Nucleic Acids Research, 1993, Vol. 21, No. 20 4711–4720

DNA-binding specificity of the S8 homeodomain

Ron de Jong, Jessica van der Heijden and Frits Meijlink*
Hubrecht Laboratory, Netherlands Institute for Developmental Biology, Uppsalalaan 8, 3584CT Utrecht, The Netherlands

Received July 15, 1993; Revised and Accepted September 3, 1993

ABSTRACT

The murine S8 homeobox gene is expressed in a mesenchyme-specific pattern in embryos, as well as in mesodermal cell lines. The S8 homeodomain is overall similar to *paired* type homeodomains, but at position 50, which is crucial for specific DNA recognition, it contains a Gln, as is found in Antennapedia (Antp)-type homeodomains. We determined the DNA-binding specificity of the purified S8 homeodomain by *in vitro* selection of random oligonucleotides. The resulting 11-bp consensus binding site, ANC/TC/TAATTAA/GC resembles, but subtly differs from, the recognition sequences of *Antp*-type homeodomains. Equilibrium binding constants of down to 6.0×10^{-10} M were found for binding of the S8 homeodomain to selected oligonucleotides. Using specific antibodies and an oligonucleotide containing an S8-site, we detected by band-shift two abundant DNA binding activities in mesodermal cell lines that correspond to S8 and two more that correspond to its close relative MHox. These S8 protein forms are differentially expressed in retinoic acid-treated P19 EC cells.

INTRODUCTION

DNA-binding regulators of transcription play a key role in embryogenesis by coordinating spatially and temporally regulated gene expression. Many of these transcription factors are characterized by the presence of a homeodomain, a sequence-specific DNA-binding domain (1). The homeodomain is encoded by the homeobox, an evolutionary conserved DNA motif initially identified in *Drosophila* developmental genes (2, 3) and subsequently found in higher organisms including mammals (4). The primary sequence of the homeodomain shows, in its C-terminal half, a striking similarity to the DNA-binding helix-turn-helix (HTH) motif present in several prokaryotic DNA-binding repressors (5). This structural resemblance was confirmed by determining the three-dimensional structure of the Antennapedia (Antp) homeodomain (6).

Of numerous full-length homeodomain proteins and homeodomain peptides, mostly of *Drosophila*, the *in vitro* DNA-binding specificity has been studied (reviewed in 7). Most homeodomain proteins recognize related sites that contain an ATTA core sequence (7), which has been shown to be critical for DNA-binding (8–11). Structural analysis of homeodomain-DNA complexes has provided insight in the occurring base-specific interactions (12, 13). The third helix of the homeodomain is homologous to the recognition helix of the prokaryotic HTH; it also makes base-specific contacts in the major groove with nucleotides in the first half of and 5′ from the ATTA, however through residues located more C-terminally than in the recognition helix of the prokaryotic HTH. In addition, the N-terminal arm of the homeodomain interacts specifically with the 3′ part of the ATTA sequence in the minor groove (12). Amino acid substitution at position 9 of the recognition helix alters the recognition of specific bases immediately preceding the ATTA sequence, indicating that this position is a major determinant of DNA-binding specificity (14, 15).

We study the murine *paired* (*prd*)-related homeobox gene *S8*. This gene is first expressed at the primitive streak stage in extra-embryonic mesoderm (16), and at later stages in a complex pattern in mesenchyme of head, limbs and trunk, in many instances in regions involved in epithelio-mesenchymal interactions (17). The mesodermal specificity of *S8* was reflected in its exclusive expression in fibroblastic cell lines (16). Recently a gene was cloned from mouse (named *MHox* in Ref. 18, and *K-2* in Ref. 19), human (*Phox;* 20) and chicken (A. Brouwer and F.M., unpublished; E.Olson, pers. comm.) that is highly homologous to *S8*. The MHox homeodomain is only different at two positions involving conservative amino acid changes. MHox from mouse and chicken is expressed in a pattern in the embryo, which is similar to that of *S8* but with notable differences (18, 19, F.M., B.Leussink and A. Brouwer, in preparation).

Although the homeobox gene *S8* has been isolated on the basis of similarity to the *Drosophila* segmentation gene *prd* (21), which is reflected at the protein level in a 67% identity in the homeodomains encoded, there are major differences at the structural level (17). First, *S8* lacks the paired box, another conserved sequence motif encoding the DNA-binding Paired domain (22, 23) that is present in several *Drosophila* segmentation genes, including *prd, gooseberry-proximal* and *gooseberry-distal* (24, 25), and the paired box gene (*Pax*) family (26). Second, whereas both the S8 and MHox homeodomains, alike the Antp-type homeodomains, contain a glutamine at position 9 of the recognition helix (17, 18), prd-class homeodomains are characterized by a serine at this position.

* To whom correspondence should be addressed

To investigate its specific DNA-binding properties we determined the consensus recognition sequence of bacterially produced S8 homeodomain by *in vitro* binding site selection using randomized oligonucleotides. This resulted in an 11-bp optimal binding site, ANPyPyAATTAPuC, that resembles the binding sites reported for *Antp*-class homeodomain proteins. The presence of S8 and MHox DNA-binding activities were examined in different cell lines. Several protein-DNA complexes containing S8 or MHox were shown to be present exclusively in fibroblast cells and in P19 cells induced to differentiate by RA.

MATERIALS AND METHODS

Oligonucleotides, PCR and plasmid constructions

The expression plasmid pET-S8HD was constructed by inserting a partial *S8* cDNA fragment encoding the S8 homeodomain into the unique *Bam*HI site of the phage T7 expression plasmid pET-3A (27). The *Bam*HI cDNA fragment was obtained from expression vector pET-S8BS (16). Using standard procedures (28), pET-S8BS was digested by *Bgl*I, treated with S1 and Klenow DNA polymerase, and the homeobox fragment was excised by *Bam*HI after addition of *Bam*HI linkers for cloning in pET-3a.

The binding site selection assay was performed with a 62-bp oligonucleotide, 5'-CAGGTCAGTTCAGCGGATCCTGTCG-$(N)_{12}$-GAGGCGAATTCAGTGCAACTGCAGC-3' (oligo R62) as designed by Pollock and Treisman (29), containing a randomized sequence of 12 nucleotides. This oligo was purchased from Isogen (Amsterdam), while all other oligos were synthesized on a Cyclone Plus DNA Synthesizer (Millipore/Biosearch). Primers A (5'-CAGGTCAGTTCAGCGGATCCTGTCG-3'), and B (5'-GCTGCAGTTGCACTGAATTCGCCTC-3'), complementary to the 3' and 5' non-random part of R62, respectively, were used in PCR amplifications of selected oligos. PCR reaction conditions were as specified by the manufacturer (Promega).

Amplifications were performed in 20 cycles of 1 min at 96°C; 1 min at 62°C; 1 min at 72°C, followed by 10 min at 72°C. After a preparative electrophoretic mobility shift assay (EMSA) step (see below) and a final PCR amplification, the oligonucleotides were extracted with phenol:chloroform (1:1), precipitated with ethanol, digested with *Eco*RI and *Bam*HI, and cloned in pBluescript (Stratagene). Plasmids containing 1 to 9 inserts were obtained and sequenced with T3 and T7 primers using an automated sequencer (Applied Biosystems 370A). Oligo PC1 (5'-CGAGTTAATTAAGCT-3') and its complement PC2 (5'-GCAGCTTAATTAACT-3') were used for EMSA; their sequences were based on a preliminary consensus binding sequence differring from the final optimal binding site only at the less constrained position 11.

