Understanding how boundaries and domains of Hox gene expression are determined is critical to elucidating the means by which the embryo is patterned along the anteroposterior axis. We have performed a detailed analysis of the mouse Hoxb4 intron enhancer to identify upstream transcriptional regulators. In the context of an heterologous promoter, this enhancer can establish the appropriate anterior boundary of mesodermal expression but is unable to maintain it, showing that a specific interaction with its own promoter is important for maintenance. Enhancer function depends on a motif that contains overlapping binding sites for the transcription factors NFY and YY1. Specific mutations that either abolish or reduce NFY binding show that it is crucial for enhancer activity. The NFY/YY1 motif is reiterated in the Hoxb4 promoter and is known to be required for its activity. As these two factors are able to mediate opposing transcriptional effects by reorganizing the local chromatin environment, the relative levels of NFY and YY1 binding could represent a mechanism for balancing activation and repression of Hoxb4 through the same site.

Key words: Hoxb4, Transcriptional regulation, NFY, YY1, Embryo, Transcription factor, Mouse

INTRODUCTION

There are 39 members of the Hox gene family in the mouse, distributed across four clusters and assigned to 13 paralogy groups (Krumlauf, 1994). One of the striking features of the genes in this family is that they display spatially restricted expression along the anteroposterior (AP) axis of the developing embryo, characterized by very precise anterior boundaries. Moreover, the physical order of genes in a cluster corresponds both to the temporal order in which they are activated and to the anterior extents of their expression, a phenomenon known as colinearity.

That Hox genes function in the determination of regional identity along the AP axis in the mouse has been firmly established by extensive mutational analyses. Thirty four of the 39 genes have now been subjected to targeted disruption and, in general, loss of an individual Hox gene results in a phenotype in the region that corresponds to the most anterior region of its expression domain. Furthermore, an increasing number of compound mutants are being generated involving paralogous and/or non-paralogous genes (Horan et al., 1995; Manley and Capecchi, 1997). These often result in synergistic phenotypes that uncover important functional roles for Hox genes that are not revealed by mutation of single genes. It is clear from these studies that an understanding of how boundaries and domains of Hox gene expression are determined is crucial to elucidating the means by which the embryo is patterned.

Further underlining the importance of transcriptional regulation, exchange of coding sequences between the Hoxa3 and Hoxd3 loci demonstrates that the proteins encoded by these two paralogous genes are functionally equivalent (Greer et al., 2000). Therefore, the specificity observed for each gene must result from differences in their patterns and/or levels of expression.

Thus far, transgenic reporter genes have been the principal strategy employed to investigate the transcriptional control of Hox genes in mice. These studies have revealed that, in general, Hox gene transcription is controlled by a modular system of enhancers located within a few kilobases up- or downstream of the gene (Maconochie et al., 1996). Typically, each enhancer directs tissue- and spatially-specific subsets of the complete expression pattern and often displays the ability to function independently of other regulatory modules. This has greatly facilitated the fine-scale analysis of these regions,
which has lead to the identification of a number of important regulators of Hox gene expression. To date, these are the retinoic acid and retinoid X receptors (Gould et al., 1998; Zhang et al., 2000); Cdx family proteins (Charite et al., 1998); Krox20 (Macanochie et al., 2001); kreasler (Manzanares et al., 1999); members of the AP-2 family of transcription factors (Macanochie et al., 1999); Sox/Oct heterodimers (Di Rocco et al., 2001); and Hox proteins themselves, in conjunction with the Pbx and Meis families of co-factors (Gould et al., 1997; Jacobs et al., 1999).

It is important to note, however, that proximal sequences have not always proved sufficient to recapitulate the endogenous expression pattern (Vogels et al., 1993; Charite et al., 1995). Indeed, for Hoxb8 there is evidence that regulatory elements located nearly 30 kb downstream of the gene (between Hoxb5 and Hoxb4) interact with local enhancers to determine the correct boundary of expression in the neural tube (Valarche et al., 1997). In addition, there are likely to be regulatory mechanisms that operate in a cluster-wide manner. Direct manipulations of the endogenous Hoxd cluster have identified enhancers that act globally, controlling expression of multiple genes in the limb and gut, and have shown that a directional insulator delineates the range of action of one such global element within the cluster (van der Hoeven et al., 1996; Kondo et al., 1998; Herault et al., 1999; Kmita et al., 2000a). Perhaps most dramatically, these experiments have revealed that release from global silencing of the cluster determines the proper timing of activation of the Hoxd genes, and is likely to be a key process controlling colinearity (Kondo and Duboule, 1999; Kmita et al., 2000b). Thus, it seems that the proper regulation of Hox gene expression may be accomplished by a number of diverse mechanisms.

