1928–1935 Nucleic Acids Research, 1995, Vol. 23, No. 11

The optimal binding sequence of the Hox11 protein contains a predicted recognition core motif

Shaojun Tang[1,2,*] and Martin L. Breitman[1,2,†]

[1]Division of Molecular and Developmental Biology, Samuel Lunenfeld Research Institute, Mount Sinai Hospital, 600 University Avenue, Toronto, Ontario M5G 1X5, Canada and [2]Department of Molecular and Medical Genetics, University of Toronto, Toronto, Ontario, Canada

Received February 21, 1995; Revised and Accepted April 24, 1995

ABSTRACT

***HOX11* is a homeobox-containing oncogene of specific T-cell leukemias. We determined the DNA binding specificity of the Hox11 protein by using a novel technique of random oligonucleotide selection developed in this study. The optimal Hox11 binding sequence, GGCGGTAAGTGG, contained a core TAAGTG motif that is consistent with a prediction based on the residues at specific positions that potentially make DNA base contacts and models of homeodomain–DNA interaction proposed from studies with other homeodomains. The specific interaction between Hox11 and the selected optimal binding sequence was further confirmed by band-shift and DNA competition assays. Given that the Hox11 homeodomain shares low homology with other well studied homeodomains, the presence of a predictable recognition core motif in its optimal binding sequence supports the notion that different homeodomains interact with DNA in a similar manner, through highly conserved residues at specific positions that allow contact with DNA.**

INTRODUCTION

Homeobox-containing genes comprise a large evolutionally conserved gene family (1,2). Various lines of evidence have suggested that this class of genes function as master control regulators in various developmental processes, including embryonic patterning and cell differentiation (3). Dysfunction or mutations of some of these genes are associated with specific malignancies (e.g. 4–6) and certain congenital human defects (e.g. 7,8). It is generally assumed that homeodomain proteins affect both normal and abnormal developmental processes by regulating the expression of other genes in a sequence-specific DNA-binding manner. The 60 amino acid homeodomain encoded by homeobox is a specific DNA binding domain (9,10) required for mediating the *trans*-regulatory activity of homeodomain proteins to specific targets (11,12). The *in vivo* functional significance of the DNA binding specificity was demonstrated by the homeodomain swap experiments (13,14).

The DNA binding sites of several homeodomain proteins have been defined (15,16). They are usually composed of 6 bp with a TAAT core (TAATNN). Structural analyses of the *Antp*, *En* and *MATα2* homeodomain–DNA complexes by NMR and X-ray crystallography have detailed the specific interactions in these DNA–protein complexes (17–19). The X-ray crystallographic model of the *En* homeodomain–DNA complex has provided a framework to understand the general characteristics of homeodomain–DNA interactions (18). According to this model, DNA binding of the *En* homeodomain is mediated by the recognition helix and the N-terminal arm. Several residues in these regions, which are most conserved in different homeodomains, play a major role in DNA recognition. As summarized in Figure 1b, the major DNA base contacts are as follows: Arg5 hydrogen bonds to the thymine at the first position; Arg3 appears to hydrogen bond to the complementary thymine of the second base pair; Asn51 interacts with the adenyl at the third position; Ile47 makes hydrophobic contacts with the thymine at the fourth position; and Gln50 can contribute to DNA binding by contacting the two bases following the TAAT core. Genetic as well as biochemical studies have shown that residue 50 determines the 2 nt following the TAAT core; Gln50 preferentially recognizes a TG dinucleotide following the TAAT core, while Lys50 contacts a CC dinucleotide (20–23). Comparable homeodomain peptide folding and DNA base contacting was also observed in the X-ray crystallographic model of α2 homeodomain–DNA complex (19). These two homeodomains have extensive differences at the amino acid level, suggesting that most homeodomains bind to DNA in a similar manner (19).

In this study, we have determined the DNA binding specificity of the Hox11 protein and further evaluated the universality of structural and genetic models of DNA–homeodomain interaction described above for divergent homeodomains. The *HOX11* gene was cloned as an oncogene involved in T-cell leukemias (24–27). During mouse embryogenesis, the murine *HOX11* homolog, *Tlx-1*, is expressed in various embryonic tissues, including branchial arches, some components of the nervous system and spleen (28–30). A mutation created by gene targeting revealed a

*To whom correspondence should be addressed at present address: Department of Genetics, Research Institute, Room 11-125, The Hospital for Sick Children, 555 University Avenue, Toronto, Ontario M5G 1X8, Canada

†Deceased

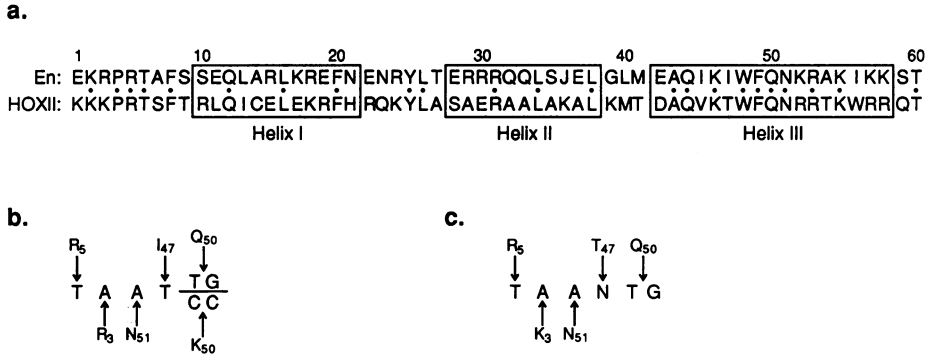

Figure 1. A prediction of the recognition motif of the Hox11 homeodomain based on the models of DNA–homeodomain interactions in other homeodomains. (**a**) A comparison of the primary structure of Hox11 homeodomain with that of *En*. Helices I, II and III are highlighted by boxes. The identical residues between the homeodomains of Hox11 and *En* are indicated by dots. The amino acid identity between them is 38%. (**b**) DNA base contacts proposed in the X-ray crystal model of *En* homeodomain–DNA complex (18) and the genetic model of homeodomain–DNA interactions (20–23). See text for details. (**c**) A prediction of the DNA binding motif of Hox11 homeodomain by analogy to the DNA base contacts by specific homeodomain residues summarized in (b). See text for details.

