The EMBO Journal vol.10 no.13 pp.4177-4187, 1991

HOX4 genes encode transcription factors with potential auto- and cross-regulatory capacities

Vincenzo Zappavigna[2], Armand Renucci, Juan-Carlos Izpisúa-Belmonte, Gilbert Urier, Cesare Peschle[1] and Denis Duboule

EMBL, Meyerhofstrasse 1, 6900 Heidelberg, FRG and [1]Department of Haematology and Oncology, Istituto Superiore di Sanità, Viale Regina Elena 299, 00161 Rome, Italy

[2]Present address: Istituto San Raffaele, Via Olgettina 60, 20132 Milan, Italy

Communicated by P.Chambon

We have looked for the binding of several HOX4 complex homeoproteins in the genomic region containing the *HOX4C* promoter, between the human *HOX4C* and *HOX4D* genes. The HOX4C, HOX4D and Hox-4.3 homeoproteins bind to a phylogenetically highly conserved DNA fragment, which is located in the proximal part of this intergenic region and contains multiple binding sites for these HOX4 proteins. Using cotransfection experiments, we show that this endogenous DNA sequence can mediate transactivation by the HOX4D and HOX4C proteins and that this effect requires the presence of TAAT-related binding sites. The Hox-4.3 protein, in contrast, is unable to activate and can repress the activation observed with the two other proteins. These results show that the *HOX4D* and *HOX4C* genes are genuine sequence-specific transcription factors and suggest that, as in *Drosophila*, cross-regulatory interactions between these genes might be essential for their proper expression.

Key words: collinearity/homeobox/homeodomain/transcription

Introduction

Genes containing a homeobox have been identified in a large number of species belonging to different phyla (reviewed in Scott *et al.*, 1989). In vertebrates, the best characterized families of homeobox-containing genes (*Hox* genes) are those found in mice and humans. As in *Drosophila*, mouse and human *Hox* genes are clustered. The comparison between the mammalian gene clusters (the *Hox* complexes) and the *Drosophila Antennapedia* and *Bithorax* complexes (ANT-C and BX-C) has revealed striking similarities both in structural organization and patterns of expression. Indeed the four *Hox* clusters identified in the mouse and human genomes (e.g. Acampora *et al.*, 1989; Kessel and Gruss, 1990) are similar in structure to the *Drosophila* ANT-C and BX-C complexes since cognate genes can be identified at equivalent positions in the four clusters which correlate with a particular *Drosophila* homeotic gene (Duboule and Dollé, 1989; Graham *et al.*, 1989; Kappen *et al.*, 1989). The analysis of the expression patterns of these genes during mouse embryonic and fetal development further supports the idea of a phylogenetic link since, in both *Drosophila* and vertebrates, the expression of these genes along the anterior-posterior (AP) axis of the developing body correlates with their position along the chromosome (the structural collinearity; Gaunt *et al.*, 1988; Akam, 1989).

The 61 amino acid peptide sequences encoded by the homeobox (the homeodomain, HD) is a DNA-binding domain structurally related to the α helix-turn-helix motif found in several prokaryotic regulatory proteins (Laughon and Scott, 1984). The presence of this motif strongly suggests that *Hox* genes are transcription factors which could act as master control genes on sets of as yet unidentified target genes. Particular combinations of such regulatory proteins could thereby select distinct classes of target genes conferring unique identity to cells or structures (e.g. a parasegment) in which they are expressed (e.g. Peifer *et al.*, 1987) although differential regulation of the same target genes could also lead to the same result. Some *Drosophila Hox* genes which encode proteins with a nuclear localization seem to play a role in direct, sequence-specific regulation of transcription. Such proteins expressed in *Escherichia coli*, both as full-length or fusion products, can bind *in vitro* to a wide variety of target sequences. Cell transfection as well as *in vitro* transcription experiments have demonstrated that *Drosophila Hox* genes are able to modulate transcription positively or negatively in a DNA binding-site dependent manner (reviewed by Hayashi and Scott, 1990). Several binding sites for *Drosophila* homeodomain proteins are found in promoters of genes belonging to the genetic network which controls development of the fruit flies. Auto- and/or cross-regulatory interactions within this network have been described by genetic analysis and partially reproduced in tissue culture transfection assays (Jaynes and O'Farrell, 1988; Han *et al.*, 1989; Krasnow *et al.*, 1989; Winslow *et al.*, 1989). Such interactions are thought to play a major role in the refinement and maintenance of the expression patterns of the homeotic genes.

The DNA-binding properties of the vertebrate *Hox* genes are similar to their cognate *Drosophila* genes with respect to target site specificities (Fainsod *et al.*, 1986; Odenwald *et al.*, 1989; Sasaki *et al.*, 1990) though their ability to act as transcriptional regulators has not been documented so far. We have studied the possible existence of regulatory interactions between mammalian *Hox* genes. Genes belonging to the murine and human HOX4 cluster have recently been cloned and characterized (Duboule *et al.*, 1990; Izpisúa-Belmonte *et al.*, 1991a; Simeone *et al.*, 1991). The developmental expression domains of these genes are collinear with their positions in the cluster with a progressive posterior restriction when moving in the 5′ part of the complex. The five last 'posterior' genes are all related to the *Drosophila AbdB* homeotic gene (Izpisúa-Belmonte *et al.*, 1991a) whose function in insects is to specify the posterior parasegments (Duncan, 1987 and references therein). In the vertebrates, these genes are also expressed

© Oxford University Press

in overlapping domains in the mesenchyme of the developing limb buds, indicating that similar molecular mechanisms are involved in the specification of positional information in all axial structures of the body (Dollé *et al.*, 1989). We proposed earlier that this patterning mechanism could be partially due to the progressive activation, in time, of this battery of genes from 3′ positions to 5′ positions in the complex (the temporal collinearity; Dollé *et al.*, 1989; Izpisúa-Belmonte *et al.*, 1991a,b). The coordinate expression (in time and space) of these vertebrate *AbdB*-related genes appears therefore to be a critical factor for the proper development of the structures along the body axes and this could be achieved in part by cross-talk between the various members of the HOX complexes.

We have used the human *HOX4C* locus as a paradigm to study possible regulatory interactions between vertebrate *Hox* genes. We report here the isolation of cDNA clones for the *HOX4C* and *HOX4D* genes and the deduced sequences of the corresponding proteins. We searched for Hox binding sites in the *HOX4C* promoter and show that a DNA sequence which is highly conserved between mammals and birds (the Hox cross-talk region, HCR) specifically binds, *in vitro*, the products of the *HOX4D*, *HOX4C* and *Hox-4.3* genes. The HCR sequence is able to mediate transactivation by the *HOX4D* and *HOX4C* gene products. It therefore represents a possible target region for cross- and auto-regulatory interactions. This work shows that mammalian *Hox* genes could be sequence-specific transcription factors and that various Hox proteins can have different effects since the product of the *Hox-4.3* gene, while unable to transactivate through HCR, represses the activation by *HOX4D* and *HOX4C* in a DNA-binding independent manner.

Results

Cloning and sequence analysis of human HOX4D and HOX4C cDNAs and genomic region

We isolated human cDNAs from the cognate genes of the mouse *AbdB*-related *Hox4* genes (Izpisúa-Belmonte *et al.*, 1991a) from a library made out of human 7 weeks p.c. embryonic spinal cord mRNAs. An ~3.4 kb cDNA clone, representing the transcript of the human homologue of the murine *Hox-4.5* gene, termed *HOX4D* (Acampora *et al.*, 1989), as well as an ~2.1 kb cDNA from the *HOX4C* gene (murine *Hox-4.4*) were thus isolated. Their structures and locations within the HOX4 complex are depicted in Figure 1. Both cDNAs appear to be nearly full length since probing poly(A)$^+$ RNA from human embryonic spinal cord reveals transcripts for *HOX4D* and *HOX4C* of ~3.5 and ~2.2 kb respectively (data not shown). The *HOX4C* and *HOX4D* genes are composed of two exons, of which the second encodes the homeodomain. Figure 2 shows the sequence of the *HOX4D* and *HOX4C* open reading frames which encode proteins of 340 and 342 amino acids, respectively. These two proteins are highly similar to their murine counterparts.

In order to characterize sequences of importance for the regulation of *HOX4C* expression, we cloned the genomic region spanning the *HOX4D* and *HOX4C* transcription units. A phage clone comprising ~15 kb of genomic DNA was isolated and shown to contain both the *HOX4D* and *HOX4C* genes (Figure 1). Comparison of the structure of this clone with its murine counterparts revealed a high conservation of the genomic organization in this part of the HOX4 cluster. A DNA fragment of ~2.5 kb extending from the end of the HOX4D transcription unit to the start of the HOX4C cDNA clone and which contains *Hox-4.4* promoter sequences (A.Renucci, V.Zappavigna and D.Duboyle, in preparation), was subcloned and sequenced. The alignment between the corresponding human and mouse sequences revealed highly conserved stretches of nucleotides (Figure 3A), the longest one being located in the most proximal part and containing a DNA sequence which is highly conserved in mammals and chicken (region A, Figure 3A and D).

