assays were analyzed by gel retardation with optimal probes for each of the four proteins. All lanes representing an individual HMG domain contain identical amounts of that protein; all lanes representing one particular probe contain identical amounts of that probe. The *arrow* indicates for each protein the corresponding specific retarded band. Specificity of individual HMG boxes for particular probes is clearly evident. The differences observed in retardation of the probe reflect differences in the size of the fusion proteins used. *S,* Ste11 optimal probe; *M,* MatMc optimal probe; *T,* Tcf1 optimal probe; *X,* Sox4 optimal probe.

thereby defined the consensus Ste11 response element (or TR box) as AACAAAGAAA. Thus, the TR box was considerably longer than the 6–8-bp motifs reported for other HMG box proteins (2). Several explanations could account for this phenomenon, such as the occurrence of an intimate interaction with another DNA-binding protein at the consensus binding site, the cooperation of the Ste11 HMG box with another DNA-binding domain in the Ste11 protein proper, or the actual recognition of an extended binding site by the single HMG box

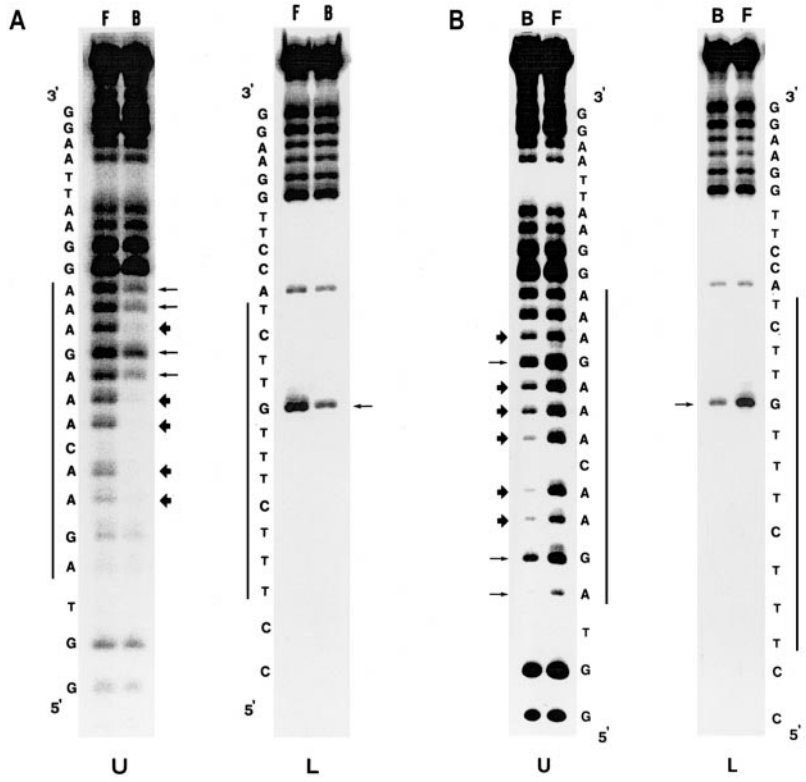

FIG. 4. **Methylation interference footprinting of Ste11.** *A,* methylation interference footprinting of the intact Ste11 HMG box (Ste113). *B,* methylation interference footprinting of Ste91, the C-terminal deletion mutant of Ste11. Note the loss of interference on the adenosines 11 and 12. *F,* cleavage 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, and the *size* of the *arrow* indicates the strength of interference. The *straight line* indicates the part of the probe identical to the optimal Ste11-binding site.

of Ste11. In order to test the hypothesis that HMG boxes can recognize extended sequence motifs, we performed PCR-mediated DNA-binding site selection. Similar selection assays have been performed previously for Sox-5 and for Sry (6, 13). In both cases, comparison of the selected sequences yielded a short sequence motif, AACAAT.

The HMG box of Ste11 (Ste113) was produced as a glutathione *S*-transferase fusion protein in *E. coli*. The integrity of the recombinant HMG box protein was confirmed in a gel retardation assay using a putative Ste11-binding site from the *mfm3* promoter (Ref. 35 and data not shown). To initiate site selection, a gel retardation probe containing a core of 18 random base pairs was prepared by PCR. In order to select the highest affinity sites, conditions in gel retardation were chosen such that less than 10% of the probe was shifted in each round. After 5 rounds of gel retardation, excision of the retarded band and regeneration of the retarded probe by PCR, the selected and amplified product was subcloned and sequenced. Examination of the sequences revealed that 31 of the 40 clones had selected a consensus GAACAAAGAAA motif directly adjacent to the fixed primer BSS. In the 9 other clones, the selected DNA motif did not directly flank the fixed primer. Out of these 9 clones, 7 selected an extra A nucleotide 5′ of GAACAAAGAAA. We concluded that the fixed part of primer BSS had likely contributed to the selection of binding sites and that the fixed A nucleotide directly adjacent to the random nucleotides constituted a genuine contact base. Fig. 1 gives the sequences of the individual clones and the location of the conserved DNA motif with respect to the fixed bases in primer BSS.