Expression and purification of the S8 homeodomain

S8 homeodomain was expressed in E.coli strain BL21(DE3) pLysS carrying the plasmid pET-S8HD. Bacteria were grown in LB medium at 37°C to OD_{600} = 0.8 and then induced by IPTG for 3 hours (27). The bacteria were harvested by centrifugation and resuspended in 0.01 volume of buffer ZTE (50 mM Tris·HCl pH 8.0, 100 mM KCl, 10 mM EDTA, 12.5 mM $MgCl_2$, 0.1% Nonidet P-40, 20% (w/v) glycerol, 1 mM DTT, 1 mM PMSF, 1μM pepstatin A, 10 μM leupeptin) essentially as described in Hoey et al. (30). All subsequent handlings were performed at 4°C or on ice. Cells were freeze-thawed once and sonicated for 3 periods of 4 min. The bacterial lysate was centrifuged for 10 min at 16,000×*g*, NaCl was added to 300 mM NaCl and nucleic acids were removeed by polyethyleneimine precipitation. The clarified extract was fractionated on a Sephacryl S-100 column (90×5.4 cm) using ZTE buffer containing 300 mM NaCl. Combined fractions from the Sephacryl S-100 column were diluted with 0.5 volume buffer A (25 mM Hepes·KOH pH 8.0, 1 mM DTT, 0.02% Nonidet P-40, 20% (w/v) glycerol, 1 mM PMSF, 1 μM pepstatin A) containing 0.25 M NaCl and applied to a 40-ml Sepharose fast-flow S column equilibrated with buffer A/0.25 M NaCl. A linear gradient of 0.5−1.0 M NaCl in buffer A eluted the homeodomain at 500 mM NaCl. Homeodomain fractions were desalted on a Sepharose G-25 gel filtration column using buffer B (25 mM Tris·HCl pH 7.8, 1 mM EDTA, 0.01% Nonidet P-40, 1 mM DTT, 20% (w/v) glycerol, 1 mM PMSF, 1 μM pepstatin A) as eluens. The flow-through fractions were directly loaded onto a 30-ml DEAE-Sepharose A-25 column equilibrated with buffer B. The homeodomain was present in the flow-through that was adjusted to 100 mM NaCl, aliquotted and stored at −80°C. Protein concentration was assayed using Bio-Rad dye reagent (Bradford method) with BSA as standard. The final homeodomain concentration was 122 μM. One liter of induced bacteria cultured yielded about 15 mg of homeodomain peptide.

Binding site selection

Affinity chromatography was carried out essentially as described (10, 31). Purified S8 homeodomain was covalently bound to cyanogen bromide-activated Sepharose 4B (Pharmacia) in coupling buffer containing 0.1 M $NaHCO_3$ pH 8.3, 0.5 M NaCl and remaining active groups were de-activated by 0.1 M Tris·HCl pH 8.0 according to the manufacturers instructions. The final concentration obtained was 400 μg S8 homeodomain per ml of resin and the coupling efficiency was >99%. Double-stranded radiolabeled oligo R62 was obtained by annealing primer R62 to a three-fold molar excess of primer A followed by filling-in with Klenow polymerase in the presence of (α-^{32}P)-dATP and cold dNTPs. The three rounds of selection were performed by loading double-stranded oligos onto a 100-μl column in buffer C (50 mM Tris·pH 8.0, 1 mM DTT, 10 μg/ml gelatin) containing 0.1 M NaCl, 10 μg/ml poly(dI·dC)/(dC·dI) and 20 μg/ml tRNA. After binding, the column was successively washed with buffer C containing 0.1 M, 0.25 M, 0.4 M, and 1 M NaCl. Prior to reloading for the third round of selection, oligos in the 1 M NaCl fraction were amplified by PCR (see Ekker *et al.*, Ref. 10). 7.5 μg, 28 ng and 5 μg, respectively, of double-stranded primer R62 were loaded on the column in the three rounds of selection, the high salt wash containing 3.1%, 53% and 3.4% of the input, respectively.

For preparative EMSA, selected oligonucleotides were PCR-amplified and labeled as described above. Then 225 ng of this oligonucleotide mixture was incubated for 1 hour at room temperature in 250 μl binding buffer (20 mM Hepes·KOH pH 7.5, 50 mM KCl, 0.05% Nonidet P-40, 50 μg/ml BSA, 5 mM DTT) containing 16 nM S8 homeodomain peptide. Upon gel electrophoresis in non-denaturing 10%-polyacrylamide gels shifted oligos (appr. 2% of the input) were excised from the gel, eluted in 0.5 M NH_4Ac, 1 mM EDTA pH 8.0 and precipitated. After PCR amplification, the oligos were cut by *Eco*RI and *Bam*HI and ligated into Bluescript.

EMSA

Selected oligos were excised from plasmids by digestion with *Eco*RI and *Bam*HI, end-labeled with (α-^{32}P)-dATP and purified by preparative polyacrylamide gel-electrophoresis using standard procedures (28). To determine the equilibrium binding constants, various amounts of homeodomain peptide (2.56×10^{-9} M-5×10^{-11} M) were incubated for 45 min at room temperature with 2×10^{-10} M, ^{32}P-labeled oligo in 20 μl binding buffer containing. Prior to further dilution and addition of DNA probe, the homeodomain peptide was incubated at a concentration of 2.56×10^{-9} M in binding buffer for 2 hours at room temperature. Upon EMSA, free and bound DNA were quantitated with a PhosphorImager (Molecular Dynamics, software from ImageQuant). Equilibrium association constants (K_a) were calculated from the midpoint of the curves of the plotted data, using the equation

$$[\text{HD-DNA}] = \frac{[\text{HD}]_0 \cdot K_a([\text{DNA}]_0 - [\text{HD-DNA}])}{1 + K_a([\text{DNA}]_0 - [\text{HD-DNA}])}$$

where $[\text{DNA}]_0$ and $[\text{HD}]_0$ are the total concentrations of DNA and of homeodomain peptide, respectively, and [DNA-HD] represents the concentration DNA-protein complexes. The equilibrium binding constant K_d is $1/K_a$.

Gel shift assays with nuclear extracts were performed with ^{32}P-labeled double-stranded oligo PC1/PC2. Nuclear extracts (5 μg) were incubated 1 hour on ice in 40 μl binding buffer (20 mM Tris·HCl pH8.0, 75 mM KCl, 5% glycerol, 50 μg/ml BSA, 0.025% Nonidet P-40, 1 mM EDTA, 5 mM DTT), 3 μg poly(dI·dC)/(dC·dI) and 0.3 ng labeled oligonucleotide. DNA-protein complexes were separated at 4°C in 5%$-$15% linear gradient polyacrylamide-gels. In competition experiments, unlabeled oligonucleotides were mixed with the binding reaction prior to the addition of nuclear extract. Affinity-purified antibodies were added after 30 min of incubation.

Antibody preparation and immunoblot analysis

We previously described the preparation and affinity-purification of α-S8(II) antibodies, raised against the C-terminus of S8 including the homeodomain (16). Anti-serum α-S8(I) was generated against oligopeptide S8I, that consisted of the sequence (NH_2)-LKSYGQEAAIEQPVAPRPTT-(COOH) located C-terminal to the homeodomain (16) and was synthesized by solid-phase coupling on a Milligen Biosearch type SAM-2. This peptide was covalently coupled to keyhole limpet hemocyanine by 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (Pierce) or by glutaraldehyde (Sigma), according to standard procedures (32). Approximately 500 μg of an equimolar mixture of both conjugates was injected with Freund's complete adjuvant into virgin New Zealand white rabbits. Booster immunizations followed after 3, 6 and 11 weeks using similar amounts of antigen mixed with Freund's incomplete adjuvant. α-S8(I) antibodies were affinity-purified by peptide coupled to reactigel H-25 (Pierce) according to the manufacturer's instructions. Immunoblot analysis of bacterial lysate was performed as described previously (16).

Preparation of nuclear extracts from cultured cells

MES-1, END-2, P19 EC, FIB-9, N1E-115, NIH-3T3 and C3H10T½ cells were cultured in DMEM/F12 medium or DMEM supplemented with 7.5% fetal bovine serum (Gibco) and buffered with CO_2 as described earlier (16). Nuclear extracts were prepared essentially as described (33).