Identification of target sequences for HOX4 gene products

Full-length proteins encoded by three neighbouring genes, *HOX4D*, *HOX4C* and *Hox-4.3*, the mouse version of the human *HOX4E*, were produced in *E.coli*. These proteins were used in gel retardation experiments in order to look for specific binding of DNA sequences located within the conserved sequences of the intergenic region. A 300 bp DNA fragment, spanning the highly conserved region A (Figure 3A and D) is retarded by the product of the *HOX4D* gene (Figure 3B). At least four complexes can be seen, three of which are specifically and gradually competed out by an oligonucleotide (SC) containing several Hox 'consensus' binding sites comprising the sequence TCAATTAAAAT (Hoey and Levine, 1988; Desplan *et al.*, 1988). This

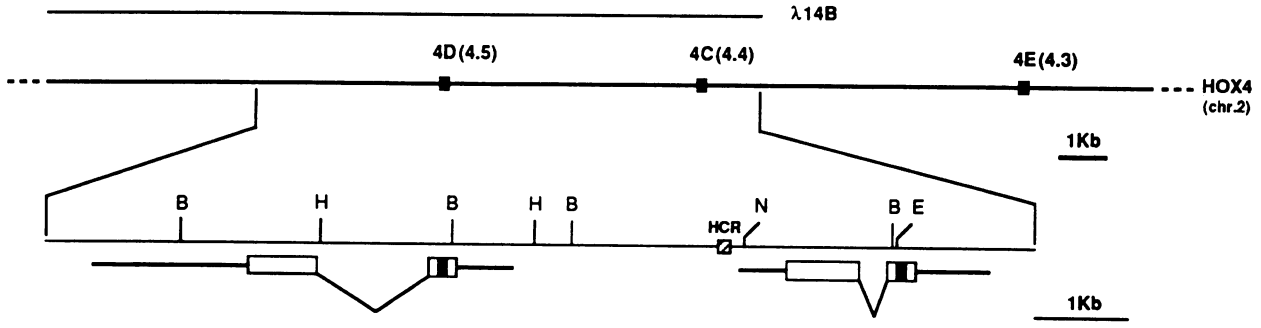

Fig. 1. Partial structure of the human HOX4 genomic region. The area depicted represents the part of the HOX4 cluster which comprises the *HOX4D*, *HOX4C* and *Hox-4.3* genes. The names of the corresponding mouse homologues are indicated in brackets (see Duboule *et al.*, 1990). The solid line on the top shows the extension of the genomic clone λ14B. The structure of the genomic subregion containing the *HOX4D* and *HOX4C* genes is enlarged at the bottom. Transcription is from left to right. Below are shown the structures of two cDNA clones corresponding to the *HOX4D* and *HOX4C* genes. Coding sequences are indicated by open boxes and the homeodomains by black boxes. The striped box shows the positions of the *Hox* genes cross-talk region (HCR, see text). B, *Bam*HI; E, *Eco*RI; H, *Hin*dIII; N, *Not*I.

oligonucleotide is bound by the HOX4D homeoprotein (not shown). No competition is observed when increasing amounts of a non-specific oligonucleotide (not bound by the *HOX4D* product) is used (NSC, Figure 3B). The formation of different complexes suggests that several HOX4D protein molecules can simultaneously occupy multiple binding sites.

A highly conserved 90 bp region contains multiple binding sites for the HOX4D, HOX4C and Hox-4.3 proteins

The sequences within region A which are specifically bound by the HOX4D, HOX4C and Hox-4.3 proteins were characterized using DNase I footprinting analysis. The same 300 bp DNA fragment was used and revealed that a sequence of ~90 bp, within region A, contains several sites protected from DNase I digestion by the HOX4D, HOX4C and Hox-4.3 proteins (Figure 3C). The patterns of HOX4D and HOX4C protection are virtually identical but partially different from that obtained with the Hox-4.3 protein. The locations of enhanced DNase I cleavage sites suggest a probable simultaneous binding of more than one protein molecule to this fragment (see also Figure 6B). No protection could be observed with an extract from bacteria transformed with the expression vector without insert (not shown). The nucleotide sequence of this 90 bp DNA region and its alignment with the corresponding mouse DNA fragment revealed a 100% conservation (Figure 3D). This region was also identified at a similar position upstream of the chicken *HOX4C* homologue (Figure 3D). Only two nucleotides differ between chicken and mammals in the footprinted part. In human and mouse, the *HOX4C* (*Hox-4.4*) gene seems to have multiple transcription start sites (A.Renucci, V.Zappavigna and D.Duboule, in preparation). The footprinted region lies ~300 bp 3′ to one of the characterized transcription start sites. Weaker start sites could be present immediately downstream and largely upstream of this sequence. The protected sequences differ both in their extent and base compositions, the only recurrent element (in three out of five footprinted parts) being a TAAT (or the complementary ATTA) motif which is a common thread of many *Drosophila* and vertebrate Hox protein binding sites (e.g. Odenwald *et al.*, 1989; Hayashi and Scott, 1990; Sasaki *et al.*, 1990). The remaining protected parts contain related (TAA or ATT) A/T-rich DNA motifs (Figure 3D).

HOX4D and HOX4C homeoproteins can stimulate reporter activity through the 90 bp binding region

We used transient cotransfection assays to test if the 90 bp footprinted region was able to mediate transcriptional regulation by the *Hox* gene products which bind to it *in vitro*. We chose a heterologous promoter−reporter system because of the difficulty of identifying a cellular system in which the complete HOX4C promoter proves to be significantly active. We therefore introduced the 90 bp region (HCR, Figure 3) upstream of the luciferase reporter plasmid (pT81luc, Nordeen, 1988; Figure 4 and Materials and methods). The eukaryotic expression vector pSG5 (Green *et al.*, 1988; Figure 4) was used to produce the HOX4C and HOX4D proteins in cotransfection assays of mouse NIH3T3 cells. Cotransfection of the pSGH4D and pSGH4C (Figure 4)

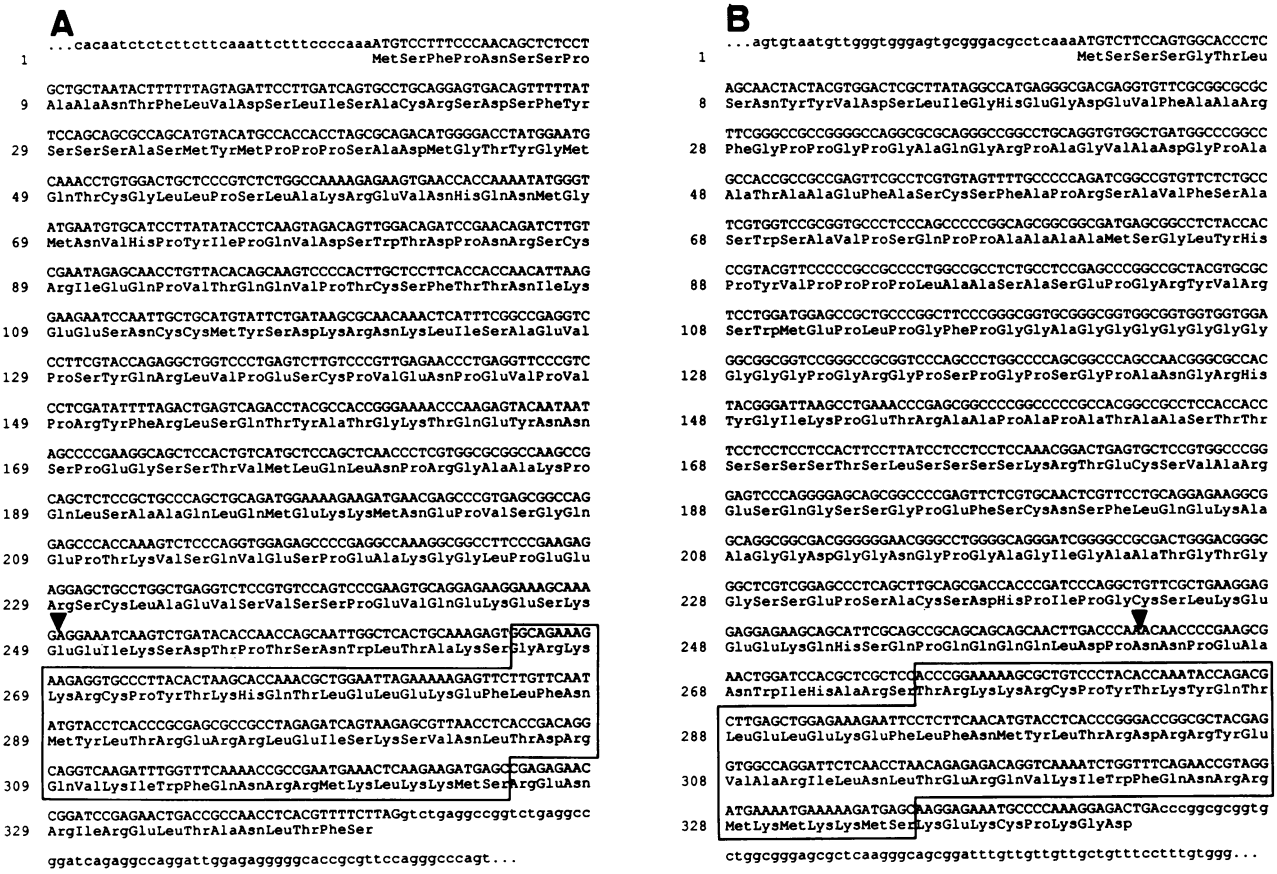

Fig. 2. Nucleotide sequences and conceptual translations of the open reading frames of the *HOX4D* (**A**) and *HOX4C* (**B**) cDNA clones. The homeobox (homeodomain) sequences are framed and the positions of the respective introns are indicated by black triangles.