To extend these observations, we performed the same site selection assay on three other HMG domains, representative of the three subclasses of the sequence-specific HMG domain family (2) as follows: the *S. pombe* mating type factor MatMc (36), the lymphoid-specific factor Tcf1 (11), and the Sry-like factor Sox4 (14, 37). Fig. 2 gives the amino acid sequences of the four HMG domain genes used in this study and compares these with the two HMG domains for which the protein/DNA structure has been determined, Lef-1 and Sry (20, 21). In addition, it gives two other HMG boxes, Rox1 and Sox-5, for which a binding site has been determined (12, 28). Highly specific DNA motifs were retrieved for all HMG domains after 5–7 selection rounds. Clear differences were noted in the length of the selected HMG domain DNA motifs, as well as in the individual consensus sequences (see Table I). Both MatMc- and Tcf1-binding sites were predominantly located directly adjacent to a stretch of fixed G residues in primer BSS. Since these G residues probably favored binding much as found for the A residue with Ste11, we performed an independent site selection for Tcf1. The selected DNA motif was forced to the center of the randomized sequence by introduction of a fixed TCAA motif in that center. These Tcf1 selections showed a clear preference for a stretch of three G nucleotides flanking the DNA motif that was selected in the initial experiment (see Table I).

Gel retardation was performed with the four HMG domains and their respective selected binding DNA motifs. This clearly demonstrated the specificity of each HMG domain for its optimal DNA motif (Fig. 3). Both Tcf1 and MatMc showed absolute preference for their own DNA motif, whereas Ste11 and Sox4 appeared to be more promiscuous.

Methylation Interference Footprinting—The existing structural and biochemical data did not provide a framework for the mode of binding to this extended DNA motif. We therefore performed a DMS methylation interference footprint analysis. Interference on adenine residues in this issue indicates a minor groove contact, because DMS directly methylates N-3 of adenine which is located in the minor groove (30). Given the expected minor groove interactions, the Ste11 motif $A^1G^2A^3A^4C^5A^6A^7A^8G^9A^{10}A^{11}A^{12}$ was particularly suited to this type of analysis because of the abundance of A residues. The outcome of the interference assay is given in Fig. 4*A*. Strong interference occurred on the adenosines $A^{10}A^{11}A^{12}$, which is outside the core-HMG box recognition sequence. This indicated that these bases, like the core recognition sequence, were contacted by the protein from their minor groove side.

