Development 121, 4339-4347 (1995)
Printed in Great Britain © The Company of Biologists Limited 1995
DEV9444

Regulation of *Hoxc-8* during mouse embryonic development: identification and characterization of critical elements involved in early neural tube expression

Cooduvalli S. Shashikant[1,*], Charles J. Bieberich[2], Heinz-Georg Belting[1], Jeanne C. H. Wang[1], Marc A. Borbély[1] and Frank H. Ruddle[1,3]

[1]Department of Biology and [3]Genetics, Yale University, Kline Biology Tower, PO Box 208103, New Haven CT 06520, USA
[2]Jerome H. Holland Laboratory, 15601 Crabbs Branch Way, Rockville, MD 20855, USA

*Author for correspondence

SUMMARY

We have characterized *cis*-acting elements that direct the early phase of *Hoxc-8* expression using reporter gene analysis in transgenic mice. By deletion we show that a 135 bp DNA fragment, located approximately 3 kb upstream of the coding region of *Hoxc-8*, is capable of directing posterior neural tube expression. This early neural tube (ENT) enhancer consists of four separate elements, designated A, B, C and D, whose nucleotide sequences are similar to binding sites of known transcription factors. Nucleotide substitutions suggest that element A is an essential component of the ENT enhancer. However element A by itself is incapable of directing neural tube expression. This element requires interactions at any two of the other three elements, B, C or D. Thus, the components of the ENT enhancer direct neural tube expression in an interdependent manner. We propose that *Hoxc-8* is activated in the neural tube by combinatorial interactions among several proteins acting within a small region. Our transgenic analyses provide a means to identify transcription factors that regulate *Hoxc-8* expression during embryogenesis.

Key words: *cis*-acting elements, homeobox gene, transgenic analysis, *lacZ* reporter gene, enhancer, transcriptional control, mouse,neural tube

INTRODUCTION

The determination of positional identities along the anteroposterior axis represents a complex and as yet unresolved problem in pattern formation. The positional cues for the generation of spatial organization are highly conserved among metazoans (Krumlauf, 1994; Lawrence and Morata, 1994; Ruddle et al., 1994). In insects, the Homeotic-Complex (HOM-C) encodes a family of regionally expressed genes that are largely responsible for the specification of segmental identity along the anteroposterior axis (Lawrence and Morata, 1994). The mammalian genome contains four separate clusters of genes, termed *Hox* genes, which share a high degree of sequence, organizational and functional similarity to the *Drosophila* HOM-C genes (Krumlauf, 1994). Furthermore, as in *Drosophila*, the mammalian *Hox* genes exhibit a collinear relationship between their order of arrangement on the chromosome and their anterior limits of expression along the embryonic axis: the more 3′ the gene is within a cluster, the more anterior is its limit of expression (Graham et al., 1989; Duboule and Dollé, 1989). Since proper spatial patterns of *Hox* gene expression are necessary for normal development, understanding how *Hox* genes are regulated in precise spatial patterns along the anteroposterior axis will help explain an important step in morphogenesis.

Hox gene expression is regulated both spatially and temporally primarily at the level of transcription. Factors that govern the spatial expression of *Hox* genes may form a hierarchy that reflects developmental decisions made prior to *Hox* gene activation. A critical first step in dissecting this molecular hierarchy is to examine transcriptional regulation of *Hox* genes. This will require isolation of *cis*-acting DNA elements that mediate spatial regulation, characterization of critical nucleotides that affect the regulatory activity of these sequences and identification of *trans*-acting factors whose activity represents a previous level of hierarchy controlling pattern formation.

Employing reporter gene analysis in transgenic mice, several investigators have identified *cis*-regulatory regions of mouse *Hox* genes (see Krumlauf, 1994). Many of these studies have demonstrated that relatively large genomic regions are capable of regulating different aspects of *Hox* gene expression but specific elements have been identified in only a few cases (Sham et al., 1993; Marshall et al., 1994; Studer et al., 1994). Many of these analyses suggest that interactions among various *cis*-regulatory regions are required for *Hox* gene expression. However, the critical nature of such combinatorial interactions has not been directly demonstrated.

We and others have previously studied the region-specific expression and function of mouse *Hoxc-8* (Awgulewitsch et al., 1986; Utset et al., 1987; Breir et al., 1988; Gaunt, 1988; Le Mouellic et al., 1988, 1992; Awgulewitsch and Jacobs, 1990;

Pollock et al., 1992, 1995). The expression of *Hoxc-8*, like that of many other *Hox* genes, can be divided into an early or 'establishment phase' and a late or 'maintenance phase' (Deschamps and Wijgerde, 1993; Gaunt and Strachan, 1994). At 8.5 days pc (post coitum), *Hoxc-8* expression extends from the base of the allantois to the segmental plate mesoderm and to a more anterior region in the neurectoderm. The anterior boundary of *Hoxc-8* at 9.5 days pc is located in the neural tube at the level of the 9th somite and in the somitic mesoderm at the level of the 13th somite. Later in development, posterior expression of *Hoxc-8* decreases while intense expression is maintained within the brachial region of the neural tube.

In previous studies of the regulation of the *Hoxc-8* gene, we described a 5.1 kb *cis*-regulatory region that mediates expression in posterior neurectoderm and mesoderm reconstituting the early but not the late phase of *Hoxc-8* expression (Bieberich et al., 1990). In the studies reported here, we have performed extensive deletion and mutational analyses of this regulatory region to delineate elements involved in the establishment of region-specific patterns of gene expression. We describe the components of a posterior early neural tube (ENT) enhancer and provide evidence that four distinct elements located within a 135 bp region interact to determine the expression pattern in the neural tube. Comparison of nucleotide sequences critical for enhancer activity reveal potential sites for interactions with families of known transcription factors including caudal, forkhead and homeodomain proteins.

MATERIALS AND METHODS

Construction of transgenes

Hsp68-lacZ gene was obtained from Dr J. Rossant. The *Bam*HI fragment of *phspPTlacZpA* was subcloned into a modified pGem (Promega) vector termed pSafyre (L. Bogarad unpublished data) at the *Bam*HI site (pHSF5). pHSF5 was used as a basic subcloning vector for all constructs. All restriction enzymes were from New England Biolabs. A 2.1 kb *Eco*RI-*Hin*dIII fragment was isolated as a *Hin*dIII fragment (*Hin*dIII site in the polylinker) from a plasmid which contained a 5.1 kb *Eco*RI fragment of the *Hoxc-8* upstream region (see Fig. 1 for restriction map). This fragment was cloned at the *Hin*dIII site in the polylinker to generate two constructs: *construct 1,* which contained the nucleotide sequence in the same orientation as that of the genomic *Hoxc-8* sequence with respect to its transcription site, and *construct 2*, which contained the nucleotide sequence in the opposite orientation. The 1.4 kb *Eco*RI-*Dra*I fragment was isolated as a *Hin*dIII-*Dra*I fragment from construct 1 and cloned at the *Hin*dIII site of pHSF5 by blunt end ligation (*construct 3*). Construct 1 was digested with *Bsp*EI and *Sma*I (in the polylinker), end filled and self-ligated to generate *construct 5. Construct 6* was generated by isolating a 0.7 kb *Dra*I-*Hin*dIII fragment from construct 1 and subcloned at the *Hin*dIII site of pHSF5 by blunt end ligation. *Construct 3* was digested with *Bsp*EI and *Sal*I, end filled and self-ligated to generate *construct 7*. Digestion of construct 7 with *Xho*I and *Pst*I (enzymes in the polylinker) released the 335 bp fragment. This fragment was further digested with *Rsa*I and the resulting fragments *Xho*I-*Rsa*I (203 bp) and *Rsa*I-*Pst*I (132 bp) were cloned directionally into pHSF5, which was rendered blunt at *Hin*dIII and staggered at *Xho*I or *Pst*I to generate *constructs 8* and *9*, respectively. *Construct 10* was made by isolating a 132 bp *Apa*I-*Pst*I (in the polylinker) fragment from construct 9, further digesting it with *Hae*III, and subcloning the resulting *Hae*III-*Pst*I fragment directionally into pHSF5, which had been digested with *Hin*dIII, rendered blunt and digested with *Pst*I. Similarly, the *Alu*I-*Pst*I fragment from the above fragment was cloned to obtain *construct 11*.