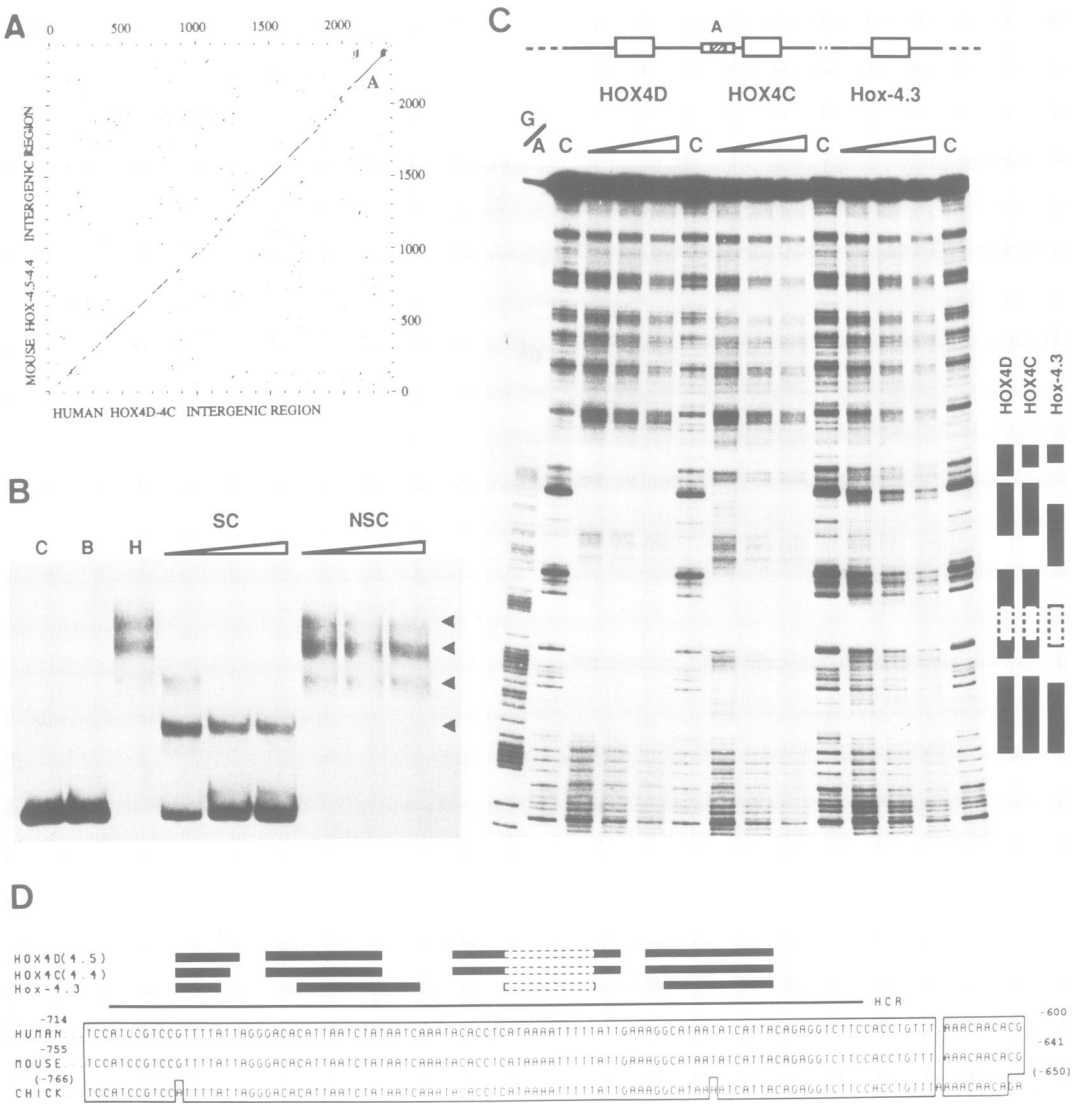

Fig. 3. Identification of a HOX4 protein binding region upstream of the *HOX4C* cDNA clone. **A**, Dot matrix plot of the alignment between the human *HOX4D–HOX4C* (*x* axis) and the corresponding mouse *Hox-4.5–Hox4.4* (*y* axis) intergenic regions. The values indicated are in base pairs. The origin represents the 3′ end of the *HOX4D* transcription unit (polyadenylation site) in both species. The comparison extends for 2323 bp and 2403 bp in mouse and human, respectively, up to the start of the *HOX4C* cDNA clone. Several regions show high sequence conservation. The region of longest and highest conservation (called region A) is shown. **B**, Gel retardation analysis of a 300 bp end-labelled fragment spanning region A. Lane C: free probe (~0.5 ng). Lane B: probe incubated with 0.5 μg (total protein) of an extract from bacteria transformed with the expression vector lacking an insert. Lane H: probe incubated with 0.5 μg (total protein) of a bacterial extract containing the HOX4D protein. Lanes SC: competition with 100-, 500- or 1000-fold molar excess of a cold 'specific' competitor (see Materials and methods). Lanes NSC: competition with 100-, 500- and 1000-fold molar excess of a 'non-specific' competitor. The black arrowheads indicate the retarded complexes. **C**, DNase I footprinting analysis of the binding by the HOX4D, HOX4C and Hox-4.3 proteins to the region A-containing probe. Lanes C: control digestions in the absence of HOX protein containing bacterial extracts. Lane G+A: purine ladder of Maxam and Gilbert sequencing reactions. Increasing amounts of bacterial extracts were used for the three proteins: (left to right) 0.5, 1.0 and 1.5 μg (total protein) for HOX4D; 1.0, 2.0 and 3.0 μg (total protein) for HOX4C and Hox-4.3. Above, schematic representation of the relative positions of the three genes (open boxes) in the human (and mouse) genome together with the position of region A (small open box) and the footprinted part within it (striped box). The patterns of protection for each protein are indicated on the right. Dashed lines indicate a region resistant to DNase I digestion. **D**, Sequence of the region containing the footprinted parts and its alignment with the corresponding regions in the mouse and chicken genomes. Conserved sequences are framed. Positions are given from the ATG of the *HOX4C* gene and its respective homologues in mouse and chicken. Values for chicken are in brackets to indicate estimated distances. On the top, positions of the protected parts (black bars) for each protein are indicated with respect to the sequence. The solid line indicates the part of the HCR used to generate the reporter construct pTHCR (see Figure 4).

expression constructs results in an average of ~8-fold and ~10-fold induction of the reporter activity, respectively, as compared with control cotransfections with the reporter vector lacking the HCR fragment (Figure 5). No significant changes in luciferase activity are observed when the pT81luc reporter vector is cotransfected with the pSG5 vector or with the pSGH4D or pSGH4C protein expression constructs (Figure 5). Similarly, the basal activity of the pTHCR reporter construct was not significantly different from that of the pT81luc vector (Figure 5). We therefore conclude that the HCR fragment can mediate transactivation by the HOX4C and HOX4D proteins.

Activation of the HOX4D protein is mediated by TAAT-containing sites

Five TAAT (ATTA) sequence motifs are present in the HCR fragment (boxed and numbered I to V in Figure 6A). Since this motif has been previously proposed as a 'core consensus'