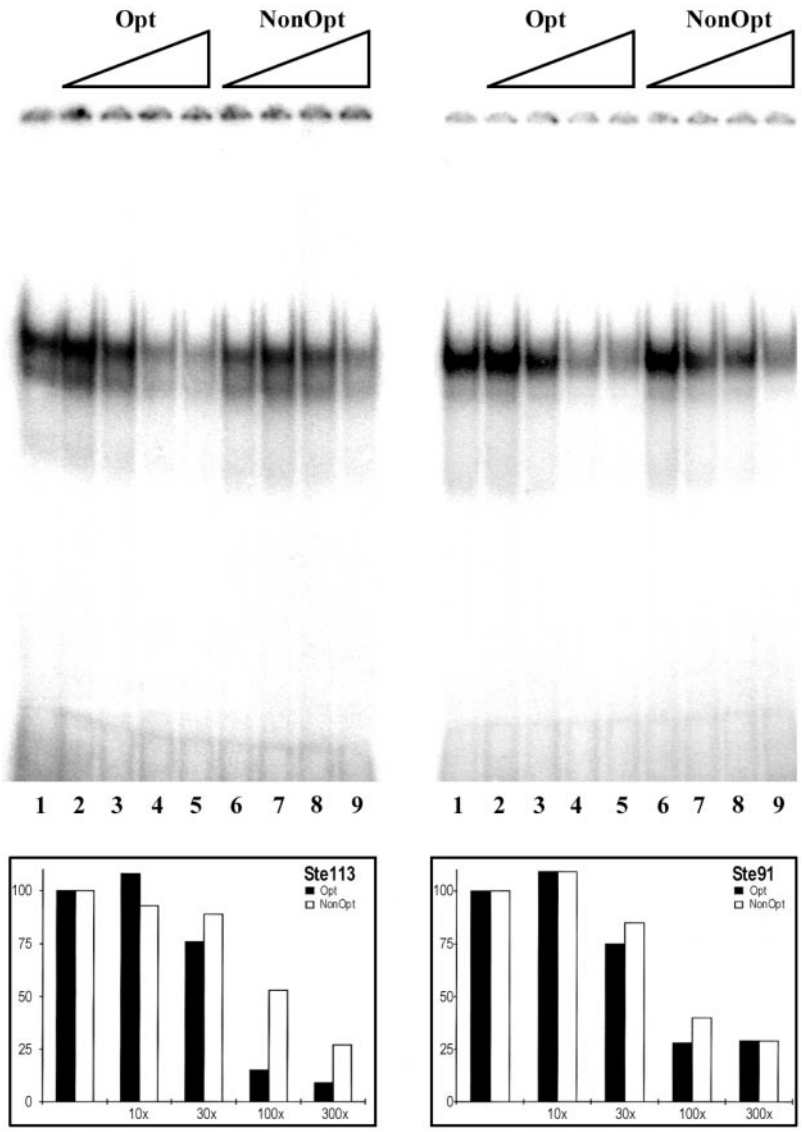

FIG. 5. **Competition for Ste11 with optimal and non-optimal DNA motifs.** The *top panels* represent the electromobility shift assays performed with either the full-length Ste11 HMG box (Ste113; *left*) or the C-terminal deletion mutant (Ste91; *right*). *Lanes 1* represent the uncompeted product. *Lanes 2–5* are competed with 10-, 30-, 100-, and 300-fold excess of optimal ($A^{11}A^{12}$) DNA motif (*Opt*) and *lanes 6–9* with similar excesses of non-optimal ($C^{11}C^{12}$) DNA motif (*NonOpt*). The *bottom panels* show a graphic representation of the gel shifts. The *black bars* represent competition by the optimal DNA motif, and the *white bars* represent the competition by the non-optimal DNA motif. *Lane 1* of either panel is set at 100%; all other bars are calculated as a percentage of the *1st lane*, subtracted by background. By using the Ste113 recombinant protein, the optimal DNA motif clearly competes at 30-fold excess and higher, whereas the non-optimal DNA motif requires at least 100-fold excess for competition. In combination with the Ste91 deletion mutant, either probe competes equally, starting at 30-fold excess.

Although we could observe interference upon methylation of the adenosine A^1, reflecting a minor groove interaction, this interference was less consistently observed (for example compare Fig. 4, *A* and *B*) and likely reflected a low contribution to the overall binding affinity.

C-terminal Analysis—In an attempt to understand the structural basis of the minor groove interactions with $A^{10}A^{11}A^{12}$, we compared the Ste11 HMG box with the NMR models of Sry and Lef-1 (16, 21). Two segments of the HMG box could theoretically interact with $A^{10}A^{11}A^{12}$. 1) According to the existing NMR models, the loop between helix 1 and helix 2 would be close to $A^{10}A^{11}A^{12}$. 2) When Arg-86 (like Arg-81 in Lef-1) would contact the backbone phosphate at A^1, then the amino acid residues C-terminal to Arg-86 could possibly reach the $A^{10}A^{11}A^{12}$ base pairs because both ends of the DNA motif are brought into proximity through the introduction of a sharp bend in the DNA motif.

The most obvious difference between Ste11 and most other HMG boxes is a 4-amino acid deletion in the loop between helix 1 and 2 in the former (see the alignment in Fig. 1). Since we felt that this would preclude the involvement of this loop in $A^{10}A^{11}A^{12}$ recognition, we investigated whether the C terminus of the HMG box of Ste11 mediated the pertinent interactions. To this end, we constructed a mutant Ste11 HMG box that was truncated C-terminally (Ste91). We performed again a DMS methylation interference footprint analysis on the selected Ste11 DNA motif. This experiment, depicted in Fig. 4*B* showed an overall increased cleavage. Comparing the bound (*B*) and free (*F*) probe fractions revealed that positions 2–10 are similarly cleaved as in Fig. 4*A*. Positions A^{11} and A^{12} exhibited a clear loss of interference, indicating that the extreme C terminus of the Ste11 HMG box indeed interacts with these bases (Fig. 4*B*). Note the obvious interference at position A^1 in this experiment.

Functional Analysis of the Extended Ste11-binding Site—In order to explore further the functionality of an extended recognition DNA motif for Ste11, we addressed the contribution of A^{11} and A^{12} in the stability of the Ste11-DNA complex. We therefore performed a competition assay. The full-length HMG box (Ste113) was prebound to the labeled optimal Ste11 DNA motif. By adding an excess of cold optimal or non-optimal DNA motif, in which $A^{11}A^{12}$ was replaced by $C^{11}C^{12}$, we attempted to disrupt the pre-existing protein-DNA complex. Fig. 5 (*left panel*) shows that competition with the optimal $C^{11}C^{12}$ DNA motif resulted in loss of binding at concentrations that were 3–10-fold lower than those of the non-optimal $C^{11}C^{12}$ DNA motif. Similar competition assays performed with the Ste91 deletion mutant revealed that now the $C^{11}C^{12}$ DNA motif was as effective in disrupting the protein-DNA complex as the optimal $C^{11}C^{12}$ DNA motif (Fig. 5, *right panel*). This indicated

that indeed the C-terminal region was involved in the recognition of positions 11 and 12 of the DNA motif.