We have previously defined the sequences required to recapitulate Hoxb4 expression in transgenic mice (Whiting et al., 1991). A 3′-flanking enhancer (region A) directs expression in an intronic enhancer (region C) mediates expression within the posterior neural tube, neural crest and mesodermal derivatives. In the absence of these regions, the Hoxb4 promoter shows only ectopic activity in the dorsal midbrain. We have focused our attention on region C because it is required to set the correct anterior limit of Hoxb4 expression in the paraxial mesoderm, at the level of somite 6/7. In this study, we have further characterized the regulatory capacity of region C on the hsp68 promoter. We demonstrate that the anterior boundary of somitic expression is initially specified correctly but in contrast to the homologous promoter we have previously shown to be essential for promoter function (Gutman et al., 1994). This element is able to bind both YY1 and an unknown factor that we named HoxTF (for Hox gene transcription factor). We now demonstrate that HoxTF is the heterotrimeric transcription factor NFY and show that the NFY/YY1 site is necessary for the mesodermal and neural activity of region C. We discuss our results in the light of evidence that NFY and YY1 play opposing roles in the determination of transcriptional states by the recruitment of chromatin modifying co-factors.

MATERIALS AND METHODS

Reporter constructs and transgenic mice

The 1.4 kb Sall-BglIII region C fragment of Hoxb4 was cloned upstream of the hsp68 promoter-lac-Z-SV40 polyA reporter gene (Whiting et al., 1991) to generate CHZ. CHZA559-599 was made by cloning an oligonucleotide into the MunI/SfiI sites of CHZ to produce a 41 bp deletion. CHZ-mNNY1, –mNFY1, –mNFY2 and –mYY1 were constructed by cloning two pairs of complementary, partially overlapping oligonucleotides (which span CR1 and carry specific mutations) into the MunI/SfiI sites of CHZ. Construct b4C-511-558-mYY1 was made by cloning an oligonucleotide into the HindIII site of p610ZAI (Whiting et al., 1991). Full details of all cloning steps are available on request.

The production, PCR diagnosis and whole-mount staining of transgenic mice were performed as described previously (Whiting et al., 1991; Gilthorpe and Rigby, 1999). Vibratome sections (70 μm) were cut after embedding specimens in 2.0% (w/v) agarose in 2.0% (w/v) formaldehyde, 0.1 M L-lysine, 0.01 M sodium m-periodate in PBS. The production, PCR diagnosis and whole-mount staining of transgenic mice were performed as described previously (Whiting et al., 1991; Gilthorpe and Rigby, 1999). Vibratome sections (70 μm) were cut after embedding specimens in 2.0% (w/v) agarose in 2.0% (w/v) formaldehyde, 0.1 M L-lysine, 0.01 M sodium m-periodate in PBS.

DNA sequence alignments

CR1-equivalent regions were identified by pair-wise DNA sequence alignment with MacVector (IBI-Kodak). The mouse Hoxb4 intron sequence was provided by R. Allemann and verified by double-stranded sequencing (dRhodamine terminator cycle sequencing kit, Perkin-Elmer). Chick Hoxa4, Hoxb4 and Hoxd4 sequences were communicated by A. Kuroiwa (Morrison et al., 1995). Upon sequencing of a chick Hoxb4 CR1 PCR fragment, three additional G residues were found (equivalent to 563, 566 and 571 of region C) that improve the degree of identity with all Hoxb4 paralogues in the alignment. The medaka Hoxa4 sequence has been published previously (Haerry and Gehring, 1997). Other sequences were obtained from the GenBank database under the following Accession Numbers: Amphioxus amphHoxA, AB028208; Fugu Hoxb4, FRU92575; horn shark (Heterodontus francisci) Hoxa4, AF224262; mouse Hoxa4, X66861; Hoxd4, MMU77364; zebralish Hoxa4a, AF071246; Hoxb4a, AF071252; Hoxd4a, AF071264. The mouse Hox4 sequence was identified from a working draft HTG sequence (AC021667) by comparison with the mouse cDNA clone NM_013553 (Geada et al., 1992). Sequences were imported in to MegAlign (DNASTAR) and aligned manually to give a best fit to the mouse Hoxb4 sequence. The resulting multiple alignment was shaded using MacBoxshade 2.1 (M. Baron, Institute for Animal Health, UK).

Electrophoretic mobility shift assays (EMSA) and supershifts

Gel retardation experiments were performed as described (Gutman et al., 1994) except poly-(dI-dC) was used as a nonspecific competitor. Whole-cell protein extracts were prepared from 10.5 dpc mouse embryos or the mouse neuroblastoma cell-line Neuro2a, and nuclear extracts from mouse F9 embryonal carcinoma (EC) stem cells. Specific competitors were added at 100-fold molar excess. For supershift experiments, an anti-NFYA antibody (gift from C. Benoist, Strasbourg, France) was added to the reaction 1 hour before addition of the labelled probe.

