The EMBO Journal vol.11 no.1 pp.265-277, 1992

The upstream region of the human homeobox gene *HOX3D* is a target for regulation by retinoic acid and HOX homeoproteins

Laura Arcioni, Antonio Simeone[1], Stefania Guazzi, Vincenzo Zappavigna, Edoardo Boncinelli[1] and Fulvio Mavilio

Department of Biology and Biotechnology, Istituto Scientifico San Raffaele, 20132 Milano and [1]International Institute of Genetics and Biophysics, CNR, 80125 Napoli, Italy

Communicated by E.Boncinelli

We studied the structure, regulation and expression of *HOX3D*, a human homeobox gene located in the HOX3 cluster on chromosome 12. *HOX3D* is developmentally regulated during embryogenesis and is activated by retinoic acid (RA) in cultured embryonal carcinoma (EC) cells. Transfection of *HOX3D* upstream genomic sequences linked to a reporter gene allowed the functional definition of its promoter, containing a canonical TATA element. This promoter directs the expression of the reporter gene in EC cells after induction with RA, and binds RA-induced nuclear factor(s) through a conserved palindromic sequence located ~100 bp upstream of the transcription start site. The *HOX3D* promoter is transactivated in both human and murine cells when cotransfected with vectors expressing the protein product of the upstream gene *HOX3C* and the paralogs of further upstream genes in the HOX4 cluster (i.e. *HOX4D*, *HOX4C* and the murine *Hox 4.3*). The HOX3D protein, and those encoded by the downstream gene *HOX3E* and its paralog *HOX4B* are instead inactive. *HOX4C* and *HOX4D* proteins synthesized in bacteria bind to the same conserved sequence located around position −120, as well as to the TATA box and immediately upstream and downstream nucleotides. These data provide evidence that cross-regulatory interactions between mammalian homeogenes take place in cultured cells, thus raising the possibility that a regulatory network may exist *in vivo*. The sequences on the *HOX3D* promoter involved in cross-regulation are different from those binding nuclear factors induced by RA.

Key words: homeobox genes/HOX3/retinoic acid

Introduction

The homeodomain is a highly conserved DNA-binding protein sequence encoded by the homeobox, a 183 bp sequence contained in a number of genes involved in pattern formation in *Drosophila melanogaster* (Akam, 1987; Ingham, 1988; Gehring *et al.*, 1990). Several families of homeobox-containing genes (homeogenes) have been identified in vertebrates (Scott *et al.*, 1989). Class I homeogenes are defined on the basis of sequence homology with the homeodomain of the archetypal *Drosophila Antennapedia* gene. In both mouse and human genomes, class I homeogenes are arranged in four homologous clusters located on different chromosomes (Acampora *et al.*, 1989; Duboule and Dollé, 1989; Graham *et al.*, 1989). Genes sharing the highest identity in their homeodomain sequences are further grouped in sub-families, or paralogy groups, each showing significant homology with a *Drosophila* prototype gene from either the Antennapedia or the Bithorax complex (Gehring and Hiromi, 1986). Paralogous genes occupy homologous positions along their respective clusters, although not all paralogs are represented in each cluster (reviewed in Kessel and Gruss, 1990; Boncinelli *et al.*, 1991).

Class I homeogenes are expressed and developmentally regulated in mouse (reviewed in Holland and Hogan, 1988; Kessel and Gruss, 1990), *Xenopus* (reviewed in Wright *et al.*, 1989) chicken (Izpisùa-Belmonte *et al.*, 1991; Nohno *et al.*, 1991) and human (Simeone *et al.*, 1986, 1987; Mavilio *et al.*, 1986; Giampaolo *et al.*, 1989) embryogenesis. In mouse embryos, *Hox 1*, *2*, *3* and *4* genes are expressed in partially overlapping, spatially restricted domains along the anteroposterior (A−P) axis of developing axial systems such as the axis skeleton and the central nervous system (CNS) (Kessel and Gruss, 1990). Position of the anterior borders of expression domains along the CNS A−P axis are colinear with positions of genes along the cluster, invariably following a 3′-anterior/5′-posterior rule (Gaunt *et al.*,1988; Akam, 1989; Kessel and Gruss, 1990). At least for the HOX2 cluster, data obtained in human embryos follow the same general concept (Giampaolo *et al.*, 1989). The colinearity rule also applies to the temporal and spatial expression of *Hox 4* genes along the A−P axis of mouse and chicken embryonic limbs (Dollé *et al.*, 1989; Izpisùa-Belmonte *et al.*, 1991; Nohno *et al.*, 1991). Perturbation of axial polarity by growth factors or retinoic acid (RA) induces parallel changes in homeogene expression patterns (Ruiz i Altaba and Melton, 1989; Cho and De Robertis, 1990; Oliver *et al.*, 1990; Izpisùa-Belmonte *et al.*, 1991; Nohno *et al.*, 1991). On the other hand, abnormal expression of homeogenes causes alteration in the specification of structures along A−P axes (Wright *et al.*, 1989; Kessel *et al.*, 1990; Cho *et al.*, 1991). These data strongly suggest that homeogenes might transduce positional information in vertebrate development by specifying positional values along the various A−P body axes (reviewed in De Robertis *et al.*, 1989; Kessel and Gruss, 1990).

The factors governing establishment and maintenance of homeogene expression patterns in vertebrates are still largely unknown. Endogenous molecules such as peptide growth factors and retinoids are instrumental in instructing axial polarity (Brokes, 1989; Melton, 1991), and are therefore natural candidates for playing a major role also in regulation of homeogene expression. Indeed, expression of homeogenes of the four clusters is differentially regulated by RA in human embryonal carcinoma (EC) cells in a concentration- and time-dependent fashion (Simeone *et al.*, 1990, 1991). This suggests that homeogene expression patterning *in vivo* may

© Oxford University Press

be at least in part directed by temporal and/or spatial modulation of the levels of endogenous retinoids. In *Drosophila* embryos, expression patterning is also achieved through multiple auto- and cross-regulatory interactions between different genes. These interactions involve transcriptional activation and repression of homeogene promoters by homeoproteins, both in embryos and cell culture (reviewed by Hayashi and Scott, 1990). Cross-regulatory interactions between homeogenes might be important in regulating homeogene expression also in vertebrate embryos, e.g. by setting correct borders of expression domains. Although some mammalian homeoproteins are known to bind DNA with sequence specificity (Fainsod *et al.*, 1986; Odenwald *et al.*, 1989) or to surrogate cognate gene function when expressed in *Drosophila* embryos (Malicki *et al.*, 1990; McGinnis *et al.*, 1990), their function as transcriptional regulators in a mammalian system is so far unproven.

We have studied a 4.5 kb human genomic region containing the *HOX3D* gene and upstream sequences, and defined its TATA-dependent promoter. This promoter directs the expression of a reporter gene in RA-induced EC cells, where it specifically binds nuclear proteins which are not present in uninduced cells. The *HOX3D* promoter is transactivated when cotransfected with vectors expressing the protein product of the upstream gene *HOX3C* and the paralogs of further upstream genes in the HOX4 cluster (i.e. *HOX4D*, *HOX4C* and *Hox4.3*), whereas the HOX3D protein, and those encoded by the downstream gene *HOX3E* and its paralog *HOX4B*, are inactive. These data indicate that a human *HOX* gene promoter is a potential target for multiple regulatory mechanisms, including activation by retinoids and cross-regulation by homeoproteins.

Results

Structure and expression of the human HOX3D gene

The human HOX3 cluster contains at least nine genes in a ~110 kb region on chromosome 12 (Acampora *et al.*, 1989). A 5 kb *Eco*RI−*Bgl*II genomic fragment containing the *HOX3D* homeobox (Figure 1) was subcloned from a human genomic cosmid library and entirely sequenced (Figure 2). A number of cDNA clones obtained from different human cDNA libraries (full-term placenta, embryonic spinal cord and RA-induced EC cells) were also sequenced to determine intron−exon boundaries and putative sequence of the HOX3D protein. Two families of *HOX3D* cDNAs were isolated, differing for alternative splicing of two different first exons to a common second exon containing the homeobox (Figure 1). The first transcript, represented by clone cp11T (Figure 1), encodes in its longest ORF a 222 amino acid protein containing the homeodomain in its most C-terminal portion (Figure 2). The conserved IYPWM pentapeptide (Mavilio *et al.*, 1986) is also present seven amino acids before the homeodomain, encoded by the first exon. A canonical polyadenylation signal is present 22 nucleotides before the poly(A) tail of cp11T, after ~0.9 kb of 3′ non-coding sequence (Figure 2). Analysis of the homeodomain sequence allows the classification of *HOX3D* in paralogy group 9, which also contains the human *HOX1C* (=murine *Hox 1.3*) and *HOX2A* (=*Hox 2.1*) (Boncinelli *et al.*, 1991). The mouse cognate of *HOX3D* is *Hox 3.4*. RNase protection and primer extension experiments using RNA from both human embryos and EC cells (not shown) allowed mapping of a major transcription start site 30 nucleotides downstream of a canonical TATA box identified in genomic DNA, and two nucleotides upstream of the 5′ end of the longest sequenced cp11T cDNA clone (see Figure 2).

The second family of *HOX3D* transcripts, represented by clone cp11 (Figure 1) contains an alternative exon ~0.6 kb in length, located on genomic DNA ~16 kb upstream of the *HOX3D* homeobox, in the intergenic region between *HOX3A* and *HOX3C*. The sequence of this exon, which is also spliced to form alternative transcripts of *HOX3C* and *HOX3E* genes, has been previously reported (Simeone *et al.*, 1988). Apparently no protein product is encoded by cp11, since a number of stop codons close the reading frame immediately upstream of the homeobox. The 3′ non-coding sequence and polyadenylation signal are the same as in cp11T (Figure 1).