We then explored whether these C-terminal base pairs directly influenced bending induced by Ste11. First, a circular permutation assay was performed using the recombinant Ste113 and Ste91 proteins on the optimal DNA motif including A^{11} and A^{12}. For each band, the distance to the free probe was determined (using the center of each band indicated with the *horizontal stripe*), and this distance was used to calculate the arbitrary bending angle according to Thompson and Landy (32). This bending angle does not reflect the actual angle as no standards were used but rather serves as a tool for comparing the observed retarded bands. Fig. 6*A* shows that the C-terminal amino acids contributed to the bending angle. The determined arbitrary bending angle decreased from 75 to 71°. Both samples were run on the same gel. We then performed a similar circular permutation assay using Ste113 protein on the optimal $A^{11}A^{12}$ and the non-optimal $C^{11}C^{12}$ Ste11-binding motifs depicted in Fig. 6*B*. *Both panels* of Fig. 6*B* were run adjacently on the same gel but were independently exposed. In this figure it is clearly shown that the inability of Ste11 to interact with base pairs 11 and 12 in a sequence-specific fashion negatively influenced bending. The calculated bending angles showed a decrease from 75 to 65°.

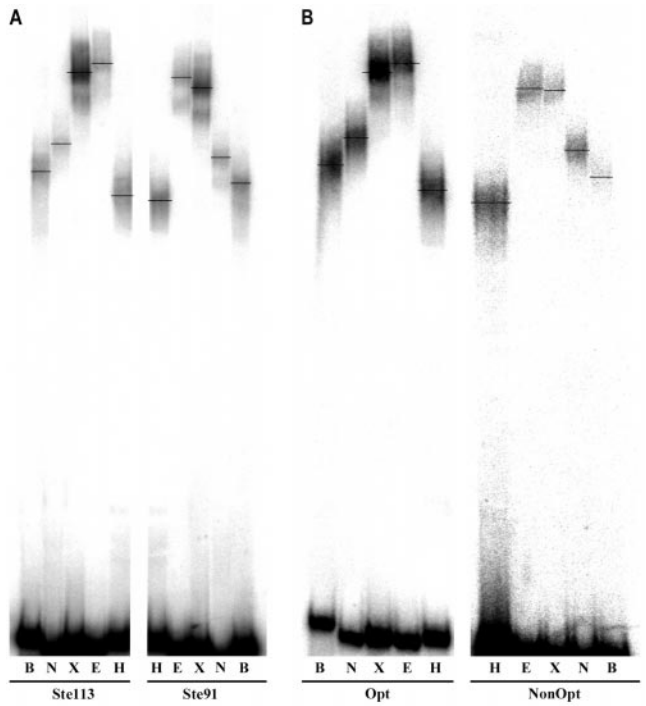

FIG. 6. **Bending by the Ste11 HMG box.** Each half of *A* and *B* originates from the same polyacrylamide gel. By using different restriction enzymes, probes were created carrying the Ste11-binding site at a different position within the probe. The bending of a probe affects its retardation within the gel. Probes bent in the middle will be more retarded than probes bent on either side. Each lane in this figure represents a probe excised from pBend2 using different restriction enzymes (see under "Experimental Procedures"). *B, Bgl*II; *N, Nhe*I; *X, Xho*I; *E, Eco*RV; and *H, Bam*HI. *Bgl*II and *Bam*HI probes position the binding DNA motif at the ends of the probe, whereas *Eco*RV positions the binding DNA motif exactly in the middle of the probe. *Nhe*I and *Xho*I fall between *Bgl*II and *Eco*RV and are expected to give an intermediate retardation pattern. *A,* bending of the optimal DNA motif by the full Ste11 HMG box (Ste113) or the C-terminal deletion mutant (Ste91). *B,* bending of the optimal or non-optimal ($C^{11}C^{12}$) DNA motif by Ste113. The optimal ($A^{11}A^{12}$) and non-optimal ($C^{11}C^{12}$) probes are loaded as mirror images. The *middle* of each retarded band is marked and is used for bending angle calculations. Note the difference in gel retardation for the *Xho*I and *Eco*RV probes of the optimal DNA motif as compared with the non-optimal DNA motif.

DISCUSSION

Sequence-specific HMG boxes constitute unusual DNA-binding domains in several aspects. Despite their high α-helical content, they interact with DNA in the minor groove. Concomitantly, they bend the DNA helix by 70–130° (see Introduction). The nature of the HMG box-DNA interaction has been unveiled through NMR analysis of the Sry and Lef-1 HMG boxes complexed with DNA (20, 21). The current study extends these insights by demonstrating that HMG domains can recognize DNA sequences of up to 12 base pairs with high specificity but that the length and composition of their binding DNA motifs can diverge for different HMG domains. In particular, Sry-type (Sox) HMG domains differ significantly in the length of the selected DNA motif. We further show that separate regions within the HMG domain are involved in the sequence-specific interaction and that the ability to bind to an optimal DNA motif as defined by site selection influences HMG domain function in terms of binding affinity and of the induced bending angle in the DNA helix.

When our current findings are combined with previously reported HMG domain/DNA structures, the following model arises. As demonstrated for four independent proteins (12, 19–21), sequence-specific HMG boxes consist of three α-helical domains and an irregularly structured C terminus. Residues in the helix 1/2 region interact with the HMG box core recognition DNA motif $A^3ACAAAGG^{10}$ or variations. As indicated by earlier biochemical studies and confirmed by the NMR analyses, all interactions with the core DNA motif occur in the minor groove. Intercalation of a hydrophobic amino acid side chain between A^6 and A^7 (as originally proposed for Sry (38)) likely represents a general feature of sequence-specific HMG boxes; it contributes considerably to the observed bending of the DNA helix. The interaction between the helix 1/2 region and the core DNA motif constitutes an important determinant of the overall affinity as single base mutations in this region are usually not tolerated (20, 39).