Size fractionation

Proteins from F9 EC nuclear extracts were separated on a 12.5% (w/v) SDS-PAGE gel. Gel slices were excised and incubated in buffer E (150 mM NaCl, 20 mM Hepes (pH 7.5), 5 mM DTT, 0.1 mM EDTA,
Peptide sequencing
Coomassie stained bands (48, 45 and 36 kDa) were excised, destained and digested in-gel with trypsin overnight at 37°C. Peptides were resolved twice with 50% (v/v) acetonitrile/1% (v/v) TFA and then concentrated prior to analysis. Peptides were resolved using tandem ion exchange and c18 reverse phase separation on a Michrom HPLC system. Edman sequencing on the collected peptides was carried out using an Applied Biosystems Procise system employing fast cycle Edman sequencing on the collected peptides was carried out using an Applied Biosystems Procise system employing fast cycle Edman sequencing on the collected peptides was carried out using an Applied Biosystems Procise system employing fast cycle Edman sequencing on the collected peptides was carried out using an Applied Biosystems Procise system employing fast cycle Edman sequencing on the collected peptides was carried out using an Applied Biosystems Procise system employing fast cycle Edman sequencing on the collected peptides was carried out using an Applied Biosystems Procise system employing fast cycle Edman sequencing on the collected peptides was carried out using an Applied Biosystems Procise system employing fast cycle Edman sequencing on the collected peptides was carried out using an Applied Biosystems Procise system employing fast cycle Edman sequencing on the collected peptides was carried out using an Applied Biosystems Procise system employing fast cycle Edman sequencing on the collected peptides was carried out using an Applied Biosystems Procise system employing fast cycle Edman sequencing on the collected peptides was carried out using an Applied Biosystems Procise system employing fast cycle Edman sequencing on the collected peptides was carried out using an Applied Biosystems Procise system employing fast cycle Edman sequencing on the collected peptides was carried out using an Applied Biosystems Procise system employing fast cycle Edman sequencing on the collected peptides was carried out using an Applied Biosystems Procise system employing fast cycle Edman sequencing on the collected peptides was carried out using an Applied Biosystems Procise system employing fast cycle Edman sequencing on the collected peptides was carried out using an Applied Biosystems Procise system employing fast cycle Edman sequencing on the collected peptides was carried out using an Applied Biosystems Procise system employing fast cycle Edman sequencing on the collected peptides was carried out using an Applied Biosystems Procise system employing fast cycle

RESULTS
The activity of region C is promoter dependent
Fig. 1A shows a time-course of the expression of the region C-

on the Hoxb4 promoter at the spinal cord/hindbrain boundary (Whiting et al., 1991). Notably, staining within the somitic and flank mesoderm corresponded to the anterior boundary of Hoxb4 expression at the level of So 6/7. However, this boundary was not maintained and by 9.5 dpc had regressed to So 13/14 (Fig. 1A, part b). This corresponds to the transition between cervical and thoracic regions, raising the possibility that different regulatory mechanisms might operate in these domains. An equivalent shift was also seen within the flank mesoderm and in the neural tube but there was no further change in this pattern until at least 12.5 dpc (data not shown).

In the central nervous system (CNS) of a 12 dpc embryo, expression was localized to the dorsal neural tube, floor plate and ventral roots of the spinal nerves, up to the first cervical nerve (Fig. 1A, parts c-e). Strong staining was evident in the dorsal root ganglia (drg) and sympathetic ganglia (sg), both neural crest derivatives.

These results demonstrate that region C, in conjunction with the hsp68 promoter, can establish the appropriate anterior boundary of Hoxb4 expression at So 6/7 but is unable to maintain it beyond 8.5 dpc. This is in contrast to constructs containing the Hoxb4 promoter (compare Fig. 1A, parts c and f), indicating that specific enhancer-promoter interactions are important in determining the boundary of somitic expression. In addition, CHZ does not display the normal graded distribution of Hoxb4 expression that is seen with larger Hoxb4 promoter constructs (Whiting et al., 1991). Strong posterior expression is evident with CHZ from the earliest stages examined.

Important regulatory elements are located in the 5'-half of CR1
We have previously shown that region C contains a short stretch of intronic sequence highly conserved between mouse and Fugu Hoxb4 and named it CR1 (Conserved Region 1) (Aparicio et al., 1995). CR1 is crucial for region C activity in transgenic mice carrying CHZ (Fig. 1A, parts j-l) or a similar chicken Hoxb4 transgene (Morrison et al., 1995). We generated multiple sequence alignments of all available paralogous group 4 (PG4) Hox gene introns from different species (Morrison et al., 1995; Haerry and Gehring, 1997; Kim et al., 2000). A CR1-like region was identifiable within the intron and close to the 5'-splice site in all of the sequences examined (Fig. 1B). The greatest degree of overall sequence identity to mouse was seen with larger Hoxb4 sequences (average=68%). Hoxa4 and Hoxc4 sequences showed an intermediate identity and Hoxd4/AmphiHox4 sequences the least. All sequences shared greatest identity over a 28 bp region corresponding to bp 573-600 of region C, previously defined as the HB-1 element. This contains several consensus homeodomain-binding sites and is able to respond to Hox family proteins in Drosophila (Haerry and Gehring, 1997; Keegan et al., 1997). Several other conserved motifs (I-IV) are also evident from this alignment.

To investigate which of the CR1 sequences are required for regulation we deleted the most conserved region including the HB1 element and motif IV from CHZ (CHZΔ599-599; Fig. 1C). In transient Fa embryos between 11.5 and 12.5 dpc staining in the flank mesoderm was absent and the anterior boundary in the neural tube was less distinct and appeared to be shifted caudally by 1 or 2 segments (Fig. 1A, parts h and i). However, somitic staining was equivalent to that of CHZ. As
detecting the two specific shifts, while m1 and m2 did not (Fig. 2B, compare lanes 5 and 8-11). This shows that the sequence TCGCCATT (mutated in m1/m2) is required for binding.