requirement of *Tlx-1* in spleen organogenesis (30). The *En* and Hox11 homeodomains are distantly related, with only 38% identity between them (Fig. 1a). However, three residues (Arg5, Gln50 and Asn51) at the DNA contacting positions in the N-terminal arms and recognition helices of both proteins are identical. Two residues at other potential DNA contacting positions are changed in the Hox11 homeodomain (Lys3 and Thr47; Fig. 1a). These changes could affect the DNA binding specificity of the Hox11 homeodomain. As both Arg3 and Lys3 have side chains with an amine group, their DNA contact behaviour is probably similar, as proposed for Arg3 in the *En* homeodomain (18). The Thr47 in Hox11, however, has both a hydroxyl and a methyl group in its side chain. Such a side chain should be able to interact with the DNA base at the fourth position through the hydrophobic methyl and the hydrophillic hydroxyl, compared with the side chain of the Ile47 in the *En* homeodomain, which only contains the hydrophobic methyl. Based on these considerations and by analogy to the DNA base recognition behaviour of the *En* homeodomain, a potential recognition sequence for the Hox11 homeodomain is predicted to be TAANTG (Fig. 1c).

This prediction was tested by defining the optimal binding sequence of the Hox11 protein, by using a random oligonucleotide selection technique. Consistently, the optimal binding sequence contained a TAAGTG motif, which is compatible with the predicted sequence, TAANTG. This confirmation suggests an universality of the structural model for *En* homeodomain–DNA complex and the recognition specificity of residue 50, and thus supports the notion that different homeodomains bind to DNA in a similar manner. This finding was discussed in the context of the result of Dear *et al.* (31), which defined a different consensus binding sequence for the Hox11 homeodomain peptide. The potential factor leading to such differences was discussed in the view of the contribution of the sequence outside the homeodomain to the DNA binding specificity.

MATERIALS AND METHODS

Construction of plasmid pGEX-2T-HOX11 cDNA

The *HOX11* cDNA fragment (24) including the full length coding sequence was cloned into the *Sma*I site of pGEX-2T plasmid (Sigma) and fused to the GST portion in the correct reading frame (Fig. 2a).

Expression and GST bead purification of GST-Hox11 fusion protein

One colony of a pGEX-2T-HOX11 cDNA recombinant was inoculated into 100 ml LB medium containing ampicillin (50 μg/ml) and was grown overnight in a 37°C shaking incubator. The overnight culture was diluted into 1 l of fresh LB/ampicillin medium and grown another 2 h at 37°C to late log phase stage. One ml of 100 mM IPTG was then added to a final concentration of 0.1 mM and incubation was continued for 0.5 h to induce the expression of the fusion protein. The GST-Hox11 fusion protein was purified by using 1 ml of the 50% slurry of glutathione–Sepharose beads according to manufacturer's instructions (Pharmacia) and subject to thrombin digestion to release the Hox11 protein from GST moiety (see below).

Release of Hox11 protein from GST moiety

The Hox11-GST beads were equilibrated by washing once with 20 ml of thrombin cleavage buffer (2.5 mM $CaCl_2$, 50 mM Tris, pH 7.5, 150 mM NaCl) and subsequently resuspended in 1 ml of thrombin cleavage buffer. Thrombin digestion was carried out as follows. Ten μl of thrombin solution (2 μg/μl in cleavage buffer) was added to the above bead slurry and incubated for 2 h in a 25°C water bath. The released Hox11 protein was recovered by washing the beads with 1 ml of wash buffer and the wash was repeated five times. The eluted fractions were analyzed on SDS–PAGE gel to verify the presence of released Hox11 protein (32). The fractions containing Hox11 protein were pooled and dialysed against 200 ml of TED solution for 2 h, with four changes of the dialysis buffer.

DNA cellulose chromatography purification of the Hox11 polypeptide

One gram of DNA cellulose (Sigma) was suspended in 10 ml of TED buffer (25 mM Tris, pH 7.5, 10 mM EDTA, 1 mM DTT, 10% glycerol, 0.1 mM PMSF) with 25 mM NaCl and used for preparation of a 5 × 0.5 mm column. The dialysed protein sample was loaded on to the column and the flow rate of the sample was

adjusted to ~0.1 ml/min. The column was then washed with 50 ml of dialysis buffer. The protein was eluted by a gradient, ranging from TED + 25 mM NaCl to TED + 1000 mM NaCl. The total gradient volume was 40 ml. The eluted protein was collected in 1.4 ml fractions and checked on a 10% SDS–PAGE gel by Coomassie staining (32). The fractions with clearest bands of full length Hox11 protein (36 kDa) were pooled and dialysed against TED with 25 mM NaCl. Protein concentration was determined by BioRad Assay and adjusted to 0.5 μg/μl. The protein was aliquoted and stored at –70°C.