In our site selection assay, we consistently observed that 2 base pairs were selected 5′ to the core recognition sequence. We will term this the 5′ flank. Assuming that the pertinent DNA-protein interactions occur within the minor groove, methylation interference would be expected on A^1. Such interference was indeed observed but was somewhat variable between assays. This observation does not exclude that the interaction may also involve major groove or phosphate backbone contacts. We assume that the interactions with the base pairs in the 5′ flank contribute only modestly to the affinity of binding. In line with this, the biological consensus DNA motif of Ste11 indicates no preference at positions 1 and 2, although site selection yielded the A^1G^2 sequence in virtually all selected DNA motifs. The biological relevance of the 5′ flank sequence might differ between HMG box factors; in the biologically defined consensus DNA motif of the *S. cerevisiae* transcription factor Rox1, a 5′ flank A^1G^2 sequence is highly conserved (28).

We do not know how base contacts with the 5′ flank sequence are established. Unfortunately, the two NMR studies that have addressed the structure of HMG box-DNA interactions have utilized DNA molecules that are either too short (20) or of a sequence that predictably would not allow the interactions with the 5′ flank to occur (21).

The current study identifies a novel third region in the HMG domain that can interact in a sequence-specific fashion with DNA. This region coincides with a stretch of basic amino acids located C-terminal to the HMG box proper, which has been shown to contribute to the affinity of binding in several studies (4, 40, 41), and more recently has been found to increase the bending angle of a DNA motif complexed to the Lef-1 HMG box

TABLE II

Natural Ste11-binding sites can be divided into two classes

The upper panel gives Ste11-binding sites in ubiquitously expressed target genes. The base pairs that adhere to the optimal motif (indicated in the top row) are depicted in capitals. The A^{11} and A^{12} base pairs, when present, are depicted in bold. The consensus of these binding sites contains the A^{11} and A^{12} base pairs of the optimally selected DNA motif. The lower panel represents Ste11-binding sites in M-cell-specific genes. The expression of these genes is not only Ste11-dependent but also MatMc-dependent. Note that the consensus DNA motif in the MatMc-dependent genes lacks specificity at the 11 and 12 positions.

Gene	DNA motif	Expression
Optimal	AGAACAAAGAAA	
fus1	AGAACAAAGA**AA**	Ubiquitous
fus1	cGAACAAAGA**AA**	Ubiquitous
map1	gtAACAAAGA**AA**	Ubiquitous
mei2	gGAACAAAGA**AA**	Ubiquitous
mei2	AaAACAAAGA**AA**	Ubiquitous
mei2	caAACAAAGA**AA**	Ubiquitous
ste4	AtAACAAAGA**AA**	Ubiquitous
ste6	taAACAAAGA**AA**	Ubiquitous
ste11	gcAACAAAGA**AA**	Ubiquitous
consensus	AACAAAGA**AA**	Ubiquitous
mam1	ctAACAAAGgcc	MatMc dependent
mam2	taAACAAAGAgg	MatMc dependent
mfm1	ccAACAAAGA**A**g	MatMc dependent
mfm2	tGAACAAAGAg**A**	MatMc dependent
mfm3	AcAACAAAGA**A**g	MatMc dependent
sxa2	gcAACAAAGAc**A**	MatMc dependent
consensus	AACAAAGA	MatMc dependent

