Development 112, 207–219 (1991)
Printed in Great Britain © The Company of Biologists Limited 1991

A chicken *caudal* homologue, *CHox-cad*, is expressed in the epiblast with posterior localization and in the early endodermal lineage

AYALA FRUMKIN[1], ZEHAVA RANGINI[2], ADI BEN-YEHUDA[1], YOSEF GRUENBAUM[2] and ABRAHAM FAINSOD[1,*]

[1]*Department of Cellular Biochemistry, Hebrew University-Hadassah Medical School, Ein Kerem, Jerusalem 91010, Israel*
[2]*Department of Genetics, Hebrew University, Givat Ram, Jerusalem 91904, Israel*

* Author for correspondence

Summary

***CHox-cad* is a chicken homeobox gene whose homeodomain is homologous to the *Drosophila caudal* and the murine *Cdx1* genes. Based on sequence analysis of a 2.5 kb *CHox-cad* cDNA clone, we deduced that the primary translation product consists of 248 amino acids. Comparison between the cDNA and genomic clones revealed the presence of an intron within the *CHox-cad* homeodomain between amino acids 44 and 45. The onset of *CHox-cad* transcription correlates temporarily with the beginning of gastrulation. During primitive streak stages *CHox-cad* exhibits a caudally localized pattern of expression restricted to the epiblast and the primitive streak. At these stages, *CHox-cad* transcripts can also be detected in the definitive endoderm cells. Later in embryogenesis *CHox-cad* is expressed in the epithelial lining of the embryonic gut and yolk sac. After four days of chicken development, no *CHox-cad* transcripts could be detected. The early *CHox-cad* posterior expression in the germ layer undergoing gastrulation and its continuous expression in the early endodermal lineage raise the possibility of *CHox-cad* involvement in the establishment of the definitive endoderm.**

Key words: homeobox, *caudal*, chicken embryogenesis, cloning, *in situ* hybridization, endoderm, *CHox-cad*, cell lineage.

Introduction

Numerous homeobox-containing genes (homeobox genes) have been cloned from vertebrate genomes and, in several instances, the same or closely related genes have been cloned from different organisms (Scott *et al.* 1989; Graham *et al.* 1989; Duboule and Dollé, 1989). The interest in this vertebrate gene family arose as a result of the discovery of the homeobox sequence in *Drosophila* developmental genes (see Gehring, 1987*a*,*b* for reviews), which was subsequently shown to have been conserved during evolution, in animals including vertebrates (McGinnis *et al.* 1984). The large number of vertebrate homeobox genes cloned to date are found in one of four major *Hox* clusters which, in mice and humans, contain 6–10 homeobox genes (Acampora *et al.* 1989; Graham *et al.* 1989; Duboule and Dollé, 1989; Njølstad *et al* 1990; Fritz *et al.* 1989; Wedden *et al.* 1989). The organization of vertebrate homeobox genes in clusters is reminiscent of *Drosophila* homeobox gene organization. Comparison of sequence and genomic organization of the clusters in mouse and man led to the suggestion that the vertebrate *Hox* clusters arose as a result of duplication of an ancestral gene complex (Hart *et al.* 1987; Kappen *et al.* 1989; Schughart *et al.* 1989). The similarity between the *Drosophila* and vertebrate homeobox gene clusters extends beyond the arrangement of homeobox sequences in a gene complex. Comparison of vertebrate genes in the clusters to the *Drosophila* genes in the *Antennapedia* and *Bithorax* complexes revealed a colinear arrangement of cognate genes (Acampora *et al.* 1989; Duboule and Dollé, 1989; Graham *et al.* 1989). Furthermore, as in *Drosophila*, the order of the genes along the cluster represents the anteroposterior order of their rostral boundary of expression (Gaunt *et al.* 1988; Graham *et al.* 1989; Duboule and Dollé, 1989).