Oligonucleotide synthesis and labelling

Sequences of synthesized oligonucleotides used in this study are as follows. (i) oligonucleotide with random sequence core: 5′-TACAAGATCCGGAATTCCTAC(N)$_{12}$-AGACGGATCCG-GCGATAAGACA-3′. The sequences underlined represent the *Eco*RI (left) and *Bam*HI (right) sites. These oligonucleotides were double-stranded and ^{32}P-labelled by Klenow reaction (32) and P2 primer (Fig. 3b) for the oligonucleotide selection experiments. (ii) P1: 5′-TACAAGATCCGGAATTCC-3′, identical to the 5′ region of random oligonucleotides. (iii) P2:5′-TGTCTTATCGCCGGATCC-3′, complementary to the 3′ region of random oligonucleotides. (iv) rF: 5′-TGGGAGCAGGTC-3′. (v) NP: 5′-TCAATTAAATGAT-3′. (vi) The optimal binding sequence of the Hox11 protein: 5′-GGCGGTAAGTGG-3′. Double-stranded oligonucleotides were ^{32}P-labelled for band-shift and DNA competition assays.

Preparation of Hox11 protein blotted nitrocellulose filters

Nitrocellulose membrane (Du Pont) was cut into 1 cm^2 squares and pretreated with hydroxyl potassium solution as follows. The membrane was soaked in 0.4 M KOH for 40 min and washed thoroughly with distilled water. Pretreated filters were then soaked in 1 × protein–DNA binding buffer (20 mM Tris, pH 7.5, 10% glycerol, 50 mM KCl, 0.1 mM DTT, 0.1 mM PMSF) briefly and placed on 3 MM Whatman paper. When the filter was briefly dried, 1 μg (2 μl) of protein was applied.

Oligonucleotide selection experiment

Five hundred ng (~10^7 c.p.m.) of the labelled double-stranded random oligonucleotide and the protein blotted filter were mixed in 100 μl of 1 × protein–DNA binding buffer in a 15 ml Falcon tube and incubated at room temperature for 20 min. A piece of BSA protein blotted filter was employed as a negative control. The filters were then washed three times with 1 ml of 1 × binding buffer, 5 min for each time. The bound oligonucleotides were then eluted from the filter by incubating the filter in 200 μl of 1 × binding buffer with 0.5 M NaCl at 55°C for 30 min and concentrated by ethanol precipitation. The collected oligonucleotides were amplified and labelled by polymerase chain reaction (PCR) and used for the next round of selection. The selections were performed for four cycles before the selected oligonucleotides were cloned into plasmid pGEM7 and sequenced.

PCR amplification of oligonucleotides

The PCR reaction conditions were according to the instruction of the kit from Cetus. The oligonucleotide amplification was performed by using the temprature cycles as follows: 30 cycles of 94°C, 30 s; 54°C, 45 s; 72°C, 10 s; followed by one cycle of: 94°C, 30 s; 54°C, 5 min; 72°C, 5 min. Five μCi of [α-^{32}P]dCTP (Du Pont) was included in the 100 μl reaction solution to label the amplified oligonucleotides.

Band shift assay and DNA competition assay

Both assays were carried out in the following buffer: 20 mM Tris, pH 7.9, 10% glycerol, 50 mM KCl, 1 mM DTT. The experimental procedure was performed according to standard methods (33). Briefly, the double-stranded oligonucleotides 5′-GGCGGTAAG-TGG-3′ were ^{32}P-labelled, and incubated with Hox11 protein in the above buffer at 37°C for 15 min. The mixture was then loaded on a 6% polyacrylamide gel and electrophoresis was carried out in 1 × TBE buffer. The shifted bands were visualized on the X-ray film by overnight autoradiography. For the DNA competition assay, various unlabelled DNA competitors were employed in the binding reaction mixture as indicated in text.

RESULTS

In order to determine the optimal DNA binding sequence of the Hox11 protein, we defined a PCR combined random oligonucleotide selection strategy to select the specific oligonucleotides that were bound by bacterially expressed Hox11. Specific interaction between the selected optimal oligonucleotide and Hox11 was confirmed by gel mobility shift and DNA competition assays.

Purification of Hox11 protein expressed in *E.coli*

The plasmid pGEX-2T (Pharmacia) was used to express Hox11 protein as a GST fusion protein in bacteria (Fig. 2a). After purification of the fusion protein by GST bead affinity chromatography, the Hox11 protein was released from the GST portion by thrombin digestion and further purified by DNA cellulose chromatography (Materials and Methods).

The full length *HOX11* coding region was used to make intact Hox11 protein. Although many experiments have shown that a homeodomain peptide is sufficient to confer the ability to bind to DNA (e.g. 34), there is also evidence suggesting that sequences outside the homeodomain can also contribute to the specificity and affinity of DNA binding. For example, Hoey *et al.* (35) have shown that the intact Even-skipped protein binds to both TCAATTAAAT and TCAGCACCG with virtually equal preference, while a C-terminal truncated protein predominantly binds to TCAATTAAAT.

The GST-Hox11 fusion protein appeared to be extremely unstable in bacterial cells and was degraded after IPTG induction (Fig. 2b). In order to minimize degradation, we varied a number of expression conditions, including bacterial strains, IPTG concentrations and induction time. However, a large proportion of the fusion protein was degraded under all conditions. Because of this proteolysis, the abundance of fusion protein in bacterial cells was low and hardly detectable in the total cell lysate. However, it was detected after GST bead purification. As shown in Figure 2b, the majority of fusion protein was degraded, resulting in a few major bands of low molecular weight. The molecular weight of the intact fusion protein was 63 kDa, as expected. In order to obtain an optimal amount of undegraded fusion protein, IPTG induction was performed at late log phase stage for only 30 min.