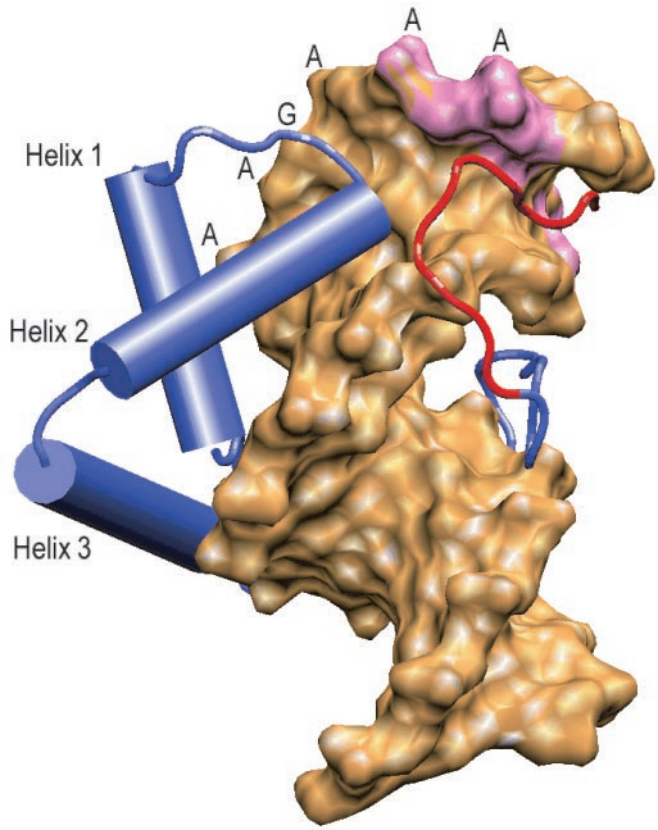

FIG. 7. **Model of Ste11/DNA binding.** This figure represents a model of the Ste11-DNA complex based on the original NMR data of the Lef-1/DNA model (21). The protein is depicted in *blue*. Cylinders represent the α-helices in the protein structure. The artificially extended C terminus is depicted in *red* (comprising residues 85–96). The loop between helices 1 and 2 is shortened as would be predicted for Ste11. The DNA surface structure is depicted in *gold*, with positions A^{11} and A^{12} in *pink*. The indicated base pairs AAGAAA are positions 7–12, the minor groove of which we observe in this figure. Note that the extended C terminus reaches positions 11 and 12.

(10). Our deletion and footprinting data on the Ste11 HMG box indicate that the C-terminal basic region mediates recognition of the $A^{11}A^{12}$ 3′ flank sequence. Examining the bending properties of the C-terminal deletion mutant, lacking 22 amino acids (Ste91), shows that the bending angle is decreased, albeit with a mere 4°. We cannot exclude that some of this decrease may be due to the smaller size of the protein. The interaction with the $A^{11}A^{12}$ positions is much more obvious in the experiments using the binding site mutant $C^{11}C^{12}$. The binding affinity of Ste113 for this mutant was clearly reduced as was DNA bending.

The 3′ flank sequence ($A^{11}A^{12}$) is conserved in biological target genes of Ste11 (26). Strikingly, two classes of Ste11-binding sites can be found in *S. pombe* target gene promoters. One class represents target genes that are ubiquitously expressed, and the second class represents genes expressed only in the mating type-specific M-cell. Binding sites in the former class adhere perfectly to the selected Ste11-binding DNA motif described in this report. Binding sites in the M-cell-specific genes are characterized as weak Ste11-binding sites and remarkably are all defective in the A^{11} and A^{12} positions (Table II). This difference confers regulation of Ste11 target genes by the HMG box mating type gene *matmc*. The ubiquitously expressed Ste11 protein can easily transactivate genes containing the perfect Ste11-binding site present in the first class of target genes. M-cell-specific Ste11 target genes can, however, only be transactivated by Ste11 in the presence of the M-cell-specific gene *matmc*. The HMG box of MatMc, in these cases, binds the DNA near the Ste11-binding site and creates a strong distortion of the DNA, which may be crucial for Ste11 binding. MatMc itself is, however, not present in the resulting DNA-protein complex (26).

NMR studies indicate that the basic C terminus of sequence-specific HMG boxes is irregularly structured. This feature of an irregularly structured sequence adjacent to the HMG box proper appears shared with non-sequence-specific HMG domain proteins. The yeast NHP6A protein contains an irregularly structured N terminus reminiscent of the C terminus of the discussed sequence-specific HMG boxes. NHP6A adopts a similar L-shaped structure in solution. When not bound to DNA, the acidic N-terminal tail is relatively unstructured. Upon binding to DNA, however, the N terminus adopts structure, wrapping around the DNA (17). It has been shown that this N-terminal tail is indispensable for the stable interaction with DNA and for proper bending (17). A similar phenomenon has been demonstrated for the B-domain of HMG-1 (42). Moreover, it has been shown for NHP6A and HMG-1D that these non-sequence specific HMG domain proteins occupy over 11 base pairs (17, 43).

By using the similarity between the tails of non-sequence-specific HMG boxes and our assumed function of the Ste11 C-terminal tail, we attempted to model Ste11 onto DNA for insight in binding. We used the existing structure of Lef-1 as determined by NMR studies (21). The most obvious difference between Lef-1 and Ste11 is the 4-amino acid deletion in the helix-1-to-2 loop. We deleted these amino acids in the Lef-1 structure. Furthermore, we added amino acids to the C terminus as the used Lef-1 HMG box fragment was too short. We aligned the longer C terminus to the minor groove at the level of the 3′ flank sequence, where we had observed sequence-specific contacts in our footprint analyses. The resulting model is depicted in Fig. 7. This model is not based on NMR studies of Ste11 in any way but reflects our view on the possible interaction of Ste11 with DNA, with respect to the 3′ flank sequence.

As most HMG boxes adopt similar structures (see Introduction), we feel that this model represents a possible structure of the Ste11-DNA complex, which should be validated by real NMR studies.