Remaining constructs were generated by the polymerase chain reaction (PCR). PCR was carried out in a 50 μl reaction containing 100 ng each of template DNA and two primers in a buffer containing 10 mM Tris, pH 8.4; 50 mM KCl, 1.5 mM $MgCl_2$, 200 μM each of the four dNTPs and 1 unit of Taq polymerase (Boehringer-Mannheim). The reaction mix was overlaid with a drop of mineral oil and placed in a thermocycler (Hybaid, Omnigene). The conditions for amplification were 94°C, 3 minutes, 1 cycle; 94°C, 1 minute; 55°C, 1 minute; 72°C, 1 minute 15 seconds; 20 cycles; 72°C 5 minutes, 1 cycle. At the end of the reaction, 10 μl of the reaction product was analyzed by agarose gel electrophoresis to confirm the amplification of correct-sized product. The fragments were diluted 3- to 5-fold and subcloned into PCR™ vector (Invitrogen) according to the manufacturer's instructions.

The nucleotide sequence of the 335 bp *Bsp*EI-*Dra*I fragment is shown in Fig. 4 and the position of the nucleotides included in the constructs is shown in Fig. 3 (constructs 12-15). The alterations in the 335 bp reporter construct were introduced by an overlapping PCR strategy, using synthetic oligonucleotide primers containing appropriate changes in the nucleotide sequence. The changes introduced are indicated in Fig. 4 (constructs 16-24).

All amplified products were sequenced by the dideoxy sequencing method (Sanger et al., 1977) using sequencing reagents (Pharmacia), 1 μCi of (^{35}S) dATP (Amersham), 3 U Sequenase (USB Biochemicals) and 10 ng T7 or SP6 primers (Promega) in each reaction. These plasmids were then digested with *Xho*I and *Pst*I, and the desired fragments were isolated and subcloned at similar sites in pHSF5. Plasmid DNAs were routinely isolated by an alkaline lysis method followed by ultracentrifugation on a CsCl gradient (Sambrook et al., 1989). Prior to injection, DNA fragments were excised from vector sequences by digesting with *Xho*I (or *Apa*I or *Sap*I) and *Not*I in the polylinkers and isolated by ultracentrifugation on a sucrose density gradient (Sambrook et al., 1989). DNA fragments were dialyzed extensively against 10 mM Tris, pH 7.5 and 0.25 mM EDTA, and their concentration was adjusted to 1-3 μg/ml.

Production of transgenic mice

Transgenic mice were generated by injecting DNA into pronuclei of fertilized oocytes of inbred FVB mice (Gordon et al., 1980). Eggs that survived injection were transplanted into the oviduct of pseudopregnant CD-1 or B6/CBAF1 fosters. The founder embryos were analyzed for the expression of the transgene on 9.5 or 10.5 days pc. Alternatively, some of the embryos were allowed to go to term and transgene expression was analyzed in the F_1 embryos obtained by timed mating. For staging embryos, the day that the plug was observed was considered to be 0.5 days pc.

Southern blot analysis

Southern blot analysis was carried out as described (Sambrook et al., 1989). Genomic DNA was isolated from placenta or tails of young mice, 4-5 weeks old as described (Hogan et al., 1986; Laird et al., 1991). 10 μg of genomic DNA was digested with *Pst*I, electrophoresed on an agarose gel and transferred to nitrocellulose (Schleicher & Schuell, Inc.) or Hybond N filters (Amersham). The filters were hybridized radioactively with random-labeled 3.7 kb *Hin*dIII-*Bam*HI fragment from pCH110 (Pharmacia). Alternatively, the probes were labeled nonradioactively with 'Genius Nonradioactive Nucleic Acid Labeling and Detection Kit' (Boehringer-Mannheim) and the hybridized products were visualized by following the manufacturer's instructions.

Detection of β-galactosidase activity

The embryos were dissected free from the maternal tissues into cold PBS, fixed for 30 minutes in 0.25% glutaraldehyde, washed in PBS and stained for β-galactosidase activity as described previously (Bieberich et al., 1990).

Detection of Hoxc-8 proteins

Hoxc-8 proteins were detected by using the whole-mount staining method (Lumsden and Keynes, 1989). The production and characterization of anti-*Hox*c-8 monoclonal antibodies will be described elsewhere (H. B. and F. H. R., unpublished data).

DNA mobility shift assay

Total protein extracts were prepared from 9.5 dpc embryos obtained from staged matings as follows. The embryos were dissected free of maternal tissue, separated from the yolk sac, washed in cold PBS and frozen by liquid Nitrogen as quickly as possible. Frozen embryos collected from 5-7 pregnant mice were thawed in 5 ml of Buffer A (50 mM Tris, pH 7.6; 5 mM $MgCl_2$, 25 mM KCl, 0.2 mM EDTA, 250 mM sucrose and protease inhibitor cocktail containing 1 mM phenylmethylsulfonylfluoride; PMSF from Sigma, 100 ng/ml aprotinin, 500 ng/ml leupeptin and 1 μg/ml pepstatin from Boehringer-Mannheim). The embryos were homogenized on ice in a tight-fitting Dounce homogenizer with 20 strokes. The extract was then brought to 0.5 M NaCl and 0.5% Triton X-100 and incubated on ice for 30 minutes with occasional mixing. At the end of incubation, the extract was repeatedly passed through a 26-gauge needle to reduce the viscosity of the solution and then centrifuged at 105,000 ***g*** for 60 minutes at 4°C. The supernatant was collected and dialyzed against Buffer A for 8-12 hours with repeated changes of the buffer. After dialysis, the extract was centrifuged at 15,000 ***g*** for 20 minutes at 4°C. The supernatant was aliquoted (100 μl) and frozen in liquid Nitrogen. The amount of protein in the extract was quantitated by Lowry's method with minor modifications (Lowry et al., 1951).

To determine protein binding to the 5′ region of the 135 bp, the following complementary oligonucleotides were synthesized: 5′AGCTTTTATGGCCCTGTTTGTCTCCCTGCTCTA3′ and 5′AGCTAGAGCAGGGAGACAAACAGGGCCATAAAA3′. To perform scanning mutations, a series of oligonucleotides, each containing three nucleotide substitutions (A to C; G to T and vice versa) were synthesized beginning from position 5. These oligonucleotides were purified by electrophoresis on a 7 M urea-15% polyacrylamide gel and isolated as described above. The complementary single-stranded oligonucleotides were then mixed in equal proportions and annealed in buffer containing 10 mM Tris, pH 7.4 and 100 mM NaCl. The reaction mixture was heated at 65°C for 15 minutes and allowed to cool gradually to room temperature. The resulting double-stranded oligonucleotides were purified on a 15% polyacrylamide gel. The oligonucleotides were then end labeled with high specific activity ^{32}P-dCTP and cold dATP, dGTP and dTTP with Klenow fragment. The labeled oligonucleotides were separated from free nucleotides on a spin column of Sephadex G-75.

Electrophoretic mobility shift assay was carried out by mixing different amounts of protein (10-40 μg) with approximately 1 ng labeled DNA (20,000 cts/minute) and 1 μg poly(dI).poly(dC) in a 50 μl reaction mix containing buffer 50 mM Tris, pH 7.6; 5 mM $MgCl_2$, 25 mM KCl, 0.2 mM EDTA, 250 mM sucrose and 6 μg/ml Bovine serum albumin (Ausubel et al., 1987). The reaction was incubated for 20 minutes on ice and then electrophoresed on a 5% polyacrylamide gel. The gels were dried and exposed to Kodak X-ray film for 12-18 hours.