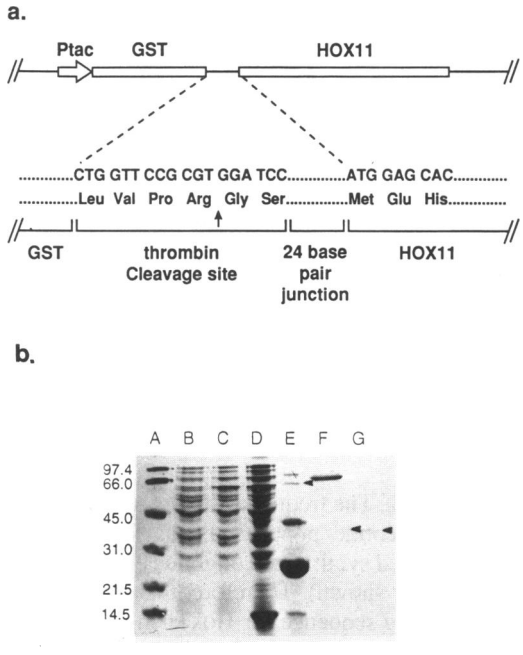

Figure 2. Purification of the Hox11 protein. (**a**) A schematic diagram of the prokaryotic expression vector of the GST-Hox11 fusion protein. The HOX11 cDNA was fused to GST in pGEX2T (Pharmacia) in correct reading frame to produce a fusion protein. The expression of this fusion protein was controlled by the IPTG inducible promoter Ptac. The junction between the GST and Hox11 is illustrated in detail in the lower nucleotide sequence and amino acid sequence. It contains a thrombin cleavage site and an additional 24 bp from the polylinker and *HOX11* upstream sequence. (**b**) The Hox11 protein was expressed in and purified from *E.coli* (for details of the purification procedure, see Materials and Methods). The protein samples were run on a 7% SDS–polyacrylamide gel and visualized by Coomassie staining. A: protein molecular weight marker. The size of each marker band is indicated on the left (Kd). B: total protein extracts from the control cells without Hox11 expression vector. C: total protein extracts from the cells with the Hox11 expression vector but before IPTG induction. D: total protein extracts from Hox11 expressing cells after IPTG induction for 30 min. The strong band of ~14 kDa is a degradation product. Compared with lane c, no prominent band of ~63 kDa, which corresponds to the expected GST-Hox11 fusion protein, can be identified. This is probably due to protein degradation. E: protein eluted from GST beads before thrombin digestion. The arrow points to the band corresponding to the undegraded fusion protein. The size of the band with the highest intensity (~27 kDa) matches that of GST. This clearly shows that the majority of the fusion protein has been degraded. F: protein eluted from GST beads after thrombin digestion. The arrow points to the band of 36 Kd Hox11 protein. The top band comes with thrombin protein. G: purified Hox11 protein after DNA cellulose chromatography.

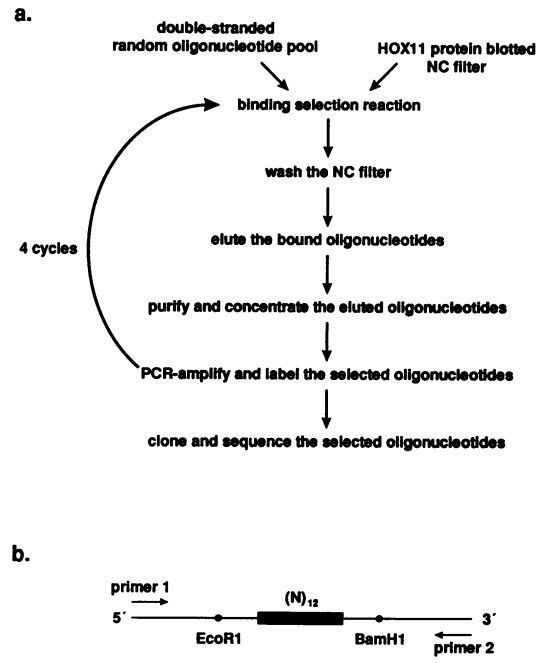

Figure 3. Nitrocellulose filter mediated random oligonucleotide selection strategy. (**a**) The experimental process of oligonucleotide selection, starting with the synthesized random oligonucleotide pool and the purified Hox11 protein expressed in *E.coli*. The selection was combined with PCR amplification and repeated for four rounds (see Materials and Methods for details). (**b**) A schematic diagram of the designed oligonucleotides. They contain a 12 bp random core sequence, which is flanked by *Eco*RI and *Bam*HI restriction sites for cloning. Primers 1 and 2 are designed for PCR amplification.

Because protein from the GST bead purification was a mixture of the intact and different truncated GST-Hox11 fusion proteins, further purification was subsequently performed to isolate the undegraded Hox11 protein from the mixture and to release it from the GST moiety. This was achieved by thrombin digestion and DNA cellulose chromatography. The fusion protein adsorbed on GST beads (Materials and Methods) was first digested by thrombin at 25 °C for 2 h. The released Hox11 peptides was eluted from the bead matrix by several washes. After elution, the Hox11 protein was separated from GST which was still on the GST affinity beads. The eluted Hox11 protein was heterogeneous and included the full length and various truncated Hox11 degradation protein. To obtain more homogenous Hox11 full length protein, DNA cellulose chromatography was performed and the undegraded Hox11 protein was eluted as a separated fraction by NaCl gradient. As shown in Figure 2b, the final purified Hox11 protein was detected as a homogenous protein on an SDS–PAGE gel. Its estimated molecular weight was ~36 kDa, consistent with the expected molecular weight deduced from the Hox11 conceptual amino acid sequence, suggesting that the purified protein was indeed encoded by the cloned *HOX11* cDNA in the correct reading frame. However, the yield was low and only 10–20 μg of purified Hox11 protein was obtained from 1 l of cell culture.