With the rapid identification of new target genes for HMG domain factors (44–51), it becomes increasingly clear that these factors need to be discriminative for DNA binding. All HMG box factors bind to a common core DNA motif roughly comprising $A^3(A/T)^4C^5A^6A^7(A/T)^8G^9$, which makes sequence-specific contacts outside the core DNA motif necessary for discrimination. The matching of the binding sequence to the optimal DNA motif, as shown here, can endow different binding capacities to an HMG domain factor, resulting in different function. A different 3′ flank will not completely abrogate binding but will increase the threshold of binding and will possibly, via altered bending, allow a different nucleoprotein complex to form at this site (52). The composition of an HMG domain binding site reflects yet another level of target gene regulation. An elegant, recent study exemplifies this level of regulation. Tcf/Lef factors transduce Wnt signals by complexing to the Wnt effector β-catenin, which then serves as a transcriptional co-activator. Sox proteins can act as inhibitors of Tcf/Lef-mediated Wnt signaling. They do so by scavenging β-catenin, rendering the Tcf/Lef factors inactive. The Sox proteins (which are potent transcriptional regulators in their own right) fail to bind to Tcf/Lef target sequences, based on the specificity considerations described in this study. Thus, the difference in sequence specificity between Sox proteins and Tcf/Lef factors combined with a shared affinity for β-catenin allows Sox proteins to down-modulate expression of Tcf/Lef target genes (53).

Sequence-specific HMG domain proteins control crucial developmental events in yeast, insects, and vertebrates. The current description of the unusual length and specificity of the cognate DNA motif may provide a basis for the high selectivity with which HMG domain proteins exert these control functions.

Acknowledgments—We thank Dr. Nick Barker for critically reading the manuscript and the members of the Clevers laboratory for helpful discussions.