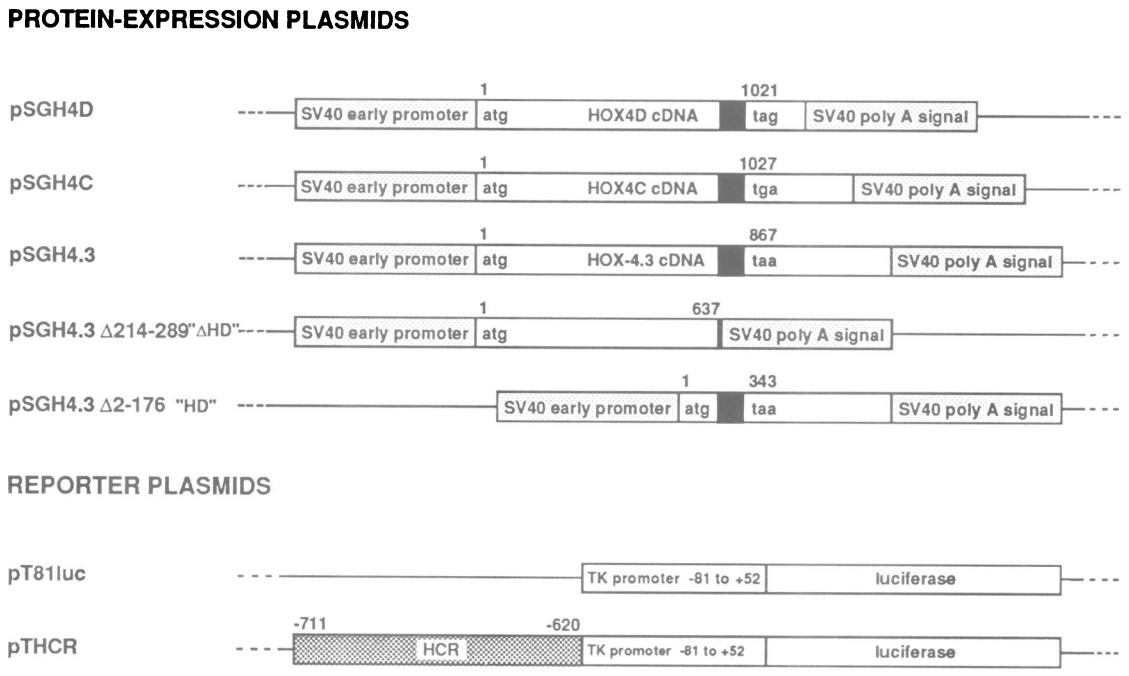

Fig. 4. Constructs used in cotransfection assays. Protein expression plasmids: the coding sequences of the *HOX4D*, *HOX4C* and *Hox-4.3* genes and the two truncated Hox-4.3 proteins were cloned into the pSG5 eukaryotic expression vector. The numbers indicate the positions of the respective ATGs and stop codons. Reporter plasmids: the reporter construct pTHCR was generated by cloning the HCR fragment (see Figure 3D) into the pT81luc reporter plasmid. Numbers above indicate the distance in base pairs of the region with respect to the ATG of the *HOX4C* gene.

for the binding of some homeodomain proteins (Müller *et al.*, 1988; Odenwald *et al.*, 1989; Scott *et al.*, 1989; Kissinger *et al.*, 1990; Otting *et al.*, 1990), the HCR region could contain at least five distinct binding sites for Hox proteins. This number rises to eight when considering sequence motifs which differ from the 'core consensus' by only one nucleotide, but which, at least in two cases, are bound by the HOX4D and HOX4C proteins (Figure 3D). We tested whether the transactivation of the reporter construct is indeed mediated by the TAAT-containing sites bound *in vitro* by selectively inactivating by mutagenesis some of the TAAT 'cores' present in HCR. T or A nucleotides were changed to G or C (Figure 6A) since such substitutions are likely to disrupt interactions with the amino acids of the homeodomain contacting the TAAT sequence (Kissinger *et al.*, 1990). The effect of these mutations on the *in vitro* binding of the HOX4D protein was analysed by DNase I footprinting analysis of the mutant sequences (Figure 6B). The results show that the disruption of TAAT motifs leads in all cases to an abolition of the binding, to the respective site, as revealed by a 'release' of the footprint in the corresponding region. These various mutants were further used in cotransfection experiments in order to evaluate their potentials to mediate transactivation (Figure 7). An alteration of the five TAAT motifs, and of one A/T-rich site (the HCR* mutant in Figure 6A), results in a drastic reduction of the transactivation levels. The small but reproducible level of activation which is still observed is possibly due to the binding of the protein to a non-mutated A/T-rich (TTATT) site, which is footprinted by the HOX4D protein (right-most panel in Figure 6B). The disruption of single TAAT motifs within HCR generally results in a reduction in the transactivation levels (e.g. HCR I*, HCR II* and HCR IV*, Figures 6A and 7) though not observed for all the sites (HCR V*). Similar results are obtained with the HOX4C protein (not shown). The analysis of HCR deletion mutants further

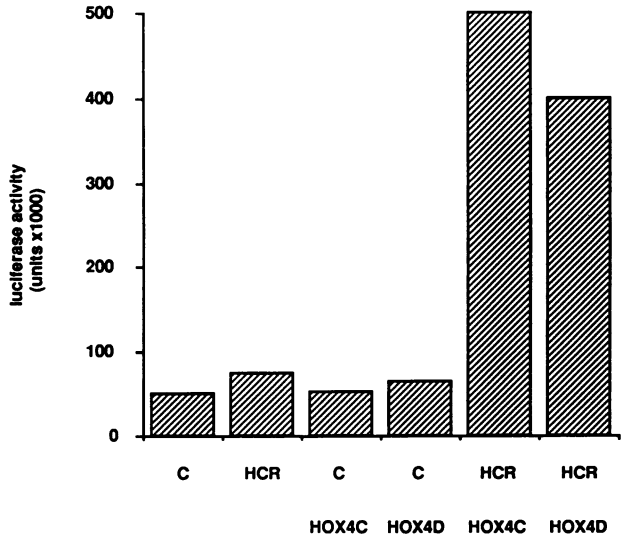

Fig. 5. HOX4D- and HOX4C-dependent stimulation of the pTHCR reporter activity. Luciferase activity, in light units (abscissa), assayed from extracts of cotransfected NIH3T3 cells. Transfected plasmids and their quantities were: (C), 10 μg of pT81luc and 5 μg of pSG5 expression vector without insert; (HCR), 10 μg of pTHCR and 5 μg of pSG5; (C/HOX4C), 10 μg of pT81luc and 5 μg of pSGH4C; (C/HOX4D), 10 μg of pT81luc and 5 μg of pSGH4D; (HCR/HOX4C), 10 μg of pTHCR and 5 μg of pSGH4C; (HCR/HOX4D), 10 μg of pTHCR and 5 μg of pSGH4D. 1 μg of pRSVβgal (see Materials and methods) was cotransfected in all the points as an internal standard. Values represent the average of the results from at least six independent experiments for each point.

supports the existence of possible differences in the relative importance of various HCR 'TAAT' sites with respect to the transactivation by HOX4D. HCR α, a deletion of 37 nucleotides in the 3′ part of HCR (Figure 6A), can mediate transactivation at a level slightly lower than HCR whereas a deletion of 55 nucleotides in the 5′ part of HCR (HCR