Nitrocellulose filter mediated random oligonucleotide selection strategy

A novel technique of random oligonucleotide selection was developed in this study. The selection process was much simplified in this technique, compared with other similar ones (for example 34,36,37). In this technique (Fig. 3a), the purified Hox11 protein was immobilized on a piece of nitrocellulose membrane, which was then used as an affinity matrix to select the DNA binding sequences from a population of 56 bp oligonucleotides containing 12 bp random core sequence. As illustrated in Figure 3b, the random core sequence within the 56mer oligonucleotide was flanked by *Bam*HI and *Eco*RI restriction sites for cloning and the complementary sequences for two 18 base primers used for the PCR amplification. The general scheme (Fig. 3a) was designed to enrich for the specific DNA sequences bound by Hox11 protein. The binding selection was carried out in a protein–DNA binding buffer at room temperature; the loosely bound and non-specifically retained oligonucleotides on the

Table 1. Oligonucleotide selection efficiency of HOX11 protein on nitrocellulose membrane

selection cycles	radioactivity on filters (CPM)			radioactivity in eluted solution (CPM)		
	BSA	HOX11	HOX11/BSA	BSA	HOX11	HOX11/BSA
first	656.6	6682.5	10.2	20.3	6004.5	295.5
second	350.4	27757.4	79.2	ND	ND	ND
third	336.0	25085.1	74.6	ND	ND	ND
fourth	236.1	43952.9	186.2	ND	ND	ND

BSA blotted membranes were employed as the negative controls. The oligonucleotides used for selection were labelled and thus the radioactivity on the membranes and eluted solution reflected the amount of oligonucleotides. As shown in the first round of selection, the majority (6004.5/6682.5 = 90%) of the bound oligonucleotides on the Hox11 membrane could be eluted while only a small fraction (20.3/656.6 = 3%) of those could be eluted from BSA control membrane, suggesting a major role of Hox11 protein in selection. ND, not determined.

membrane were washed away; tightly bound DNA was then eluted with high salt concentration and high temperature (55°C). The specifically selected DNA was subjected to PCR amplification, thus facilitating either the next round of selection or cloning (Materials and Methods). Four cycles of selection were performed and the selected oligonucleotides from the fourth round of selection were then amplified, cloned and sequenced.

Another feature of this procedure is that labelled DNA was used throughout the selection procedure and therefore the DNA bound to the filter was easily monitored. As shown in Table 1, Hox11 bound to DNA in each cycle of selection, as compared with the background in column A. This result indicated that Hox11 selected its binding sequence during each cycle. However, the precise degree of enrichment for binding sequences in each round of selection could not be determined because of differences in quantity of DNA and the specific radioactivity used in each round. Enrichment between rounds one and two was suggested by the increased ratio of the radioactivity (DNA) on the Hox11 membrane and control BSA membrane in round two (Table 1). A similar increase was also observed between rounds three and four. However, the ratio between the Hox11 and BSA membranes did not increase between rounds two and three, possibly because of overwashing of the Hox11 membrane in round three. However, this observation does not necessarily suggest that this particular round of selection did not achieve any enrichment.

The optimal Hox11 DNA binding sequence

The selected oligonucleotides from the fourth round of selection were amplified, cloned and sequenced. To compare the enrichment of the selection, the original population of synthesized oligonucleotides and the oligonucleotides from the first round of selection were also sequenced. Strikingly, all of the 12 sequenced oligonucleotide clones from the fourth round of selection contained a single core sequence, GGCGGTAAGTGG (sequence 1). As shown in Table 2, this sequence occurred in the oligonucleotide population after the first round of selection at a

Table 2. Hox11 protein selected sequences and their frequencies

selected oligos	frequency	
	first round of selection	fourth round of selection
GGCGGTAAGTGG	23/28	12/12
GGCGGTAAGCCG	5/28	0/12

Two categories of oligonucleotides selected by Hox11 protein were detected in the 28 oligonucleotide clones from the first round of selection, with frequencies of 82% (23/28) and 18% (5/28). However, only the first sequence was found in the 12 oligonucleotide clones from the fourth round of selection.

frequency of 78.2%. The frequency of this oligonucleotide in the original oligonucleotide population was low, as none of 10 randomly sequenced synthesized oligonucleotides contained this sequence (data not shown). Therefore, this sequence represents the optimal binding sequence for Hox11 in this oligonucleotide pool. Another species found in the oligonucleotide population from the first round of selection was GGCGGTAAGCCG (sequence 2), which differed from sequence 1 only at the two nucleotide positions underlined. Although its frequency in this population was 21.8%, it was not present at a high frequency (0/12) in the fourth round of selection (Table 2). The co-occurrence of these two similar sequences after the first round of selection indicates that they were enriched by Hox11 specific selection. Both of them contain a TAAG core which is a variant of the TAAT motif recognized by many homeodomains. The differences between these two sequences in the nucleotides located next to TAAG could influence the DNA binding affinity by the Hox11 (see Discussion).