REFERENCES

1. Jantzen, H. M., Admon, A., Bell, S. P., and Tjian, R. (1990) *Nature* **344,** 830–836
2. Laudet, V., Stehelin, D., and Clevers, H. (1993) *Nucleic Acids Res.* **21,** 2493–2501
3. Grosschedl, R., Giese, K., and Pagel, J. (1994) *Trends Genet.* **10,** 94–100
4. Giese, K., Amsterdam, A., and Grosschedl, R. (1991) *Genes Dev.* **5,** 2567–2578
5. van de Wetering, M., and Clevers, H. (1992) *EMBO J.* **11,** 3039–3044
6. Denny, P., Swift, S., Connor, F., and Ashworth, A. (1992) *EMBO J.* **11,** 3705–3712
7. Dooijes, D., van de Wetering, M., Knippels, L., and Clevers, H. (1993) *J. Biol. Chem.* **268,** 24813–24817
8. Ferrari, S., Harley, V. R., Pontiggia, A., Goodfellow, P. N., Lovell-Badge, R., and Bianchi, M. E. (1992) *EMBO J.* **11,** 4497–4506
9. Giese, K., Cox, J., and Grosschedl, R. (1992) *Cell* **69,** 185–195
10. Lnenicek-Allen, M., Read, C. M., and Crane-Robinson, C. (1996) *Nucleic Acids Res.* **24,** 1047–1051
11. van de Wetering, M., Oosterwegel., M., Dooijes, D., and Clevers, H. (1991) *EMBO J.* **10,** 123–132
12. Connor, F., Cary, P. D., Read, C. M., Preston, N. S., Driscoll, P. C., Denny, P., Crane-Robinson, C., and Ashworth, A. (1994) *Nucleic Acids Res.* **22,** 3339–3346
13. Harley, V. R., Lovell-Badge, R., and Goodfellow, P. N. (1994) *Nucleic Acids Res.* **22,** 1500–1501
14. van de Wetering, M., Oosterwegel, M., van Norren, K., and Clevers, H. (1993) *EMBO J.* **12,** 3847–3854
15. Read, C. M., Cary, P. D., Crane-Robinson, C., Driscoll, P. C., and Norman, D. G. (1993) *Nucleic Acids Res.* **21,** 3427–3436
16. Weir, H. M., Kraulis, P. J., Hill, C. S., Raine, A. R. C., Laue, E. D., and Thomas J. O. (1993) *EMBO J.* **12,** 1311–1319
17. Allain, F. H.-T., Yen, Y.-M., Masse, J. E., Schultze, P., Dieckman, T., Johnson, R. C., and Feigon, J. (1999) *EMBO J.* **18,** 2563–2579
18. Jones, D. N., Searles, M. A., Shaw, G. L., Churchill, M. E., Ner, S. S., Keeler, J., Travers, A. A., and Neuhaus, D. (1994) *Structure* **2,** 609–627
19. van Houte, L. P. A., Chuprina, V. P., van der Wetering, M., Boelens, R., Kaptein, R., and Clevers, H. (1995) *J. Biol. Chem.* **270,** 30516–30524
20. Werner, M. H., Huth, J. R., Gronenborn, A. M., and Clore, G. M. (1995) *Cell* **81,** 705–714
21. Love, J. J., Li, X., Case, D. A., Giese, K., Grosschedl, R., and Wright, P. E. (1995) *Nature* **376,** 791–795
22. Dubin, R. A., and Ostrer, H. (1994) *Mol. Endocrinol.* **8,** 1182–1192
23. Giese, K., and Grosschedl, R. (1993) *EMBO J.* **12,** 4667–4676
24. Molenaar, M., van de Wetering, M., Oosterwegel, M., Peterson-Maduro, J., Godsave, S., Korinek, V., Roose, J., Destree, O., and Clevers, H. (1996) *Cell* **86,** 391–399
25. Yen, Y.-M., Wong, B., and Johnson, R. C. (1998) *J. Biol. Chem.* **273,** 4424–4435
26. Kjaerulff, S., Dooijes, D., Clevers, H., and Nielsen, O. (1997) *EMBO J.* **16,** 4021–4033
27. Sugimoto, A., Iino, Y., Maeda, T., Watanabe, Y., and Yamamoto, M. (1991) *Genes Dev.* **5,** 1990–1999
28. Balusubramanian, B., Lowry, C. V., and Zitomer, R. S. (1993) *Mol. Cell. Biol.* **13,** 6071–6078
29. van Houte, L., van Oers, A., van de Wetering, M., Dooijes, D., Kaptein, R., and Clevers, H. (1993) *J. Biol. Chem.* **268,** 18083–18087
30. Siebenlist, U., and Gilbert, W. (1980) *Proc. Natl. Acad. Sci. U. S. A.* **77,** 122–126
31. Kim, J., Zwieb, C., Wu, C., and Adhya, S. (1989) *Gene* (*Amst.*) **85,** 15–23
32. Thompson, J. F., and Landy, A. (1988) *Nucleic Acids Res.* **16,** 9687–9705
33. Evans, S. V. (1993) *J. Mol. Graph.* **11,** 134–138
34. Humphrey, W., Dalke, A., and Schulten, K. (1996) *J. Mol. Graph.* **14,** 33–38
35. Kjaerulff, S., Davey, J., and Nielsen, O. (1994) *Mol. Cell. Biol.* **14,** 3895–3905
36. Kelly, M., Burke, J., Smith, M., Klar, A., and Beach, D. (1988) *EMBO J.* **7,** 1537–1547
37. Schilham, M., Oosterwegel, M., Moerer, P., Jing, Y., de Boer, P., van de Wetering, M., Verbeek, S., Lamers, W., Kruisbeek, A., Cumano, A., and Clevers, H. (1996) *Nature* **380,** 711–714
38. King, C.-Y., and Weis, M. A. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90,** 11990–11994
39. Deckert, J., Khalaf, R. A., Hwang, S.-M., and Zitomer, R. S. (1999) *Nucleic Acids Res.* **27,** 3518–3526
40. Carlsson, P., Waterman, M. L., and Jones, K. A. (1993) *Genes Dev.* **7,** 2418–2430
41. Read, C. M., Cary, P. D., Preston, N. S., Lnenicek-Allen, M., and Crane-Robinson, C. (1994) *EMBO J.* **13,** 5639–5646
42. Teo, S. H., Grasser, K. D., and Thomas, J. O. (1995) *Eur. J. Biochem.* **230,** 943–950
43. Churchill, M. E. A., Jones, D. N. M., Glaser, T., Hefner, H., Searles, M. A., and Travers, A. A. (1995) *EMBO J.* **14,** 1264–1275
44. Brannon, M., Gomperts, M., Sumoy, L., Moon, R. T., and Kimelman, D. (1997) *Genes Dev.* **11,** 2359–2370
45. He, T.-C., Sparks, A. B., Rago, C., Hermeking, H., Zawel, L., da Costa, L. T., Morin, P. J., Vogelstein, B., and Kinzler, K. W. (1998) *Science* **281,** 1509–1512
46. Laurent, M., Blitz, I. L., Hashimoto, C., Rothbacher, U., and Cho, K. W. (1997) *Development* **124,** 4905–4916
47. McKendry, R., Hsu, S.-C., Harland, R. M., and Grosschedl, R. (1997) *Dev. Biol.* **192,** 420–431
48. Riese, J., Yu, X., Munnertyn, A., Eresh, S., Hsu, S.-C., Grosschedl, R., and Bienz, M. (1997) *Cell* **88,** 777–787
49. Roose, J., Huls, G., van Beest, M., Moerer, P., van der Horn, K., Goldschmeding, R., Logtenberg, T., and Clevers, H. (1999) *Science* **285,** 1923–1926
50. Yamaguchi, T. P., Takada, S., Yoshikawa, Y., Wu, N., and McMahon, A. P. (1999) *Genes Dev.* **13,** 3185–3190
51. Tetsu, O., and McCormick, F. (1999) *Nature* **398,** 422–426
52. Giese, K., Kingsley, C., Kirshner, J. R., and Grosschedl, R. (1995) *Genes Dev.* **9,** 995–1008
53. Zorn, A. M., Barish, G. D., Williams, B. O., Lavender, P., Klymkowsky, M. W., and Varmus, H. E. (1999) *Mol. Cell* **4,** 487–498

Sequence-specific High Mobility Group Box Factors Recognize 10–12-Base Pair Minor Groove Motifs

Moniek van Beest, Dennis Dooijes, Marc van de Wetering, Søren Kjaerulff, Alexandre Bonvin, Olaf Nielsen and Hans Clevers

J. Biol. Chem. 2000, 275:27266-27273.
originally published online August 25, 2000