The *HOX3D* gene is expressed in human embryonic spinal cord and placenta as a heterogeneous, 1.7 kb RNA band as assayed by Northern blotting (Figure 3A). This band contains transcripts of both cp11 and cp11T type, which have approximately the same size and can be distinguished only

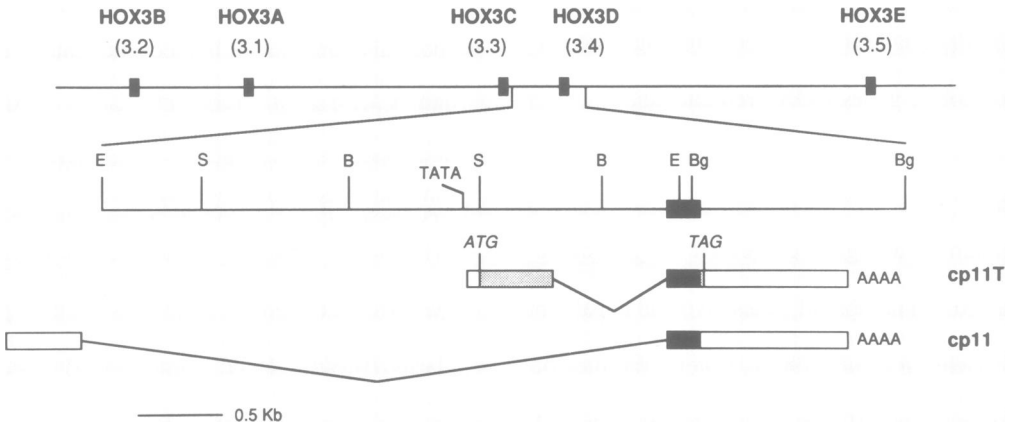


Fig. 1. Restriction map of the *Eco*RI−*Bgl*II human genomic region on chromosome 12 containing the *HOX3D* gene, and the two major alternative transcripts represented by cDNA clones cp11T and cp11. Names of cognate mouse genes are indicated in parentheses. Solid boxes represent homeoboxes, shaded boxes coding regions, and open boxes transcribed regions. B, *Bam*HI; Bg, *Bgl*II, E, *Eco*RI; S, *Sst*I. TATA, ATG and TAG indicate position of the TATA element and the translation initiation and termination codons respectively. AAAA represents poly(A) tails.

by differential hybridization or RNase protection (not shown). *HOX3D* is constitutively active in a number of human neuroblastoma cell lines, such as SK-N-BE (Figure 3B), and is activated by RA in the EC cell line NT2/D1 (Figure 3C). Only the protein-coding, cp11T-like type of transcript is expressed in neuroblastoma cells and induced by RA in EC cells (Simeone *et al.*, 1991, and unpublished data). High levels of RA (10^{-6} to 10^{-5} M) are required to induce maximal accumulation of *HOX3D* transcripts (Figure 3C), as previously observed also for the *HOX3D* paralog *HOX2A* and the other 'posterior' *HOX2* genes (Simeone *et al.*, 1990). Activation of *HOX3D* takes place between 4 and 8 days of continuous RA induction (Simeone *et al.*, 1991).

Run-on nuclear transcription assay showed that *HOX3D* is not transcribed in uninduced cells and is transcribed at low levels 8 days after induction, whereas only a barely detectable signal was observed at day 3. Conversely, transcription of the early induced *HOX1F* gene (Simeone *et al.*, 1991) is activated at virtually the same levels at day 3 or 8 (Figure 3D).

Translation in a cell-free rabbit reticulocyte extract of a capped RNA derived by T7 polymerase-dependent *in vitro* transcription of cp11T gave rise to protein products with an M_r of 26–28 kDa (Figure 3E), in good agreement with the mol. wt predicted from its conceptual translation (i.e. 24.9 kDa).

Definition of the HOX3D promoter

A genomic fragment spanning the whole *HOX3C–HOX3D* intergenic region, extending from position −2219 to +46 from the transcription start site (Δ−2219), was cloned in front of the reporter human growth hormone (hGH) gene into the p0GH vector (Figure 4A) and transfected into SK-N-BE cells, which express the endogenous *HOX3D* gene constitutively (see above). This fragment was able to direct the synthesis of hGH at levels of ~50% of those directed by the mouse metallothionein (MoMT) promoter in the same vector, as assayed by RIA in the culture medium (Figure 4B). Progressive 5′ deletions of the original fragment down to position −114 showed approximately the same activity, with the only exception of Δ−326, which consistently directed hGH synthesis at levels ~1.5-fold higher than Δ−2219. The Δ−64 fragment was virtually inactive, as well as the Δ−1223 fragment cloned in the reverse orientation with respect to the hGH gene (Figure 4B). These data indicate that the *HOX3D* upstream region extending to position −114 from the cap site contains a functional promoter, as defined by its ability to direct transcription of a reporter gene into permissive cells in the direct orientation. The Δ−1223–GH construct showed no activity when transfected into a number of cells which do not express *HOX3D*, i.e. HeLa, HepG2 and Hep3B hepatoma, CaCo intestinal epithelial, COS7 kidney epithelial, HL60 myeloblastic, Raji and MOLT4 lymphoblastic and NIH3T3 fibroblastic cell lines (not shown), indicating that tissue-specific factor(s) present in neuroblastoma cells are probably necessary for the *HOX3D* promoter basal activity.

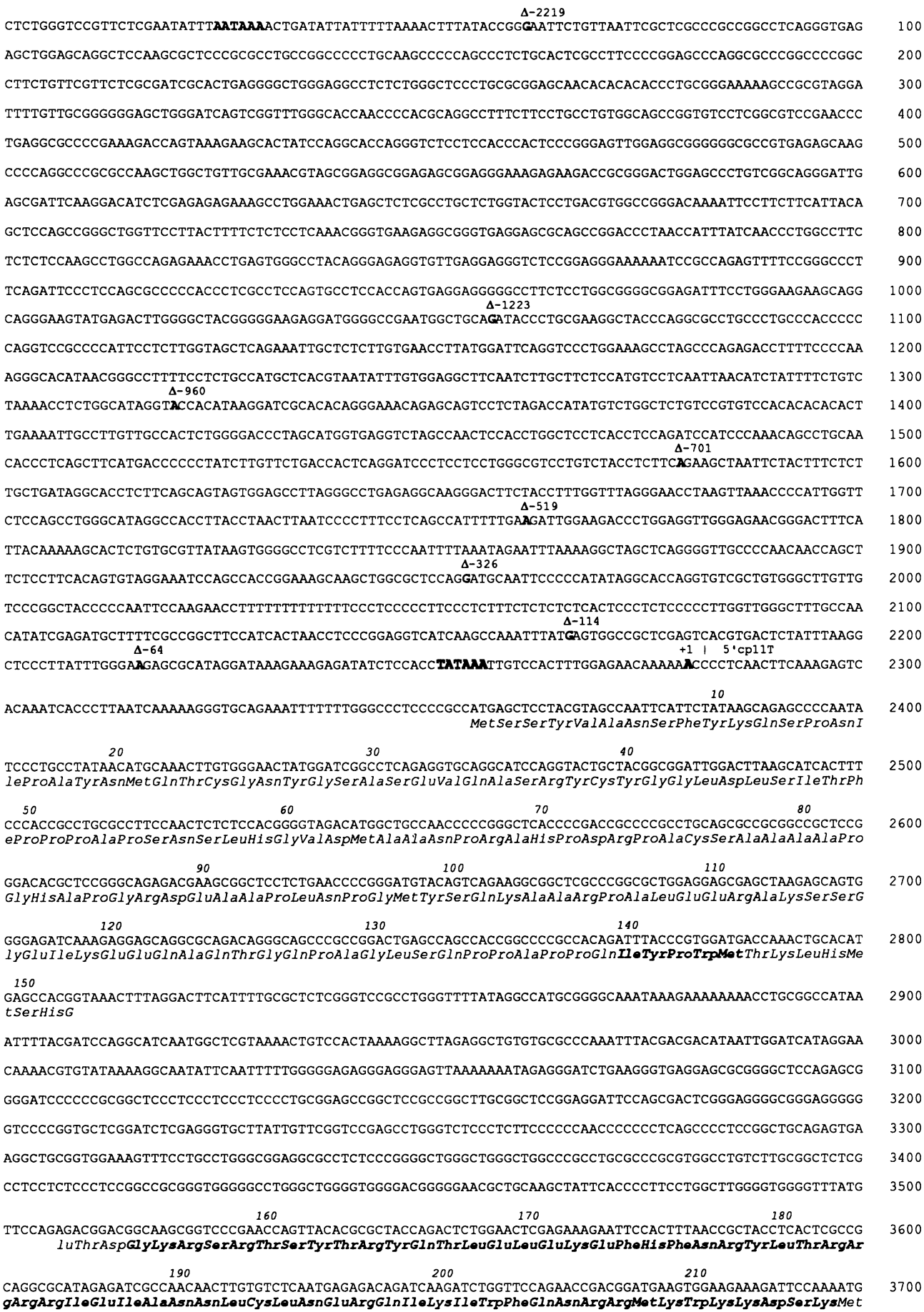
Activation of the HOX3D promoter in RA-induced EC cells

The *HOX3D* promoter–GH constructs were transfected into the EC cell line NT2/D1, both as stem cells and after induction with 10^{-5} M RA for 3 and 8 days. The Δ−1223, Δ−701 and Δ−519 fragments showed virtually no activity in stem cells, whereas the Δ−326 and Δ−114 fragments were able to direct the synthesis of low but detectable hGH levels. The Δ−64 fragment was inactive (Figure 4C). All fragments, except the Δ−64, showed a 2- to 3-fold increase in activity when transfected in cells induced by RA for 8 days (Figure 4C), whereas no increase was observed in cells induced for 3 days (not shown). No significant variation was observed in the activity of a control HSV–TK promoter–hGH construct in uninduced versus induced cells (Figure 4C). Activation by RA occurs late during induction (8–10 days) whereas in earlier phases (3–5 days) the promoter activity is at the same level as in uninduced cells. Furthermore, although the activity of all deletion fragments increases as a result of RA induction, deletion of sequences upstream of position −324 allows function of the *HOX3D* promoter at a detectable level in uninduced cells.

Binding of RA-induced nuclear factors to the HOX3D promoter

Crude protein extracts were prepared from nuclei of NT2/D1 cells before and after induction with 10^{-5} M RA for 10 days and assayed in gel mobility shift experiments for the presence of activities binding to different fragments of the extended *HOX3D* promoter. Sequences from position −519 to −326 were retarded after incubation with uninduced cell nuclear extract, whereas much less binding activity was present in nuclei of induced cells (Figure 5A). Sequences from −326 to −114 were apparently unable to bind specific proteins in both uninduced and induced cells. The −114 to +46 fragment, i.e. the minimal promoter in both neuroblastoma and EC cells, showed two faint retarded bands when tested with uninduced cell extracts, and a strong specific extra band of higher mobility when tested with nuclear extracts of RA-induced cells. No binding was observed with the shortest fragment, from −64 to +46 (Figure 5A). Double-stranded oligonucleotides containing sequences from −109 to −64 and −104 to −84 were synthesized and tested with nuclear extracts from uninduced cells and cells induced for 3–8 days with 10^{-5} M RA. The latter oligonucleotide was able to reproduce the binding pattern in both uninduced and induced cells shown by the entire Δ−174 region (Figure 5B). The binding activity specific for RA-induced cells appeared between 7 and 8 days of induction whereas two slowly migrating bands were constantly present in uninduced and induced cells (Figure 5B). The uppermost band was also detected in experiments carried out with nuclear extracts of SK-N-BE and a large number of other cells (not shown) and probably represents a ubiquitous factor. Specificity of binding was confirmed in all experiments by competition with a cold excess of either specific or unrelated fragments or oligonucleotides (Figure 5A and B). These experiments show that a nuclear factor binding to the *HOX3D* promoter is induced by RA in EC cells at a time when these cells express the endogenous gene and become competent for activation of the transfected promoter in a reporter construct.