RESULTS

Purification of S8 homeodomain

We expressed in *E.coli* a 74-amino acid S8 polypeptide containing the homeodomain flanked by 11 N-terminal and 3 C-terminal residues. The N-terminal extension shows some sequence similarity with the corresponding portion of *prd*-class homeoproteins and consists of mainly basic amino acids (17). A cDNA sequence encoding this polypeptide was inserted in the expression vector pET-3a that harbours a T7-RNA polymerase promoter (27). This resulted in a construct, pET-S8HD, that encodes a polypeptide containing 13 N-terminal and 4 C-terminal additional vector-derived residues. After IPTG-induction of *E.coli* carrying pET-S8HD, S8 homeodomain peptide made up 10%$-$20% of total protein in the bacterial lysate (Fig. 1, lane 3). The homeodomain peptide exhibited an apparent molecular weight of 15 kD in SDS-polyacrylamide gels, which is higher than calculated from the amino acid sequence (10.5 kD). Expression of S8 protein was confirmed by immunoblot analysis using anti-S8 antibodies (α-S8II; Ref.16) raised against a large moiety of S8 (data not shown). S8 homeodomain was purified to homogeneity by conventional chromatography, as judged from a Coomassie Brilliant Blue-stained SDS-polyacrylamide gel (Fig. 1). Soluble proteins in the bacterial lysate were separated by gel filtration (Sepharose S-100) and the S8 homeodomain was further purified by ion-exchange chromatography (fast-flow S and DEAE) as described in detail in Material and Methods.

Selection of S8-binding sites

Our approach to determining the optimal binding site of S8 consisted of the combined use of affinity chromatography and preparative EMSA (see Fig. 2). The purified S8 homeodomain peptide was used to select oligonucleotides by affinity chromatography, as previously described for the transcription

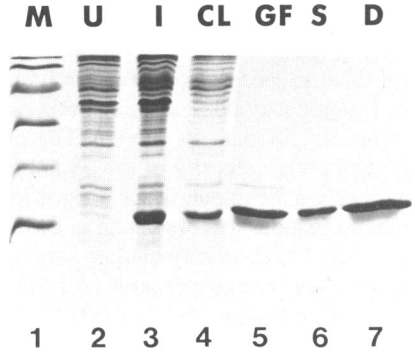


Figure 1. Purification of S8 homeodomain from *E.coli* extract. Purification was followed by SDS-PAGE in a 20%-polyacrylamide gel stained with Coomassie Brilliant Blue. Lane 1, Molecular weight marker proteins with molecular masses of 94, 67, 43, 30, 20.1 and 14.4 kD; lanes 2 and 3, 10 μg of total bacterial proteins from uninduced (U) and induced (I) cells carrying the S8 homeodomain expression construct; lanes 4$-$7, protein analysis of different purification stages. CL, cleared lysate; GF, pooled fractions of Sephacryl S-100 gel filtration column; S, Sepharose fast-flow S column; D, DEAE-Sephadex column. Samples containing 5 μg protein were loaded in lanes 4$-$7.

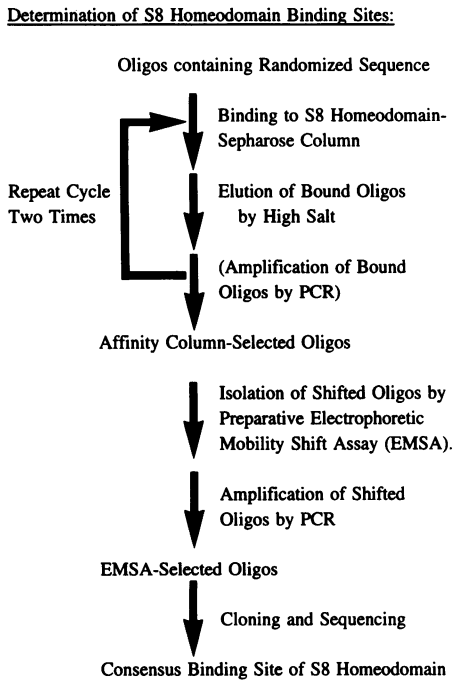
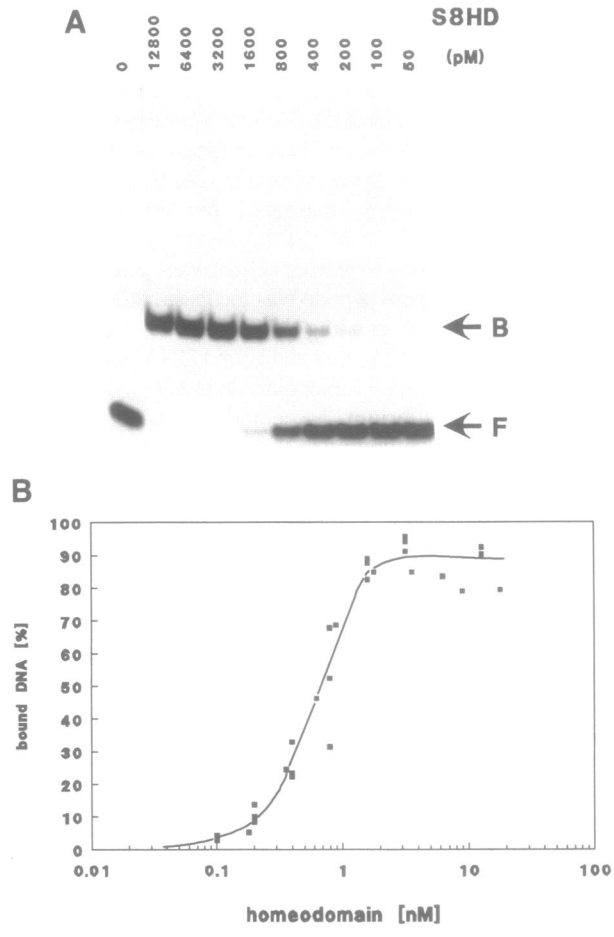


Figure 2. Determination of optimal S8 homeodomain recognition sequence. Outline of strategy employed to select oligonucleotides containing binding sites for S8.

factors GCN4 and Ubx (10, 31). Labeled random-sequence oligonucleotides were loaded on a S8 homeodomain-Sepharose column in a low salt buffer containing non-specific DNA competitor. The tightness of the interaction between bound oligonucleotides and S8 homeodomain was challenged by washing with increasing NaCl concentrations (0.1 M, 0.25 M, 0.4 M, respectively). Oligonucleotides retained on the column were then eluted with high salt buffer (1 M NaCl), and re-chromatographed on the S8 homeodomain affinity column in two additional rounds, the third selection round being preceded by PCR amplification. The enrichment of oligonucleotides containing S8 binding sites was estimated by EMSA. Selected oligonucleotides formed considerably more stable complexes with S8 homeodomain in the presence of the non-specific competitor poly(dI·dC)/(dC·dI) than unselected oligonucleotides, demonstrating an enrichment in sequences bound by S8 (data not shown). This enrichment also shows that the functional integrity of the S8 homeodomain was retained upon matrix-attachment. Since a significant fraction of the isolated oligonucleotides was not bound in EMSA, we decided to increase the yield of high-affinity binding sites by an additional selection step. We performed preparative EMSA using a S8 homeodomain concentration of 1.6×10^{-8} M, a binding condition that allowed no detectable DNA-protein complex formation with unselected oligonucleotides. Oligonucleotides specifically bound by S8 homeodomain were isolated from gel, cloned and sequenced. The results of the sequence analysis are shown in Table I.