RESULTS

Early activation of *Hoxc-8* is mediated by a 2.1 kb enhancer

Previous studies identified a 5.1 kb DNA fragment from the *Hoxc-8* locus that was capable of directing a region-specific pattern of gene expression in transgenic mice (Bieberich et al., 1990). This fragment included the *Hoxc-8* promoter (Awgulewitsch et al., 1990) and was shown to direct a pattern of reporter gene expression that was consistent with the early phase of *Hoxc-8* expression (Bieberich et al., 1990). In this

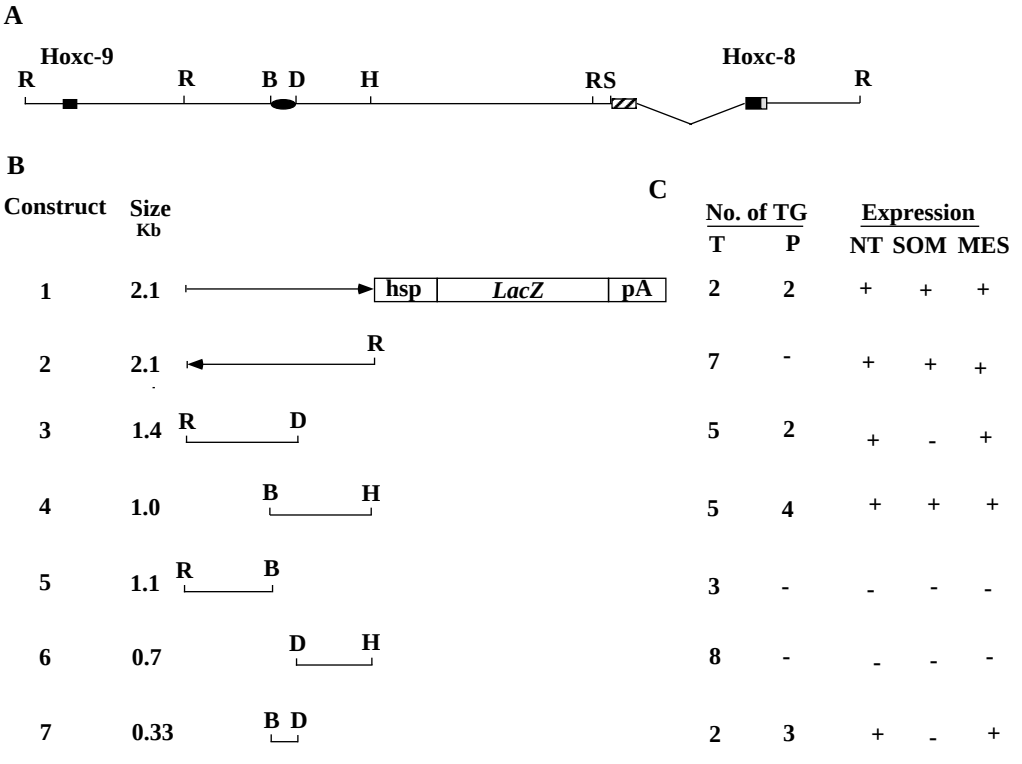

Fig. 1. Characterization of regulatory regions of the *Hoxc-8* gene in transgenic mouse embryos. Schematic of reporter gene constructs and sites of expression. (A) Partial restriction map of *Hox*c-9-*Hoxc-8* intergenic region. R, *Eco*RI; B, *Bsp*EI; D, *Dra*I; H, *Hin*dIII; S, *Sac*I. Black boxes indicate homeoboxes of *Hox*c-9 and *Hoxc-8*; Black sphere is the position of the ENT enhancer. Shaded box represents the first exon of the *Hoxc-8*. (B) Schematic of various regions of the *Hoxc-8* genomic loci tested for enhancer activity in transgenic mice; hsp, heat-shock promoter 68; *lacZ*, *β-galactosidase* gene (Rossant et al., 1993); pA, polyadenylation signal derived from SV40. The direction of the arrow in constructs 1 and 2 indicate the orientation of the fragment with respect to the genomic sequences. (C) Results of transgenic expression analysis. The first column represents the number of transgenic embryos (TG) that express *β-galactosidase* gene: T, founder generation embryos; P, established transgenic lines; expression (+); and absence of expression (–) is indicated in neural tube (NT), somites (SOM) and lateral plate mesoderm (MES). The embryos not expressing in any of these tissues do show expression of the β-galactosidase gene at ectopic sites that are unique to each transgenic embryo for the same construct.

study, deletion of the distal 2.1 kb *Eco*RI-*Hin*dIII fragment (Fig. 1) completely abolished activity, suggesting that elements within this fragment are required for *Hoxc-8* transcriptional activity (data not shown). This 2.1 kb DNA fragment was linked to a heterologous promoter, mouse heat-shock promoter, *hsp68*, that by itself does not confer region-specific expression on a β-galactosidase reporter gene (Kothary et al.,1989; Rossant et al., 1991). In either orientation, the 2.1 kb enhancer conferred on the *hsp68* promoter an expression pattern that was very similar, though not identical, to the early phase of *Hoxc-8* expression (Fig. 2A,B). Like the endogenous *Hoxc-8* gene and the previously described 5.1 kb transgene, the 2.1 kb enhancer became transcriptionally active around 8.0 days pc, established appropriate anterior boundaries of expression in the neural tube and paraxial mesoderm in 8.5 days pc embryos and directed expression in most cells of the tail bud region at this stage. However, the expression pattern of the 2.1 kb enhancer-containing construct differed from the 5.1 kb transgene in that expression was stronger in the somites than in the lateral plate mesoderm and was down regulated at later stages of development (i.e., after 10.5 days pc; data not shown). Thus the 2.1 kb fragment, which is located 3 kb from the transcriptional start sites of the *Hoxc-8*, can function as an enhancer directing posterior early neural tube and mesoderm expression in the context of a heterologous promoter.

Regulatory elements for neural tube expression of the transgene reside within a 335 bp fragment

Fragments derived from the 2.1 kb enhancer were tested for activity in transgenic mice as a prelude to a detailed mutational analysis. Two partially overlapping fragments, a 1.4 kb *Eco*RI-*Dra*I and a 1.0 kb *Bsp*EI-*Hin*dIII fragment (constructs 3 and 4 respectively; Fig. 1) were capable of driving reporter gene expression in the neural tube and mesoderm (Fig. 2C,D), whereas fragments lacking the region of overlap (constructs 5 and 6) were inactive. The two active constructs direct similar anterior boundaries of expression in the neural tube at somite levels 12 and 11 in 9.5 days pc. embryos, respectively. Practically all cells in the tail bud region including those in the allantois showed intense staining for β-galactosidase activity. However, the level of β-galactosidase activity in the somites differed between the two constructs. Construct 3 showed strong expression of the transgene in somites in the most posterior region but weaker expression more anteriorly. In contrast, somitic expression was uniformly strong in embryos containing construct 4. These differences were verified at the cellular level in serial sections (data not shown). A comparison between the expression pattern of construct 4 and the *Hoxc-8* protein distribution, shown in Fig. 2E and F, demonstrates the similar anterior boundaries of expression in the neural tube and mesoderm in 8.5 days pc embryos.

The above results suggest that the region of overlap between construct 3 and construct 4, a 335 bp *Bsp*EI-*Dra*I fragment, contains elements that support the expression of the transgene in the posterior neural tube. The reporter gene construct containing only this 335 bp region overlap (construct 7) was shown to drive the expression in the posterior neural tube and mesoderm surrounding the caudal regions of the neural tube (Fig. 2G,H). Compared to larger constructs (constructs 3 and 4), the enhancer