To delineate more precisely the sequence bound by the induced-cell specific factor, we analyzed by DNase I footprinting the complexes obtained by incubation of a fragment spanning positions −174 to +46 with nuclear extracts of NT2/D1 cells uninduced and induced for 8 days with RA. 4–8 μg (total protein content) of induced-cell


Δ-2219
Δ-1223
Δ-960
Δ-701
Δ-519
Δ-326
Δ-114
Δ-64
+1 | 5'cp11T
10
20
30
40
50
60
70
80
90
100
110
120
130
140
150
160
170
180
190
200
210

```
          220    222
AAAAGCAAAGAGGCTCTTTAGAGGCAGCGGGGGAGGCCCGCAGAGCGCGCCCCTAGCCGGTTCCTGTCCCTGCGCCTTTCCTTTTCGCCTTTCCTCTCTA   3800
LysSerLysGluAlaLeuEnd

TATTTCGGGTCGGGGGCAGGTGCTGGAGCACTGGGCTCCCGGGCCCCACAGACAAAAGCGCTTTTCCTTGGCATTCCGCATCCCTACCGACCCAGGGTTC   3900

CCGCGGGGCTGTCGGCGCTGCCCCATCTCCCCTCAGCTCGGCTCAGCTCGGTACCCGGGGCCCAGGGAAGCTCCGTAGGACTTCCCCGGAGGGCTGCGGC   4000

GTACAGGCTGGCGCAGAACGAACCTTGGCCTGGGCCGTATCTCCGGCTCCCAGCCTCAGCGCGGCCCTCCCGAGTTAAGGTGGGCCCGGCCCGCGCCACA   4100

GGACCCTCGCCGGACCCTCTAACCTCGCCCTCTCCTTTGTTCCTGGCTGGACGGGTTAGACAGCCAAAGGCTGGCGAGAGTCTGGCCCTAAACTCGGGGT   4200

GCTTCCTTGTAGCGACTAAACTAGATTTTCACTTATGAATGATTTGCATATGAAAGGAGAGCATCGGCCTAGGGCCCCCACAGTTGCTCTATGCTTTCCA   4300

AACCTTATCTCCACAACCTCTTCCCCCCAAAACCCGGGAACCTCCCCAGCCTGCGCCTGCTGCATGCCCTCTCAGGCCGGCAGCCCCAGCCTGCTAGCTA   4400

GCTCAACTAGTGGGGTTTCCTGGCACTGGACCCCAGCAAGTGGTCCTAGAGGCCCTTTGCTGTCCCATAGTCCCTGCCACGAATTTCTGTGCCCTCCTGA   4500
                                                                                  3'cp11T  |
CCCATTGCTGTTGTCCAACTATTTATTGACTCTGGGTCCTTCCTGAAACTATATTTTGTCATATCAAATAAAGAGAGAACAGGACTAAAGATGCAGTGGC   4600
```

Fig. 2. Nucleotide sequence of the genomic region containing the *HOX3D* gene and its conceptual translation into the HOX3D homeoprotein. Nucleotides are numbered on the right, amino acids are numbered above the sequence. The TATA element and the mapped major transcription start site are outlined. The poly(A) addition signals of *HOX3C* and *HOX3E* are both indicated in bold type in the nucleotide sequence, as well as the first nucleotides in the deletion fragments ($\Delta-2219$ to $\Delta-64$) shown in Figure 4. The IYPWM homeopeptide (Mavilio *et al.*, 1986) and the homeodomain are indicated in bold type in the protein sequence. The 5′ and 3′ ends of the longest isolated cDNA clone cp11T are indicated by vertical bars.

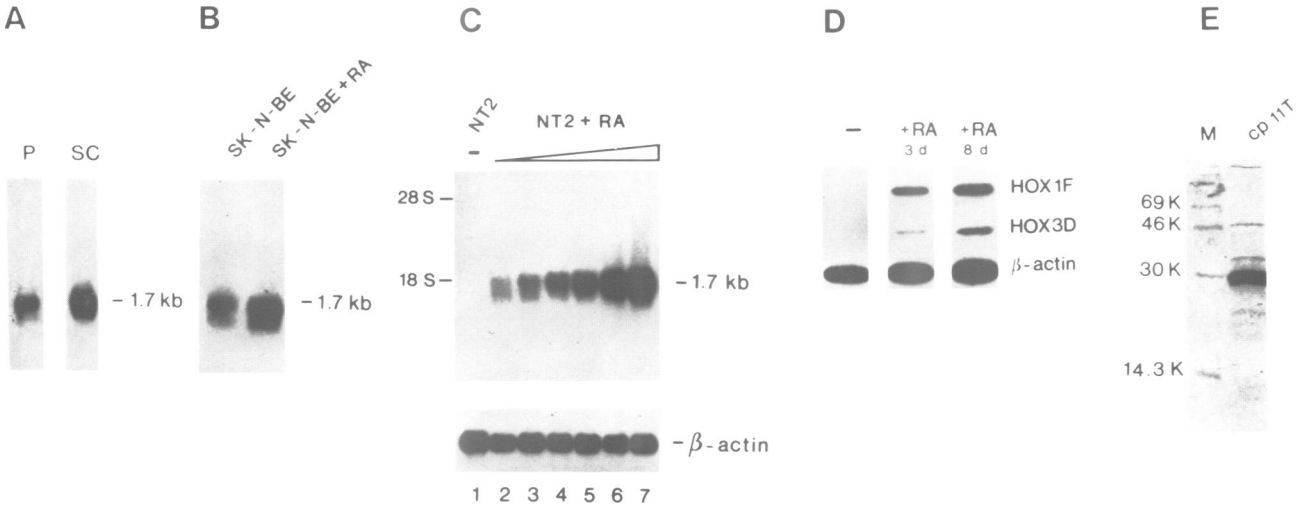


Fig. 3. A–C. Northern blot analysis of 2 μg of poly(A)$^+$ RNA from human full-term placenta (P) and 7-week-old embryo spinal cord (SC), 5 μg of poly(A)$^+$ RNA from the human neuroblastoma cell line SK-N-BE before and after 3 days induction with 10^{-5} M retinoic acid (+RA) and 5 μg of poly(A)$^+$ RNA from the human embryonal carcinoma cell line NT2/D1, before (−) and after 14 days induction with retinoic acid (+RA) at molar concentrations of 10^{-8} (lane 1), 5×10^{-8} (2), 10^{-7} (3), 5×10^{-7} (4), 10^{-6} (5), 5×10^{-6} (6) and 10^{-5} (7). Northern blots were hybridized to a 3′ non-coding region of the cDNA clone cp11T and a human β-actin cDNA probe for normalization. Mobilities of 28S and 18S rRNAs are shown as size markers. **D.** Run-on nuclear transcription assay of *HOX3D* and the early-activated *HOX1F* (Simeone *et al.*, 1991) genes in NT2/D1 cells, uninduced or induced with 10^{-5} M RA for 3 or 8 days. Transcription of β-actin is shown as normalization. **E.** Cell-free translation in rabbit reticulocyte lysates of synthetic RNA generated by T7 polymerase-directed transcription of cp11T cDNA (see Figure 1) in pCT vector (see Materials and methods). Proteins were labeled by [^{35}S]Met, resolved on SDS–PAGE and fluorographed. A ^{14}C-labeled protein ladder (M) is shown as size marker. Molecular mass is given in kilodaltons (K).

nuclear extract specifically protected a 21-bp sequence encompassing positions −105 to −84, which was not protected by 16 μg of nuclear extract from uninduced cells (Figure 6A). The core of the protected region is a 14-bp palindromic sequence (GAGTCACGTGACTC) which contains an inverted repeat of a sequence (GAGTCA) closely related to the AP-1 binding site consensus (TGA*G/C*TCA, Angel *et al.*, 1987; Lee *et al.*, 1987). However, neither the gel shift nor the footprint is competed by a 2000-fold molar excess of an AP-1 consensus oligonucleotide (not shown). It is noteworthy that a very similar, imperfect AP-1 inverted repeat was reported to mediate RA-induced negative regulation of the rat stromelysin gene (Nicholson *et al.*, 1990). The 6-bp CACGTG sequence encompassing the palindrome dyad symmetry axis perfectly matches the core of the ubiquitous USF/MLTF transcription factor binding site (Sawadogo and Roeder, 1985; Carthew *et al.*, 1985) and is identical to the binding site for the Myc–Max heterodimer (Blackwell *et al.*, 1990). This sequence is entirely conserved at corresponding locations upstream of transcriptional start sites in the mouse cognate *Hox 3.4* gene and two paralogous genes *Hox 2.1* and *Hox 1.3* (Figure 7), and is at the core of a transcription factor binding site described in the mouse *Hox 2.3* gene promoter (Zwartkruis *et al.*, 1991). The minimal sequence required for binding of both constitutive and RA-induced factors was established by gel shift