Alignment of isolated oligonucleotides for definition of a consensus recognition sequence

Sequence analysis showed that all 59 oligonucleotides selected were different, indicating absence of artefactual distortion towards particular sequences caused by the PCR amplification. Strikingly,



C

oligo	sequence/position (1 2 3 4 5 6 7 8 9 10 11)	K_d (M)
22	cgcctcTATACCCAATTAcgacag	6.0×10^{-10}
7	cgcctcTAAACCTAATTAcgacag	6.6×10^{-10}
9	cgcctcTTGACCTAATTAcgacag	6.8×10^{-10}
21	ctgtcgTTTATATCATTAgaggcg	9.6×10^{-10}

Figure 3. Determination of the equilibrium binding constant. The affinity of four different selected oligos was tested in gel-shift assays (see Material and Methods). **(A)** Typical titration experiment with a fixed concentration (2×10^{-10} M) of oligo 22 (see Table I and Fig. 3C) and increasing concentrations of S8 homeodomain peptide as indicated. Bound (B) and free (F) DNAs were visualized by autoradiography and quantitated with a PhosphorImager. **(B)** Plotted data of four saturation experiments with oligo 22. The equilibrium binding constant (K_d) was calculated from the point of half-saturation of the DNA. **(C)** Table containing the equilibrium binding constants of different oligos tested. Values shown are the averages of at least three independent experiments. Sequence numbers and numbering of the nucleotides within the binding site are as in Table I.

all cloned sequences contained at least a single ATTA motif (see Table I), demonstrating the stringency of the selection. Furthermore, it confirms the importance of an ATTA motif for interaction of the S8 homeodomain with its binding site. Two further observations were made during a first alignment by visual

Table I. Determination of optimal S8 homeodomain recognition sequence

Partial sequence of randomized oligo R62: Orientation:

EcoRI *BamHI*

5′-GAATTCGCCTC N_{12} CGACAGGATCC-3′ +

3′-CTTAAGCGGAG GCTGTCCTAGG-5′ -

1	CAGTGCCAATTA	+	31	TTCTTAATTAGT	+
2	GGACCTAATTAA	+	32	TTGATCAATTAA	+
3	TTATCCAATTAA	+	33	ATGCGCAATTAC	-
4	AAGCCTTAATTA	+	34	CAATTATGTACT	+
5	CGAACAATTAG	+	35	ATCAATTATCCA	-
6	AATAGCCAATTA	+	36	GCACTCAATTAT	+
7	TAAACCTAATTA	+	37	CAATTAACTTTT	+
8	TAGTCATGATTA	-	38	AATTAGCCTGGA	-
9	TTGACCTAATTA	+	39	CCAGCCAATTAG	+
10	TAACTCAATTAG	+	40	CCCAATTAGGTT	+
11	TATGTTAATTAT	-	41	CGACCTAATTAC	+
12	AATTAATTAATC	-	42	GATAATTATCGC	-
13	ACTATTTCATTA	-	43	ATTAACGGATTG	+
14	AAGTTAATTAAT	+	44	ACCGCCAATTAA	+
15	AGTCTTCAATTA	-	45	AATTAGCGTATT	+
16	TATTGTTAATTA	+	46	TTAATTACCCA	-
17	GCTAAGTAATTA	-	47	TCGACTAATTAG	+
18	TAGTGATAATTA	-	48	CCAGTAATTAG	+
19	TGAACCAATTAA	+	49	TAATGCGATTAT	-
20	CATTATAATTAT	+	50	GGTACCAATTA	+
21	TTTATATCATTA	-	51	AATTAGCCGTAT	+
22	TATACCCAATTA	+	52	TGATTCAATTAT	+
23	AGAGCCAATTAA	-	53	CGGACTAATTAA	+
24	CCAATTTGATTA	-	54	AGCCCAATTAGC	+
25	TAATCAATTAGT	+	55	ATTAGTATCTAA	+
26	ATCAAATTAGGG	+	56	TAGGCCAATTAA	-
27	TTAAGCAATTAA	+	57	ATCTGTCAATTA	+
28	AATTAAACATTA	+	58	GTATTAATTAACA	+
29	AACCTTAATTA	-	59	GTAATCTAATTAA	-
30	ATTAATCAATTA	-			

	-3	-2	-1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
A	6	12	11	25	14	6	1	52	59	0	0	59	16	1	2	4	2	1	3	1
C	2	7	7	4	14	20	27	2	0	0	0	0	3	10	6	1	2	0	0	0
G	2	4	13	8	12	5	0	3	0	0	0	0	14	3	4	2	1	1	0	1
T	7	15	14	10	10	20	25	0	0	59	59	0	8	5	3	5	3	6	3	0
	17	38	45	47	50	51	53	57	59	59	59	59	41	19	15	12	8	8	6	2

CONSENSUS SEQUENCE:

		C	C						A		
A	N	–	–	A	A	T	T	A	–	C	
		T	T						G		
1	2	3	4	5	6	7	8	9	10	11	

Listing of the aligned sequences of oligos that were cloned and sequenced after the *in vitro* selection procedure. *On top*, a part of the oligo R62 sequence is shown, including the 12-bp random core and the flanking non-random sequences containing the *Bam*HI and *Eco*RI sites. (+) and (−) denote the orientation of the binding site sequence relative to the restriction sites. *Below*, the nucleotide occurrence at each position of the binding site is shown, positions being numbered at the top line where +1 is the first position of the consensus sequence.

inspection, using the ATTA motif as a landmark. First, the sequence PyAATTA is present in 48 clones. Second, frequently (38 times) more than one ATTA was found. It is unlikely that this reflects a bias of our selection procedure for simultaneous binding by two peptides, since in 21 of these cases the two ATTAs occurred in the small palindrome TAATTA. This led to ambiguity in the orientation of the sequences, which hampered alignment. We proceeded by deriving a preliminary consensus sequence by aligning the sequences not containing a double ATTA motif, using the hexamer sequence PyAATTA as a reference. The remaining sequences were then optimally ordered to define an optimal binding site (Table I), which only slightly differed from the preliminary consensus sequence. The aligned sequences were tabulated for base occurrence at each position (Table I). Statistical analysis by χ^2 test was performed to determine whether the observed patterns at each position were significantly skewed from random expectation. Only if a nucleotide at a given position showed a probability score of $P \leq 0.05$, it was included in the consensus sequence. This resulted in the 11-bp consensus sequence, A-N-C/T-C/T-A-A-T-T-A-G/A-C. Equal preference for C or T occurred at two positions 5′ to the ATTA core (Table I, positions 3 and 4). No significant selection was observed of a particular dinucleotide combination involving these two positions. Nucleotides at position 10 demonstrated equal preference for A and G.

Determination of the equilibrium binding constant

We examined a set of selected sequences in EMSA to determine quantitatively the affinity of the S8 homeodomain for its binding sites. Their equilibrium binding constants (K_ds) were measured by saturation experiments, in which increasing concentrations of S8 homeodomain peptide were incubated with a limited, fixed amount of DNA (Fig. 3). Three oligonucleotides (oligos 22, 7, 9) used in this assay contained sequences matching the consensus binding site except for the nucleotides at position 10 and 11, that correspond to the non-random sequence of oligo R62 (Fig. 3C). The S8 homeodomain showed the highest affinity for the binding site in oligo 22 with a calculated K_d of 6.0×10^{-10} M . The K_ds of oligos 7, 9 and 22 were similar, indicating no influence on binding of the (random-) sequences flanking the binding site (Fig. 3C, compare oligo 7 to 9) and, in addition, showing no significant preference for either a C or T base at position 4 (compare oligo 22 with 7 and 9). We tested a sequence which was more divergent from the consensus (oligo 21) in containing unfavourable nucleotides at positions 3, 5 and 11. The K_d value found for this sequence is higher compared to the other oligos tested. This is in accordance with the results of the *in vitro* binding site selection, that suggest that positions 3 and 5 are considerably constrained with respect to base preference.