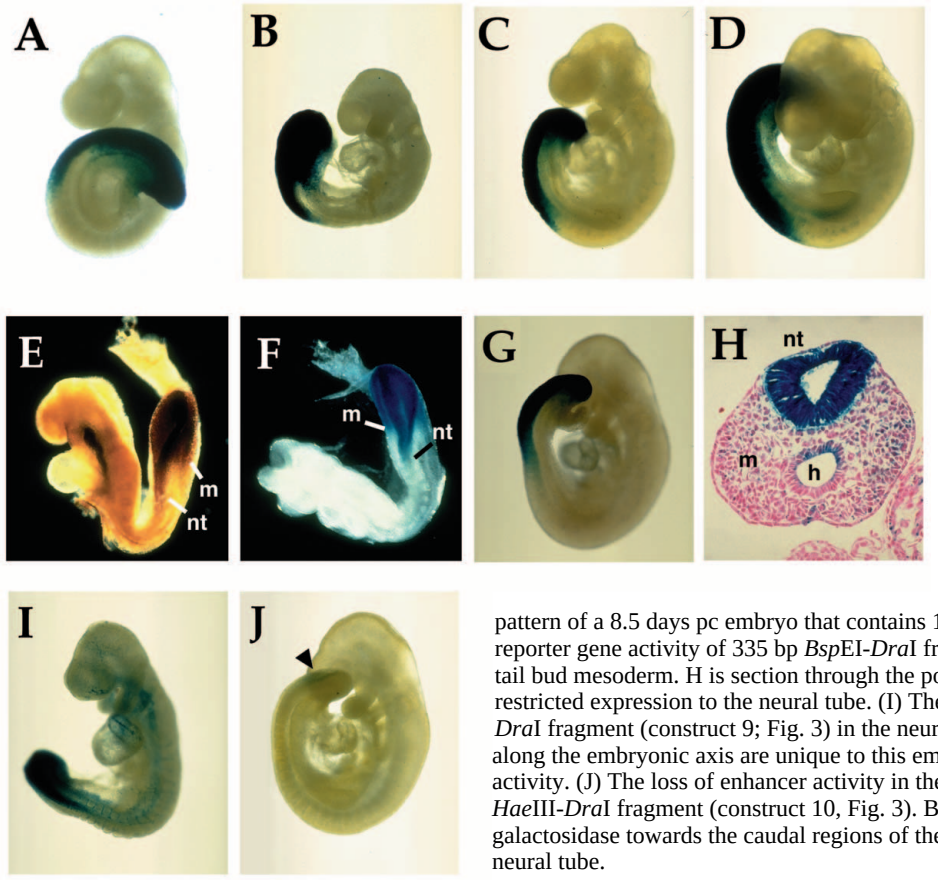

Fig. 2. Expression pattern of reporter constructs carrying various genomic regions of *Hoxc-8* in transgenic mice. Expression of the β- galactosidase gene is detected in the posterior regions of the neural tube and mesoderm. Restriction fragments of the reporter constructs (1-7) are shown in Fig. 1. The transgenic embryos are staged between 9.0 and 9.5 days pc. (A,C-G) Embryos sired from established transgenic lines; (B,I,J) founder generation embryos. (A,B) Enhancer activity of 2.1 kb *Eco*RI-*Hin*dIII fragment in direct and inverted orientations (constructs 1 and 2). (C,D) Differential staining patterns in the somites of 1.4 kb *Eco*RI-*Dra*I fragment and 1.0 kb *Bsp*EI-*Hin*dIII fragments (constructs 3 and 4) respectively. (E) Whole-mount staining for Hoxc-8 protein with anti-Hoxc-8 monoclonal antibodies of 8.5 days pc embryos. (F) β-galactosidase-staining pattern of a 8.5 days pc embryo that contains 1 kb *Bsp*EI-*Hin*dIII transgene. (G,H) The reporter gene activity of 335 bp *Bsp*EI-*Dra*I fragment (construct 7) in the neural tube and tail bud mesoderm. H is section through the posterior region of the embryo showing restricted expression to the neural tube. (I) The reporter gene activity of the 135 bp *Rsa*I-*Dra*I fragment (construct 9; Fig. 3) in the neural tube and the mesoderm. The other stainings along the embryonic axis are unique to this embryo and unrelated to the ENT enhancer activity. (J) The loss of enhancer activity in the transgenic embryo that contains a 98 bp *Hae*III-*Dra*I fragment (construct 10, Fig. 3). Black arrow indicates limited staining of β-galactosidase towards the caudal regions of the neural tube. h, hind gut; m, mesoderm; nt, neural tube.

activity of construct 7 was less pronounced as reflected by the longer time required to stain the embryo for β-galactosidase activity, as well as by a slight temporal delay in the onset of expression (data not shown). In addition, the anterior boundary of expression with construct 7 was 1- to 2-somite levels posterior to that observed with the larger constructs. Moreover, construct 7 showed no activity in somites when assayed by either whole-mount staining or by histological analysis of tissue sections (Fig. 2H). However, some mesoderm in the tail bud was weakly positive for β-galactosidase activity.

It can be concluded that the 335 bp fragment contains essential regulatory elements for early expression in the posterior neural tube. Although sequences surrounding this fragment may contribute to enhanced neural tube expression, especially at the anterior border of *Hoxc-8* expression, no independent *cis*-acting elements were detected outside of this region by reporter gene analysis.

Deletion analysis defines two interacting regulatory regions in the 335 bp enhancer

Deletion analyses were performed to delineate regulatory regions within the 335 bp enhancer (*Bsp*EI-*Dra*I fragment) that direct early neural tube expression. Assays for β-galactosidase activity of these deletion constructs were performed on founder generation embryos at 9.5 days pc because of the high level of β-galactosidase activity detected at this stage. Construct 8, containing the 5′ 200 bp of the *Bsp*EI-*Dra*I fragment, did not contain specific enhancer activity (Fig. 3) and the inclusion of increasing amounts of 3′ sequence (*Bsp*EI-237, 272, 284 and 314 in constructs 12-15 respectively) did not reconstitute the neural tube expression pattern completely, suggesting the presence of additional element(s) at the 3′ terminus of the 335 bp fragment. In contrast, construct 9, containing the 3′ 135 bp of the *Bsp*EI-*Dra*I fragment, showed β-galactosidase activity in the posterior neural tube and mesoderm in the tail bud region (Fig. 2I). Although construct 9 clearly showed region-specific expression, the expression was somewhat posterior to that driven by the 335 bp fragment.

Further deletion within the 135 bp region from the 5′ end resulted in loss of enhancer activity in the neural tube. Construct 10, containing a 98 bp *Hae*III-*Dra*I fragment, showed expression in only a few cells at the caudal neuropore (Fig. 2J). Hence a 32 bp region between the *Rsa*I and a *Hae*III sites (see Fig. 4 for nucleotide positions) contains a regulatory element essential for early enhancer activity. These results, taken together, indicate the presence of regulatory elements