Hox11 specifically binds to the optimal sequence

The specific interaction between Hox11 and the optimal sequence was further analyzed by gel mobility shift and DNA competition assays. In a band shift assay, 10 ng (~10^5 c.p.m.) of labelled oligonucleotide, GGCGGTAAGTGG (sequence 1), 100 ng of poly dI.dC and 100 ng of BSA protein were employed in each DNA binding reaction. As shown in Figure 4a, there was a stoichiometric band shift that did not occur in the absence of Hox11. The binding specificity was further studied by DNA competition experiments. The competitor DNAs used in this study were rF, NP and sequence 1 (for their sequences see Materials and Methods). rF was used as a non-specific competitor while sequence 1 was used as a self-competitor. NP DNA contains a TAAT core sequence, which is required for DNA binding of many other homeodomain proteins, but was not present in either sequence 1 or 2. Although Hox11 can bind to a sequence without this TAAT core, because the precise frequency of any specific oligonucleotide in the original population was not known, this observation does not necessarily imply that the Hox11 cannot bind to the TAAT containing sequence. To test this possibility, NP DNA which contained the TAAT core was employed as a special DNA competitor to test whether it can compete with sequence 1. As shown in Figure 4b, a 20-fold excess of self-competitor DNA (unlabelled sequence 1) almost completely abolished binding of Hox11 to sequence 1; in contrast,

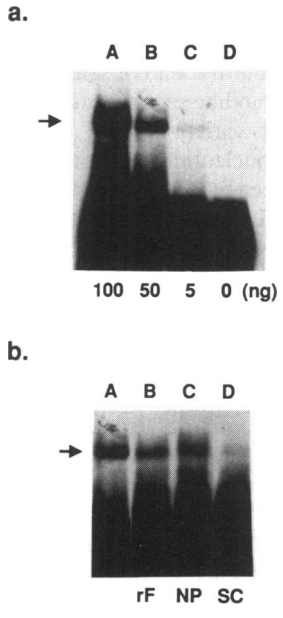

Figure 4. Band mobility shift and DNA competition assays. (**a**) Band shift assay, showing the binding between purified Hox11 protein and its *in vitro* selected binding sequence GGCGGTAAGTGG. The numbers below the figure indicate the amount of protein used in each binding reaction. Ten ng (10^5 c.p.m.) of DNA probe was employed in each reaction. The arrow points to the shifted bands. (**b**) DNA competition assay, showing the specific interaction between Hox11 protein and its *in vitro* selected binding sequence. Ten ng (10^5 c.p.m.) of DNA probe and 50 ng of Hox11 protein were used in each reaction. 200 ng (a 20-fold excess) of different competition DNAs were employed in different reactions. A: no competitor DNA; B: rF oligonucleotide as competitor; C: TAAT site containing oligonucleotide (NP) as competitor; D: self-competitor (SC). The arrow points to the shifted bands.

the rF and NP competitors were only able to decrease the intensity of the shifted band by ~1/2. These data suggest that the interaction between Hox11 and the selected sequence 1 was specific. It also indicates that Hox11 has a marked specificity for binding to a core TAAG sequence, as compared to the TAAT preferred by many other homeodomain proteins.

DISCUSSION

In this paper we have described the optimal DNA binding sequence for the Hox11 protein, by using a novel technique of random oligonucleotide selection developed in this study. Other different experimental procedures have been described to determine the DNA binding specificities of proteins, including selection mediated by immunoprecipitation (37), protein affinity chromatography (34) and gel retardation (36). The procedure defined in this study has several advantages, compared with these techniques. It is simple and easy to manipulate and the whole process of selection can be clearly monitored (Table 1). More importantly, only a small amount of protein (~4 μg) is required for four cycles of selection. This feature should be extremely helpful when large quantities of purified protein are not available (as in this study). Based on this technique, it may be possible to develop an even simpler procedure to define the DNA binding sites of proteins, by combining Western blotting with this random oligonucleotide selection technique, thereby eliminating the protein purification step.

That only the sequence GGCGGTAAGTGG repeatedly (12/12) appeared in the enriched oligonucleotides after four rounds of selection (Table 2) strongly suggests that this sequence is the optimal binding sequence for Hox11 in this oligonucleotide pool. The great enrichment implies that only the most strongly bound oligonucleotides survived through all four rounds of selection in this study. Similar but less extensive enrichment of single specific random oligonucleotides has also been described in other studies (for example, 31). Rather than the theoretical possibility of clonal selection during cloning procedure, which was unfavoured by the fact that all 10 sequenced oligonucleotide clones from the original pool contained different sequences sharing no obvious homology with the selected oligonucleotide (data not shown), the striking enrichment of the selected sequence might attribute to the following factors or combination of them. The first practical possibility could be that the synthesized oligonucleotide pool was partially biased and thus the variety of the 'randomized' sequence was limited, although sequencing of a small collection of them failed to detect such bias. Secondly, the sequence surrounding the recognition core was contacted by the protein sequence outside the homeodomain and therefore was also selected. Thirdly, given the fact that the selected sequences (Table 2) happened to markedly resemble a sequence GGCGATAAGACA at the 3′ downstream of the randomized core (Materials and Methods), these two sequences may serve as a cooperative binding site of Hox11 and thus the selection favoured this particular sequence. However, because the selected sequence was highly restricted, the 3′ downstream sequence may be only a weak binding site on its own. On the other hand, as shown in the band-shift and DNA competition assays, the selected sequence was a fairly strong binding site as compared with NP sequence.