This motif is similar to the HoxTF/YY1 site (TGGCCATT) that we have characterized in the Hoxb4 promoter (Gutman et al., 1994). Specific mutational analysis demonstrated that HoxTF is essential for the transcriptional activity of a minimal Hoxb4 promoter construct in transfected cells. In EMSAs, the promoter site (HoxPwt) efficiently competed for the two complexes formed by b4Cwt (Fig. 2C, lane 2). Furthermore, HoxTF sites from the myogenin and MyoD1 genes were also able to compete effectively for HoxTF but not YY1 binding (Fig. 2C, lanes 4 and 5) (Gutman et al., 1994). By contrast, a mutated MyoD1 site failed to compete (Fig. 2C, lane 6). These results show that the CR1 motif is comparable with the site in the Hoxb4 promoter and is able to bind both HoxTF and YY1 in a non-cooperative manner.

Differences in the relative affinities of HoxTF and YY1 for the different sites are evident from these experiments and were investigated further by direct binding assays. Each of the Hoxb4 probes produced YY1 shifts of comparable intensities, indicating that it binds with a similar affinity. However, the HoxTF shift with b4Cwt was five- to 10-fold weaker than that observed for HoxPwt (Fig. 2D, lanes 1-2). YY1 had the lowest affinity for the MyoD1 site (Fig. 2D, lane 5), which competed only weakly for YY1 binding to b4Cwt (Fig. 2C, lane 5). However, this site had the highest affinity for HoxTF and for this reason we subsequently used it to detect HoxTF.

To define the polypeptide factor(s) that constitutes HoxTF, we performed EMSAs with renatured proteins that had been size fractionated by SDS-PAGE. A shift corresponding to HoxTF was evident when the 31-51 kDa fraction was used (Fig. 3A). This shift was abolished by competition with an excess of unlabelled probe, but not of a mutated probe, indicating that binding was sequence specific (Fig. 3A, lane 10-12). A retarded band was also observed with the 98-116 kDa fraction (lane 6) but was nonspecific as it was competed for by both wild-type and mutated probes (data not shown). The 59-66 kDa fraction produced a shift corresponding to YY1 (molecular mass=65 kDa) (Shi et al., 1997).

NFY is the HoxTF binding activity
HoxTF was purified from F9 EC extracts using wheat germ agglutinin and sequence-specific DNA affinity chromatography (Fig. 3B). Three major bands (36, 45 and 48 kDa) were observed when purified proteins were analysed by SDS-PAGE. Peptide sequencing of the 36, 45 and 48 kDa species identified them as the A-subunit of NFY, a transcription factor composed of three subunits (NFY A to NFYC), all of which are identified them as the A-subunit of NFY, a transcription factor composed of three subunits (NFY A to NFYC), all of which are NFY A to NFYC, all of which are bound to DNA. The 36 kDa band was shown to be NFYB by western blotting. In the absence of an NFY A-specific antibody we could not determine the relative molecular weight. In the absence of an NFY A-specific antibody we could not investigate this point further.

We confirmed the binding of NFY to the b4Cwt site by

CHZΔ559-599 did not exhibit the dramatic loss of activity that we have previously observed with a larger deletion of CR1 (CHZΔ515-607; Fig. 1A, parts j-l), we focused on the presence of positive regulatory sequences located between bp 515-558.

HoxTF and YY1 bind to overlapping sites in CR1
To search for transcription factor-binding sites in CR1 we conducted DNA-electrophoretic mobility shift assays (EMSAs) with a series of oligonucleotide probes. Under various reaction conditions we did not detect any specific binding with probes that represented the 3’ sequences of CR1, including the HB1 element (bp 563-617; data not shown). However two discrete complexes were detected with a probe representing bp 509-568 (Fig. 2A,B, lanes 1-4). We further localized the sequences required for binding to bp 525-551 (b4Cwt; Fig. 2A,B, lanes 5-7) and mutated this region (b4Cm1-4, Fig. 2A). Probes m3 and m4 behaved similarly to b4Cwt