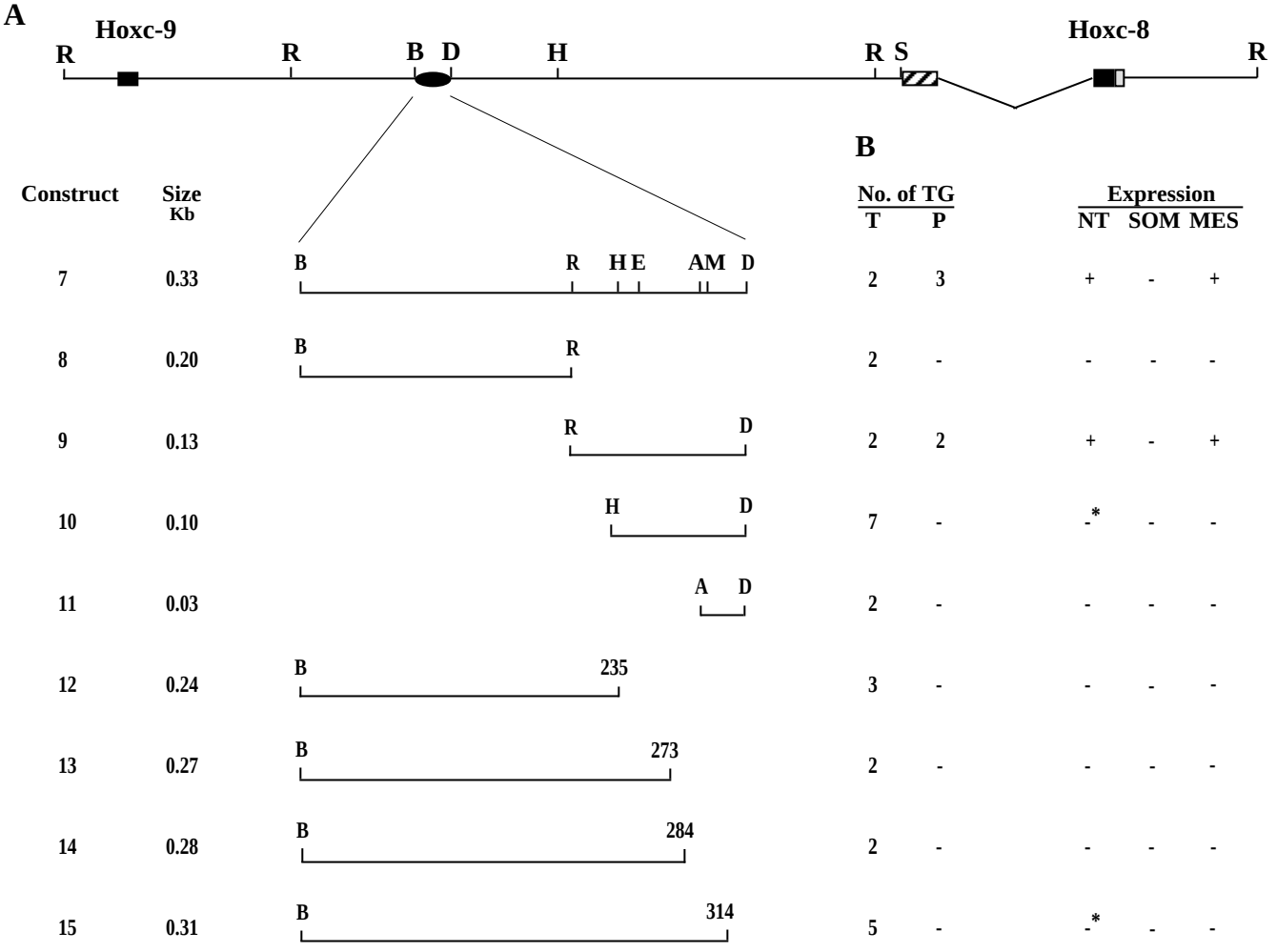

Fig. 3. Deletion analysis of ENT enhancer region. (A) Schematic of reporter gene constructs containing different regions of the 335 bp *Bsp*EI-*Dra*I fragment. B, *Bsp*EI; R, *Rsa*I, H, *Hae*III, E, *Eco*NI, A, *Alu*I, M, MaeII, D, *Dra*I. The nucleotide positions correspond to the mouse sequence in Fig. 4. (B) Results of transgenic analysis. See legend to Fig. 1 for the description. * indicates consistent but limited expression in the neural tube.

located at the 5′ and 3′ termini of the 135 bp fragment that are essential for neural tube expression.

Mutational analysis defines four distinct elements in the ENT enhancer

To gain insight into critical nucleotides involved in the posterior neural tube enhancer activity, we next identified elements that serve as specific binding sites for proteins in the mouse embryo extracts. DNA fragments from termini of the 135 bp fragment (a 52 bp *Rsa*I-*Eco*NI fragment and a 30 bp *Mae*I-*Dra*I fragment; see Fig 4) bound proteins present in 9.5 dpc mouse embryo extracts in DNA-mobility shift assays (data not shown). Based on comparisons of nucleotide sequences to the binding sites for known transcription factors, each terminus was identified to contain two distinct sites, designated as A and B at the 5′ end, and C and D at the 3′ end (Fig. 4). The nucleotide sequences of sites A and D represent potential binding sites for caudal-related proteins (Dearolf et al., 1989; Margalit et al., 1993; Suh et al., 1994), site B represents a potential binding site for HNF-3/forkhead-related proteins (Liu et al., 1991; Jackson et al., 1993; Overdier et al., 1994) and site C is a typical homeodomain protein-binding sequence (Triesman et al., 1992).

To identify the nucleotides involved in the binding activity at the 5′ terminus, a synthetic oligonucleotide containing sequences 227-259 was tested for binding to proteins using a 9.5 dpc mouse embryonic extract in a DNA-mobility shift assay (Fig. 5 lanes 1-3; I). We further synthesized a series of oligonucleotides, each containing three nucleotide substitutions and tested them for protein binding. Mutations in the 5′TTTTATGGC3′ sequence resulted in reduced protein-binding activity (Fig. 5 lanes 1-3; II-IV), whereas mutations in the adjacent sequences (Fig. 5 lanes 1-3; V) did not affect protein binding. This sequence is identical to the protein-binding sites for the caudal-related proteins from different organisms (Dearolf et al., 1989; Margalit et al., 1993; Suh et al., 1994). Preliminary results indicate that hamster Cdx-3 protein can bind to both sites A and D of the ENT enhancer (A. Carr and C. S. S., unpublished observations).

To determine if these elements were critical for enhancer activity in vivo, three nucleotide substitutions were introduced in sites A and D individually (TTT<u>TAT</u>GG to TTT<u>GCG</u>GG) in the 335 bp reporter construct. Nucleotide substitutions in site A (construct 16; Fig. 4) dramatically reduced enhancer activity. Weak expression of the transgene was noticeable in a few embryos in cells at the caudal neuropore (Fig. 6A) in a pattern similar to that observed with construct 10 (Fig. 2J) in which element A is deleted. These results indicate that site A is essential for enhancer activity. In contrast, mutating site D (construct 19) did not affect the neural tube expression of the transgene but did result in weaker expression in the mesoderm (Fig. 6D). This demonstrates that sites A and D, although identical in their core nucleotide sequences, are functionally distinct.

Individual alterations at the sequences adjacent to sites A and D (designated as B and C; Fig. 4) did not affect the neural tube expression of the enhancer. Deletion of site B (18 bp *Hae*III-*Eco*NI sequence) in the 335 bp enhancer (construct 17) or mutations in site C (TT<u>AA</u>TTG to TT<u>CC</u>TTG) in the 335 bp enhancer (construct 18) showed no observable changes in the neural tube expression of the reporter genes in transgenic embryos (Fig. 6B,C). However, mutations introduced simultaneously at sites B, C and D did result in pronounced changes in the neural tube expression patterns as discussed below.

Pair-wise alterations at sites B, C and D were introduced in various combinations. The construct containing a deletion at site B and substitutions at site C (construct 20) led to a loss of neural tube expression except in the caudal limit (Fig. 5E), as did construct 21, which contained substitutions at sites B and C (data not shown).

When nucleotide substitutions were introduced at both site B and site D (construct 22) or at both site C and site D (construct

```
BspEI                                                                                      90
TCCGGATTAGATTCGATTTTATTCATTACCATTTATGAATTAAAACAACATGACACACAGCTGGGGAGAGAAATGAAAGCAGGCTAGGAG

                                                                                          180
AGAGGCTAGCAAACGCCCTCTTCCCCTCATCTTCCCTCCCACCTCTCCTCTGCTCCTTTGTCGGAATCACAAAACCCTAAAGGTTGTCTC

                                      A             B                                     
                    RsaI                     HaeIII      EcoNI                            270
ACTTGGGAAGGCAGCAGATAGGTACATTTCCGTAGCCCAGAAATGCCACTTTTATGGCCCTGTTTGTCTCCCTGCTCTAGGTTCTGAATG

                                         C      D
                         AluI                    DraI    335
GGGCTGAACAAAACAGCAGTGCAGAGCTGGCTAGACGTCTGGGCTTAATTGTTTTATGGTTTAAA
```

Construct No.	Nucleotide substitutions	No. of Tg. Embryos	Expression
16	Site A: TTT<u>TAT</u>GG to TTT<u>CGC</u>GG	7	limited to caudal neuropore or ectopic sites
17	Site B: CCTGTTTGTCTCCCTGCT, deletion	7	unaffected
18	Site C: TTAATT to TTCCTT	3	unaffected
19	Site D:TTT<u>TAT</u>GG to TTT<u>CGC</u>GG	6	unaffected
20	Site B:deletion; Site C:substitutions	3	limited to caudal neuropore
21	Site B:TG<u>TTT</u>GT to TG<u>GGG</u>GT Site C:substitutions	2	limited to caudal neuropore
22	Site B:deletion Site D:substitutions	6	limited to posterior neural tube
23	Site A:substitutions Site D:substitutions	2	limited to ectopic sites
24	Site C:substitutions Site D:substitutions	2	limited to posterior neural tube

Fig. 4. Mutational analysis of ENT enhancer region. The nucleotide sequence of the 335 bp *Bsp*EI-*Dra*I fragment is shown at the top. Underlined sequences represent recognition sequences for restriction enzymes. The nucleotide sequence of sites A, B, C and D are indicated in the bold. The results of mutations at these sites are indicated at the bottom. Tg, transgenic embryos that express *lacZ.*

24), reduced levels of neural tube expression were seen (construct 22, Fig. 6F; construct 24, not shown) and the anterior extent of the neural tube expression was at least 6 somite levels more posterior than that observed in embryos carrying an unaltered 335 bp construct. The level of neural tube expression obtained with these constructs was much higher than that obtained with constructs that contained alterations at sites B and C.