Expression of S8 and the S8-related MHox gene

In a previous paper we have documented that *S8* acts as a mesodermal marker in cultured cells. *S8* transcripts, detected by northern blot analysis, and several immuno-reactive proteins detected by anti-S8 antibodies on western blots, were exclusively present in mesodermal cell lines (16). Expression of *MHox* transcripts and DNA-binding activity is also restricted to established cell lines of mesodermal origin (18). We now examined the S8 protein distribution in established cell lines by analyzing nuclear extracts in band shift assays for the presence of S8 DNA-binding activity. Using a double-stranded oligonucleotide containing an S8 binding site as defined above, four DNA-binding complexes (complex 1–4 in Fig. 4A and B) were observed in the mesodermal cell lines, NIH-3T3, C3H10T½, FIB-9 (a C1003 EC-clonal derivative) and MES-1 (a P19 EC-derived cell line), but not in other cell types including mouse neuroblastoma N1E-115, undifferentiated P19 EC and endoderm-like P19 clonal derivative END-2 (Figs. 4A and 5). The non-mesodermal cells either contain no retarded bands at all, or, in the case of N1E-115 a relatively slow migrating set of complexes (see below). The complexes in Mes-1 were

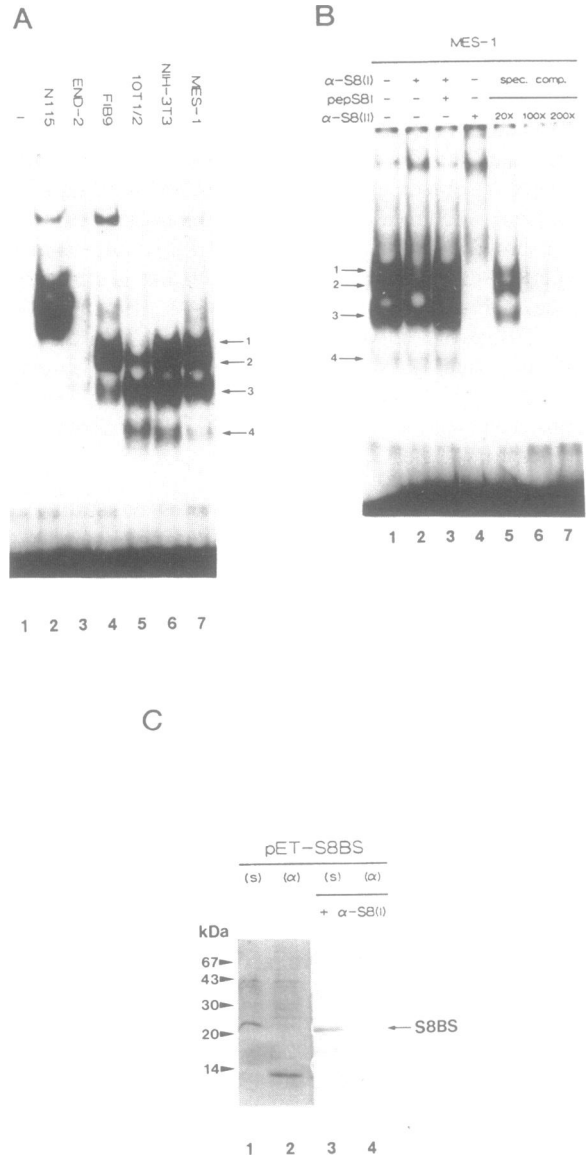


Figure 4. Determination of S8 DNA-binding activity in nuclear extracts of different cell lines in gel shift assay using S8-specific antibodies. **(A)** Binding reactions with oligo PC containing no extract (lane 1) or 5 μg nuclear extracts of N1E-115 (lane 2), END-2 (lane 3), fibroblast cell lines: FIB-9 (lane 4); C3H10T½ (lane 5); NIH-3T3 (lane 6); MES-1 (lanes 7). Arrows and numbers at the right denote the four mesoderm-specific DNA-complexes representing S8 (1 and 2) and Mhox (3 and 4). **(B)** Detection of S8 protein in the fibroblast-specific complexes. Binding reactions containing oligo PC and MES-1 nuclear extract were incubated with antibodies as indicated on top. Oligo S8I was added as competitor in the binding reaction shown in lane 3. Numbered arrows indicate complexes as in panel A. Specificity of the fibroblast-specific DNA-complexes was tested by competition with 20-fold, 100-fold and 200-fold molar excess of unlabelled oligo PC (lanes 5–7). **(C)** Immunoblot detection of S8 by anti-S8 anti-peptide antibodies. 10 μg of soluble bacterial lysate from *E.coli* carrying the sense (lanes 1 and 3, S) or anti-sense (lanes 2 and 4, α) expression constructs were transferred to nitrocellulose. The sense-expression construct produces the C-terminal part of the S8 protein including the homeodomain and exhibits an apparent molecular weight of 22 kD. The anti-sense construct gives rise to an unrelated peptide of 13 kD. Lanes 1 and 2, proteins of the *E.coli* lysate visualized by amido-black staining. Lanes 3 and 4, same samples incubated with α-S8(I).

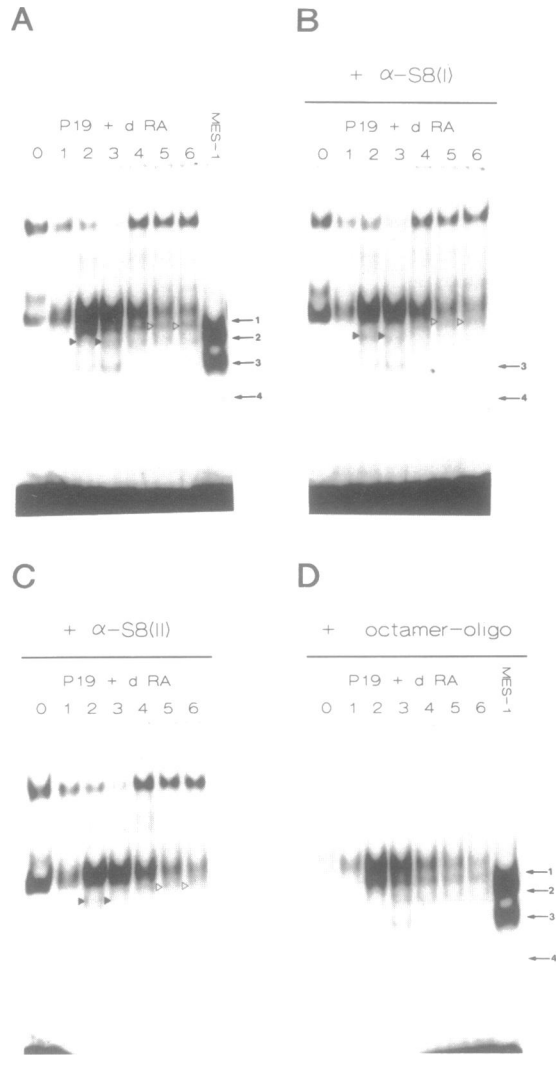


Figure 5. Induction of different S8 protein forms during RA-induced differentiation of P19 EC cells. **(A)** Mobility shift assay with oligo PC and nuclear extracts from P19 EC cells grown in presence of RA for the number of days indicated on top. Lane 8, MES-1 nuclear extract. Binding reactions similar as shown in (A) were incubated with anti-S8 anti-peptide antibodies (panel **B**, α-S8(I)), anti-S8/MHox antibodies (panel **C**, α-S8(II)) or a 50-fold molar excess of non-labeled oligonucleotides containing an octamer-site (D). Arrows in (A), (B) and (D) point to positions of S8-specific (1 and 2) and MHox-specific (3 and 4) DNA-complexes. Closed and open arrow heads in (A)–(C) indicate non-S8 DNA-complexes co-migrating with S8 complexes 1 and 2, respectively.

specifically prevented to form in the presence of competing non-labeled double-stranded oligonucleotide (Fig. 4B, lanes 5–7).