The high degree of homology between the vertebrate homeobox sequences and their *Drosophila* counterparts has resulted in the organization of the genes into families based on their homeodomain sequences (Scott *et al.* 1989). In some instances, a homeobox gene family contains several sequences from the same organism due to the duplication of the *Hox* complexes. The homology between cognate genes is not restricted to the homeobox region but it extends throughout the genes

(Graham *et al.* 1988; Njølstad *et al* 1988, 1990). In addition to the genes in the *Hox* clusters, a few vertebrate homeobox genes have been cloned for which there is no evidence that they belong to a homeobox gene complex. In mice, the reported genes that belong to this category are the *Hox 7.1* (Hill *et al.* 1989; Robert *et al.* 1989), *Cdx1* (Duprey *et al.* 1988), *Evx1* and *Evx2* (Bastian and Gruss, 1990), *En-1* and *En-2* (Joyner and Martin, 1987) genes. The *Drosophila* genome also contains a number of homeobox genes that are not members of the known clusters. Some of these genes have been identified by genetic analysis, as in the case of *even-skipped* (*eve*; Macdonald *et al.* 1986), while others, cloned solely by homeobox homology, had no known mutations for them, such as *caudal* (*cad*; Mlodzik *et al.* 1985).

Of the *Drosophila* homeobox genes that do not belong to one of the complexes, several are of particular interest as they appear to have vertebrate homologues. The *Drosophila* homeobox genes are *msh* (Robert *et al.* 1989), *eve* and *cad* whose vertebrate homologues are *Hox 7.1*, *Xhox3* together with *Evx1* and *Evx2* (Bastian and Gruss, 1990; Ruiz i Altaba and Melton, 1989) and *Cdx1*, respectively. The *cad–Cdx1* pair is of great interest as *Cdx1* appears to be expressed in a pattern that resembles, in part, the *cad* expression pattern as it relates the organ and tissue layer in which they are expressed. In the course of embryonic development, the *Drosophila cad* gene transcripts are first detected as maternal transcripts that form a gradient whose maximal levels are localized to the posterior pole of the embryo (Mlodzik *et al.* 1985). This gradient is replaced at later stages of development by posterior transcript localization; in this case, it is made of transcripts of zygotic origin. In developmental stages after gastrulation, the *cad* mRNA can be seen to be localized to the posterior midgut and Malpighian tubules, the posterior midgut being of endodermal origin (Mlodzik and Gehring, 1987). Expression of the *Cdx1* homeobox gene during the course of murine development is first detected by *in situ* hybridization in embryos 14 days *post coitum* (*p.c.*), and from this stage on its transcripts are localized to the epithelial lining of the intestine, which in mice is of endodermal origin (Duprey *et al.* 1988). Expression of the *Cdx1* gene could not be detected in earlier embryos or in adult ovaries. *Cdx1* and *cad* not only share high homology in their homeodomain sequences but *Cdx1* is expressed in what appears to be part of the *cad* expression pattern.

In this paper, we report the isolation of a new chicken homeobox gene that contains a homeobox sequence belonging to the *cad* family of homeodomains and was thus termed *CHox-cad*. Transcripts of this gene are first detected as gastrulation begins and, by day five of incubation, no transcripts are detectable. The onset of *CHox-cad* transcription correlates with the onset of gastrulation suggesting a role for this gene during the establishment of the three germ layers. *In situ* hybridization analysis revealed that the early *CHox-cad* expression is localized mainly to the caudal region of the embryo and is restricted to the epiblast, primitive streak and definitive endoderm, while at later stages the expression is localized to the endodermal lining of the developing gut.

Materials and methods

Genomic and cDNA library screening

A genomic clone of the *CHox-cad* gene, λGG4, was isolated by screening a chicken oviduct genomic library in EMBL4 kindly provided by Dr B. W. O'Malley as previously described (Rangini *et al.* 1989).

The cDNA library of stage 12–13 chicken embryos (2 days of incubation) was constructed from poly $(A)^+$ RNA that was purified twice through oligo (dT)-cellulose columns. cDNA was prepared according to the procedure of Okayama and Berg (1982) as modified by Gübler and Hoffman (1983). After *Eco*RI methylation of the cDNA, *Eco*RI linkers were added and the cDNA was cloned into the λgt10 (Huynh *et al* 1985). About 10^6 phage were screened with a genomic fragment containing the *CHox-cad* homeobox. Screening was performed under high-stringency conditions Briefly, the filters were prehybridized at 65 °C in a solution containing 50 mM phosphate buffer pH 6.9, 5×SSC, 5×Denhardt's solution, 0.1 mg ml^{-1} Salmon sperm DNA as carrier and 0.1 % SDS. After prehybridization the probe was added to a similar solution and the filters were hybridized under the same conditions for a further 20 h. The final washes of the library were performed in 0 1×SSC, 0.1 % SDS at 65 °C