In summary, interactions at four distinct sites are critical for the expression of the transgene in the neural tube (Fig. 7). Site A is essential but not sufficient for enhancer activity. Interactions at sites B, C and D are required for the determination of the anterior extent of transgene expression driven by the ENT enhancer. Further, the nucleotide sequences of these sites implicate members of caudal, forkhead and homeodomain-protein families as transcription factors involved in the regulation of the *Hoxc-8* ENT enhancer.

DISCUSSION

Combinatorial interactions involving the ENT enhancer

We have characterized an early-acting enhancer from the 5′ flanking region of the mouse *Hoxc-8* gene. This enhancer, termed the early neural tube (ENT) enhancer, directs expression in the posterior neural tube. The key elements that mediate ENT transcriptional regulation have been mapped to a 135 bp fragment, although other elements located elsewhere may also contribute towards neural tube expression (C. S. S., unpublished observations). Within this fragment, we have identified four distinct but interdependent elements designated AB and CD that occur in pairs at its termini (Fig. 6). The ENT enhancer requires combinatorial interactions between these elements and presumably specific transcription factors for appropriate transcriptional regulation. Element A is essential but not sufficient to direct ENT enhancer activity. At least two other elements in conjunction with A are required to reconstitute enhancer activity (i.e., ABC, ABD and ACD). Elements A and D are identical with respect to sequence

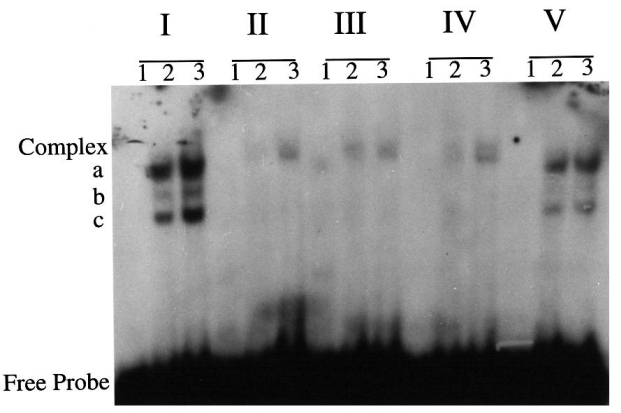

Fig. 5. DNA mobility shift assay for proteins binding to altered sequences from the ENT enhancer region. Oligonucleotide I contains a 32 bp sequences spanning from positions 233-260 (see Fig. 4). Oligonucleotides II-V each contain three nucleotide substitutions at positions TTT, TAT, GGC or CCT, respectively. The positions of the free probe and the DNA-protein complexes, a, b and c are indicated. Lane 1 contains no protein, lanes 2 and 3 contain 20 and 40 μg of proteins, respectively.

but differ functionally: mutations at site D alone do not eliminate expression. In contrast, elements B and C represent dissimilar nucleotide sequences, yet mutations at these sites individually or in conjunction with mutations at site D lead to very similar reporter gene expression patterns, suggesting that these elements are functionally equivalent. These results, taken together, implicate site A as having primary importance in activating expression, whereas the other sites play a secondary role as modulators. Moreover, element B in conjunction with C and D appears to influence the anterior limit of expression of the reporter in the neural tube. Thus, ENT enhancer activity depends on combinatorial interactions among its components. Multiple *cis*-regulatory regions have been implicated in the regulation of *Hox* gene expression in other analyses performed to date. However, these studies were confined to regions of DNA significantly larger than the ENT enhancer (see Krumlauf, 1994, for references).

Candidate factors mediating ENT enhancer activity

Characterization of critical nucleotides required for ENT enhancer activity presents us with an opportunity to identify *trans*-acting factors acting upstream of *Hoxc-8*. Such factors have been detected in mouse embryo extracts (C. S. S., unpublished observations) but their exact identity remains to be determined. A comparison of the nucleotide sequence of the ENT enhancer with the binding sites for known transcription factors reveals that three of the four elements are potential sites for interactions with homeodomain proteins and that the fourth element is a potential binding site for forkhead-related proteins. Two classes of homeodomain proteins, caudal and Hox proteins, may interact with the nucleotide sequences of the ENT enhancer.

A comparison with the zebra stripe enhancer of the *Drosophila ftz* gene suggests that caudal-related proteins can interact with elements A and D. Three murine caudal-related genes have been identified to date: *Cdx-1*, *Cdx-2* (homologous to hamster *Cdx-3*) and *Cdx-4* (Duprey et al., 1988; Hu et al, 1993; James et al., 1994; Gamer and Wright, 1993; German et al., 1992; Suh et al., 1994). The sequences to which caudal proteins bind are similar to the 'TTTATG' sequence found in sites A and D of the ENT enhancer (Dearolf et al., 1989; German et al., 1992; Margalit et al., 1993; Suh et al., 1994). Significantly, these binding sites are grouped in pairs or associated with binding sites for other transcription factors (Dearolf et al., 1989; German et al., 1992; Suh et al., 1994). Mouse Cdx proteins are expressed during gastrulation, prior to *Hoxc-8* gene expression and their spatial and temporal distribution makes them candidates for the regulation of the ENT enhancer (Meyer and Gruss, 1993; Gamer and Wright, 1993; F. Beck, personal communication).

The second class of proteins that may interact with the ENT enhancer includes members of the *Hox* gene family. Element C (TTAATTGT) contains a common TAAT core shared by most *Hox*-binding sites (Triesman et al., 1992). Since *Hox* genes are expressed in overlapping expression patterns, many *Hox* proteins may interact with this element of the ENT enhancer at different developmental stages. In *Drosophila*, several auto- and cross-regulatory interactions among homeodomain proteins have been demonstrated by biochemical and genetic studies. In mammals, binding of *Hox* proteins to several promoters in vitro and their transcriptional regulation in cell transfection assays have been demonstrated (for example, see Goomer et al., 1994; Guazzi et al., 1994). Circumstantial evidence suggests that the *H0XD4* gene is autoregulated similarly to its *Drosophila*

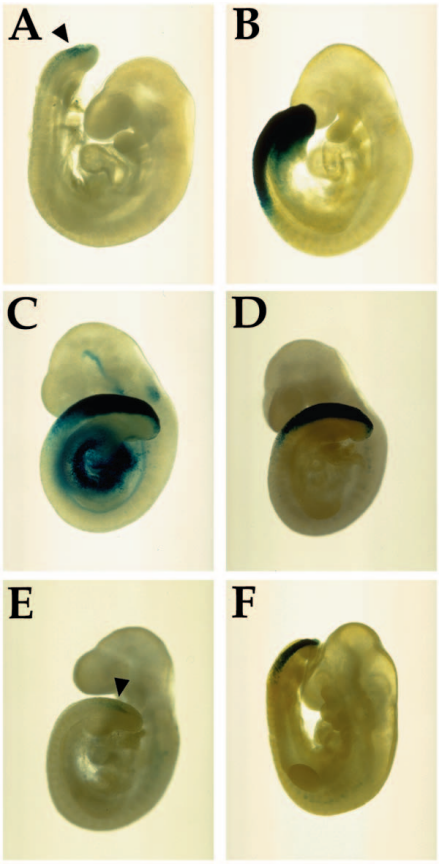

Fig. 6. Expression patterns of reporter genes in transgenic embryos carrying mutations within the 335 bp fragment. (A-F) Enhancer activity of the construct 16-20 and 22, respectively (see Fig. 3). The expression of the *β-galactosidase* gene is found in the posterior regions of the embryo (black arrow). Additional expression in the viscera of the transgenic embryo in C is due to the site of integration of the transgene, unrelated to the ENT enhancer activity.