Fig. 1. (A) Expression patterns of region C constructs. (a-e) CHZ. (l-g) Hoxb4 promoter-region C construct. (h-l) CR1 deletions. (A) Lateral and dorsal views of a 10 somite (So) stage embryo, showing anterior boundaries of expression: neural tube adjacent to So 4 (yellow arrow); somitic and flank mesoderm at So 6/7 (black arrow). Typical of Hox gene expression domains, staining is weaker in So 7 (light blue triangle) than in So 8 (dark blue). (b) 9.5 dpc: somitic boundary regresses to So 13/14 (triangles); neural boundary regresses to So 6/7 (yellow arrow). (c) 10.5dpc: staining in spinal ganglia up to the first cervical nerve (cn1) and ventral neural tube, extending anteriorly (open triangle). (e) Transverse section (TS), forelimb; limb bud of the left forelimb. (f) 14.5dpc embryo (Hoxb4 promoter, construct 5) (Whiting et al., 1991): note strong anterior domains of expression in the neural tube up to the spinal cord/hindbrain boundary (yellow arrow) and in the somitic mesoderm between the anterior So 6/7 boundary and So14 (white arrows). (g) TS forelimb level of the same embryo: note strong staining in sclerotomal derivatives (open triangles) and the dorsal aorta (d). (h,i) Lateral and dorsal views of a 12 dpc embryo (CHZΔ559-599): boundaries of expression in the neural tube (yellow arrow) and in the somitic mesoderm (blue triangle) are indicated. (j,k) Lateral and dorsal views of a 12dpc embryo (CHZΔ515-607): consistent expression is restricted to a domain in the ventral neural tube (vnt). (l) TS thoracic level of the same embryo. Scale bars: 100μm. (B) Sequence alignment showing a comparison between CR1 of the mouse Hoxb4 intron and those of other paralogous group 4 Hox genes, identical bases are highlighted in black. Numbering is with respect to that of region C (bp 1 is the first base of the Sa/I site in exon 1, +321 of Hoxb4). The number of base pairs (bp) in each aligned sequence is shown on the right. The extents of the two deletions in constructs CHZΔ559-599 and CHZΔ515-607 are marked with blue lines and the boundaries of a possible cis-positive regulatory element (bp 515-558), identified by these deletions, are marked by red triangles. The margins of the 28bp HB-1 element (bp 574-601) are shown (black triangles). The locations of four conserved motifs are marked below (I-IV). (C) Schematic diagram of the transgenes in A. The hsp68 promoter-lacZ reporter, which is common to each construct, is not shown to scale. Exons are shaded grey. CR1 is represented by a black rectangle within the intron (white rectangle) flanked by MmuI and SfI restriction sites. CHZΔ515-607 carries a deletion of the entire CR1 region (Aparicio et al., 1995). An example is shown for comparison. Exp. # denotes the total number of independent transgenic F0 embryos and lines generated with each construct giving a consistent pattern of expression.
EMSA competition and super-shift experiments (Fig. 3C). A known NFY binding site from the MHC class II Eα gene (Dorn et al., 1987) competed efficiently for the binding of HoxTF to b4Cwt without affecting the binding of YY1. A site bearing a mutation that impairs the binding of NFY failed to compete (Fig. 3C, lanes 2 and 3). A consensus binding site for the CCAA T/enhancer binding protein (C/EBP), which is also able to bind to CCAA T box containing sequences, though unrelated to NFY (Mahoney et al., 1992), also failed to compete (Fig. 3C, lane 4). Finally, the retarded band observed with b4Cwt was completely super-shifted with an anti-NFY A antibody (Fig. 3C, lane 6), verifying that NFY is the HoxTF binding activity. We also noted the presence of an additional site located within the 3¢-half of the intron (GCCATGG; 944-952 of region C). NFY and YY1 were also able to bind to CCAAT box containing sequences, though unrelated to NFY (Mahoney et al., 1992), also failed to compete (Fig. 3C, lane 4). Finally, the retarded band observed with b4Cwt was completely super-shifted with an anti-NFYA antibody (Fig. 3C, lane 6), verifying that NFY is the HoxTF binding activity. We also noted the presence of an additional site located within the 3′-half of the intron (GCCATTGG; 944-952 of region C). NFY and YY1 were also able to bind to this site, although the affinity of NFY for it is three- to fivefold less than for b4Cwt (data not shown; Fig. 6A).

The NFY-binding site is essential for enhancer activity in vivo

Based on our previous analysis of the Hoxb4 promoter site (Gutman et al., 1994) we engineered specific mutations able to impair the binding of NFY and YY1. Because the NFY-specific mutation had 15-20% residual binding activity (b4CmNFY1, data not shown), we designed a second mutation (b4CmNFY2) that completely abolished NFY binding. We tested the capacity of oligonucleotides carrying the various mutations to compete for the binding of NFY and YY1 to b4Cwt (Fig. 4A). The double mutant (b4CmN+Y, same as b4Cm2, Fig. 2A) failed to compete for the binding of either NFY or YY1 (Fig. 4A, lane 5). The b4CmYY1 mutant competed for NFY binding but not for that of YY1 (Fig. 4A, lane 3), while the b4CmNFY2 mutant competed for the binding of YY1 but not for that of NFY (Fig. 4A, lane 4).

We introduced each mutation into CHZ and analysed the patterns of expression in transgenic embryos (Fig. 4B). The double NFY/YY1 mutation (b4C-mN+Y) had a drastic effect on the expression of the transgene. At 12.5 dpc, all of the mesodermal components of CHZ expression were consistently absent and expression within neural tissues was greatly reduced (Fig. 4B, parts a and b). However, the b4CmN+Y mutant construct exhibits residual expression within the nervous system that may indicate a positive contribution by other elements within CR1, or that the b4C-mN+Y mutation retains some level of binding indiscernible by our EMSAs. Interestingly, of the 3 F0 embryos obtained at 9.5-10.5 dpc, all displayed much stronger relative levels of expression than at later stages, including staining in the paraxial and flank mesoderm that was spatially equivalent to the wild-type construct, though weaker (Fig. 4B, parts c and d). This suggests that there is both an early requirement for
the NFY/YY1 site to achieve the proper level of enhancer-mediated activation, and a later requirement to maintain expression within the mesoderm and nervous system.

Specific mutation of the NFY site (b4C-mNFY2) resulted in a staining pattern that was similar to that of the double mutation (Fig. 4B, parts g and h). Staining was detected at a low level in the neural tube, ventral roots, dorsal root ganglia and sympathetic ganglia. However, there was an absence of mesodermal activity and a reduction in the level of neural staining compared to mN+Y and this was mosaic in all 10 embryos. Interestingly, the b4C-mNFY1 construct produced a pattern of activity that was intermediate between those of CHZ and mNFY2 (Fig. 4B, parts j-l). Staining was present in the same tissues but was not mosaic.