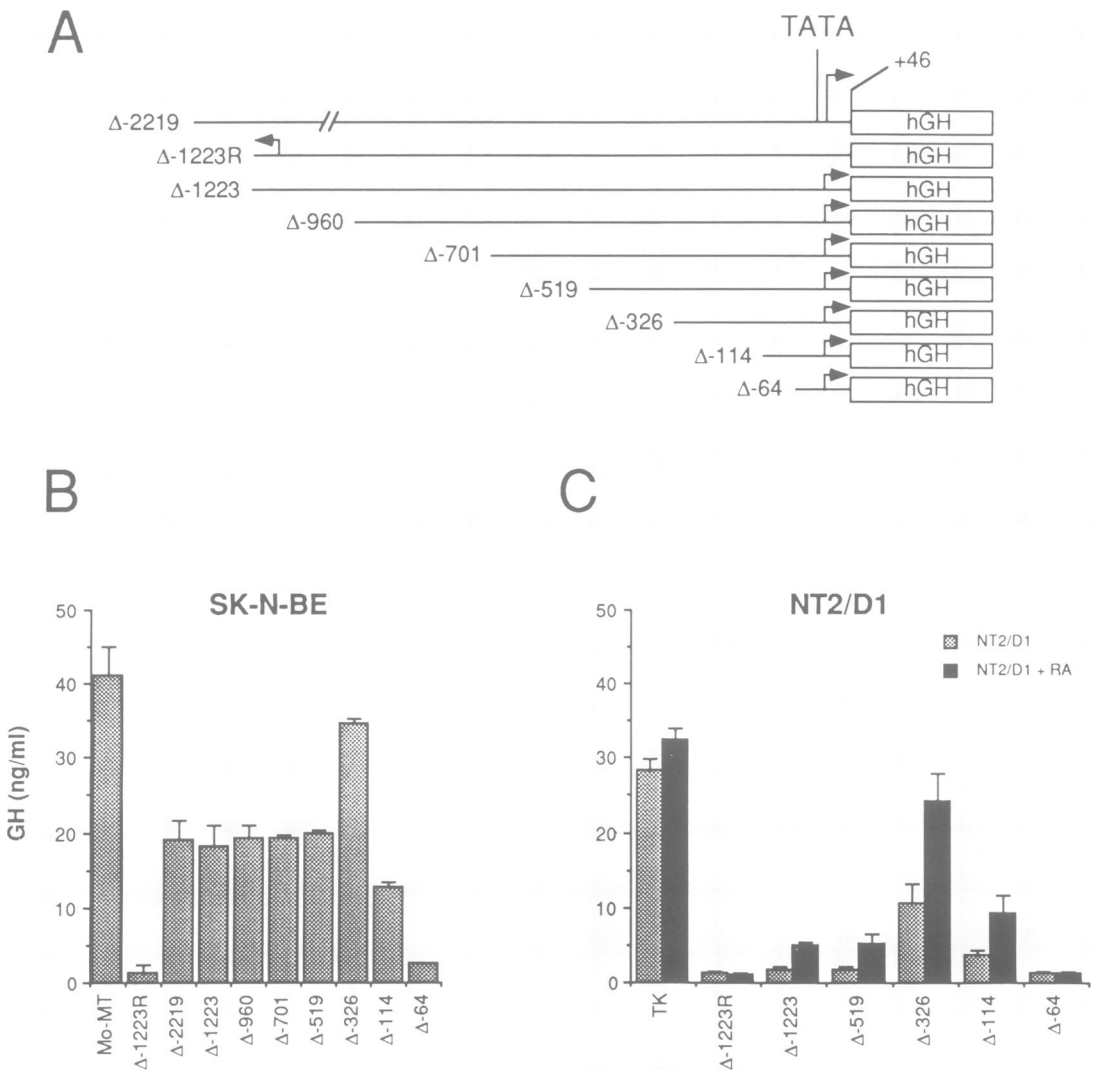


Fig. 4. A. Map of the *HOX3D*−hGH reporter constructs. Progressive 5′ deletions of the *HOX3D* upstream genomic region were cloned upstream from the human growth hormone (hGH) gene into the p0GH vector (see Materials and methods). The TATA element, the transcription initiation start site (arrow) and the 3′ end of all fragments, corresponding to position +46, are indicated. **B.** Activity of the reporter constructs transfected in the human neuroblastoma cell line SK-N-BE, which constitutively expresses the *HOX3D* gene (see Figure 3B). Activity of the mouse metallothionein-I promoter (Mo-MT) in the same construct is shown for comparison. Activity is expressed in ng/ml of hGH secreted in the culture medium 48 h after transfection. Each column represents the mean ± SE of three to six independent determinations. **C.** Activity of Δ−1223, Δ−519, Δ−326, Δ−114 and Δ−64 reporter constructs transfected in NT2/D1 cells, uninduced (stippled bars) or induced for 8 days with 10^{-5} M RA (solid bars), expressed as mean ± SE of four independent determinations. GH assay was done 48 h after transfection. The activity of the HSV-TK promoter in the same construct (see Materials and methods) is shown for comparison.

experiments with synthetic oligonucleotides; it is GTCACGTGACTC (Table I). Mutation of the internal CG to TA completely abolished binding of both factors.

The causal role of the RA-induced factor in mediating *HOX3D* promoter activation in 8-day-induced cells is difficult to establish. In fact, a nested deletion of the −99 to −84 sequence in the Δ−701−GH construct caused a 10-fold decrease in the basal promoter activity as assayed by transfection in SK-N-BE cells (not shown). We interpret these findings by suggesting that the footprint A palindrome is a binding site for both ubiquitous factors necessary for transcriptional activity (possibly the high molecular weight bands observed in gel retardation experiments, see Figure 5) and the RA-induced factors, thus preventing us from assessing the relative role of the latter by mutagenesis of target sequences.

Transactivation of the HOX3D promoter by HOX3 and HOX4 homeoproteins

The Δ−1223−GH construct was transfected in HeLa cells together with expression vectors in which the cp11T cDNA and cDNAs for the upstream gene *HOX3C* and the downstream gene *HOX3E* were put under the control of the CMV immediate-early promoter (pCTH3D, pCTH3C and pCTH3E, see Materials and methods). Cotransfection with the *HOX3C* cDNA induced a strong activation of the *HOX3D* promoter, which directed the synthesis of GH at almost twice the level produced by the MoMT−GH construct in the same cells. No activation was observed with the construct in which the Δ−1223 fragment was in the reverse orientation, or by cotransfecting a *HOX3C* cDNA in which a frameshift mutation causes synthesis of a protein lacking the homeodomain (Figure 8A). Cell-free translation of a

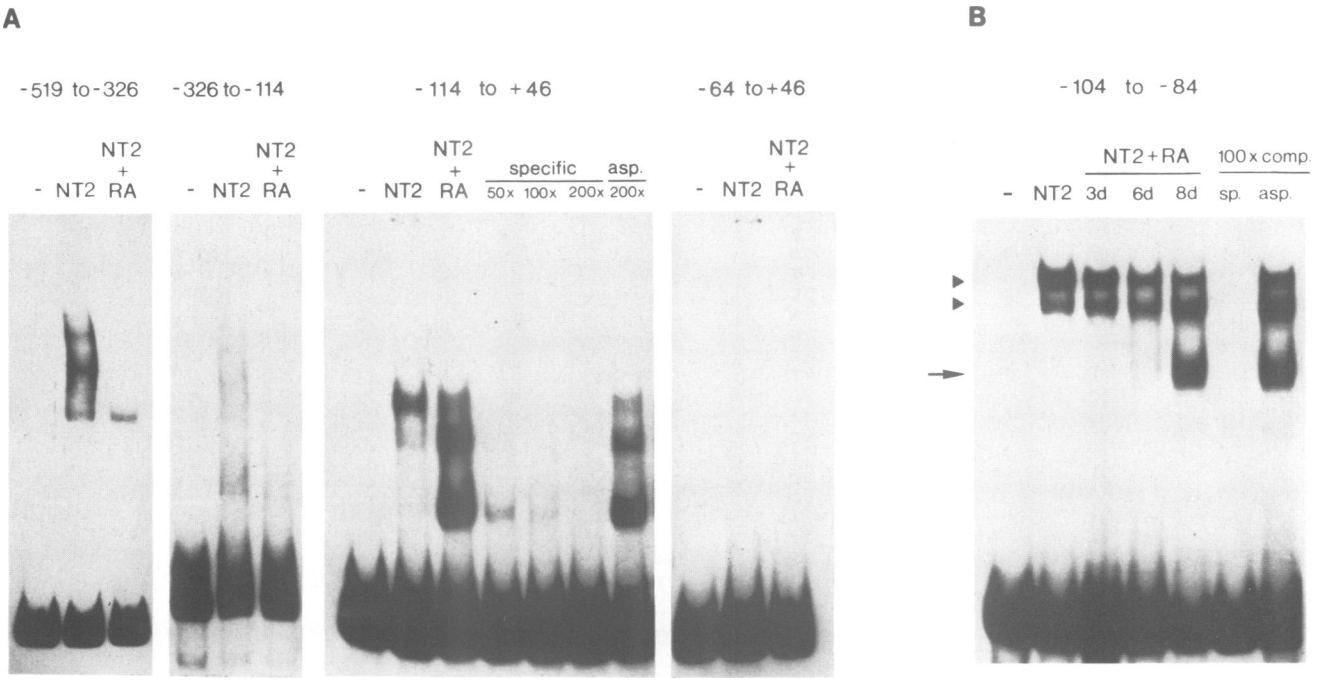


Fig. 5. A. Gel retardation analysis of *HOX3D* 5′ upstream regions spanning position −519 to −326, −326 to −114, −114 to + 46 and −64 to + 46 with respect to the transcription start site (Figure 2) with nuclear extracts from NT2/D1 cells, uninduced (NT2) or induced with 10^{-5} M RA for 10 days (NT2 + RA). Where indicated, binding was competed with a 50-, 100- or 200-fold molar excess of the cold fragment (specific), or a cold 250 bp restriction fragment of pUC19 plasmid (asp.). **B.** Gel retardation analysis of a double-stranded synthetic oligonucleotide corresponding to nucleotides −104 to −84 with nuclear extracts from NT2/D1 cells, uninduced (NT2) or induced with 10^{-5} M RA (NT2 + RA) for 3, 6 and 8 days. The two arrowheads indicate retarded bands present before and after induction, whereas the arrow indicates the band specific for RA-induced cells. Binding was competed with a 100-fold molar excess of either cold oligonucleotide (sp.) or a cold unrelated 25-mer double-stranded oligonucleotide (asp.). Lanes containing free probes are indicated by a − sign.

synthetic RNA transcribed by T7 polymerase confirmed synthesis of a shorter protein of the expected M_r by the mutated cDNA (not shown). Conversely, cotransfection of *HOX3D* and *HOX3E* cDNAs caused no activation of the *HOX3D* promoter (Figure 8A). A second series of experiments was carried out with constructs in which cDNAs for *HOX4* genes belonging to further upstream, paralogy groups, i.e. *HOX4D*, *HOX4C* and *Hox 4.3*, the mouse cognate of *HOX4E*, were under the control of the SV40 early promoter (pSGH4D, pSGH4C and pSGH4.3; Zappavigna *et al.*, 1991). Cotransfection with these cDNAs caused strong transactivation of the *HOX3D* promoter, at levels comparable to those observed with the *HOX3C* cDNA. Cotransfection with a construct containing the cDNA for *HOX4B*, the paralog of *HOX3E* on the *HOX4* cluster, had instead no effect on the activity of the *HOX3D* promoter (Figure 8A). All transfections were repeated in NIH3T3 cells and produced comparable results (not shown).

The whole series of 5′ deletion constructs was then cotransfected in HeLa cells together with the pCTH3C construct. All constructs were transactivated at comparable levels, with the exception of Δ−64−GH, which directed synthesis of 4- to 5-fold lower levels of hGH (Figure 8B). Comparable results were obtained with pSGH4D, pSGH4C and pSGH4.3 vectors, in both HeLa and NIH3T3 cells (not shown). The extent of the transcriptional increase of the *HOX3D* promoter in HeLa cells after cotransfection with the homeoprotein expression constructs was also measured by RNase protection, using a probe for the fifth exon of the hGH gene. As shown in Figure 8C, accumulation of hGH RNA in cells transfected with only the reporter plasmid was virtually undetectable (it became detectable in 2 day overexposures of the gels), whereas a strong signal was obtained when the reporter plasmids were cotransfected with the pCTH3C expression construct. The increase in mRNA accumulation was estimated as >20-fold.