RNA preparation and northern blot analysis

For the preparation of embryonic RNA, fertilized chicken eggs were obtained from local farms and incubated for the desired time. Embryos were extracted in ice-cold PBS and staged according to Hamburger and Hamilton (1951) Preparation of RNA was according to Chirgwin *et al* (1979) as previously described (Rangini *et al* 1989) Northern blot analysis was performed under high stringency hybridization conditions as described (Rangini *et al.* 1989). In all RNA gels, 28S, 23S, 18S and 16S ribosomal RNAs were used as size markers and the estimation of transcript sizes was performed according to Lehrach *et al* (1977)

Sequence analysis

Sequencing was performed on both strands of the DNA by the dideoxy chain-termination method (Sanger *et al* 1977) using the Sequenase II kit (US Biochemicals) The genomic sequence was determined on specific subclones prepared using known restriction sites. Sequencing of the C33 cDNA clone was accomplished by performing a series of nested unidirectional deletions as described by Henikoff (1984) using the Erase-a-Base kit (Promega). In some instances, specific oligonucleotide primers were used to sequence through regions where no deletions were obtained.

In situ *hybridization*

The protocol employed for *in situ* hybridization was adapted to chicken embryos from Wilkinson *et al.* (1987) Briefly, embryos at the appropriate developmental stages were extracted onto ice-cold PBS and staged. Fixation was performed in ice-cold 4 % paraformaldehyde in PBS for time periods ranging from 30 min to 2 h depending on their size. Following fixation, the embryos were embedded in paraffin and 5–8 μm sections were collected on TESPA (3-aminopropyltriethoxysilane)-treated glass slides. The slides were treated with xylene to remove the wax and then rehydrated

through an ethanol series. Further treatments included saline for 5 min, PBS 5 min, fixation 20 min, twice PBS for 5 min, 20 $\mu g\,ml^{-1}$ proteinase K in 50 mM Tris–HCl, 5 mM EDTA pH 8.0 for 5 min, PBS 5 min and fixation 20 min. After acetylation for 10 min, PBS and saline for 5 min each, the slides were dehydrated through an ethanol series and air dried. The slides were hybridized in 50% formamide, 0.3 M NaCl, 20 mM Tris–HCl, 5 mM EDTA pH 8.0, 10% dextran sulphate, 1×Denhardt's solution, 0.5 $mg\,ml^{-1}$ yeast RNA, 10 mM dithiothreitol and 2×10^5 cts $min^{-1}\,\mu l^{-1}$ of ^{35}S-UTP labelled probe. The hybridizations were performed for about 18 h at 50°. Washes of the slides included 5×SSC, 10 mM DTT at 50°C for 60 min, 50% formamide, 2×SSC, 20 mM DTT at 65°C for 30 min, three 10 min washes in 0.5 M NaCl, 10 mM Tris–HCl, 5 mM EDTA at 37°C, an extra 30 min wash in the same buffer containing 20 $\mu g\,ml^{-1}$ RNAse A followed by a 15 min wash in the same buffer excluding RNAse. The slides were washed in 50% formamide, 2×SSC, 20 mM DTT at 65°C for 30 min and the final washes were for 15 min each at 65°C in 2× and 0.1×SSC, respectively. After dehydration, the slides were dipped in photographic emulsion for exposure.

Results

Cloning of CHox-cad

A chicken oviduct genomic library was screened under low-stringency hybridization conditions as described by Rangini *et al.* (1989). From this library screen about 15 independent homeobox-containing phage were isolated. One that appeared to hybridize preferentially to the *scr* probe was selected for further analysis. This phage, λGG4, was restriction mapped and the position of the homeobox within this genomic fragment was established (Fig. 1A). The analysis of λGG4 by hybridization to a number of homeobox probes revealed that this phage probably contains only one homeobox sequence. This result suggests that the homeobox gene contained in λGG4 might not be a member of one of the known vertebrate *Hox* clusters unless the distance to the neighboring homeobox genes is greater than the DNA contained in the cloned phage.