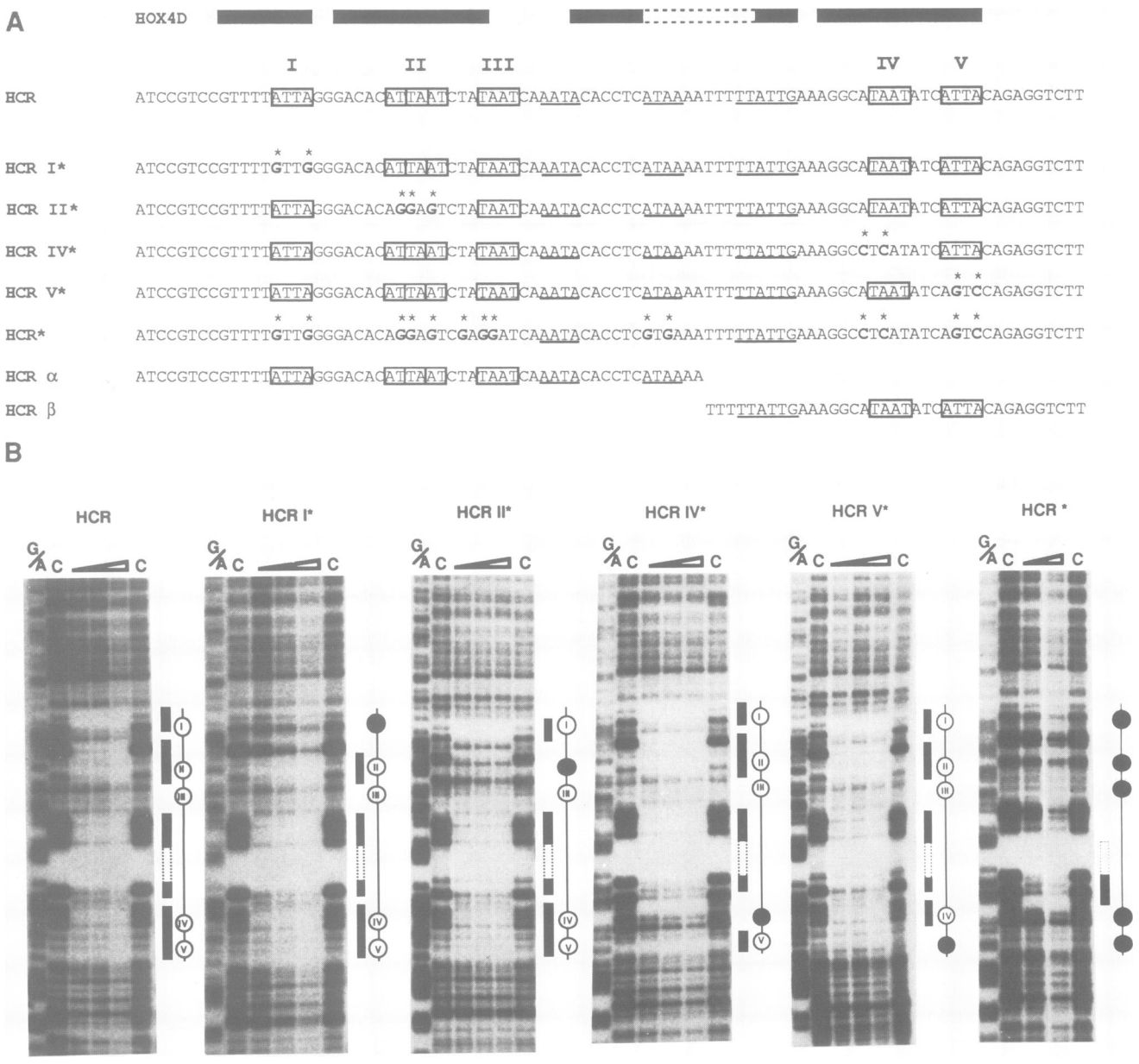

Fig. 6. Mutational analysis of the HCR sequence. **A**, Mutant forms of HCR were generated by substituting Ts or As in the TAAT (or ATTA) core sequence (boxed and numbered on top from I to V) with Gs or Cs (indicated by bold letters). Positions of the substitutions are marked by asterisks. Underlined are indicated motifs related to the TAAT 'core consensus'. Mutants HCRα and HCRβ represent deletions of the 3′ and 5′ parts of HCR, respectively. On the top the pattern of protection by the HOX4D protein on the non-mutated HCR. All mutants were cloned in the pT81luc reporter plasmid (see Materials and methods). **B**, DNase footprinting analysis of the binding of the HOX4D protein to the mutated forms of HCR. On the left, as a comparison, the footprinting of the non-mutated HCR. Open numbered circles represent the TAAT cores as shown in A, filled circles represent their mutated forms. Patterns of protection are indicated by black vertical bars. (C): control DNase I digestions of the probes in the absence of the bacterial extract containing the HOX4D protein. (G/A): Maxam and Gilbert purine ladder. Increasing amounts of bacterial extract containing the HOX4D protein (0.5, 1.0 and 1.5 μg of total protein, left to right) were used in all the experiments shown with the exception of the HCR* mutant where only 0.5 μg and 1.5 μg (total protein) of bacterial extract were used. In all experiments, an end-labelled *Bgl*III–*Nde*I fragment of the respective mutant pTHCR reporter constructs was used as probe.

β, Figure 6A), causes a more substantial reduction in the transactivation levels (Figure 7).

The Hox-4.3 homeoprotein represses activation by the HOX4D and HOX4C proteins

Cotransfection experiments using the pSGH4.3 expression construct (Figure 4) show that this protein is unable to exert any significant transactivating effect on the pTHCR reporter construct (Figure 8), even though a weak binding to HCR is observed (Figure 3). To investigate possible combinatorial actions of the HOX4 proteins through the HCR sequence, we performed cotransfections using various combinations of equal amounts of pSGH4D or pSGH4C and pSGH4.3 expression constructs (Figure 8). Cotransfection of the Hox-4.3 expressor mixed to pSGH4D or pSGH4C resulted in a reduction in the levels of activation of ~50% in both cases. Cotransfection of the same amounts of the pSGH4D and pSGH4C gives an average activity which is comparable to that obtained with either pSGH4D or pSGH4C alone. No significant variations of the reporter activity were observed when double amounts of the HOX4D and HOX4C expressors were transfected, which indicates that a saturation

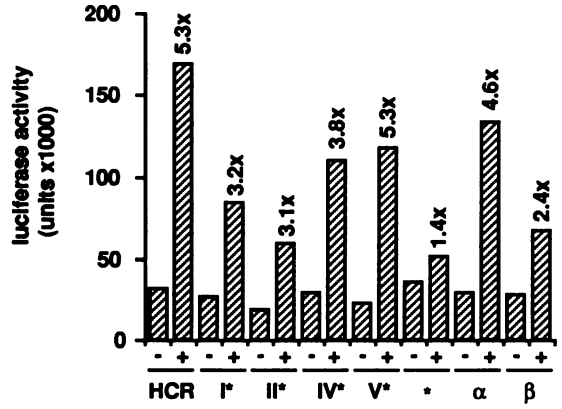

Fig. 7. Effects of the HCR mutations on the HOX4D-dependent stimulation of the reporter construct. Luciferase activity assayed from cotransfection experiments with the mutated form of HCR (Figure 6A). (−): cotransfection of 10 μg of the indicated reporters with 5 μg of pSG5 expression vector without coding sequences. (+): cotransfection of 10 μg of the reporters with 5 μg of the HOX4D expression construct (pSGH4D). Activation values (with respect to the basal reporter activity) are indicated at the top of the respective bars. Values were obtained by averaging the results of at least four independent experiments for each point. 1 μg of the pRSVβgal plasmid was included as an internal control in each cotransfection.

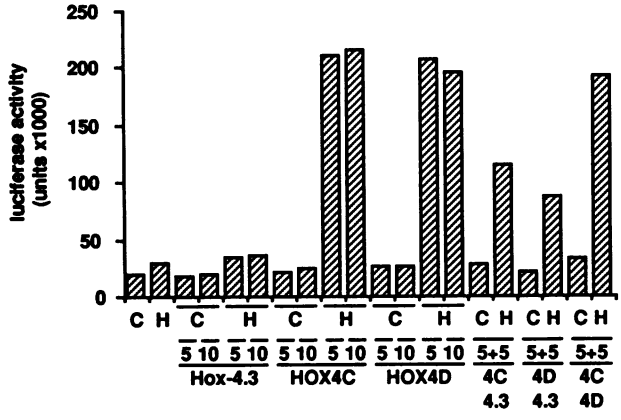

Fig. 8. Effects of the Hox-4.3 protein on the reporter activity and on the activation of HOX4D and HOX4C. Cotransfections were performed using either 10 μg of pT81luc reporter plasmid (C) or 10 μg of pTHCR (H) with pSGH4.3 (*Hox-4.3*), pSGH4C (*HOX4C*), and pSGH4D (*HOX4D*), each one in two different amounts, 5 μg and 10 μg. Cotransfections of combinations of two different protein expressors were performed as follows: (4C/4.3; 5+5), 5 μg of pSGH4D and 5 μg of pSGH4.3; (4D/4.3), 5 μg of pSGH4D and 5 μg of pSGH4.3; (4C/4.3), 5 μg of pSGH4C and 5 μg of pSGH4D. Values represent the average of the results of at least five independent experiments.

of the responding capacity of the reporter construct is reached (Figure 8). Cotransfections of various combinations of HOX4C and HOX4D, in non-saturating amounts, led to additive levels of activation (not shown). We further analysed to what extent the HCR-mediated activation by HOX4D and HOX4C could be antagonized by the Hox-4.3 protein. Different amounts of pSGH4.3 expressor were transfected with fixed amounts of either pSGH4D or pSGH4C (Figure 9A and B). The HCR-dependent activations by HOX4D and HOX4C are decreased in a similar way by increasing the amount of cotransfected Hox-4.3 expressor, which suggests that similar mechanisms are involved in both cases. This effect is reproducibly observed even with a ratio of activator versus Hox-4.3 expressor of 1:10. As a comparison, when increasing amounts of HOX4D expressor were transfected with a fixed amount of pSGH4C, no significant variation of the activation levels was observed (Figure 9A).

Hox-4.3 repression does not require DNA binding

Since Hox-4.3 binds, *in vitro*, the same and/or adjacent sites within HCR, a competition for the binding sites by a non-activating Hox-4.3 homeoprotein or, alternatively, quenching, appeared as possible explanations for this negative effect. We used two truncated versions of the Hox-4.3 protein to investigate these possibilities. The first one (pSGH4.3Δ214-289 in Figure 4 or 'ΔHD') is a Hox-4.3 protein with its homeodomain deleted. In the second one, the protein produced by the pSGH4.3Δ2-176 expressor (or 'HD') consists essentially of the homeodomain (Figure 4). Surprisingly, cotransfection of increasing amounts of the 'ΔHD' construct with a fixed amount of either HOX4D or HOX4C expressors led to a decrease in the activation levels comparable to that observed when transfecting equivalent amounts of the construct expressing the complete Hox-4.3 protein (Figure 9C). When equal amounts of 'ΔHD' and pSGH4D or pSGH4C are cotransfected, reproducible reductions of 60% to 50% of the activation are observed. No significant effect on the activation levels of HOX4D or HOX4C are observed when cotransfected with increasing amounts of the construct 'HD' expressing the Hox-4.3 homeodomain. The repression can thus be mediated by the Hox-4.3 protein domains from amino acids 1−214 and therefore does not appear to rely on DNA binding.

Discussion

The five genes located in the 5′ part of the HOX4 cluster are all related to the *Drosophila AbdB* homeotic gene (Izpisúa-Belmonte *et al.*, 1991a and references therein). They show highly coordinated expression patterns, in space and time, along both the cranio-caudal axis of the embryo and the developing limbs (Dollé *et al.*, 1989; Izpisúa-Belmonte *et al.*, 1991b). In both cases, their overlapping but displaced domains of expression suggest that these genes might encode positional information. In order to explore the possibility that cross-regulatory interactions between these genes could play a role in the establishment and maintenance of their expression patterns, as well as to show that these vertebrate *Hox* genes are genuine sequence-specific transcription factors, we searched for possible HOX4 homeoprotein binding sites within the HOX4D−HOX4C intergenic region. *In vitro* binding assays using various bacterially produced HOX4 proteins and sequence comparisons between human, mouse and chicken *HOX4* (*Hox-4.4*) promoter regions have led to the identification of an ~90 bp region in the proximal part of the *HOX4C* transcription unit which contains multiple sites specifically bound by the protein products of the *HOX4C*, *HOX4D* and *Hox-4.3* genes. This region (HCR) shows an almost complete conservation between mammals and birds and this intriguing observation thus suggests a critical function for this sequence.