These data indicate that homeoproteins encoded by genes upstream from *HOX3D*, like *HOX3C*, or belonging to upstream paralogy groups, like *HOX4D*, *HOX4C* or *Hox 4.3*, are strong activators of the *HOX3D* promoter in either human or murine cells, where the promoter is otherwise virtually inactive. Conversely, neither the HOX3D protein nor those encoded by the downstream *HOX3E* gene and its paralog *HOX4B* show any activity in the same cells. The ability of the *HOX3D*, *HOX3E* and *HOX4B* expression construct to synthesize RNA was controlled by RNase protection in transfected cells and all RNAs were able to translate protein of the correct M_r when tested in cell-free extracts (Figure 3E and results not shown). Furthermore, nuclear extracts from HeLa cells transfected with *HOX3D*, *HOX3E* and *HOX4B* expression constructs showed specific bands in a gel retardation assay using an oligonucleotide containing multiple copies of the consensus homeodomain binding sequence NP (TCAATTAAAT, Hoey and Levine, 1988; Desplan *et al.* 1988) as probe (Figure 8D). These results indicate that co-transfection experiments were not biased by inefficient synthesis of some HOX proteins in transfected cells.

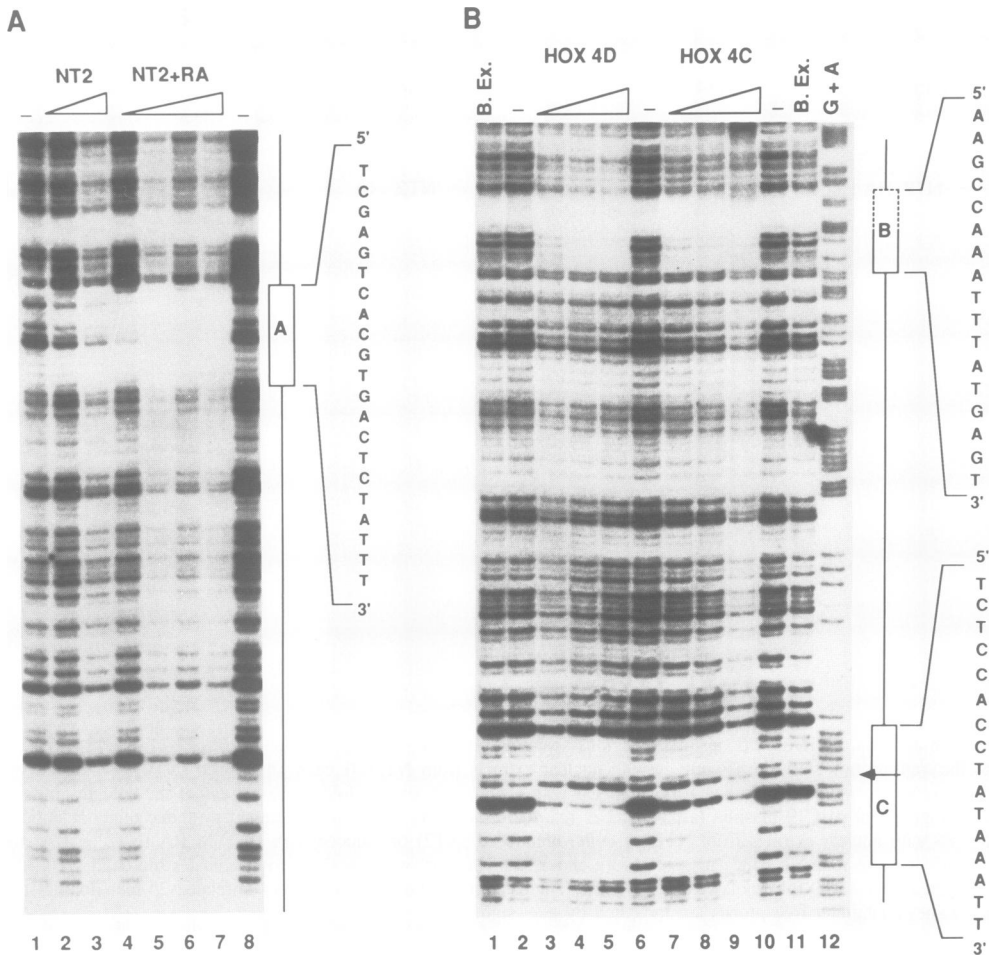


Fig. 6. DNase I footprint analysis of the *HOX3D* upstream genomic region spanning positions −174 to + 46 with respect to the transcription start site (Figure 2). **A.** The 3′ end-labeled fragment was incubated with 8 and 16 μg of NT2/D1 uninduced cell nuclear extracts (lanes 2 and 3) or 4, 8, 12 and 16 μg of extracts from NT2/D1 induced by 10^{-5} M RA for 10 days (lanes 4−8), before digestion with DNase I. Unprotected DNA is shown in lanes 1 and 8. The footprint observed with RA-induced cells is indicated by box A together with the corresponding protected sequence. **B.** Incubation with 1, 2 and 3 μg of crude extracts of *E.coli* cells engineered to synthesize HOX4D and HOX4C proteins (see Materials and methods; lanes 3−5 and 7−9 respectively). Protected sequences are shown by boxes B and C. The upper part of the B box (dotted line) indicates an uncertainty region, due to the presence of sequences very resistant to DNase I digestion (see A for further comparison) The arrow indicates a DNase I hypersensitive site generated by binding to HOX-containing extracts. Unprotected DNA is shown in lanes 2, 6 and 10, and DNA incubated with control bacterial extracts (B. Ex.) in lanes 1 and 11. The G + A Maxam and Gilbert sequence reaction is shown in lane 12 as size marker.

Binding of HOX4D and HOX4C homeoproteins to the HOX3D promoter

The full-length cDNAs for *HOX4D* and *HOX4C* were cloned in a bacterial expression vector and used to produce the respective homeoproteins in *Escherichia coli* (Zappavigna *et al.*, 1991). Crude bacterial extracts were positive in gel mobility shift assays for binding to the −174 to +46 *HOX3D* fragment, and were then used in DNase I footprinting experiments. Two sequences were protected from DNase I digestion by extracts containing both *HOX4D* and *HOX4C* (Figure 6B). The first (footprint B) is an A/T-rich sequence located upstream from position −115, whose 5′ border could not be exactly established because of the presence of a long DNase hyper-resistant stretch. The AATTTATGA sequence is highly conserved at corresponding positions in the upstream region of *Hox 3.4*, *Hox 2.1* and *Hox 1.3* (Figure 7), and corresponds to the 3′ half of a long stretch of sequence on the *Hox 1.3* promoter protected from DNase I digestion by the Hox 1.3 protein (Odenwald *et al.*, 1989). The sequence TTTATG is also a binding site for the activating *cad* homeoprotein in the promoter of the *Drosophila ftz* gene (Dearolf *et al.*, 1989).

The second footprinted sequence (footprint C, Figure 6B) includes the TATA element and a few upstream and downstream nucleotides, which are conserved in *Hox 3.4* but not in *Hox 1.3* and *Hox 2.1* (see Figure 7). Binding of *HOX4C* and *HOX4D* to this sequence generates a DNase I hypersensitive site immediately upstream of the TATA box (Figure 6B). The relative role of the two footprinted sequences in the homeoprotein-mediated activation of the *HOX3D* promoter was determined by site-directed mutagenesis. Mutation of the AATTTAT core sequence into CCCTCCT caused a 3- to 5-fold decrease in transactivation of the Δ−519−GH construct by *HOX4C* and *HOX4D* (Figure 9), comparable to that observed in experiments with the Δ−64 construct (Figure 8B), in which a much larger region is deleted. Conversely, removal of the TATA element virtually abolished the basal promoter activity as assayed by transfection in the SK-N-BE cell line (not shown). It was therefore impossible to demonstrate by mutagenesis any role of the TATA box in mediating *HOX* transactivation activity.

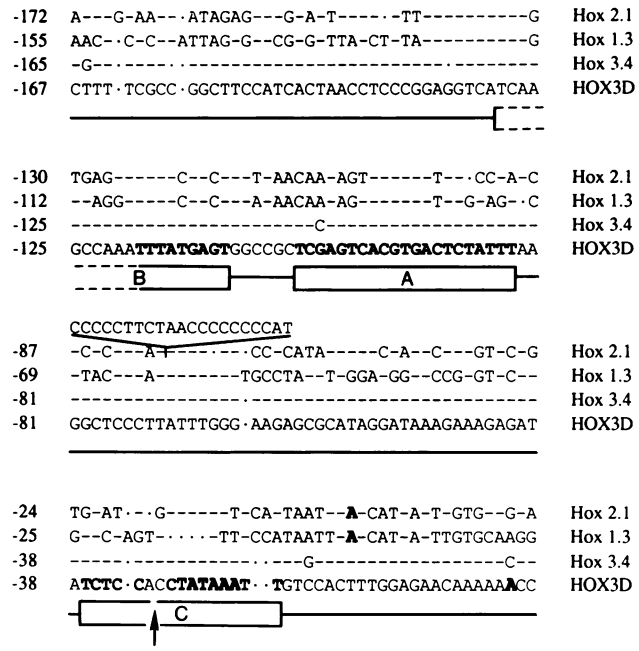


Fig. 7. Best-fit alignment of nucleotide sequences of the upstream region of *HOX3D*, its mouse cognate *Hox 3.4* (courtesy of P.T.Sharpe, unpublished) and the mouse *Hox 2.1* (Krumlauf *et al.*, 1987) and *Hox 1.3* (Odenwald *et al.*, 1987) genes belonging to the same paralogy group. Sequences are numbered on the left, starting from the transcription start sites (outlined). *Hox 3.4* sequence is conventionally numbered assuming the transcription start site to be the same as in *HOX3D*. Dashes indicate nucleotides identical to those of *HOX3D*, dots indicate gaps. The A, B and C boxes indicate observed DNase I-protected sequences (Figure 6), shown in bold face.

Table I. Summary of gel retardation experiments

Sequence	Binding	
	Constitutive	RA-induced
−104 TCGAGTCACGTGACTCTATTT −84	+++	+++
GTCACGTGACTC	+++	+++
TCACGTGA	±	±
GTCATATGACTCT	−	−

The table shows *in vitro* binding of constitutive and RA-induced nuclear factors from NT2/D1 embryonal carcinoma cells to normal or mutated sequences contained in footprint A (see Fig. 6). All sequences were tested as part of 25-mer, double-stranded oligonucleotides.

Discussion

Regulation of HOX3D expression

The human *HOX3D* homeogene is expressed in both embryos and extraembryonic tissues in at least two types of transcripts, the first containing a distal, non-coding exon contained also in *HOX3C* and *HOX3E* transcripts, the second containing a proximal, *HOX3D*-specific exon encoding the amino-terminal domain of the 222 amino acids HOX3D protein. As previously described, the HOX3 cluster is extensively transcribed in a very heterogeneous family of RNAs starting at multiple sites. Usage of a still undefined upstream promoter leads to the generation of *HOX3C*, *HOX3D* and *HOX3E* transcripts containing a common 5′ exon (Simeone *et al.*, 1988; Boncinelli *et al.*, 1990). In the case of *HOX3C* and *HOX3E*, these transcripts encode putative homeoproteins (Simeone *et al.*, 1988), whereas in *HOX3D* the homeodomain reading frame is closed immediately upstream of the splice junction. The functional significance of this alternative, non-coding transcript, which is at least as abundant in embryos and placenta as the coding one, is so far unclear. It is plausible that apparently 'useless' RNAs derive from inaccuracies in the complex transcription pattern of HOX clusters, which is typically highly redundant. In fact, non-coding homeobox-containing transcripts are not uncommon in early embryogenesis of vertebrates (Condie *et al.*, 1990). Transcription of *HOX3D* and other *HOX3* genes is much less complex in adult cells and cultured cell lines, notably in EC cells where only the protein-coding transcript is induced by RA. This observation led us to investigate the genomic region immediately upstream of the major transcription start site in order to find regulatory regions.