The YY1 mutant construct (b4C-mYY1) displayed a reduced level of staining in the somites, relative to the neural tube, and a caudal shift in the anterior boundary of strong expression, to the level of the hind limb (Fig. 4B, parts e and f). Interestingly, staining within the flank mesoderm appeared normal. In several embryos (4/7), neuroectodermal staining extended to regions more anterior than the normal extent of region C activity. While this shows that the b4C-mYY1 mutation deregulates the activity of region C, this single bp change also causes a slight reduction in NFY affinity (Fig. 4A, lane 3). Hence the observed changes in expression may result from either the loss of YY1 binding or a subtle effect on the level of NFY binding.

These results demonstrate that the NFY/YY1 site in CR1 is important for region C enhancer activity in both mesodermal and neural domains and that positive regulation is largely mediated through the binding of NFY. Owing to the overlapping nature of the NFY/YY1 motif and the difficulty of identifying YY1-specific mutations that do not also interfere with NFY binding, we are unable to definitively assign any function to the YY1 interaction.

**A single NFY site is not sufficient to confer spatially specific expression on a heterologous promoter**

To address whether NFY is sufficient to direct tissue-specific expression we tested two versions of the NFY binding site (incorporating the YY1 mutation) on the hsp68 promoter. The staining patterns obtained with a construct containing a single copy of the b4CmYY1 sequence (bp 525-551) were essentially random (seven cases) and indicative of nonspecific integration-site effects on the transgene. With a longer version of the NFY site (bp 511-558 of region C), which also included conserved elements I and III of CR1 (Fig. 1B), weak but consistent expression was detected in a subset of the spinal and cranial ganglia and the neural tube at 11.5 dpc (construct b4C-511-558-mYY1. Fig. 5A-C). These results show that while a single NFY-binding site is unable to recapitulate any aspect of region C activity, the inclusion of flanking sequences can lead to reproducible expression. This suggests that NFY may cooperate with factors that bind in close proximity to it.

**DISCUSSION**

**The role of NFY in Hoxb4 regulation**

In this study, we demonstrate that our combined approach of transgenesis and biochemistry is capable of identifying hitherto
Regulation of Hoxb4 by NFY

Fig. 4. In vivo requirement for the NFY site in CR1. (A) EMSA competition experiments showing the ability of specific mutations to interfere with the binding of YY1 and NF-Y to b4Cwet. On the right is a schematic of the various mutations and reporter constructs. Mutated nucleotides are underlined. The binding characteristics of the probes are summarized on the right (b4Cwet, wild type; b4C-mN+Y, double mutation; b4C-mYY1, YY1 specific mutation; b4C-mNFY1 and b4C-mNFY2, NFY specific mutations; +, binding; +/-, partial binding; -, no binding). (B) Transgenic mouse embryos stained for β-galactosidase activity showing the expression patterns derived from the mutant NF-Y/YY1 constructs. (a) Lateral and (b) dorsal views of a 12.5dpc embryo carrying construct CHZ-mN+Y. Residual staining was consistently observed in nervous system (open triangles). (c,d) Two different 9.5-10dpc embryos carrying the same construct. Weak somitic expression is visible at the level of So 13/14 (blue triangle in c) or in the most caudal somites (blue arrow in d), as is weak expression in the flank mesoderm (red arrow in d). (e) Lateral and (f) dorsal views of an 11.5dpc embryo carrying construct CHZ-mYY1. Black arrows indicate ectopic neural expression and blue arrowhead the anterior limit of somitic expression. Flank mesoderm staining is unaffected (red arrow). (g) Lateral and (h) dorsal view of a 12.5 dpc embryo carrying construct CHZ-mNFY2. (i) TS at the forelimb level of a similar embryo. (j) Lateral and (k) dorsal views of similar 12.5 dpc embryos carrying construct CHZ-mNFY1. (l) TS at the forelimb level of a similar embryo. drg, dorsal root ganglion; v, ventral root; f, floorplate; sg, sympathetic ganglia. Scale bars: 100 μm.

The role of YY1 in Hoxb4 regulation

Mutation of the YY1-binding site in CR1 results in both a reduction in the level of somitic expression. As this is in contrast to constructs containing the Hoxb4 promoter (Whiting et al., 1991), specific enhancer-promoter interactions are implicated in maintaining the boundary of Hoxb4 expression after 8.5 dpc. It is an intriguing possibility that the NFY/YY1-binding sites might play a key role in mediating such interactions.

A wide variety of eukaryotic genes have been shown to contain NFY-binding CCAAT boxes in their promoters, although Hox genes have previously been cited as a large gene family that do not (Mantovani, 1999). Interestingly, the Hoxb4 NFY sites are atypical for several reasons. Known NFY-binding CCAAT boxes are almost exclusively located upstream of the transcriptional start site in proximal promoter regions. The three Hoxb4 sites are all positioned downstream of the two major transcriptional start sites of the gene and the intron sites are distally located. Furthermore, the presence of an overlapping YY1 site is novel.