The optimal binding sequence for Hox11 contains a TAAGTG motif, which is close to the potential binding site TAANTG predicted on the basis of structural consideration (see Introduction). The consistency between this result and the prediction supports an universality of the *En* model of homeodomain–DNA interaction (18). Because the Hox11 homeodomain shares only 38% amino acid identity with the *En* homeodomain (Fig. 1a), the presence of a potential binding motif predicted by the *En* model in the optimal binding sequence for Hox11 is consistent with the notion that different homeodomains interact with DNA in a conserved way determined by a combination of three-dimensional structural features and specific amino acid residues (18,19). Significantly, the difference between the Hox11 and *En* homeodomains at position 47 was associated with a change at the corresponding nucleotide (fourth nucleotide T→G), suggesting that, like the *En* homeodomain, this residue at position 47 of the Hox11 homeodomain plays a critical role in determining the specific DNA base at the fourth position in the binding motif. This correlation has also been established recently by mutagenesis analysis (31). Interestingly, the consensus binding sequences of the Hox11 homeodomain peptide determined in another study contained either a TAAC or TAAT core (31), instead of the TAAG observed in this study (see below). Comparing the binding sequences determined by these two studies revealed a requirement of the TAA trinucleotide and that the nucleotide at the fourth position is variable. As proposed in Figure 5, threonine can use its hydroxyl or methyl in the side chain to bind guanine, cytosine and thymine. In addition, the consensus binding sequences of the Hox11 homeodomain peptide did not show a preference at the dinucleotide following the TAAC/T core (31). The differences in the binding sequences observed in these two studies could reflect

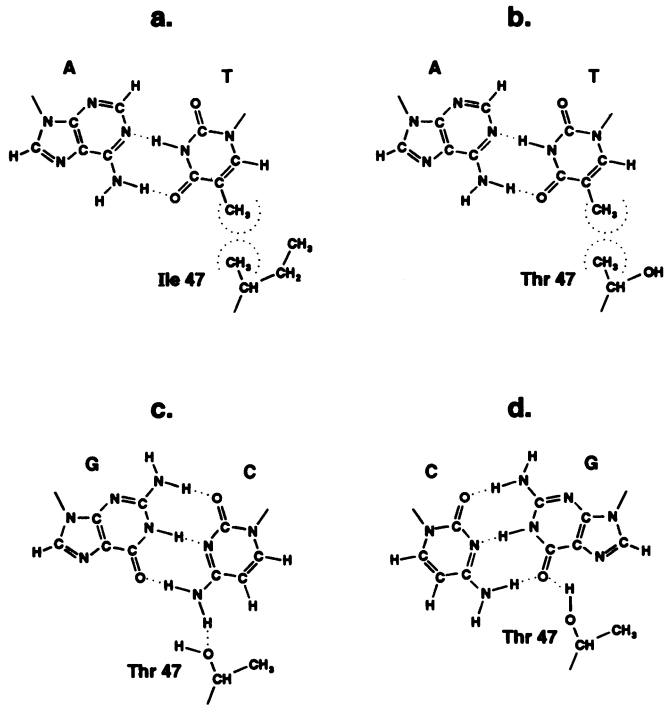

Figure 5. The potential interaction between the threonine residue at position 47 of the Hox11 homeodomain and a nucleotide at the Hox11 recognition motif. It was shown (18) that the isoleucine in the position 47 of the *En* homeodomain contacts a thymine by making a hydrophobic methyl-methyl contact (**a**). Threonine at this equivalent position in the Hox11 homeodomain could have more flexibility to interact with different nucleotides due to its side chain which contains both a hydrophobic methyl group and a polar hydroxyl group. Theoretically, it could interact with thymine by making hydrophobic methyl–methyl interaction (**b**) and with cytosine (**c**) as well as guanine (**d**) by making a hydrogen bond.

different experimental conditions, especially the protein peptide, employed in the experiments; TAAC and TAAT containing sequences could represent the most suitable binding sites for the homeodomain peptide, whereas the TAAGTG containing sequence revealed in this study represents the best binding motif for the full length Hox11 protein. Consistent with this interpretation, the contribution of sequences outside the homeodomain in DNA binding specificity has been reported in other homeodomain proteins (35). In addition, the finding that a TAAT competitor sequence (NP) was not able to compete efficiently with sequence 1 indicates that the TAAT-containing NP sequence is not the optimal binding sequence for the intact Hox11 protein.

Several studies have shown the important role of residue 50 (the ninth residue in the recognition helix) in discriminating the DNA binding sites of different homeodomain proteins, by determining the dinucleotide following the TAAT core (20–23). A Gln at position 50 in the *Antp*, *En* and *Ftz* homeodomains preferentially recognizes a TG dinucleotide, whereas Lys50 in the *Bcd* homeodomain favours a CC dinucleotide (15). Consistent with the proposed DNA recognition specificity of the Gln50 in other homeodomain, there is a TG dinucleotide following the TAAG core in Hox11 optimal binding sequence, which corresponds a Gln50 in the Hox11 homeodomain. Intriguingly, this dinucleotide was changed to a CC dinucleotide in sequence 2 GGCGGTAAGCCG which appeared after the first round of selection (Table 2) and this change represented the only difference between these two binding sequences. However, this initially selected sequence disappeared by the fourth round of selection, and is presumably, therefore, a weaker binding sequence for the Hox11 protein. This result suggests that variation from the optimal recognition sequence in this dinucleotide affects binding affinity. Similar observations have been reported for other homeodomain proteins (22). Because the TG dinucleotide preference was not detected in the consensus binding sequence of the Hox11 homeodomain peptide in the position following the TAAC or TAAT core (31), this preferential recognition by Gln50 could be influenced by sequences outside the homeodomain.

In summary, the following conclusions can be drawn from the results of this study. First, although the Hox11 homeodomain shares a low (38%) amino acid identity with the *En* homeodomain, the DNA binding specificities of Hox11 can be predicted by the DNA binding model of the *En* homeodomain, suggesting that these two homeodomains bind to DNA in a similar mode and that they share similar three-dimensional structural features. Given the great divergence between these two homeodomains, this result suggests a general applicability of the *En* model. Secondly, the presence of the predicted binding motif in the optimal binding sequence suggests that, as for the *En* homeodomain, the Hox11 homeodomain also uses residues at comparable positions to contact DNA bases. Residue change in these position (e.g. Thr47) can cause corresponding DNA base substitutions.