The HOX3D promoter is inducible by RA

Transfection of the *HOX3D* upstream region linked to a reporter gene into neuroblastoma cells led to identification and mapping of a minimal promoter for the protein-coding *HOX3D* transcript. The *HOX3D* promoter contains a canonical TATA element, and is at least to some extent tissue- or cell-specific, since it does not work in most of the cell lines tested. A fragment extending 1223 bp upstream from the cap site was unable to direct the synthesis of the reporter gene when transfected into uninduced EC cells and up to 3−5 days of RA induction. However, a slight but readily detectable activation was observed when the reporter construct was transfected in cells induced for 8 days. Activation of the endogenous gene, as measured by run-on transcription, is undetectable 3 days after RA induction, and is only 2- 3-fold higher than the background level at 8 days (Figure 3D). Accordingly, accumulation of *HOX3D* RNA proceeds slowly, starting at day 4 and reaching levels detectable by poly(A)$^+$ RNA hybridization around days 6−7 (Simeone *et al.*, 1991). The weak and late induction of the isolated *HOX3D* promoter by RA therefore agrees relatively well with the activation pattern of the endogenous gene. The existence of additional control elements located elsewhere in the genome, or the contribution of post-transcriptional mechanisms to the overall gene expression levels, is nevertheless not ruled out by our data. Interestingly, however, a specific factor binding at a short distance from the TATA box appears in the EC cell nuclei at approximately the same time when the cells become competent to activate the transfected promoter. This factor binds to a palindromic sequence mostly conserved at the corresponding position in the mouse cognate gene *Hox 3.4* and paralogous genes *Hox 2.1* and *Hox 1.3* (Figure 7). A similar element, although with a much longer spacer between the two half-sequences, has been found to negatively regulate the activity of the rat stromelysin gene upon RA induction (Nicholson *et al.*, 1990). We were unable formally to rule out the identity of the RA-induced binding factor to a RA receptor, although a series of control experiments, including competition with known synthetic RAR binding sites, gel shift assays with nuclear extracts of RAR-overproducing cells and transactivation of the reporter constructs with RAR

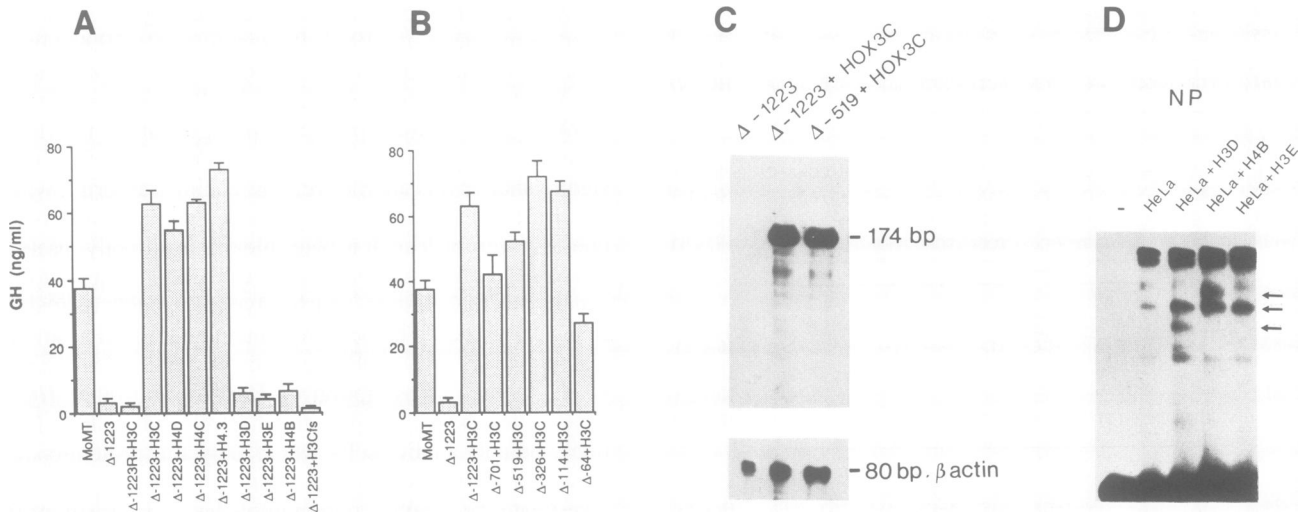


Fig. 8. Transactivation of the upstream region of *HOX3D* by homeoproteins. **A.** Activity of the Δ−1223−GH reporter construct (Figure 4) transfected in HeLa cells together with expression constructs for the homeogenes *HOX3C* (+H3C), *HOX3D* (+H3D), *HOX3E* (+H3E), a frameshift mutant of *HOX3C* (+H3Cfs), *HOX4D* (+H4D), *HOX4C* (+H4C), *HOX4B* (+H4B) and the mouse *Hox 4.3* (+H4.3). **B.** Activity of 5′ *HOX3D* deletion reporter constructs (Figure 4) transfected in HeLa cells together with the expression construct for *HOX3C*. Activity of the MoMT−GH construct is shown for comparison. Values are expressed as ng/ml of secreted hGH, and as mean ± SE of three to eight independent experiments. **C.** Transactivation of the Δ−1223−GH and Δ−519−GH reporter constructs transfected in HeLa cells together with the *HOX3C* expression construct, as evaluated by RNase protection. A 174 nucleotide RNA probe corresponding to a region in the hGH fifth exon (see Materials and methods) was used together with an 80 nucleotide fragment of the human β-actin cDNA for normalization. **D.** Gel retardation analysis of nuclear extracts from HeLa cells transfected with *HOX3D*, *HOX4B* and *HOX3E* expression plasmids, using a synthetic oligonucleotide containing a direct repeat of the homeobox-binding consensus sequence NP (i.e. TCAATTAAAT) as probe. The arrows indicate retarded bands undetectable in control HeLa cells extracts.

expression vectors (not shown) yielded negative results. The kinetics of appearance of the binding factor in nuclear extracts also argues against such a possibility. In fact, most known RARs are either present in uninduced EC cells (RAR-α and RAR-γ) or induced very shortly after RA exposure (RAR-β) (Simeone *et al.*, 1990), and would be expected to bind the relevant sequence already in the 2−3 days RA-induced cell nuclear extracts. The existence of cell-type specific RAR co-regulators capable of modulating affinity, and therefore binding, of RARs to different responsive elements has been recently suggested (Glass *et al.*, 1990). RA-induced synthesis of a specific factor directing late RAR binding to the *HOX3D* promoter is entirely compatible with our findings. Altogether, the kinetics of *HOX3D* activation and the temporal coincidence between appearance of the binding activity and activation of both the transfected promoter and the endogenous gene suggests that *HOX3D* activation is not an early response to RA directly mediated by an RAR but rather by RA-induced *HOX*-activating factor(s). This could either synergize or replace the action of apparently ubiquitous factors (MLTF/USF, Myc) possibly binding the CACGTG core of the palindrome.

Transfection of progressive deletion of the extended *HOX3D* promoter in EC cells showed that removal of sequences between −519 and −326 increases the promoter activity in both uninduced and induced cells. These sequences are apparently able to bind nuclear factors in uninduced cells, which might be involved in down-modulation of the promoter activity. We previously showed that inhibitors of protein synthesis are able to activate a large number of *HOX* genes, including *HOX3D*, in EC cells, although at lower levels than those achieved by RA induction and with loss of the correct timing and activation sequence (Simeone *et al.*, 1991). A combination of active, reversible repression in uninduced cells and time-regulated appearance of activating factors in RA-induced cells is therefore likely to underlie the sequential, RA-specific activation of *HOX* genes in EC cells.

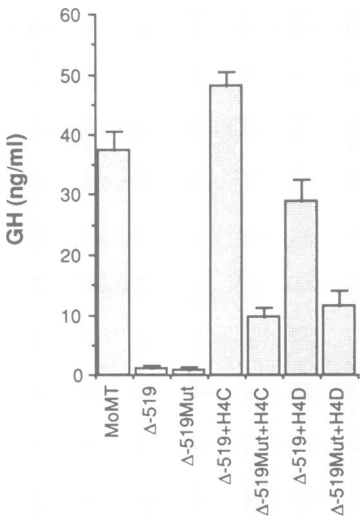


Fig. 9. Transactivation of the Δ−519−GH and Δ−519Mut−GH constructs by HOX4C and HOX4D homeoproteins. In Δ−519 Mut-GH, nucleotides from −121 to −116 (footprint B in Figure 6) were mutated from AATTTA to CCCTCC. Values are expressed as ng/ml of secreted hGH, and as mean ± SE of three independent experiments.

The HOX3D promoter is transactivated by HOX3 and HOX4 homeoproteins

Auto- and cross-regulatory interactions between homeoproteins and homeogene promoters have been reported

for a number of *Drosophila* genes, both in embryos and in cell culture systems (reviewed by Hayashi and Scott, 1990). These interactions are likely to play a significant role in the establishment and/or maintenance of well defined expression domains in the embryo, even though doubts have been raised about their real significance in the control of pattern formation (Gonzales-Reyes *et al.*, 1990). The expression domains of vertebrate homeogenes in developing embryos are largely overlapping and the existence of four highly homologous clusters considerably increases the complexity of the system, and also its redundancy. We addressed the issue of cross-regulation of human homeogenes by analyzing the activity of a number of homeoproteins on the *HOX3D* promoter by cotransfection experiments in cultured cells. We show that the products of at least four homeogenes, i.e. *HOX3C*, *HOX4D*, *HOX4C* and *Hox 4.3* (the mouse cognate of *HOX4E*), are able to transactivate at high levels the otherwise silent *HOX3D* promoter in both human HeLa and murine NIH3T3 cells. All these genes belong to upstream, or more 'posterior' paralogy groups with respect to *HOX3D*. Conversely, the HOX3D protein and those encoded by *HOX3E* and *HOX4B*, belonging to a downstream, or more 'anterior' group have no activating effect. At least in the case of *HOX3C*, deletion of the homeodomain completely abolishes the activating function, which is therefore likely to be mediated by direct protein–DNA interaction. Binding of HOX4D and HOX4C homeoproteins synthesized in bacteria to at least two sites in the *HOX3D* promoter was demonstrated *in vitro* by DNase I footprinting. These results indicate that (i) human homeoproteins behave as genuine transcription factors, in that they are able to recognize promoter regulatory sequences and activate transcription; (ii) specificity of binding is degenerate, since at least four proteins from two different clusters may activate the same promoter, and at least two of these proteins apparently bind the same sequences; and (iii) cross-regulation has a directional component, at least on *HOX3D*, since 'posterior' genes appear to activate a more 'anterior' one. The latter two conclusions, which are based on results obtained by cell transfection, are at the moment speculative. Subtle differences in affinity for different targets or correct protein dosage, which are likely to be important *in vivo*, are either masked or impossible to achieve in transfection experiments due to the overloading implicit in the technique. Nevertheless, the occurrence of a 'posterior' dominance in homeogene cross-regulation, which might have a role *in vivo* in the most likely combinatorial definition of expression domains, is suggested by our findings.