HOX4C and HOX4D genes encode sequence-specific transcription factors

The sites within HCR which are bound by the Hox proteins, while slightly different from each other, contain a common TAAT (or ATTA) sequence. This motif has been proposed

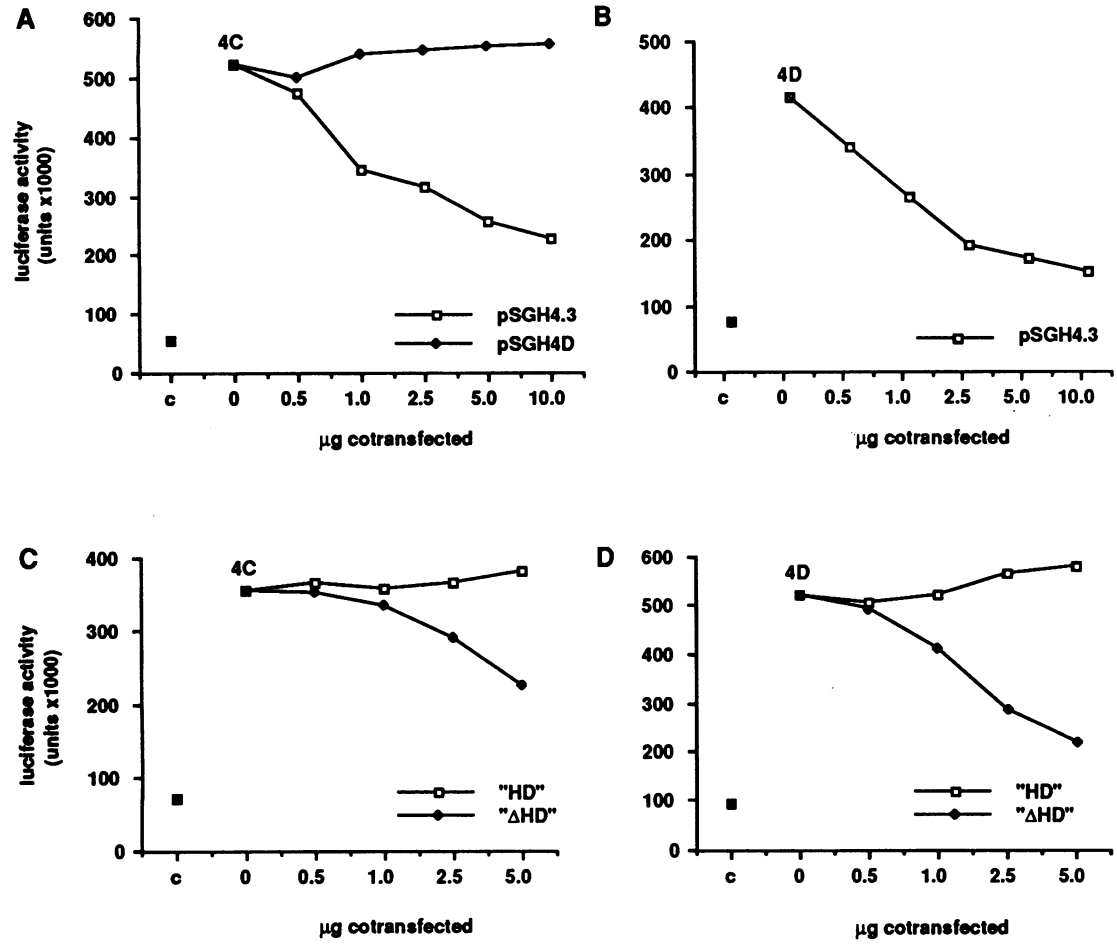

Fig. 9. Analysis of the repression by the Hox-4.3 protein of the activation by HOX4D and HOX4C. **A**, Different amounts of Hox-4.3 (pSGH4.3) or HOX4D (pSGH4D) expressors (abscissa) were cotransfected with a constant amount (5 μg) of HOX4C expressor (pSGH4C). The total amount of transfected DNA was kept constant by complementing up to 10 μg with the pSG5 expression vector without coding sequences. (c): transfection of 10 μg of the pTHCR reporter with 10 μg of pSG5. **B**, Different amounts of Hox-4.3 expressor (abscissa) were cotransfected with a constant amount (5 μg) of HOX4D expressor (pSGH4D). (c): same as in A. **C**, Increasing amounts of the expressors pSGH4.3Δ214-289 ('ΔHD') (abscissa) or pSGH4.3Δ2-176 ('HD') were cotransfected with a constant amount (5 μg) of HOX4C expressor (pSGH4C). **D**, Cotransfections with the same amounts of ('ΔHD') and ('HD') as in C with a constant amount (5 μg) of HOX4D expressor (pSGH4D). (c): cells transfected with 10 μg of the pTHCR reporter with 10 μg of the pSG5 expression vector not containing *Hox* coding sequences.

as a 'core consensus' for sequences bound by the class I homeobox proteins, since it can be identified as a sub-element in nearly all the sites characterized so far using *Drosophila* or vertebrate HOM/Hox proteins *in vitro* (reviewed in Scott *et al.*, 1989 and Hayashi and Scott, 1990; Ekker *et al.*, 1991). Furthermore, methylation and ethylation interference or X-ray crystallographic studies and NMR spectroscopy of homeobox proteins bound to DNA have shown that the TAAT sequence is a critical point of contact between the DNA and the homeodomains. This is further supported by mutagenesis of the TAAT 'core' which reduces binding drastically (Desplan *et al.*, 1988; Jaynes and O'Farrell, 1988; Odenwald *et al.*, 1989). Sequences flanking the TAAT motif could also influence the binding and/or recognition affinities of a Hox protein for a given site. The HCR fragment contains from five to eight individual sites able to be bound by Hox proteins. The remarkable phylogenetic conservation in base composition of the HCR sequence suggests that the number, DNA sequences and spatial array of the binding sites are essential for the regulation of the *HOX4C* gene expression. Together with the results obtained with mutant HCR sites, this indicates that such repetition of Hox binding sites might be used, *in vivo*, for qualitative discrimination rather than for increasing the strength of a potential regulatory function.

We have thus identified sequences in the HOX4C promoter region which are potentially responsible for mediating regulatory interactions by other members of the *HOX* gene family. We show that the HCR region can mediate transactivation when cotransfected with the product of the *HOX4C* gene itself, suggesting the existence of a positive autoregulatory circuit for *HOX4C*. Activation through the HCR fragment is also observed by the product of *HOX4D*, the gene located immediately 5′ to *HOX4C*, which shows that the similarities between the homeodomain sequences and the *in vitro* binding capacities of *HOX4C* and *HOX4D* are also observed in this *in vivo* situation. It further suggests a possible control of *HOX4C* expression by the neighbouring gene *HOX4D*. The activation potential of these homeoproteins is dependent upon the integrity of the *in vitro* identified binding sites within HCR since alterations of their TAAT cores led to both abolition of binding (as tested by DNase footprinting) and decrease in the levels of activation (in cotransfection assays). These experiments demonstrate that mammalian *Hox* genes are able to regulate transcription directly in a DNA binding dependent manner.

HCR and HOX cross-regulatory interactions

In *Drosophila*, genetic analyses have revealed auto- and/or cross-regulatory interactions in the network of homeotic genes and more recently, some of these interactions have been successfully reproduced in tissue culture transfection assays (Jaynes and O'Farrell, 1988; Han *et al.*, 1989; Krasnow *et al.*, 1989; Winslow *et al.*, 1989). These regulatory interactions are believed to play a major role in the maintenance and refinement of the expression patterns of the homeotics (reviewed in Hayashi and Scott, 1990), even though some of these genetically defined interactions might not have a real functional significance (González-Reyes *et al.*, 1990; González-Reyes and Morata, 1990). As for *Drosophila*, vertebrate class I homeobox genes are expressed in partially overlapping domains. The hypothesized combinatorial action of such proteins to regulate target genes in insects (e.g. Jaynes and O'Farrell, 1988 and references therein), can therefore be applied to the vertebrate Hox network. The HCR fragment can be bound simultaneously by several protein molecules and could thus allow various combinations of Hox proteins to interact with it. Such an array of proteins could influence transcription of a target gene according to the resulting functional property of a given combination of proteins. The HCR fragment might therefore be an example of a structural element required as a basis for a combinatorial type of regulation, as for the 'multi-switch' model proposed by Han *et al.* (1989). The fact that HOX4C and HOX4D can bind to most if not all of these sites (within HCR) suggests, however, that a relatively high amount of one homeoprotein could probably saturate the HCR sites, indicating that concentrations of homeoproteins may be at least as important as their specific affinities in establishing this 'combinatorial' regulatory strategy (see also references above).