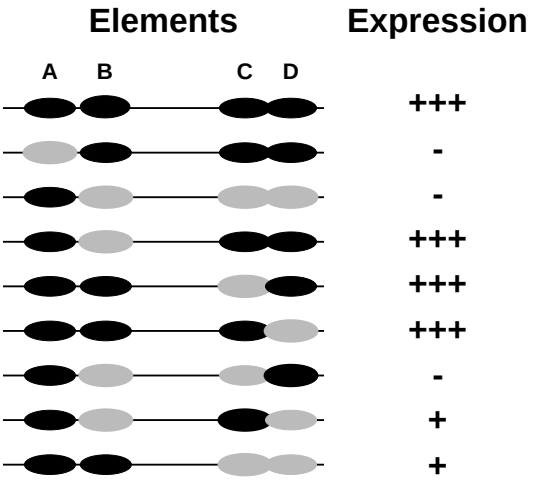

Fig. 7. Summary of the mutational analysis of the ENT enhancer. Dark circles represent wild-type sequences whereas the grey circles represent mutated sequences. +++, wild type; +, reduced; –, no expression of the transgene in the neural tube.

ortholog, *deformed* (Pöpperl and Featherstone, 1992). However, examples of in vivo regulation of enhancers by Hox proteins are few (Pöpperl et al., 1995). Although Hox proteins are expected to demonstrate exquisite in vivo specificity, it is difficult to demonstrate such specificity in vitro, as most Hox proteins bind to related sequences with very similar affinities. It has been suggested that the specificity of *Hox* gene interactions is provided in part by protein-protein interactions as in the case of *Drosophila* extradenticle and ultrabithorax proteins, which both regulate the activity of the downstream *decapentaplegic* gene (Chan et al., 1994; van Dijk and Murre, 1994). The ENT enhancer provides an excellent opportunity to investigate protein-protein interactions involving mammalian Hox proteins.

The third type of nucleotide sequence, defined by element B, includes a binding sequence for HNF3/forkhead-related proteins (Liu et al., 1991; Jackson et al., 1993; Overdier et al., 1994). Many of these proteins are expressed early during gastrulation (Ang et al., 1993; Sasaki and Hogan, 1993; Pierrou et al., 1994; Kaestner et al., 1994).

Thus, a number of developmentally important proteins appear to have the potential to regulate the activity of the ENT enhancer. The elucidation of these interactions will aid in the identification of networks of transcription factors necessary for *Hox* gene regulation. To date, two types of upstream factors have been implicated in the regulation of *Hox* genes. These are retinoids and the zinc finger protein, Krox 20, in the regulation of *Hox*b-1 and *Hox*b-2 genes, respectively (Sham et al., 1993; Marshall et al., 1994; Studer et al., 1994). On this basis, it has been argued that the factors acting upstream of mammalian *Hox* genes are distinct from those acting in *Drosophila*, reflecting differences in their modes of early development. However, the nature of the ENT enhancer described here suggests that some of the regulators of mammalian *Hox* genes may still be evolutionarily derived from *Drosophila* counterparts such as caudal and other homeodomain proteins.

This work was supported by a grant (GM 09966-33) from the National Institutes of Health. We wish to acknowledge the following for critical discussions on experiments and/or writing of the manuscript: Drs Michael Salbaum, Suzanne Bradshaw, Kevin Bentley, Wendy Bailey, Manuel Utset, Alexander Awgulewitsch and Trevor Williams. We also thank Dr Aruna Pawashe for advice on overlapping PCR and Zhiling Jiang for oligonucleotide synthesis.

REFERENCES

Ang, S.-L., Wierda, A., Wong, D., Stevens, K. A., Casico, S., Rossant, J. and Zaret, K. S. (1993). The formation and maintenance of the definitive endoderm lineage in the mouse: involvement of HNF3/*forkhead* proteins. *Development* **119**, 1301-1315.

Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D. Seidman, J. G., Smith, J. A. and Struhl, K. (1987). *Current Protocols in Molecular Biology*. New York: Wiley and Sons.

Awgulewitsch, A., Utset, M. F., Hart, C. P., McGinnis, W. and Ruddle, F. H. (1986). Spatial restriction in expression of a mouse homeo box locus within the central nervous system. *Nature* **320**, 328-335.

Awgulewitsch, A. and Jacobs, D. (1990). Differential expression of *Hox* 3.1 protein in subregions of the embryonic and adult spinal cord. *Development* **108**, 411-420.

Awgulewitsch, A., Bieberich, C., Bogarad, L., Shashikant, C. and Ruddle, F. H. (1990). Structural analysis of the *Hox-3.1* transcription unit and the *Hox-3.2-Hox-3.1* intergenic region. *Proc. Natl. Acad. Sci. USA* **87**, 6428-6432.

Bieberich, C. J., Utset, M. F., Awgulewitsch, A. and Ruddle, F. H. (1990). Evidence for positive and negative regulation of the *Hox* 3.1 gene. *Proc. Natl. Acad. Sci. USA* **87**, 8462-8466.

Breir, G., Dressler, G. R. and Gruss, P. (1988). Primary structure and developmental expression pattern of *Hox* 3.1, a member of the murine *Hox*3 complex homeobox gene cluster. *EMBO J.* **7**, 1329-1336.

Chan, S.-K., Jaffe, L., Capovilla, M., Botas, J. and Mann, R. S. (1994). The DNA binding specificity of Ultrabithorax is modulated by cooperative interactions with extradenticle, another homeoprotein. *Cell* **78**, 603-615.

Dearolf, C. R., Topol, J. and Parker, C. S. (1989). The caudal gene product is a direct activator of *fushi tarazu* transcription during *Drosophila* embryogenesis. *Nature* **341**, 340-343.

Deschamps, J. and Wijgerde, M. (1993). Two phases in the establishment of HOX expression domains. *Dev. Biol.* **156**, 473-480.

Duboule, D. and Dollé, P. (1989). The structural and functional organization of the murine HOX gene family resembles that of *Drosophila* homeotic genes. *EMBO J.* **8,** 1497-1505.

Duprey, P., Chowdhury, K., Dressler, G. R., Balling, R., Simon, D., Guenet, J. and Gruss, P. (1988). A mouse gene homologous to the *Drosophila* gene caudal is expressed in epithelial cells from embryonic intestine. *Genes Dev.* **2**, 1647-1654.

Gamer, L. W. and Wright, C. V. E. (1993). Murine cdx-4 bears striking similarities to the *Drosophila caudal* gene in its homeodomain sequence and early expression pattern. *Mech. Dev.* **43**, 71-81.

Gaunt, S. J. (1988). Mouse homeobox gene transcripts occupy different but overlapping domains in embryonic germ layers and organs: a comparison of *Hox*-3.1 and *Hox*-1.5. *Development* **103**, 135-144.

Gaunt, S. J. and Strachan, L. (1994). Forward spreading in the establishment of a vertebrate *Hox* expression boundary: The expression domain separates into anterior and posterior zones, and the spread occurs across implanted glass barriers. *Dev. Dyn.* **199**, 229-240.

German, M., Wang, J., Chadwick, R. B. and Rutter, W. J. (1992). Synergistic activation of the insulin gene by a LIM-homeo domain protein and a basic helix-loop-helix protein: building a functional insulin minienhancer complex. *Genes Dev.* **6**, 2165-2176.

Goomer, R. S., Holst, B. D., Wood, I. C., Jones, F. S. and Edelman, G. M. (1994). Regulation *in vitro* of an L-CAM enhancer by homeobox genes *HoxD9* and *HNF-1. Proc. Natl. Acad. Sci. USA* **91**, 7985-7989.

Gordon, J. W., Scangos, G. A., Plotkin, D. J., Barbosa, J. A. and Ruddle, F. H. (1980). Genetic transformation of mouse embryos by microinjection of purified DNA. *Proc. Natl. Acad. Sci. USA* **77**, 7380-7384.