The sequences recognized by the two HOX4 proteins in the *HOX3D* promoter do not contain the ATTA core almost invariably present in *Drosophila* class I homeodomain-binding sites (reviewed in Hayashi and Scott, 1990). The more distal, A/T-rich sequence is conserved in paralogous gene upstream regions, and it has previously been reported to bind the Hox 1.3 homeoprotein in the *Hox 1.3* promoter (Odenwald *et al.*, 1989). It is suggestive that this sequence also matches entirely the TTTATG site present in the *Drosophila ftz* gene promoter, where it binds the activating and more posterior *cad* homeoprotein (Dearolf *et al.*, 1989). The more proximal HOX4 binding site is instead conserved only in the cognate *Hox 3.4* gene, and includes the TATA element. This could either reflect an *in vivo* artifact—the *HOX3D* TATA box contains the same TAAATT sequence contained in the opposite strand of footprint B (see Figure 7)—or suggest a close interaction of the homeoproteins with the TATA box and the TFIID transcriptional complex. Interaction of a homeoprotein with TATA boxes in competition with TFIID has been reported at least in the case of the *Drosophila engrailed* gene (Ohkuma *et al.*, 1990).

In conclusion, we have shown that the *HOX3D* promoter is a potential target for multiple regulatory mechanisms. The first is induced by RA, apparently through the action of a specific nuclear factor, and causes a moderate transcriptional activation in EC cells. The second is exerted by 'posterior' homeoproteins and is quantitatively more significant. The two regulations are independent of each other and are mediated by different target sequences. This is in good agreement with our previous findings showing that activation of homeogene clusters after RA induction follows a striking 3′ to 5′ polarity (Simeone *et al.*, 1990, 1991). This is unlikely to be mediated by homeogene cross-activation apparently following, if any, an opposite polarity. If the regulation observed in cell culture indeed operate in embryonic development, the final patterning of homeogene expression domains could result from the combination of activating signals provided by retinoids, or their 'downstream' intermediates, and cross-regulation induced by coexpression of a number of proteins with differential specificity of action and affinity for their targets.

Materials and methods

Cloning and analysis of HOX3D expression

A human genomic cosmid library in pcos2EMBL vector was kindly provided by A.M.Frischauf. cDNA libraries were prepared in λgt11 from poly(A)$^+$ RNA of full-term placenta, 7-week embryonic spinal cord and NT2/D1 EC cells induced by RA for 14 days (Simeone *et al.*, 1988). All libraries were screened according to standard procedures (Sambrook *et al.*, 1989). Fragments of interest were subcloned in pEMBL8 or pUC19 plasmids and sequenced according to either Sanger or Maxam and Gilbert techniques (Sambrook *et al.*, 1989). Total cellular RNA was extracted from tissues or cells by the guanidine isothiocyanate technique (Chirgwin *et al.*, 1979), poly(A)$^+$ was selected by one passage on an oligo(dT)–cellulose column, run in 5 μg aliquots on 1% agarose–formaldehyde gels, transferred to nylon filters (Amersham Hybond-N) by Northern capillary blotting (Thomas, 1980) and UV cross-linked. Filters were pre-hybridized and hybridized to 10^7 c.p.m. of radiolabeled DNA probe (a 0.4 kb *Sma*I–*Sma*I fragment of the 3′ non-coding portion of cp11T) in buffers containing 50% formamide (Sambrook *et al.*, 1989), then washed under high stringency conditions (0.1×SSC, 60°C) and exposed for 16–72 h to Kodak X-AR5 films at −70°C in intensifying cassettes. Run-on nuclear transcription assay was carried out according to Greenberg and Ziff (1984) using cDNA fragments of 5′ and 3′ non-coding regions of *HOX1F* (Simeone *et al.*, 1991), *HOX3D* and human β-actin.

Reporter plasmids and cDNA expression vectors

The 2.3 kb *Eco*RI–*Apa*I (Δ−2219), 1.3 kb *Pst*I–*Apa*I (Δ−1223) and 1.0 kb *Hin*cII–*Apa*I (Δ−960) fragments of the *HOX3D* genomic upstream sequence, all extending at their 3′ ends to position +46 from the transcription start site, were blunt-end cloned in either orientation into the *Hin*cII site of the p0GH expression vector, containing the whole human growth hormone gene (Selden *et al.*, 1986). Deletion fragments Δ−701 to Δ−64 were generated by polymerase chain reaction (PCR) using a common 3′ reverse primer corresponding to sequences +30 to +46, and cloned into the same vector. Control reporter plasmids contained the mouse metallothionein-I (pXGH5) or the herpes simplex virus thymidine kinase (pTK-GH) promoters (Selden *et al.*, 1986).

Nested deletions of nucleotides −99 to −84 in the Δ−701–GH construct were generated by PCR. A fragment was amplified using the common 3′ reverse primer and a 5′ primer introducing an *Xho*I site just before position −88. The −104 to +46 fragment was then removed from Δ−701–GH by *Xho*I and *Xba*I digestion and replaced by the amplified fragment. Site-directed mutagenesis at positions −121 to −116 into the Δ−517–GH construct was carried out similarly by PCR amplification and recloning of the −517 to −100 fragment using a mismatched 3′ primer.

cDNA fragments containing only the coding regions of *HOX3C* (c8,

Simeone *et al.*, 1988), *HOX3D*, (cp11T, this paper), *HOX3E* (cp19, Simeone *et al.*, 1988) and *HOX4B* (c13, Mavilio *et al.*, 1986) were blunt-end cloned between the Klenow-filled *Xba*I and *Bam*HI sites of the pCT expression vector under the control of the cytomegalovirus immediate-early promoter (Thali *et al.*, 1988); they were named pCTH3C, pCTH3D, pCTH3E and pCTH4B respectively. A frameshift mutation was introduced nine codons before the homeobox in the *HOX3C* sequence by digestion, fill-in and re-ligation of a unique *Eco*RI site. The modified cDNA was then cloned in the same vector (pCTH3Cfr). The coding regions of *HOX 4.3* (Izpisùa-Belmonte *et al.*, 1990), *HOX4C* and *HOX4D* were cloned into the *Bam*HI site of the pSG5 expression vector (Green *et al.*, 1988) after addition of *Bam*HI synthetic linkers and named pSGH4.3, pSGH4C and pSGH4D (Zappavigna *et al.*, 1991).

Cell-free translation

Capped sense RNAs were transcribed *in vitro* from the T7 phage promoter present in both pCT and pSG5 using 1 μg of linearized template, 0.5 mM NTP + 0.5 mM G(5′)ppp(5′)G and 20 U T7 polymerase (Promega Biotec). After two ethanol precipitations, 1 μg of synthetic RNA was translated in the presence of 40 μCi [^{35}S]Met and RNasin in cell-free rabbit reticulocyte lysates using a commercial kit (Promega Biotec). Translation products were analyzed by SDS−PAGE and fluorography.

Cell culture and transfection

All cell lines were cultured in high glucose Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal calf serum and 20 mM HEPES (all from GIBCO) in 5% CO_2 in air humidified atmosphere. Induction of NT2/D1 cells with 10^{-5} M retinoic acid was carried out as previously described (Simeone *et al.*, 1990). SK-N-BE neuroblastoma cells were transfected at 80% confluency in 6-cm Petri dishes with a mixture of plasmid DNA (10−20 μg) and cationic liposomes (Lipofectin, Bethesda Research Laboratories) according to the manufacturer's specifications. HeLa, NIH3T3, NT2/D1 and all other cell lines were transfected at 50% confluency by overnight exposure to 9−15 μg of plasmid DNA in calcium phosphate/BES-buffered saline solution in 3% CO_2 in air atmosphere, as described by Chen and Okayama (1987). RA was present throughout the procedure when RA-induced NT2 cells were transfected. In cotransfection experiments 2−5 μg of target plasmid and 5 μg of expression vector were used. Transient expression of hGH was evaluated as amount of protein secreted in the medium 48 h after transfection by a commercial RIA assay kit (Nichols Institute). In some cases, expression of hGH RNA was studied by RNase protection. Briefly, RNA was extracted from two to three 10 cm Petri dishes of confluent cells 48 h after transfection, and hybridized overnight in 20 μg aliquots to radiolabeled sense or antisense hGH RNA (a 174 bp *Bgl*II−*Pvu*II fragment of the hGH fifth exon subcloned in pBluescript vector (Stratagene) and transcribed from either the T3 or the T7 promoter or β-actin RNA at the appropriate temperatures (Simeone *et al.*, 1990). Mixtures were then digested with RNase A and T1 (2 h at 33°C) and proteinase K (all from Promega Biotec), extracted with phenol−chloroform and ethanol precipitated. Electrophoresis was carried out on 7% urea−polyacrylamide sequencing gel, dried and autoradiographed for 8−48 h at −70°C.

Preparation of nuclear and bacterial extracts

Crude nuclear extracts were prepared from 10^6−10^8 uninduced or RA-induced NT2 cells as described by Dignam *et al.* (1983), aliquoted and stored in liquid nitrogen. Crude bacterial extracts containing HOX4D and HOX4C homeoproteins were prepared from *E.coli* BL21 (DE3) cells transformed with pET3a expression vectors (Studier *et al.*, 1990), containing full-length cDNAs, then induced with IPTG (Zappavigna *et al.*, 1991) and stored at −80°C. Total protein contents were estimated by the Bio-Rad protein assay kit.