ACKNOWLEDGEMENTS

This work was supported by the Medical Research Council (MRC) of Canada to MLB. MLB was a MRC scientist. We thank Dr Alan Bernstein for his critical reading of this manuscript and Dr Ian D. Dube for providing HOX11 cDNA.

REFERENCES

1 Hayashi, S. and Scott, M.P. (1990) *Cell* **63**, 883–894.
2 Kappen, C.. Schughart, K. and Ruddle, F. (1993) *Genomics* **18**, 54–70.
3 McGinnis, W. and Krumlauf, R. (1992) *Cell* **68**, 283–302.
4 Blatt, C., Lotem, J. and Sachs, L. (1992) *Cell Growth Differ.* **33**, 671–676.
5 Nourse, J., Mellentin, J. D., Galili, N., Wilkinson, J., Stabridge, E., Smith, S.D. and Cleary, M.C. (1990) *Cell* **60**, 535–545.
6 Kamps, M.P., Murre, C., Sun, X.H. and Baltimore, D. (1990) *Cell* **60**, 547–555.
7 Jabs, E.W., Muller, U., Li, X., Lou, W., Haworth, I.S., Klisak, I., Sparkes, R., Warman, M.L., Mulliken, J.B., Snead, M.l., and Maxson, R. (1993) *Cell* **75**, 443–450.
8 Tassabehji, M., Read, A.P., Newton, V.E., Harris, R., Balling, R., Gruss, P., and Strachan, T. *Nature* **355**, 635–636.
9 Shepherd, J.C., McGinnis, W., Carrasco, A.E., De Robertis, E.M. and Gehring, W.J. (1984) *Nature* **310**, 70–71.
10 Laughon, A. and Scott, M.P. (1984) *Nature* **310**, 25–31.
11 Desplan, C., Theis, J. and O'Farrell, P.H. (1986) *Nature* **328**, 630–635.
12 Janes J. and O'Farrell, P.H. (1988) *Nature* **336**, 744–749.
13 Kuziora, M.A. and McGinnis, W. (1989) *Cell* **59**, 563–71.
14 Lin,L. and McGinnis, W. (1992) *Genes Dev.* **6**, 1071–1081.
15 Treisman, J., Harris, E., Wilson, D., and Desplan, C. (1992) *BioEssays* **14**, 145–150.
16 Gehring, W.J., Qian, Y.Q., Billeter, M., Furukubo-Tokunaga, K., Schier, A.F., Resendez-Perez, D., Affolter,M., Otting, G., and Küthrich, K. (1994) *Cell* **78**, 211–223.
17 Otting, G., Qian, Y.Q., Billeter, M., Muller, M., Affolter, M., Gehring, W.J. and Wüthrich, K. (1990) *EMBO J.* **9**, 3085–3092.
18 Kissinger, C.R., Liu, B., Martin-Blanco, E., Kornberg, T.B. and Pabo, C.O. (1990) *Cell* **63**, 579–590.

19 Wolberger, C., Vershon, A.K., Liu, B., Johnson, A.D. and Pabo, C.O. (1991) *Cell* **67**, 517–528.
20 Treisman, J., Gonczy, P., Vashishtha, M., Harris, E. and Desplan, C. (1989) *Cell* **59**, 553–562.
21 Hanes, S.D. and Brent, R. (1989) *Cell* **57**, 1275–1283.
22 Hanes, S.D. and Brent, R. (1991) *Science* **251**, 426–430.
23 Percival-Smith, A., Müller, M., Affolter, M. and Gehring, W.J. (1990) *EMBO J.* **9**, 3967–3974.
24 Dube, I.D., Kamel-Reid, D., Yuan, C.C., Lu, M., Wu, X., Corpus, G., Raimondi, S.C., Crist, W.M., Carroll, A.J. and Minowada, J. (1991) *Blood* **78**, 2996–3003.
25 Hatano, M., Robert, C.W.M., Minden, M., Crist, W.M. and Korsmeyer, S. (1991) *Science* **253**, 79–82.
26 Kennedy, M.A., Gonzalez-Sarmiento, R., Kess, U.R., Lampert, F., Dear, N., and Rabbitts, T.H. (1991) *Proc. Natl. Acad. Sci. USA* (1991) **88**, 8900–8904.
27 Lu, M., Gong, Z.Y., Shen, W.F. and Ho, D.A. (1991) *EMBO J.* **10**, 2905–2910.
28 Raju, K., Tang, S., Dube, I.D., Kamel-Reid, S., Bryce, D.M., and Breitman, L.M. (1993) *Mech. Dev.* **44**, 51–64.
29 Cheng, S.H. and Mak, T.W. (1993) *Develop. Growth Differ.* **35**, 655–663.
30 Robert, C.W.M., Shutter, J.R. and Korsmeyer, S.J. (1994) *Nature* **368** 747–749.
31 Dear, T.N., Sanchez-Garcia, I. and Rabbitts, T.H. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 4431–4435.
32 Sambrook, KJ., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
33 Ausubel, F.M., Brent, R., Kingston R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. (1989) *Current Protocols in Molecular Biology.* Wiley Intersceince.
34 Ekker, S.C., Young, K.E., von Kessler, D.P. and Beachy, P.A. (1991) *EMBO J.* **10**, 1179–1186.
35 Hoey, T., Warrior, R., Manak, J. and Levine, M. (1988) *Mol. Cell. Biol.* **8**, 4598–4607.
36 Blackwell, T.K. and Weintraub, H. (1990) *Science* **250**, 1104–1110.
37 Woods, D.B., Ghysdael, J. and Owen, M.J. (1992) *Nucleic Acid Res.* **20**, 699–704.