Initial characterization of the cloned homeobox gene was performed by sequencing the genomic fragment from λGG4 that contains the homeobox itself. This region was sequenced on both strands after subcloning the appropriate restriction fragments. The sequence of the homeobox and flanking regions is shown in Fig. 2. Comparison of *CHox-cad* with homeoboxes from other organisms revealed 83.6% homology to *Cdx1* (Duprey *et al.* 1988), 72.6% to *caudal* (Mlodzik *et al.* 1985) and 61.7% to the *antp* (Garber *et al.* 1983) homeoboxes at the nucleotide level. When the comparisons were performed between the putative protein translations, the extent of identity was 95% with the Cdx1, 80.3% with caudal and 62.3% with the antp homeodomains. Due to its homology to the *caudal* and *Cdx1* homeodomains, we call this novel chicken homeobox gene *CHox-cad*. In addition, the homology between Cdx1, CHox-cad and caudal extends five amino acids upstream of the homeodomain (Fig 3). Downstream from the homeobox, five out of six amino acids are shared between Cdx1 and CHox-cad but not with caudal. It is interesting to note that the homology upstream from the homeobox ceases at the position where the *caudal* gene is interrupted by an intron. Analysis of the genomic sequence upstream from the *CHox-cad* homeobox revealed the sequence 5′ CTCTCTCTGCCAGG (Fig. 2, overlined) which is a good consensus sequence for a splice acceptor site (Shapiro and Senapathy, 1987). This sequence in *CHox-cad* localizes the intron–exon boundary to the same position relative to the homeobox as in the case of the *Drosophila caudal* gene.

Interestingly, the genomic sequence of the *CHox-cad* homeobox reveals that the homeodomain itself is interrupted by an intron. The intron is localized 118 bp from the beginning of the homeobox sequence thus breaking the homeodomain sequence between amino