NFY could regulate transcription of Hoxb4 in various ways. It is known to stabilize the binding of other proteins to regulatory elements close to the CCAAT box and to interact directly with other transcription factors (Mantovani, 1999). These properties of NFY could be important for the recruitment of additional proteins to region C and/or the Hoxb4 promoter in order to establish complexes that are capable of activating transcription. It can also interact with proteins of the general transcriptional machinery (Frontini et al., 2002), which could be important for initiation of transcription at the Hoxb4 promoter or for facilitating enhancer/promoter interactions. In addition, NFY activates transcription by modifying local chromatin architecture via two distinct mechanisms. First, NFY binding has been shown to reposition nucleosomes, thus resetting the promoter for activation of transcription (Li et al., 1998). Secondly, it is able to indirectly alter the chromatin environment by recruiting transcriptional coactivators that possess histone acetyl transferase (HAT) activity (Currie, 1998; Jin and Scotto, 1998). Either or both of these properties could be essential to the role of NFY in regulating Hoxb4 expression via the promoter or region C.

The role of YY1 in Hoxb4 regulation

Mutation of the YY1-binding site in CR1 results in both a reduction in the level of somitic expression, and a posteriorization of the anterior somitic boundary. However,
interpretation of these effects is complicated by the slight reduction in NFY binding that accompanies the b4C-mYY1 mutation. It remains possible that this, rather than the loss of YY1 binding, is responsible for the observed effects in expression. However, the potential role of YY1 in Hoxb4 regulation is intriguing. YY1 is a multifunctional protein that can act as an activator, repressor or initiator of transcription (Shi et al., 1997), but it is of particular interest to us that the DNA-binding domain of YY1 is structurally similar to that of the protein encoded by the Drosophila gene pleiohomeotic (pho) (Brown et al., 1998). pho is a member of the Polycomb-Group (PcG) of genes that are required to maintain the transcriptionally inactive state of Hox genes in appropriate regions of the embryo (Pirrotta, 1998). It has recently been shown that YY1 interacts with EED, a vertebrate homologue of the fly PcG protein Extra Sex Combs (ESC) (Satijn et al., 2001), which itself associates with another PcG protein EZH2, related to Drosophila Enhancer of Zeste (E(Z)) (van Lohuizen et al., 1998), and with histone deacetylases (HDACs) (van der Vlag and Otte, 1999). The Drosophila ESC and E(Z) proteins are also components of a complex that contains an HDAC (Tie et al., 2001), suggesting that YY1 may be part of an important complex involved in transcriptional repression. Such a complex may act during the early stages of embryogenesis as mice homozygous for a null allele of YY1 die at implantation (Donohoe et al., 1999) and embryos lacking functional Eed or Ezh2 do not survive beyond gastrulation (Faust et al., 1995; O’Carroll et al., 2001). Furthermore, mutation of another mouse PcG gene M33 results in early activation of Hoxd11 (Bel-Vialar et al., 2000). A similar effect is seen for Hoxd4 and Hoxd10 expression when the global repression of the Hoxd cluster is disrupted by targeted genomic deletions (Kondo and Duboule, 1999). In both of these cases expression at later stages appears normal. It is possible that YY1 recruits a repressive complex including PcG proteins and HDACs to the Hoxb4 locus to prevent early expression of this gene.

The overlapping NFY/YY1 site
YY1 and NFY do not bind cooperatively to the Hoxb4 sites. The CR1 NFY site contains a degenerate CCAAT box on each strand: on the sense strand with the core sequence CCAAT; and in the reverse orientation with the core sequence gCAAT. The other Hoxb4 sites both contain a CCAAT motif on the sense strand and a perfect CCAAT motif in the reverse orientation. As the mutant YY1 site (sense=CCAgT) is still able to bind NFY efficiently, it is likely that NFY binds to the site in the reverse orientation and YY1 to the CCAAT motif on the sense strand. However, as mutations that alter the core of the reverse site do not prevent NFY binding (b4Cm3 and b4CmNFY1; Fig. 2A,B and Fig. 4A), it seems likely that NFY can also bind to the forward sites at some level.