Similar clusters of homeobox protein DNA-binding sites are found in *cis*-regulatory regions of several *Drosophila* homeotic genes (references in Hayashi and Scott, 1990). In the *Antennapedia* (*Antp*) P1 promoter, two regions, of ~50 and ~60 bp respectively, and located at ~300–400 bp downstream of the transcription start site were shown to bind the *Ultrabithorax* (*Ubx*) homeotic protein product (Beachy *et al.*, 1988) and analogous sequences are present immediately downstream of the transcription start site of the *Ubx* gene (Beachy *et al.*, 1988). In both cases these sequences appear to be essential to mediate activation or repression by the *Ubx* and *Antp* gene products (Krasnow *et al.*, 1989; Winslow *et al.*, 1989). Moreover, one of the elements identified in the *Ubx* promoter is required for the maintenance of *Ubx* expression in visceral mesoderm, *in vivo*, through an autoregulatory pathway (Müller *et al.*, 1989). The vertebrate HCR sequence we report in this study resembles very much these *Drosophila* clusters of homeodomain binding sites. This observation suggests that the described structural and functional similarities between these phylogenetically related genetic systems (e.g. Akam, 1989) could possibly be extended to regulatory mechanisms within both gene networks.

Cotransfection experiments revealed no particular synergistic or antagonist effects between the products of *HOX4D* and *HOX4C*. In contrast to HOX4D and HOX4C, the Hox-4.3 homeoprotein (encoded by the gene located 3′ to *HOX4C*) is unable to activate through this region in cotransfection experiments. Moreover, when cotransfected with either HOX4C or HOX4D expressors, the Hox-4.3 protein is able to interfere with their activating capacities. This effect seems to be independent of the binding of the Hox-4.3 product to HCR since a Hox-4.3 protein whose homeodomain has been deleted shows the same efficiency, when compared with the full-length protein, in causing repression. In contrast, the Hox-4.3 homeodomain alone does not seem to compete for binding, maybe due to its weaker affinity for the HCR sites. This raises the possibility that the repressing effect is due to protein–protein interactions between the activators and the *Hox-4.3* product which could lead, for example, to a masking of the HOX4C and HOX4D activation domains or to the formation of inactive complexes. In both cases, DNA binding would not be necessary for Hox-4.3 to establish and maintain these interactions. Such a DNA-binding independent repression effect has recently been reported for a POU homeodomain transcription factor (I-POU; Tracy *et al.*, 1991). The product of this gene, whilst unable to bind DNA due to an alteration in its homeodomain sequence, can nevertheless form stable complexes with a related POU transcription factor. This interaction seems to prevent the POU transcription factor from exerting its DNA binding and transactivating properties. Our results suggest that this novel type of protein–protein interactions between POU proteins (see Herr, 1991) could also take place between class I homeodomain proteins. This could provide an explanation for the unexpected finding of alternatively spliced transcripts, in the *Xenopus Xhox2* gene, which encode proteins with or without the homeodomain (Wright *et al.*, 1987). Alternatively, the repression effect could involve squelching, an interaction between domains upstream of the *Hox-4.3* homeodomain with a cofactor necessary, but not sufficient, for activation, which could also take place independently of DNA binding. Interestingly, the Hox-4.3 protein can transactivate transcription in the same system when co-transfected with a reporter plasmid containing a multimerized artificial 'Hox' consensus binding site (G.Urier, V. Zappavigna and D.Duboule, not shown) which suggests that the Hox-4.3 function (through a positive or negative effect) could be partially dependent on the recognition sequence.

Materials and methods

Isolation of human embryonic cDNA clone

A cDNA library was constructed in λgt10 vector from 10 μg of poly(A)$^+$ RNA isolated from spinal cord of 7 weeks p.c. human embryos (Mavilio *et al.*, 1986). Mouse genomic clones corresponding to parts of the *Hox-4.4* and *Hox-4.5* transcription units, a 0.5 kb *Bam*HI–*Hin*dIII fragment and a 0.8 kb *Pst*I–*Pst*I fragment, respectively, both located 3′ to the homeobox (Duboule and Dollé, 1989) were used as probes to screen under reduced stringency conditions (4 × SSC, 42% formamide, 0.1% SDS, 1 × Denhardt's, 5 mg/ml salmon sperm DNA, 37°C). Two cDNA clones of ~3.2 kb and ~2.4 kb corresponding to transcripts from the *HOX4D* and *HOX4C* genes were isolated. Restriction fragments were subcloned into the pUC18 plasmid (Boehringer) and sequenced according to conventional methods (Sanger *et al.*, 1977; Maxam and Gilbert, 1980).

Isolation and mapping of the human HOX4D and HOX4C genes

A human genomic library constructed in λEMBL3 phages was screened under high stringency conditions (Maniatis *et al.*, 1982) using the human *HOX4D* and *HOX4C* cDNA clones as probes. Various clones were isolated, one of which (λ14B) contained an ~15 kb region comprising both the *HOX4D* and *HOX4C* transcription units. The published sequences have received accession numbers X59372 and X59373 from the EMBL data library.

Constructions

The pSGH4D expression construct contains the complete open reading frame of the *HOX4D* cDNA clone. The leader sequence 5′ to the translation start

site was deleted by digestion at an *Alu*I site located 16 bp 3′ to the ATG and replaced by the synthetic linker 5′-GGATCCATGTCCTTTCCC-AACAG-3′, which contains a *Bam*HI site 5′ to the ATG. The construction was cloned into the *Bam*HI site in the polylinker of the pSG5 expression vector (Green *et al.*, 1988). Similarly, pSGH4C was constructed by replacing the leader sequence, after digestion with *Dde*I, 19 bp 3′ to the ATG, with the synthetic linker 5′-GGATCCATGTCTTCCAGTGGCACCT-3′ containing a *Bam*HI site 5′ to the ATG. To generate the pSGH4.3 construct the linker 5′-AGATCTGGATCCATGAGCT-3′ was used to replace the leader sequence, after its removal by digestion with *Sca*I, 3 bp 3′ to the ATG and to introduce a *Bam*HI site 5′ to the translation start site. The truncated forms of the Hox-4.3 proteins were constructed as follows: pSGH4.3Δ214-289 ('ΔHD') was obtained by digestion at two *Eco*RI sites at position +637 of the coding sequence (Izpisúa-Belmonte *et al.*, 1990) and in the polylinker of the pSG5 vector, followed by reclosure of the plasmid with T4 DNA ligase. This results in a deletion of 75 amino acids containing the homeobox and the carboxy-terminal end of the protein. A stop codon (TAA) located 16 bp 3′ from the *Eco*RI site in pSG5 was put in-frame with the deleted Hox-4.3 protein. The pSGH4.3Δ2-176 ('HD') construct was generated by digestion at two *Sac*I sites located at positions +3 and +520 with respect to the ATG, resulting in an in-frame internal deletion of amino acids 2–176. This leads to the production of a protein containing essentially the homeobox and the carboxy-terminal end of the Hox-4.3 protein. The pTHCR reporter construct was constructed by PCR amplification of a 93 bp genomic region, located at positions −711 to −620 from the ATG of HOX4C (HCR, see Figure 3D), using a genomic subclone of the λ14B phage as template. The primers, containing artificial construction sites, used in the reaction were 5′-GGATCCGTCCGTTTTATTAG-3′ and 5′-AAGCTTAAGACCTCTGTAATG-3′. The amplified fragment was cloned in 5′–3′ orientation into the *Bam*HI–*Hin*dIII sites in the polylinker of the pT81luc luciferase reporter vector (Nordeen, 1988). Mutated forms of HCR, as shown in Figure 6A, were generated through PCR using primers with the desired mutations. The products of the PCR reactions were cloned in the polylinker of pT81luc. All amplifications were verified by sequence analysis of the corresponding reporter constructs.

Protein production

The cDNA clones corresponding to the *HOX4D*, *HOX4C* and *Hox-4.3* genes were modified, as described in the previous section, using synthetic linkers to introduce a *Bam*HI site 5′ to the ATG. They were subsequently cloned into the pET3A T7 prokaryotic expression vector (Studier *et al.*, 1990). Crude bacterial extracts were prepared from *E.coli* BL21 (DE3) strain transformed with the expression construct or, as a control, with the vector alone. The proteins were extracted according to a modified version of the method described in Kadonaga *et al.* (1987). Cells were grown at 37°C in 400 ml LB medium to an A_{600} = 0.8. Isopropyl-β-D-thiogalacto-pyranoside was then added to a final concentation of 0.4 mM and the cells further cultured for 2 h. The bacteria were pelleted, washed in 20 ml cold LB medium, repelleted and resuspended on ice with 2 ml of buffer A (40 mM Tris–HCl, pH 7.7, 25% sucrose, 0.2 mM EDTA, 1 mM DTT, 1 mM PMSF, 1 mM sodium metabisulphite). 0.5 ml buffer A + 1 mg/ml lysozyme were then added and the mix was incubated for 1 h. Urea was added to a final concentration of 4 M and the mix was subjected to sonication (2 × 15 s, grade 2, Branson sonifier) followed by a 1 h incubation on ice. The samples were further centrifuged at 63 000 *g* for 1 h. The supernatant was dialysed against buffer B (20 mM Tris–HCl, pH 7.7, 50 mM KCl, 10 mM $MgCl_2$, 1 mM EDTA, 20% glycerol, 1 mM DTT, 0.2 mM PMSF, 1 mM sodium metabisulphite) + 1 M urea for 90 min. The mix was further dialysed twice (2 h and overnight) against buffer B. The extracts were stored at −80°C.