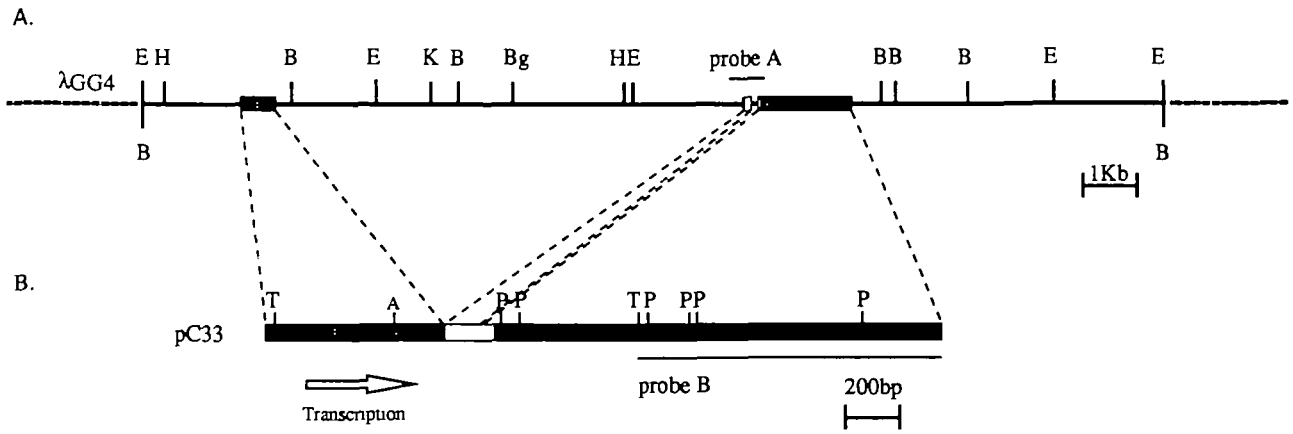

Fig. 1. Genomic and cDNA clones of *CHox-cad*. (A) Restriction map of λGG4 the genomic clone that contains the *CHox-cad* gene. The positions of the different exons of the *CHox-cad* gene are marked. (B) Restriction map of the *CHox-cad* cDNA, pC33, is shown. The direction of transcription is marked by an arrow. In both maps, the homeobox is marked as a white box, the rest of the protein coding sequences are marked as grey boxes and the 5′ and 3′ untranslated regions are black boxes. The genomic probe A and the cDNA probe B are indicated. A, *Ava*I, B, *Bam*HI; Bg, *Bgl*II; E, *Eco*RI; H, *Hind*III, K, *Kpn*I; P, *Pst*I, T, *Taq*I.

```
                                        CTTCTGCTCCTCCGTCAGACCAGCAGATGGGTCGATGCAGGACGCGATC
    CCGCGTTGCCGCAGCAGCAGCAGCGGGGCCGTAACGCGGCGCCGGGCCGCTCTCCAGGCCATGGCGGTGCGGGAGGCGG
    CCGCTGCGTGACACCCGGCCGCTTCGGCACGGAGCGGCGCAGCGCGGTGCATCGGCACAGCATGAGCACAGAGCCAGGT

    GGAGAAGATGTATGTAGGCTATCTTTTGGATAAGGACACCAAC ATG TAT CCC AGT CCC GTC CGG CAT CCC
1                                               Met Tyr Pro Ser Pro Val Arg His Pro

    GGC CTC AAC CTC AAC CCC CAG AAC TAC GTG CCG GGG CCA CCC CAG TAC TCG GAC TTC GCC
10  Ser Leu Asn Leu Asn Pro Gln Asn Tyr Val Pro Gly Pro Pro Gln Tyr Ser Asp Phe Ala

    AGC TAC CAC CAT GTG CCA GGG ATT AAC AAC GAC CCT CAC CAT GGG CAG CCC GCA GCC GCC
30  Ser Tyr His His Val Pro Gly Ile Asn Asn Asp Pro His His Gly Gln Pro Ala Ala Ala

    TGG GGC TCA CCC TAC ACC CCT GCC AAG GAG GAC TGG CAC TCA TAC GGC ACC GCT GCC GCC
50  Trp Gly Ser Pro Tyr Thr Pro Ala Lys Glu Asp Trp His Ser Tyr Gly Thr Ala Ala Ala

    TCC GCC GCC ACC AAC CCG GGG CAG TTT GGA TTT AGC CCC CCG GAT TTT AAC CCC ATG CAG
70  Ser Ala Ala Thr Asn Pro Gly Gln Phe Gly Phe Ser Pro Pro Asp Phe Asn Pro Met Gln

    CCC CAT GCG GGC TCT GGA CTC CTG CCT CCA GCC ATC AGC AGC TCG GTG CCA CAG CTG TCC
90  Pro His Ala Gly Ser Gly Leu Leu Pro Pro Ala Ile Ser Ser Ser Val Pro Gln Leu Ser

    CCT AAT GCA CAG AGG CGC ACC CCG TAC GAA TGG ATG AGG CGC AGT ATT CCC AGC ACC AGC
110 Pro Asn Ala Gln Arg Arg Thr Pro Tyr Glu Trp Met Arg Arg Ser Ile Pro Ser Thr Ser
            .......
    AGC AGC G gtaaggaaaccccagccctacacctgggtcttctcactggagcctcagttggagtctgcagggccgga
130 Ser Ser G

    tcctgctctgggacctcccaatggagagctgtggggctggcgaggtggcatggattgaggatggcatggagagcattgg
    tgtgcggtgacccg....(Intron)... gctcgtgcagctcctccatgaccatcaagcgatagctcctgatcccccct
                                 ttttttttttt ttttttttttttt
    agccctcttgtctcaccctgcactctgctctctctgccag GC AAG ACG AGG ACA AAG GAC AAG TAC CGG
132                                          ly Lys Thr Arg Thr Lys Asp Lys Tyr Arg