Previously, using affinity-purified polyclonal antibodies (α-S8(II)) generated against a bacterially produced C-terminal portion of the S8 protein, we detected on a western blot at least four distinct immuno-reactive proteins in extracts from fibroblast cell lines (16). Some of these bands probably represented MHox due to cross-reactivity of the antibodies. In EMSA experiments, the four mesoderm-specific DNA-complexes were abolished upon mixing the binding reactions with either α-S8(II) antibodies (Fig. 4B, lane 4) or a polyclonal anti-serum raised against the entire MHox protein (a gift of E. Olson, Houston; data not shown). These observations suggest that these complexes contain either MHox or S8 proteins. We can however not exclude the existence of one or more S8-related proteins that have remained unidentified so far and are also recognized by our antisera. Such a protein should have very similar expression patterns in the cell lines investigated. The formation of the bands in N1E-115 nuclear extracts was not affected by adding any of these antibodies (not

Table II. Comparison of optimal homeodomain binding sites

	position 1	2	3	4	5	6	7	8	9	10	11	Ref.
S8	A	N	C/T	C/T	A	A	T	T	A	A/G	C	
Ftz	A	A	G	C	A	A	T	T	A	A	G	11
Ubx		G	G	C	C	A	T	T	A	A		47
Dfd		G	T	T	C	A	T	T	A	A/G		47
Oct-1 POU_{HD}				T	N	A	T	T	A	C/T		48

shown). The presence of antibodies in the EMSA did not cause clear super-shifts, presumably due to their polyclonal nature leading to high molecular complexes unable to enter the gel. Two complexes formed by END-2 nuclear proteins co-migrated with complex 2 and 3 (Fig. 4A, lane 3), but were unaffected by α-S8(II) antibodies, indicating absence of S8 and MHox (data not shown).

To discriminate between S8 and MHox-containing DNA-complexes, a specific polyclonal anti-S8 antiserum (α-S8(I)) was raised against the oligopeptide S8I that encompasses a less conserved S8 amino acid sequence. Affinity-purified α-S8(I) antibodies detected S8 protein only in the two slower migrating complexes 1 and 2 but were less efficient than α-S8(II) in preventing the appearance of the gel shifts (Fig. 4B, lane 2). Oligopeptide S8I could block the binding of the antibodies (Fig. 4B, lane 3). In addition, α-S8(I) antibodies recognized in immunoblot analysis a truncated S8 protein containing the oligopeptide sequence (Fig. 4C). Both these observations confirmed the specificity of the α-S8(I) antibodies. This suggests that the two slower migrating complexes 1 and 2 contained S8 proteins and the two faster migrating complexes 3 and 4 are MHox protein specific complexes. Assuming that this assignment is correct, the gel-shift experiment shown in Fig. 4A would indicate that S8 protein is more abundant than Mhox in FIB-9 nuclear extract, whereas the opposite is true in C3H10T½ nuclear extract. In conclusion, while it is likely that other homeodomain proteins can recognize the S8 binding site due to the promiscuity of DNA-binding by homeodomain proteins *in vitro* (1), our results suggest that S8 and MHox proteins are abundant homeodomain proteins specifically present in mesodermal cell lines.

Differential expression of S8 protein forms in RA-treated P19 cells

The existence of multiple S8 and MHox proteins prompted us to investigate their expression during differentiation. EC cells resemble the cells in the inner cell mass of the murine blastocyst stage embryo in their capacity of differentiating into various cell types. P19 EC cells grown in monolayer, differentiate into endodermal and mesodermal cell types under influence of retinoic acid (RA) (34). Presence of S8 and MHox binding activity proteins were monitored in P19 EC cells grown in monolayer and treated with RA for several days (Fig.5). Both α-S8(I) and α-S8(II) antibodies were used to detect S8- and MHox-containing complexes (Fig.5B and C). Despite two co-migrating non-S8 DNA-complexes that obscured parts of the gelshift pattern, the following expression profiles could be discerned: S8 complex 1 was transiently induced reaching a peak level at day 2. remaining present at a lower level after day 2, while S8 complex 2 appeared later, around day 3, and remained constant afterwards. Expression of MHox proteins, presumably detected in complex 3 and 4, followed a similar pattern as complex 1, but reached their peaks at day 3. Complex 4 was only detectable after prolongued exposure of the autoradiogram. These observations agree with the expression of multiple proteins followed by immunoblot analysis of extracts of RA-treated P19 using α-S8(II) antibodies (16). Noteworthy is a partially transient induction of the slowest migrating immuno-reactive protein that presumably corresponds to complex 1. Competition with an oligo containing the canonical octamer-site abolished a slow migrating band (Fig.5D) but not the S8 or MHox bands. The slow band therefore may correspond to an Oct-protein complex. Both S8 protein forms are thus differentially expressed during RA-induced differentiation of P19, whereas no difference between the MHox protein forms was observed. This suggests separate functional roles of S8 proteins during formation of mesodermal cell types *in vitro*.

DISCUSSION

In this paper we describe the determination of the S8 homeodomain binding site preference from sequence analysis of oligonucleotides selected *in vitro* from a pool of random sequences. This is a useful strategy to determine the binding site preference of a transcription factor in absence of its known target sequences, which has been successfully employed for various homeodomain-containing and other sequence-specific DNA-binding proteins (10, 11, 29, 31, 35). The analysis of the recovered binding sites in our selection procedure revealed the presence of the ATTA sequence motif in all cloned oligos. Alignment of these sequences resulted in the definition of an 11-bp optimal binding site which is similar to those specifically recognized by *Antp*-class homeodomain proteins (1). The S8 optimal recognition site also resembles a sequence TTATAATTAAC, present in the muscle creatine kinase promoter, which was utilized by Cserjesi *et al.* (18) to isolate the S8-gene relative MHox from an expression library. This site matches 9 of the 11 bp of the S8 consensus sequence, the changes being at less constrained positions.

Purified S8 homeodomain peptide bound cloned oligonucleotides with high-affinity in EMSAs, confirming the successful selection of S8 binding sites. The lowest K_d measured, 6.0×10^{-10} M for binding to oligo 22, indicates a tight interaction between the S8 homeodomain and this site. Higher K_d-values were reported for binding of some homeodomain peptides to sites in putative target promoter sequences (9, 36, 37) and for certain optimal binding sites (38). When compared to the affinities of Ubx and Ftz for their respective optimal binding sites, however, the S8 binding site affinity is almost one order of magnitude lower (10, 11). Some of the variation in binding affinities reported is possibly due to differences in experimental conditions and methods to estimate protein concentrations.

We found in the gel shift assays no intermediate complex in binding reactions with a S8 protein containing the complete C-terminus (17) and the S8 homeodomain (data not shown), indicating that, like most homeodomain proteins, S8 binds as a monomer (9, 11, 12).

Many homeodomain binding sites identified up till now are characterized by the presence of an ATTA motif (reviewed in 7 and 39). The ATTA sequence has been shown to be essential for DNA-binding of ATTA-preferring homeodomains by methylation interference assays and base-substitutions experiments (10, 11, 36, 40). Base-specific contacts occurring in DNA-homeodomain interaction were determined by structural analysis for the Antp-, engrailed (*en*)- and yeast Matα2-homeodomains (12, 13, 41). The first two are ATTA-preferring homeodomains whose base-specific interactions have become the paradigm for other homeodomain-DNA interactions. Although the S8 homeodomain shares only approximately 33% of its amino acids with *Antp*-class homeodomains (17), it contains nearly identical amino acids at the key positions for homeodomain-DNA recognition, as will be explained below.