Gel retardation and DNase I footprinting analysis

Upstream genomic sequences of *HOX3D* were subcloned in pBluescript vectors, released by restriction enzyme digestion, end-labeled using a standard T4 polynucleotide kinase reaction (Sambrook *et al.*, 1989) and purified by polyacrylamide gel electrophoresis. Synthetic double-stranded oligonucleotides were labeled and purified by the same techniques. 10 000 c.p.m. of labeled fragments were incubated for 90 min on ice with 2 μg of nuclear or bacterial extracts and 7 μg of poly(dI−dC) in 20 μl of binding buffer (20 mM HEPES pH 7.9, 120 mM KCl, 8 mM $MgCl_2$, 0.2 mM EDTA, 0.5 mM DTT, 0.2 mg/ml BSA and 20% glycerol). Cold competitors were added in molar proportion together with labeled fragments or oligonucleotides. The DNA−protein complexes were resolved on 4−6% polyacrylamide gels in 1×TBE buffer at 10 V/cm. For DNase I footprint analysis, an *Xba*I−*Bam*HI fragment corresponding to the −174 to +46 region was 3′ end-labeled, purified on a polyacrylamide gel and incubated in 20 μl of gel retardation buffer with 4−16 μg of NT2/D1 cell nuclear extracts, or 1−3 μg of bacterial extracts, together with 7 μg of poly(dI−dC) for 90 min on ice. DNase I (Boehringer Biochemia) was then added to the mixture at a 8.5 μg/ml final concentration and the incubation continued for 60 min at 20°C. The reaction was terminated by addition of 200 μl stop solution [20 mM Tris−HCl pH 7.5, 0.1 M NaCl, 1% (v/v) SDS, 5 mM EDTA, 50 μg/ml proteinase K], incubated for 30 min at 45°C, extracted with phenol−chloroform and precipitated in ethanol with excess tRNA as carrier. Pellets were resuspended in 80% formamide in 1×TE, denatured for 5 min at 95°C and run on an 8% urea−polyacrylamide sequencing gel. Gels were dried and exposed overnight at −70°C to Kodak X-AR5 films.

Acknowledgements

Sequences described in this paper appear in the EMBL Nucleotide Sequence Data Library under the accession number X61755. Thanks are due to Denis Duboule, Claude Desplan, Sylvie Mader and Roberto Di Lauro for criticisms and suggestions, Juan Carlos Izpisùa-Belmonte for gift of the *Hox 4.3* cDNA clone, Paul T.Sharpe for sharing *Hox 3.4* genomic sequences prior to publication and Jacopo Meldolesi for continual support. The expert assistance of Nadia Matti and Valeria Ganazzoli in typing the manuscript is hereby gratefully acknowledged. This work was supported by grants from the Italian Association for Cancer Research (AIRC) and the Italian National Research Council (CNR), Special Projects 'Ingegneria Genetica' (to F.M.) and 'Biotecnologie e Biostrumentazione' (to E.B.). L.A. is a recipient of a San Raffaele fellowship.

References

Acampora,D., D'Esposito,M., Faiella,A., Pannese,M., Migliaccio,E., Morelli,F., Stornaiuolo,A., Nigro,V., Simeone,A. and Boncinelli,E. (1989) *Nucleic Acids Res.*, **17**, 10385−10402.
Angel,P., Imagawa,M., Chiu,R., Stein,B., Imbra,R.J., Rahmsdorf,H.J., Jonat,C., Herrlich,P. and Karin,M. (1987) *Cell*, **48**, 729−739.
Akam,M. (1987) *Development*, **101**, 1−22.
Akam,M. (1989) *Cell*, **57**, 347−349.
Blackwell,T.K., Kretzer,L., Blackwood,E.M., Eisenman,R. and Weintraub,H. (1990) *Science*, **250**, 1149−1151.
Boncinelli,E., Simeone,A., Acampora,D. and Mavilio,F. (1991) *Trends Genet.*, **7**, 329−334.
Brokes,J.P. (1989) *Neuron*, **2**, 1285−1294.
Carthew,R.W., Chodosh,L.A. and Sharp,P.A. (1985) *Cell*, **43**, 439−448.
Chen,C. and Okayama,H. (1987) *Mol. Cell. Biol.*, **7**, 2745−2752.
Chirgwin,J.M., Przybyla,A.E., MacDonald,R.J. and Rutter,W.J. (1979) *Biochemistry*, **18**, 5294−5300.
Cho,K.W.Y. and De Robertis,E.M. (1990) *Genes Dev.*, **4**, 1910−1916.
Cho,K.W.Y., Morita,E., Wright,C.V.E. and De Robertis,E.M. (1991) *Cell*, **65**, 55−64.
Condie,B.G., Brivanlou,A.H. and Hartland,R.M. (1990) *Mol. Cell. Biol.*, **10**, 3376−3385.
Dearolf,C.R., Topol,J. and Parker,C.S. (1989) *Nature*, **341**, 340−343.
De Robertis,E.M., Oliver,G. and Wright,C.V.E. (1989) *Cell*, **57**, 189−191.
Desplan,C., Theis,J. and O'Farrell,P.H. (1988) *Cell*, **54**, 1081−1090.
Dignam,J.D., Lebowitz,R.M. and Roeder,R.G. (1983) *Nucleic Acids Res.*, **11**, 1475−1489.
Dollé,P., Izpisùa-Belmonte,J.C., Falkenstein,H., Renucci,A. and Duboule,D. (1989) *Nature*, **342**, 767−772.
Duboule,E. and Dollé,P. (1989) *EMBO J.*, **8**, 1497−1505.
Fainsod,A., Bogarad,L.D., Ruusala,T., Lubin,M., Crothers,D.M. and Ruddle,F.H. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 9532−9536.
Gaunt,S.J., Sharpe,P.T. and Duboule,D. (1988) *Development*, **104** (suppl.), 71−82.
Gehring,W.J., Müller,M., Affolter,M., Percival-Smith,A., Billeter,M., Qian,Y.Q., Otting,G. and Wuethrich,K. (1990) *Trends Genet.*, **6**, 323−329.
Gehring,W.J. and Hiromi,Y. (1986) *Annu. Rev. Genet.*, **20**, 147−173.
Giampaolo,A. *et al.* (1989) *Differentiation*, **40**, 191−197.
Glass,C.K., Devary,O.V. and Rosenfeld,M.G. (1990) *Cell*, **63**, 729−738.
Gonzales-Reyes,A., Urquia,N., Gehring,W.J., Struhl,G. and Morata,G. (1990) *Nature*, **344**, 78−80.
Graham,A., Papalopulu,N. and Krumlauf,R. (1989) *Cell*, **57**, 367−378.
Green,S., Isseman,I. and Sheer,E. (1988) *Nucleic Acids Res.*, **16**, 369−373.
Greenberg,M.E. and Ziff,E.B. (1984) *Nature*, **311**, 433−438.
Hayashi,S. and Scott,M.P. (1990) *Cell*, **63**, 883−894.
Hoey,T. and Levine,M. (1988) *Nature*, **332**, 858−861
Holland,P.W.H. and Hogan,B.L.M. (1988) *Development*, **102**, 159−174.

Ingham,P.W. (1988) *Nature*, **335**, 25–34.
Izpisùa-Belmonte,J.C., Tickle,C., Dollé,P., Wolpert,L. and Duboule,D. (1991) *Nature*, **350**, 585–589.
Kessel,M., Balling,R. and Gruss,P. (1990) *Cell*, **61**, 301–308.
Kessel,M. and Gruss,P. (1990) *Science*, **249**, 374–379.
Krumlauf,R., Holland,P.W.H., McVey,J.H. and Hogan,B.L.M. (1987) *Development*, **99**, 603–617.
Lee,R.F., Concino,M.F. and Weinmann,R. (1987) *Cell*, **49**, 741–752.
Malicki,J., Schughart,K. and McGinnis,W. (1990) *Cell*, **63**, 961–967.
Mavilio,F., Simeone,A., Giampaolo,A., Faiella,A., Zappavigna,V., Acampora,D., Poiana,G., Russo,G., Peschle,C. and Boncinelli,E. (1986) *Nature*, **324**, 664–668.
McGinnis,M., Kuziora,M.A. and McGinnis,W. (1990) *Cell*, **63**, 969–976.
Melton,D.A. (1991) *Science*, **252**, 234–241.
Nicholson,R.C., Mader,S., Nagpal,S., Leid,M., Rochette-Egly,C. and Chambon,P. (1990) *EMBO J.*, **9**, 4443–4454.
Nohno,T., Noji,S., Koyama,E., Ohyama,K., Myokai,F., Kuroiwa,A., Saito,T. and Taniguchi,S. (1991) *Cell*, **64**, 1197–1205.
Odenwald,W.F., Taylor,C.F., Palmer-Hill,F.J., Friedrich,V., Tani,V., Tani,M. and Lazzarini,R.A. (1987) *Genes Dev.*, **1**, 482–496.
Odenwald,W.F., Garbern,J., Arnheiter,H., Tournier-Lasserve,E. and Lazzarini,R.A. (1989) *Genes Dev.*, **3**, 158–172.
Ohkuma,Y., Horikoshi,M., Roeder,R.G. and Desplan,C. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 2289–2293.
Oliver,G., De Robertis,E.M., Wolpert,L. and Tickle,C. (1990) *EMBO J.*, **10**, 3093–3099.
Ruiz i Altaba,A. and Melton,D.A. (1989) *Nature*, **341**, 33–38.
Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning. A Laboratory Manual*, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Sawadogo,M. and Roeder,R.G. (1985) *Cell*, **43**, 165–175.
Scott,M.P., Tamkun,J.W. and hartzell,G.W. III (1989). *Biochim. Biophys. Acta Rev. Cancer*, **989**, 25–48.
Selden,F.S., Burke Howie,K., Rowe,M.E., Goodman,H.M. and Moore,D.D. (1986) *Mol. Cell. Biol.*, **6**, 3173–3179.
Simeone,A., Mavilio,F., Bottero,L., Giampaolo,A., Russo,G., Faiella,A., Boncinelli,E. and Peschle,C. (1986) *Nature*, **320**, 763–765.
Simeone,A. *et al.* (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 4914–4918.
Simeone,A., Pannese,M., Acampora,D., D'Esposito,M. and Boncinelli,E. (1988) *Nucleic Acids Res.*, **16**, 5379–5390.
Simeone,A., Acampora,D., Nigro,V., Faiella,A., D'Esposito,M., Stornaiuolo,A., Mavilio,F. and Boncinelli,E. (1991) *Mech. Dev.*, **33**, 215–228.
Simeone,A., Acampora,D., Arcioni,L., Andrews,P.W., Boncinelli,E. and Mavilio,F. (1990) *Nature*, **346**, 763–766.
Studier,F.W., Rosenberg,A.H., Dunn,J.J. and Dubendorff,J.W. (1990) *Methods Enzymol.*, **185**, 60–89.
Thali,M., Müller,M.M., De Lorenzi,M., Matthias,P. and Bienz,M. (1988) *Nature*, **336**, 598–601.
Thomas,P. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 5201–5205.
Wright,C.V.E., Cho,K.W.Y., Hardwicke,J., Collins,R.H. and De Robertis,E.M. (1989) *Cell*, **59**, 81–93.
Wright,C.V.E., Cho,K.W.Y., Oliver,G. and De Robertis,E.M. (1989) *Trends Biochem. Sci.*, **14**, 52–56.
Zappavigna,V., Renucci,A., Izpisùa-Belmonte,J.C., Urier,G., Peschle,C. and Duboule,D. (1991) *EMBO J.*, **10**, 4177–4187.
Zwartkruis,F., Höijmakers,T., Deschmps,J. and Meijlink,F. (1991) *Mech. Dev.*, **33**, 179–190.

Received on July 29, 1991; revised on October 17, 1991