    GTG GTG TAC ACG GAC CAC CAA CGC CTG GAG CTG GAG AAG GAA TTT CAC TAC AGC CGC TAC
142 Val Val Tyr Thr Asp His Gln Arg Leu Glu Leu Glu Lys Glu Phe His Tyr Ser Arg Tyr
                                                                    ......
    ATC ACC ATC CGC CGT AAG GCC GAG CTG GCT GCT GCC CTG GGG CTC ACT GAA CGG CAG gtg
162 Ile Thr Ile Arg Arg Lys Ala Glu Leu Ala Ala Ala Leu Gly Leu Thr Glu Arg Gln
    ....
    agtgctggggatggatttgagagaggcagtgccttcccattctgcttccagcccccatggtgcccttgggtgctcttcc
                                      ++++++++++++++
     tcgtccacatccccttctctgcatggcgatctgtgtcttcccacag GTG AAA ATC TGG TTT CAG AAC CGG
181                                             Val Lys Ile Trp Phe Gln Asn Arg

    AGG GCG AAA GAG AGG AAG GTG AAT AAA AAG AAG CTG CAG CAG CAG AGC CAG CCC ACC AGC
189 Arg Ala Lys Glu Arg Lys Val Asn Lys Lys Lys Leu Gln Gln Gln Ser Gln Pro Thr Ser

    ACC ACC ACG CCA ACC CCC CCA GCC GTG GGC ACG CCA GGG CCC ATG GGG ACC CTC TGC AGT
209 Thr Thr Thr Pro Thr Pro Pro Ala Val Gly Thr Pro Gly Pro Met Gly Thr Leu Cys Ser

    GGC AGT GCC CCA AGC CTT GTC TCC TCA TCT CCG CTC ACC ATC AAG GAG GAG TTC ATG CCT
229 Gly Ser Ala Pro Ser Leu Val Ser Ser Ser Pro Leu Thr Ile Lys Glu Glu Phe Met Pro

    TAG CCCCCTCCACCAGCCATAGACACTGAGAGATAAGCACTACACGGCAGAGCTTGTTGCTGAGCGCCCAAAGACACC
    ter

    ACGTCCCCTGGGGCATGGTGGCTCTGTTGCCTTCCCAAGAGGCTGATGAGCTCTGTCCTACACAAAGCCCATACGTTAG
    GTAGTGCTGTTCTCCTGTAGGAAACGTGCGCTGGTCTTCGGTATTGGTTTTGGAGCAGCTTTAGTGAGGCTGAACTAGG
    GGAAAGGCTTGGCTGGACTCGAAGCAGAATAAATCTCAGCTGAGATTGTCTGCAGGATCTGCCATACCTGGAGCGAGTT
    GGACAACAGGAGCTGGGTAGGACCCTCAGCAGTGGGGCTGGAGGGGGAATGGGGGCTCCTTAAACTGAGACAATGACAG
    TGAGTGTCCCCCATGTCCCCCAGCCACCATCCCCAGCTCCATCCCTGCAGCAACTCAGGGCCATCGCGAGACCTGCAGA
    CATCTGTCTCACAGATGGACCATCCTGAGGAGGTTTTGAGACCACCTTCCAGCATGGAGGAACAAGAACAAGGTTTGCA
    GTTAGGACACTGGACAGCTCTAGGTTGCCCAGAGGGGCCACTCATCTACCGATGGCATTGCAGAACCTGCACAGCTCCA
    TGCTGCACCAAGGCACTGCCAGCTGCCAGGACCATCATGGTGATGGGACATGTCCGGCACAAGATGATGGATAGCTGGC
    TTCATAGAGAGGGATGGGGTGGATTTGTGCTGTTTTCTCTGAAACACAAAAGGCTCTGGGTTTTGCCCACTTTGTACTC
    CTCCATGTGCTGCACCATTGGTAACCAATGGCTTTTTTTTCCTGCTGGAAATCCACAGTTAAAATTCCCCTTAATCCTG
    CTGACTCCTGGGATCTCTACTGGCCCAGCTACCTTACAGCCCAGTTTTGGATGCCCCAAAGGGCGTGCTGGGACACAGG
    AGGGTTTGAGGCTCCTGCTTGGTTCTGTGGCTTCCCAAGACCAAGGTGAGGGTGCAGAGGCTCTGATGGTTCTCTCCTA
    GCACCATCATGCTGAGCCCCAGCAGGAACAAGAAGTGGAGGTGTCCCTGTTGCCTGCAGAGCCCCTCACCCATTCCCAG
    CCCATTCTGGCCCAGAACTCACAGTGTTAGGGGTTTATACTGGGATTTAAGCTCCTTGACACCCACTGCACCTACAGCC
    CCTTAATAGGGTCTGCATCCCCCACCATGGCGACCAGTGACACCTCCCTGGGGCCACCACTGCTACTTTGCAGCAAAGG
    CAGACACCGGTGTTGGCTTGGGACTTCAGCTCAGGGCTTGTGTACAGAAGGCAGAACGTTTTATTTTCTAGGCTTTGGT
    CAA**AATAAA**AGCGAGCAGGAGCCCTGTGAAA
```