According to the reported *en*-DNA structure (12), the residues Arg-3 and Arg-5 in the N-terminal arm make minor-groove contacts with the third and fourth base of the ATTA motif corresponding to positions 8 and 9 of the S8 consensus site (Table I). The recognition helix docks with the major groove in such a way that residues Ile-47, Gln-50 and Asn-51 can make critical contacts. The conserved residue Ile-47 and the invariant residue Asn-51 interact with the first two base pairs (positions 6 and 7) of the ATTA (12). All amino acids described above are present in the S8 homeodomain, except for position 47 that is occupied by a Val. Val-47 is also present in several other ATTA-recognizing homeodomain types including prd and even-skipped (eve) and would be expected to contribute similarly to binding specificity by making a hydrophobic thymine contact (12). A recent study on the homeodomain protein HOX11 suggests that a Thr at position 47 makes it possible to bind a GTTA instead of an ATTA motif (42).

The eve homeodomain, which is 45% similar to S8 at amino acid level, can bind with similar affinity a second, totally unrelated sequence, TCAGCACCG (30); we found no evidence for such a second specificity.

The ninth residue in the recognition helix, equivalent to position 50, is the major determinant for DNA-binding specificity (14, 15). Residue 50 may specifically interact with one or both positions 5′ to the ATTA, as inferred from genetic and biochemical analysis (36, 40). S8 contains a Gln-50 that is present in Antp-class and numerous other homeodomains. Gln-50 in the engrailed homeodomain makes hydrophobic contacts with a thymine at position 4 (binding site position as in Table II), located at 5′ side of the ATTA core, but may also contact the base at position 5 (12) while Gln-50 in *Antp* interacts with a cytosine at position 5 (13). Position 54 may be involved in specific DNA interaction: Met-54 in *Antp* makes base-specific contact with the base pair at position 5 (13). *S8* contains an Ala-54, which, in the engrailed homeodomain (12), is unable to contribute to DNA binding due to its short side chain.

Bases at positions 1, 2, 3, 10 and 11 of the S8 consensus sequence are presumably not contacted by amino acids in the S8 homeodomain and their contribution to DNA binding affinity must be indirect. The Ftz homeodomain also recognizes an 11-bp optimal binding site (see Table II) and unfavourable substitutions at identical positions impair to some extent the binding affinity (11). The phosphate-backbone of the majority of these bases is specifically contacted by certain homeodomain residues (9, 12, 36) and could possibly give an indirect readout of the DNA sequence that establishes the conformational structure for optimal binding. The N-terminal arm of the homeodomain is close to the DNA (12) raising the possibility that the N-terminal extension of the S8 homeodomain participates in DNA-binding. One possibility we cannot entirely exclude, is that, since the core of the binding site found is an imperfect palindrome, some of the tabulated sequences are in reverse orientation. This would have placed residues 3′ from the centre instead 5′ resulting in an artifactual extension of the consensus site by probably two bases.

We show that the optimal binding site of S8 is similar to *Antp*-class binding sites, in keeping with the notion that position 50 dictates to a large extent the DNA-binding specificity (15, 23). The prd-like S8 homeodomain contains a Gln-50 and binds *Antp*-sites, whereas the prd homeodomain with a Ser-50 does not. The group of homeodomain proteins consisting of S8, Mhox (18), *X. laevis* Mix-1 (43), *D.melanogaster* aristaless (44), *S. mansonia* Smox-3 (45) and *C.elegans* CEH-10 (46) most likely form a distinct category of *prd*-related proteins that lack the DNA-binding paired domain (22, 23) and contain a Gln-50, although the absence of a paired box has not been proven for the latter two genes. This leaves the question of the biological significance of the sequence conservation between the homeodomains of the members of this sub-family and prd. An obvious possibility is the requirement of certain homeodomain amino acids for several types of protein-protein interactions with basal or specific transcription (co-)factors or other chromatin proteins. It should be noted that the presence of a *prd*-type homeodomain containing Ser-50 is restricted to homeoproteins containing in addition a paired domain.

Table II lists a number of optimal homeodomain binding sites that resulted from studies involving *in vitro* selection. The closely related *Antp*-class homeodomain proteins Ftz, Ubx and Dfd all share the key amino acids that make base-specific contacts. The Oct-1 POU homeodomain is part of the bi-partite DNA-binding POU domain and is very divergent from the classic homeodomain proteins. Nevertheless, the Oct-1 homeodomain when tested as a separate domain, needs an ATTA motif for optimal binding (48).

Examination of the optimal binding sites shows that several dinucleotide combinations occur in the contact region of residue 50, 5′ from the ATTA. Previous compilations of homeodomain binding sites showed that the amino acid at position 50, either a Gln or a Lys, can recognize a limited set of dinucleotide combinations, and therefore is flexible in making base-specific contacts (7, 40). Furthermore, nucleotides flanking the hexamer sequence NNATTA, to which base-specific contacts by the homeodomain are restricted, show significant differences. Therefore, DNA-binding specificity does not depend entirely on the residue at position 50. Dfd and Ubx homeodomains have distinct binding site preferences that were determined by testing their affinity for certain known homeodomain binding sites (49) and by defining the optimal recognition sequence (47). The sequence specificity of Dfd and Ubx for bases preceding and following the ATTA core recognition sequence resides in the C- and N-terminal regions of the homeodomain, respectively. In addition, the terminal regions of the TTF-1 homeodomain were also shown to contribute to the binding specificity of Antp-TTF-1 chimeric proteins (50).

Minor differences in target site preference are thought to have large impact on their regulatory specificities *in vivo*. Differential binding site preference of Ubx and Dfd was shown, in a yeast system, to be the basis for differential target gene expression driven from a promoter fragment containing multiple optimal binding sites (47). Furthermore, the DNA-binding preferences

of Dfd and Ubx as specified by the homeodomain terminal regions correlate well with the target specificities of (chimeric) Dfd and Ubx proteins in the *Drosophila* embryos as measured by their capacity to auto-regulate from the Dfd enhancer and to repress *Antp* transcription, respectively (51, 52). The S8 optimal binding site is dissimilar in the nucleotides flanking the ATTA motif when compared to the optimal recognition sequences of Ubx, Dfd and Ftz. It is tempting to speculate that S8, too, possesses a distinct target specificity *in vivo* that is reflected *in vitro* by small differences in binding site preference.

We described earlier the mesoderm-specific expression of several proteins, that immuno-reacted with S8 antibodies in western blot analysis (16). In the present report we extend our analysis of protein expression by investigating the presence of S8 DNA-binding activity in mesodermal cell lines. Our data indicate that, as expected, S8- and MHox-specific DNA-complexes are abundant in fibroblast cell lines. Of both proteins two specific complexes were observed, but their exact nature remains to be explained. Many transcription factors are post-translationally modified either by phosphorylation or glycosylation (53) that can cause changes in migration behaviour in non-denaturing or SDS-PAGE (8, 53). Alkaline-phosphatase treatment of proteins in an NIH-3T3 nuclear extract did not result in the disappearance of S8 or MHox bands as detected by immunoblot analysis, suggesting that the slower migrating protein forms are not the product of phosphorylation (data not shown). Other possible explanations include alternative splicing and alternative usage of translation start codons. Alternative splicing of *MHox/K-2* results in two transcript forms, *K-2a* and *K-2b*, that would encode polypeptides of 245 and 217 amino acids, respectively (19). The C-terminus of the K-2a protein is much more similar to S8 than K-2b. On the other hand, *MHox* cDNA is identical to the *K-2b* transcript (18). MHox *in vitro* translation products corresponding to the full-length protein and a shorter polypeptide starting at the second in-frame methionine, have been compared to DNA-complexes of C3H10T½ nuclear proteins and were shown to co-migrate with two unique mesoderm-specific DNA-complexes (18). However, using nuclear extract of C3H10T½, we present evidence for the presence of two additional complexes containing S8 protein, though present in lesser amounts. Our EMSA experiments did not provide confirmation for the existence of a *K-2a*-encoded protein and it remains to be solved if this larger protein is produced in the examined cell lines.