Gel retardation and DNase I footprinting analysis

A pUC18 plasmid containing the 300 bp *Ava*I restriction fragment comprising region A (see Figure 3A) was linearized by digestion with *Hin*dIII, end-labelled using the Klenow fragment of *E.coli* DNA polymerase I and [^{32}P]dCTP and digested with *Eco*RI to release the insert which was purified by elution after polyacrylamide gel electrophoresis. Gel retardation analysis was performed by preincubating the bacterial extracts for 15 min on ice in 18 μl of binding buffer (100 mM KCl, 2 mM $MgCl_2$, 10% glycerol, 4 mM spermidine, 100 μg/ml BSA, 3 μg poly(dI–dC), 0.1 mM EDTA, 0.25 mM DTT). 2 μl of probe (0.5 ng, 2 × 10^4 c.p.m.) were then added and the incubation was continued for another 30 min. The mix was electrophoresed at 10 V/cm in 0.25 × TBE through a 6% polyacrylamide gel. After electrophoresis the gel was dried and exposed to a Kodak X-AR film. The sequence of the oligonucleotide used as a 'specific' competitor (SC) is 5′-AAATATCAATTAAATCTTAATTATAA-3′. The sequence of the oligonucleotide used as a 'non-specific' competitor (NSC) is 5′-CTAGAGGATCCATGTCCTTTCCCAACAG-3′. DNase I footprinting was performed essentially as described by Xiao *et al.* (1987).

Cell cultures, transfections and luciferase assays

Mouse fibroblast NIH3T3 cells were maintained in Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum (Gibco), 100 IU/ml of penicillin and 100 μg/ml streptomycin. Transfections were performed using the calcium phosphate precipitation method (Di Nocera and Dawid, 1983). In a typical transfection experiment 10 μg of reporter plasmid, 5 μg of protein expression construct and, as an internal control, 1 μg of pRSV-β-gal (Bonnerot *et al.*, 1987) or pCH110 (Promega) plasmids were used per 9 cm dish. Cells were harvested 48–60 h after transfection, lysed with three cycles of freezing and thawing and centrifuged at 4°C in an Eppendorf microfuge for 15 min. Luciferase assays were performed as described in deWet *et al.* (1987). β-Galactosidase assays were done according to Maniatis *et al.* (1982).

Acknowledgements

We would like to thank H.Kulessa for his help in sequencing parts of the chicken *Hox-4.4* promoter, Drs P.Chambon and M.Yaniv for critical reading of the manuscript and suggestions, and H.Davies and the EMBL photolab for preparing the manuscript. A.R. and G.U. are recipients of long term fellowships from HFSP and EMBO, respectively.

References

Acampora,D., D'Esposito,M., Faiella,A., Pannese,M., Migliaccio,E., Morelli,F., Stornaiulo,A., Nigro,V., Simeone,A. and Boncinelli,E. (1989) *Nucleic Acids Res.*, **17**, 385.
Akam,M. (1989) *Cell*, **57**, 347–348.
Beachy,P.A., Krasnow,M.A., Gavis,E.R. and Hogness,D.S. (1988) *Cell*, **55**, 1069–1081.
Biggin,M.D. and Tjian,R. (1989) *Cell*, **58**, 433–440.
Bonnerot,C., Rocancourt,D., Briand,P., Grimbier,G. and Nicholas,J. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 6795–6799.
DeWet,J.R., Wood,K.V., DeLuca,M., Helsinki,D.R. and Subramani,S. (1987) *Mol. Cell. Biol.*, **7**, 725–737.
Desplan,C., Theis,J. and O'Farrell,P.H. (1989) *Cell*, **54**, 1081–1090.
Di Nocera,P.P. and Dawid,I.B. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 7095–7098.
Dollé,P., Izpisúa-Belmonte,J.-C., Falkenstein,H., Renucci,A. and Duboule,D. (1989) *Nature*, **342**, 767–772.
Duboule,D. and Dollé,P. (1989) *EMBO J.*, **8**, 1497–1505.
Duboule,D., Boncinelli,E., DeRobertis,E., Featherstone,M., Lonai,P., Oliver,G. and Ruddle,F.H. (1990) *Genomics*, **7**, 458–459.
Duncan,I. (1987) *Annu. Rev. Genet.*, **21**, 285–319.
Ekker,S.C., Young,K.E., Von Kessler,D.P. and Beachy,P. (1991) *EMBO J.*, **10**, 1179–1186.
Fainsod,A., Bogard,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.
González-Reyes,A. and Morata,G. (1990) *Cell*, **61**, 515–522.
González-Reyes,A., Urquia,N., Gehring,W.J., Struhl,G. and Morata,G. (1980) *Nature*, **344**, 78–80.
Graham,A., Papalopulu,N. and Krumlauf,R. (1989) *Cell*, **57**, 367–378.
Green,S., Issemann,I. and Sheer,E. (1988) *Nucleic Acids Res.*, **16**, 369–373.
Han,K., Levine,M.S. and Manley,J.L. (1989) *Cell*, **56**, 573–583.
Hayashi,S. and Scott,M. (1990) *Cell*, **63**, 883–894.
Herr,W. (1991) *Nature*, **350**, 554–555.
Hoey,T. and Levine,M. (1988) *Nature*, **332**, 858–861.
Izpisúa-Belmonte,J.-C., Dollé,P., Renucci,A., Zappavigna,V., Falkenstein, H. and Duboule,D. (1990) *Development*, **110**, 733–745.
Izpisúa-Belmonte,J.-C., Falkenstein,H., Dollé,P., Renucci,A. and Duboule,D. (1991a) *EMBO J.*, **10**, 2279–2289.
Izpisúa-Belmonte,J.-C., Tickle,C., Dollé,P., Wolpert,L. and Duboule,D. (1991b) *Nature*, **350**, 585–589.
Jaynes,J.B. and O'Farrell,P.H. (1988) *Nature*, **336**, 744–749.
Jaynes,J.B. and O'Farrell,P.H. (1991) *EMBO J.*, **10**, 1427–1433.
Kadonaga,J.T., Carner,K.R., Masiarz,F.R. and Tjian,R. (1987) *Cell*, **51**, 1079–1090.
Kappen,C., Schughart,K. and Ruddle,F.H. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 5459–5463.
Kessel,M. and Gruss,P. (1990) *Science*, **249**, 374–379.

Kissinger,C.R., Liu,B., Martin-Blanco,E., Kornberg,T.B. and Pabo,C.O. (1990) *Cell*, **63**, 579–590.
Krasnow,M.A., Saffman,E.E., Kornfeld,K. and Hogness,D.S. (1989) *Cell*, **57**, 1031–1043.
Laughon,A. and Scott,M.P. (1984) *Nature*, **310**, 24–31.
Licht,J.D., Grossee,M.J., Figge,J. and Hausen,V.M. (1990) *Nature*, **346**, 76–79.
Maniatis,T., Fritsh,E.F. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Mavilio,F. *et al.* (1986) *Nature*, **324**, 664–666.
Maxam,A.M. and Gilbert,W. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 560–564.
Müller,M., Affolter,M., Leupin,W., Otting,G., Wüthrich,K. and Gehring,W.J. (1988) *EMBO J.*, **8**, 4299–4304.
Müller,J., Thüringer,F., Biggin,M., Züst,B. and Bienz,M. (1989) *EMBO J.*, **8**, 4143–4151.
Nordeen,S.K. (1988) *BioTechniques*, **6**, 454–457.
Odenwald,W.F., Garbern,J., Arnheiter,H., Tournier-Lasserve,E. and Lazzarini,R.A. (1989) *Genes Dev.*, **3**, 158–172.
Otting,G., Qian,Y.Q., Billeter,M., Müller,M., Affolter,M., Gehring,W.J. and Wüthrich,K. (1990) *EMBO J.*, **9**, 3085–3092.
Peifer,M., Karch,F. and Bender,W. (1987) *Gene Dev.*, **1**, 891–898.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
Sasaki,H., Yokoyama,E. and Kuzoiwa,A. (1990) *Nucleic Acids Res.*, **18**, 1739–1747.
Scott,M.P., Tamkun,J.W. and Hartzel,G.W. (1989) *Biochim. Biophys. Acta*, **989**, 25–48.
Simeone,A., Acampora,D., Nigro,V., Faiella,A., D'Esposito,M., Stronaiulo.A., Mavilio,F. and Boncinelli,E. (1991) *Mech. Dev.*, **3**, 215–228.
Studier,F.W., Rosenberg,A.H., Dunn,J.J. and Dubendorff,J.W. (1990) *Methods Enzymol.*, **185**, 60–89.
Tracy,M.N., He,X. and Rosenfeld,M.G. (1991) *Nature*, **350**, 577–584.
Winslow,G.M., Hayashi,S., Krasnow,M., Hogness,D.S. and Scott,M.P. (1989) *Cell*, **57**, 1017–1030.
Wright,C.V.E., Cho,K.W.Y., Fritz,A., Bürglin,T.R. and De Robertis,E.H. (1987) *EMBO J.*, **6**, 4083–4094.
Xiao,J.H., Davidson,I., Ferrandon,D., Rosales,R., Vigneron,M., Macchi,M., Ruffenach,F. and Chambon,P. (1989) *EMBO J.*, **6**, 3005–3013.

Received on August 21, 1991; revised on September 18, 1991