**Fig. 6.** Comparison of NFY/YY1 sites. (A) An alignment of NFY/YY1 binding sites from the Hoxb4 intron and promoter. The sequence of the forward strand is shown above aligned with the consensus for YY1 (Hyde-DeRuyscher et al., 1995). The sequence of the reverse strand is shown below aligned with the consensus for NFY (Mantovani, 1998). In each case, the core nucleotides of the site are shown in bold. Nucleotides that are unfavourable for binding based upon either consensus are represented in lower case. Variant nucleotides within the core sequences that are thought to abolish YY1 or NFY binding are shown in red. (B) The consensus sequence (forward strand) for the Hoxb4 NFY/YY1 motif. Shown below is a representation of the CR1 NFY/YY1 site. Nucleotides over which YY1 makes base-specific contacts in the major groove are highlighted on the upper strand. Similarly, nucleotides over which NFY makes base-specific contacts in the minor and major grooves are shown on the lower strand. Nucleotides that are predicted to enhance (green) or reduce (red) the affinity of either factor for its site are shown on the appropriate strand. Core nucleotides are shown in bold. (C) Predicted NFY sites from the Hox8 early enhancer and the Hox8/Hox7-Hox6 four cluster sequence (H8/7-6 FCS). Sites are shown in the reverse orientation with respect to transcription. The acquisition of alternative CCAAT boxes in the mouse and human HoxC cluster is underlined. Bold, lowercase and red nucleotides have equivalent meanings to those in A.
YY1 makes base specific contacts in the major groove (Hyde-DeRuysscher et al., 1995) while NYF makes key major groove interactions with the –2 to +2 bases and minor groove contacts over the +3 to +7 region (Ronchi et al., 1995). Based on the sequence of the CR1 site, it seems highly unlikely that both NYF and YY1 can bind simultaneously because of the requirement for common major groove interactions at positions 10/11 of the YY1 site and +1/+2 of the NYF site (Fig. 6B). By analysis of base preferences, it appears that the NYF/YY1 binding sites are highly interdependent (Fig. 6B). In the overlapping region, bases that are predicted to reduce the affinity of NYF interactions (G at +1, C at +8 and G at +9) are favoured by YY1. The converse is also true, as highlighted by the suboptimal G at position 10 of the YY1 site, and similar relationships exist for all three Hoxb4 sites (data not shown). It would appear, therefore, that the NYF/YY1 site is a specialized motif that is able to bind either factor in a mutually exclusive fashion.

The NYF/YY1 motif and the global regulation of Hox genes

We sought to find other examples where NYF sites of this class are conserved amongst Hox family members and uncovered a compelling example. Kim et al. (Kim et al., 2000) describe a highly conserved region in the intergenic region between PG-8/7 and PG-6 (H8/7-6 FCS, for Four Cluster Sequence) in all four vertebrate Hox clusters. We noted the presence of a highly conserved NYF site in the same transcriptional orientation as the Hoxb4 sites (Fig. 6C). This strongly suggests that the NYF motif is, in general, an important regulatory feature of Hox genes. In the mouse and human Hoxc clusters the conserved CCAAT box is altered (CCAAAT to CCAAAg), which should abolish, or drastically impair, NYF binding. However, immediately adjacent to this is an alternative CCAAT box that is not present in any of the other clusters, including Fugu HoxC. This suggests that the mammalian HoxC cluster has acquired this variation in the H8/7-6 FCS NYF site after its ancestral divergence from Fugu, approximately 430 million years ago, and argues that conservation of a CCAAT box at this position is important.

The mouse and chicken Hoxc8 genes show a heterochronicity in their activation along the rostrocaudal axis. This correlates with a transposition of Hox gene boundaries in conjunction with morphological ones (Burke et al., 1995). The anterior boundaries of Hoxc8 expression in mouse are at the level of So10 and So14 in the neural tube (nt) and paraxial mesoderm (pm), respectively (Belting et al., 1998). In a chick embryo at the same relative developmental stage, they lie at So14 (nt) and So20 (pm). A 399 bp regulatory region located 3 kb upstream of the mouse Hoxc8 gene (the Early Enhancer) directs the early phase of Hoxc8 expression in transgenic mice (Shashikant and Ruddle, 1996). When coupled to the hsp68 promoter this enhancer, like region C, is able to direct expression in the neural tube and somites of transgenic embryos to levels that are several segments more posterior than the endogenous gene at 9.5 dpc. The replacement of a highly conserved segment with the corresponding chicken sequences (151 bp, 80% identity) results in a caudal shift of the anterior boundaries of enhancer activity by several further segmental units. In spite of this there appeared to be no significant difference in the sequences of five elements that have been shown to be important for the activity of the mouse enhancer (Shashikant et al., 1995; Shashikant and Ruddle, 1996). We have noticed, however, that the mouse enhancer sequence contains a consensus NYF-binding site containing a CCAAT motif that is mutated in the chick enhancer to CTAT (Fig. 6C).

We propose that this subtle base alteration that removes an NYF binding site may provide an explanation for the caudal shift in expression.

In conclusion, we have shown that NYF is required for the regulated expression of Hoxb4 in the mesoderm by binding to a site that is a common feature of multiple Hox gene regulatory regions. NYF is thus a newly identified upstream Hox gene regulatory factor. We suggest that NYF could act to stabilize and enhance the activating effects of other cell-type specific transcriptional regulators and maintain this state by the recruitment of chromatin modifying enzymes. The relative levels of NYF and YY1 binding could represent a mechanism for balancing activation and repression through the same site. This is intriguing in light of the recent discovery that the proper timing of Hoxd gene activation is controlled by a release from global repression of the entire cluster (Kondo and Duboule, 1999; Kmita et al., 2000b).

We thank Alistair Sterling for assistance with peptide sequencing, Christoph Benoist (IGBMC, Strasbourg) for the NYF and NYF antibodies, and Alistair Morrison/Robb Krumlauf (NIMR, Mill Hill) for the region C-Hoxb4 promoter construct transgenic embryos. We are grateful to Hannah Boyes, Lorraine Jones and Zoë Webster for animal husbandry, and to members of the Rigby and Krumlauf laboratories for helpful discussion. J. G. and T. B. were supported by MRC Studentships, and M. V. by a Fellowship from the International Human Frontier Science Program. This work was funded by the Medical Research Council.

REFERENCES


Regulation of Hoxb4 by NFY