Graham, A., Papalopulu, N. and Krumlauf, R. (1989). The murine and *Drosophila* homeobox gene complexes have common features of organization and expression. *Cell* **57**, 367-378.

Guazzi, S., Lonigro, R., Pintonello, L., Boncinelli, E., Di Lauro, R. and Mavillo, F. (1994). The thyroid transcription factor-1 gene is a candidate target for regulation by *Hox* proteins. *EMBO J.* **13**, 3339-3347.

Hogan, B., Constantini, F. and Lacy. E. (1986). *Manipulating the Mouse Embryo A Laboratory Manual.* Cold Spring Harbor, New York: Cold Spring Harbor Press.

Hu, Y., Kazenwadel, J. and James, R. (1993). Isolation and characterization of the murine homeobox gene, *cdx-1*.: regulation of expression in intestinal epithelium. *J. Biol. Chem.* **268**, 27214-27225.

Jackson, D. A., Rowader, K. E., Stevens, K, Jiang, C., Milos, P. and Zaret, K. S. (1993). Modulation of liver-specific transcription by interactions between hepatocyte nuclear factor 3 and nuclear factor 1 binding DNA in close apposition. *Mol. Cell Biol.* **13**, 2401-2410.

James, R., Erler, T. and Kazenwadel, J. (1994). Structure of the mouse homeobox gene *cdx-2. J. Biol. Chem.* **269**, 15229-15237.

Kaestner, K. H., Lee,K.-H., Schlöndorff, J., Hiemisch, H., Monaghan, P. and Schütz, G. (1993). Six members of the mouse forkhead gene family are developmentally regulated. *Proc. Natl. Acad. Sci. USA* **90**, 7628-7631.

Kothary, R., Clapoff, S., Darling, S., Perry, M. D., Moran, L. A. and Rossant J. (1989). Inducible expression of an *hsp68-lacZ* hybrid gene in transgenic mice. *Development* **105**, 707-714.

Krumlauf, R. (1994). *Hox* genes in vertebrate development. *Cell* **78,** 191-201.

Laird, P. W., Zijderveld, A., Linders, K., Rudnicki, M. A. and Jaenisch, R. A. (1991). Simplified mammalian DNA isolation procedure. *Nucleic Acids Res.* **19**, 4293-4294.

Lawrence, P. A., and Morata. G. (1994). Homeobox genes: Their function in *Drosophila* segmentation and pattern formation. *Cell* **78**, 181-189.

Le Mouellic, H., Condamine, H. and Brûlet, P. (1988). Pattern of transcription of the homeo gene *Hox-3.1* in the mouse embryo. *Genes Dev.* **2**, 125-135.

Le Mouellic, H., Lallemand, Y. and Brûlet, P. (1992). Homeosis in the mouse induced by a null mutation in the *Hox-3.1* gene. *Cell* **69**, 251-264.

Liu, J. K., DiPersio, C. M. and Zaret, K. S. (1991). Extracellular signals that regulate liver transcription factors during hepatic differentiation *in vitro*. *Mol. Cell. Biol.* **11**, 773-784.

Lowry, O.H., Rosebrough, N. J., Farr, A. L. and Randall, R. J. (1951). Protein measurement with Folin phenol reagent. *J. Biol.Chem.* **193**, 265-275.

Lumsden, A. and Keynes, R. (1989). Segmental patterns of neuronal development in the chick hindbrain. *Nature* **337**, 424-428.

Margalit, Y., Yarus, S., Shapira, E. Gruenbaum, Y. and Fainsod, A. (1993). Isolation and characterization of target sequences of the chicken *CdxA* homeobox gene. *Nucleic Acids Res.* **21**, 4915-4922.

Marshall, H., Studer, M., Pöpperl, H., Aparicio, S., Kuroiwa, A., Brenner, S. and Krumlauf. R. (1994). A conserved retonoic acid response element required for early expression of the homeobox gene *Hox b-1*. *Nature* **370**, 567-571.

Meyer, B. I., and Gruss, P. (1993). Mouse *cdx-1* expression during gastrulation. *Development* **117**, 191-203.

Overdier, D.G., Porcella, A. and Costa, R.H. (1994). The DNA-binding specificity of the hepatocyte nuclear factor 3/forkhead domain is influenced by amino acid residues adjacent to the recognition helix. *Mol. Cell. Biol.* **14**, 2755-2766.

Pierrou, S., Hellquist, M., Samuelsson, L., Enerbäck, S. and Carlsson, P. (1994). Cloning and characterization of seven human forkhead proteins: binding site specificity and DNA bending. *EMBO J.* **13**, 5002-5012.

Pollock, R.A., Jay, G. and Bieberich, C. J. (1992). Altering the boundaries of *Hox-3.1* expression: evidence for antipodal gene regulation. *Cell* **71**, 911-924.

Pollock, R. A., Sreenath, T., Ngo, L. and Bieberich, C. J. (1995) Gain of function mutations for paralogous *Hox* genes: implications for the evolution of *Hox* gene function. *Proc. Natl. Acad. Sci. USA* **92**, 4492-4496.

Pöpperl, H. and Featherstone, M. S. (1992). An autoregulatory element of the murine *Hox-4.2* gene. *EMBO J.* **11**, 3673-3680.

Pöpperl, H., Bienz, M., Studer, M., Chan, S-K., Aparicio, S., Brenner, S., Mann, R. S. and Krumlauf, R. (1995). Segmental expression of *Hoxb-1* is controlled by a highly conserved autoregulatory loop dependent upon *exd/pbx*. *Cell* **81**, 1031-1042.

Rossant, J., Zirngibl, R., Cado, D., Shago, M. and Giguére, V. (1991). Expression of a retinoic acid response element-*hsplacZ* transgene defines specific domains of transcriptional activity during mouse embryogenesis. *Genes Dev.* **5**, 1333-1344.

Ruddle, F. H., Bartels, J. L., Bentley, K. L., Kappen, C., Murtha, M. T. and Pendleton, J. W. (1994). Evolution of *Hox* genes. *Ann. Rev. Genet.* **28**, 423-442.

Sambrook, J., Fritsch, E. F. and Maniatis. T. (1989). *Molecular cloning: A Laboratory Manual.* Cold Spring Harbor, New York: Cold Spring Harbor Laboratory.

Sanger, F., Nicklen, S. and Coulson., A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467.

Sasaki, H. and Hogan, B. L. M. (1993). Differential expression of multiple fork head related genes during gastrulation and axial pattern formation in the mouse embryo. *Development* **118**, 47-59.

Sham, M. H., Vesque, C., Nonchev, S., Marshall, H., Frain, M., Gupta, R. D., Whittig, J., Wilkinson, D., Charnay, P. and Krumlauf, R. (1993). The zinc finger gene *Krox20* regulates *HoxB2 (Hox 2.8)* during hindbrain segmentation. *Cell* **72**, 183-196.

Studer, M., Pöpperl, H., Marshall, H., Kuroiwa, A. and Krumlauf, R. (1994). Role of a conserved retinoic acid response element in rhombomere restriction of *Hoxb-1*. *Science* **265**, 1728-1732.

Suh, E., Chen, L., Taylor, J. and Traber, P. (1994). A homeodomain protein related to *caudal* regulates intestine-specific gene transcription. *Mol. Cell Biol.* **11**, 7340-7351.

Triesman, J., Harris, E., Wilson, D. and Desplan, C. (1992). The homeodomain: a new face for the helix-turn-helix. *BioEssays* **14**, 145-150.

Utset, M. F., Awgulewitsch, A., Ruddle, F. H. and McGinnis, W. (1987). Region-specific expression of two mouse homeo box genes. *Science* **235**, 1379-1382.

van Dijk, M. A. and Murre, C. (1994). *extradenticle* raises the DNA binding specificity of homeotic selector gene products. *Cell* **78,** 617-624.

(Accepted 15 September 1995)

Note added in proof:

The nucleotide sequence reported in the paper has previously been described (Awgulewitsch et al., 1990; GeneBank accession no. M35603).