Our data (16, 17, this paper) and those of Cserjesi *et al.* (18) indicate that S8 and MHox are abundant homeodomain proteins present in established mesodermal cell lines. Studying their interactions with promoter sequences of target genes is the next step in unraveling their biochemical roles in specifying the identity of mesodermal cells. General insight in their function *in vivo* awaits the analysis of morphogenesis in mice in which these genes have been either inactivated or are mis-expressed.

ACKNOWLEDGEMENTS

We are grateful to E.Olson for MHox antibodies. We wish to thank J.Vervaart for DNA-sequencing, G.Folkers for oligonucleotide synthesis, E.Freund for oligopeptide synthesis, and H.-J.van Mazijk for technical assistance. We also thank S.Ekker and P.Verrijzer for helpful suggestions and J.Deschamps and S.de Laat for critical reading of the manuscript.

REFERENCES

1. Hayashi, S. and Scott, M.P. (1990) Cell **63**, 883–894.
2. McGinnis, W., Levine, M.S., Hafen, E., Kuroiwa, A. and Gehring, W.J. (1984) Nature **308**, 428–433.
3. Scott, M.P. and Weiner, A.J. (1984) Proc. Natl. Acad. Sci. USA **81**, 4115–4118.
4. Kessel, M. and Gruss, P. (1990) Science **249**, 374–379.
5. Laughon, A. and Scott, M.P. (1984) Nature **310**, 25–31.
6. Qian, Y.Q., Billeter, M., Otting, G., Müller, M., Gehring, W.J. and Wüthrich, K. (1989) Cell **59**, 573–580.
7. Laughon, A. (1991) Biochemistry **30**, 11357–11367.
8. Odenwald, W.F., Garbern, J., Arnheiter, H., Tournier-Lasserve, E. and Lazzarini, R.A. (1989) Genes & Dev. **3**, 158–172.
9. Affolter, M., Percival-Smith, A., Müller, M., Leupin, W. and Gehring, W.J. (1990) Proc. Natl. Acad. Sci. USA **87**, 4093–4097.
10. Ekker, S.C., Young, K.E., Von Kessler, D.P. and Beachy, P.A. (1991) EMBO J. **10**, 1179–1186.
11. Florence, B., Handrow, R. and Laughon, A. (1991) Mol. Cell. Biol. **11**, 3613–3623.
12. Kissinger, C.R., Lui, B., Martin-Blanco, E., Kornberg, T.B. and Pabo, C.O. (1990) Cell **63**, 579–590.
13. Otting, G., Qian, Y.Q., Billeter, M., Müller, M., Affolter, M., Gehring, W.J. and Wüthrich, K. (1990) EMBO J. **9**, 3085–3092.
14. Hanes, S.D. and Brent, R. (1989) Cell **57**, 1275–1283.
15. Treisman, J., Gonczy, P., Vashishtha, M., Harris, E. and Desplan, C. (1989) Cell **59**, 553–562.
16. De Jong, R. and Meijlink, F. (1993) Dev. Biol. **157**, 133–146.
17. Opstelten, D.E., Vogels, R., Robert, B., Kalkhoven, E., Zwartkruis, F., de Laaf, L., Destrée, O.H., Deschamps, J., Lawson, K.A. and Meijlink, F. (1991) Mech. Dev. **34**, 29–42.
18. Cserjesi, P., Lilly, B., Bryson, L., Wang, Y., Sassoon, D.A. and Olson, E.N. (1992) Development **115**, 1087–1101.
19. Kern, M.J., Witte, D.P., Valerius, M.T., Aronow, B.J. and Potter, S.S. (1992) Nucl. Acids Res. **20**, 5189–5195.
20. Grueneberg, D.A., Natesan, S., Alexandre, C. and Gilman, M.Z. (1992) Science **257**, 1089–1095.
21. Kongsuwan, K., Webb, E., Housiaux, P. and Adams, J.M. (1988) EMBO J. **7**, 2131–2138.
22. Chalepakis, G., Fritsch, R., Fickenscher, H., Deutsch, U., Goulding, M. and Gruss, P. (1991) Cell **66**, 873–884.
23. Treisman, J., Harris, E. and Desplan, C. (1991) Genes & Dev. **5**, 594–604.
24. Bopp, D., Burri, M., Baumgartner, S., Frigerio, G. and Noll, M. (1986) Cell **47**, 1033–1040.
25. Baumgartner, S., Bopp, D., Burri, M. and Noll, M. (1987) Genes & Dev. **1**, 1247–1267.
26. Gruss, P. and Walther, C. (1992) Cell **69**, 719–722.
27. Studier, F.W., Rosenberg, A.H., Dunn, J.J. and Dubendorff, J.W. (1990) Meth. of Enzymology **185**, 60–89.
28. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular cloning. A laboratory manual, 2nd edn., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
29. Pollock, R. and Treisman, R. (1990) Nucl. Acids Res. **18**, 6197–6204.
30. Hoey, T., Warrior, R., Manak, J. and Levine, M. (1988) Mol. Cell. Biol. **8**, 4598–4607.
31. Oliphant, A.R., Brandl, C.J. and Struhl, K. (1989) Mol. Cell. Biol. **9**, 2944–2949.
32. Harlow, E. and Lane, D. (1988) Antibodies. A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
33. Lee, K.A.W., Bindereif, A. and Green, M.R. (1988) Gene Anal. Tech. **5**, 22–31.
34. Jones-Villeneuve, E.M.V., McBurney, M.W., Rogers, K.A. and Kalnins, V.I. (1982) The journal of cell biology **94**, 253–262.
35. Blackwell, T.K. and Weintraub, H. (1990) Science **250**, 1104–1110.
36. Percival-Smith, A., Müller, M., Affolter, M. and Gehring, W.J. (1990) EMBO J. **9**, 3967–3974.
37. Corsetti, M.T., Briata, P., Sanseverino, L., Daga, A., Airoldi, I., Simeone, A., Palmisano, G., Angelini, C., Boncinelli, E. and Corte, G. (1992) Nucl. Acids Res. **20**, 4465–4472.
38. Catron, K.M., Iler, N. and Abate, C. (1993) Mol. Cell. Biol. **13**, 2354–2365.
39. Treisman, J., Harris, E., Wilson, D. and Desplan, C. (1992) Bioessays **14**, 145–150.
40. Hanes, S.D. and Brent, R. (1991) Science **251**, 426–430.
41. Wolberger, C., Vershon, A.K., Lui, B., Johnson, A.D. and Pabo, C.O. (1991) Cell **67**, 517–528.

42. Dear, T.N., Sanchez-Garcia, I. and Rabbits, T.H. (1993) Proc. Natl. Acad. Sci. USA **90**, 4431–4435.
43. Rosa, F.M. (1989) Cell **57**, 965–974.
44. Schneitz, K., Spielmann, P. and Noll, M. (1993) Genes & Dev. **7**, 114–129.
45. Webster, P.J. and Mansour, T.E. (1992) Mech. Dev. **38**, 25–32.
46. Hawkins, N.C. and McGhee, J.D. (1990) Nucl. Acids Res. **18**, 6101–6106.
47. Ekker, S., Von Kessler, D.P. and Beachy, P.A. (1992) EMBO J. **11**, 4059–4072.
48. Verrijzer, C.P., Alkema, M.J., Van Weperen, W.W., Van Leeuwen, H.C., Strating, M.J.J. and Van der Vliet, P.C. (1992) EMBO J. **11**, 4993–5003.
49. Dessain, S., Gross, C.T., Kuziora, M.A. and McGinnis, W. (1992) EMBO J. **11**, 991–1002.
50. Damante, G. and Di Lauro, R. (1991) Proc. Natl. Acad. Sci. USA **88**, 5388–5392.
51. Kuziora, M.A. and McGinnis, W. (1989) Cell **59**, 563–571.
52. Lin, L. and McGinnis, W. (1992) Genes & Dev. **6**, 1071–1081.
53. Berk, A.J. (1989) Biochim. Biophys. Acta. **1009**, 103–109.