Fig. 2. cDNA sequence and putative protein of *CHox-cad*. The *CHox-cad* cDNA sequence is shown in uppercase letters, genomic intronic sequences are shown in lowercase letters. The putative *CHox-cad* protein sequence is shown underneath. The consensus DNA sequence for the initiation of translation is underlined. The homeobox sequence is double underlined 5′ and 3′ splice sites are shown as dotted overlined. The polyadenylation signal is shown in bold. In the amino acid sequence only the hexapeptide is underlined. Numbering is for the amino acid sequence.

acids 44 and 45. The length of the intron is 128 bp and it is flanked by the sequences 5′ AGGTGAGT and 5′TCTTCCCACAGG, which are in good agreement with the consensus splice sequences for the donor and acceptor splice sites, respectively (Shapiro and Senapathy, 1987).

cDNA cloning and putative protein product of CHox-cad

In order to study the organization of the *CHox-cad* gene as well as its putative protein product, cDNA clones were isolated. A cDNA library prepared from embryos at stage 12–13 (H and H, 2 days of incubation) was

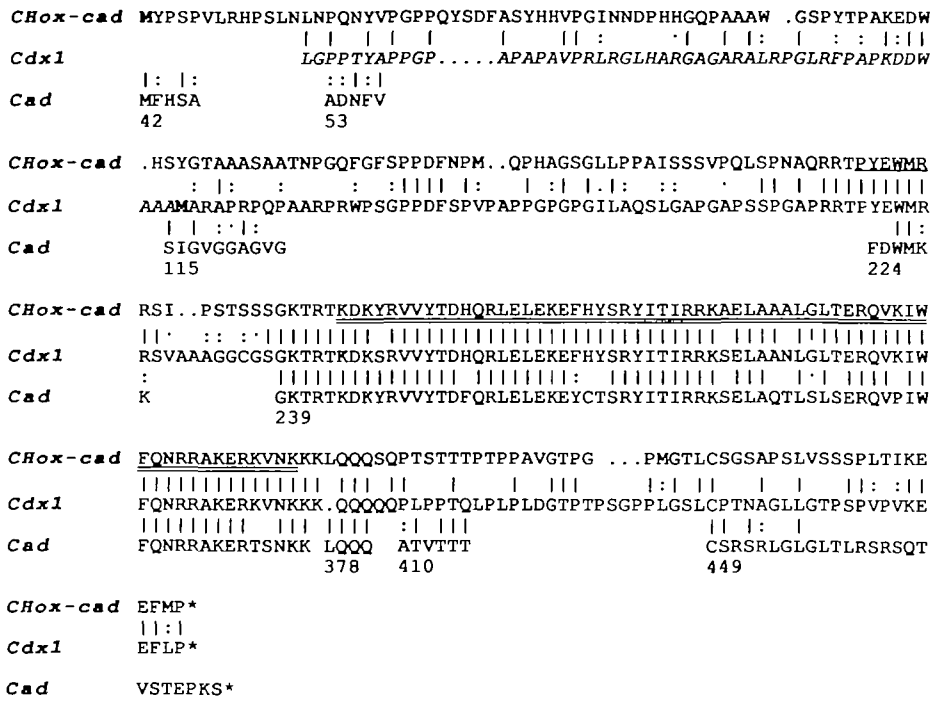

Fig. 3. Comparison of the *CHox-cad* protein to Cdx1 and caudal. The whole *CHox-cad* and *Cdx1* putative proteins are aligned, a number of small gaps had to be introduced From the putative *caudal* protein only the small relevant regions are shown, the numbering represents the position of the first amino acid of the region shown in the whole *caudal* protein. The CHox-cad hexapeptide is underlined and the homeodomain is double underlined. A line represents identity to CHox-cad, conservative changes are marked with a colon The suggested initiator methionines of Cdx1 and CHox-cad are shown in bold The amino acid sequence upstream to the suggested initiator methionine in Cdx1 is shown in italics.

prepared in λgt10. About $5\times10^5-10^6$ recombinant phage were screened with probe A. Of the 7 recombinant phage isolated, the one with the largest insert of approximately 2.6 kb (C33), was selected for sequencing and further analysis (Fig. 1B). The insert in C33 is 2486 bp long (Fig. 2), suggesting that this cDNA clone is almost full length in relation to the estimated size of the *CHox-cad* transcript from northern analysis. The cDNA contains a 744 bp long open reading frame capable of coding for a 248 amino acid protein which includes the *CHox-cad* homeodomain. The AUG of the putative initiator methionine is present in the sequence 5′CCAACAUG at position 246–253 of the cDNA which is a good consensus sequence for the initiation of translation (Kozak, 1986). This AUG is present 61 bp downstream from an in-frame stop codon. The stop codon in frame with the homeodomain that marks the end of the *CHox-cad* protein is present 51 amino acids downstream from the homeodomain (Fig. 2). The position of the putative initiator methionine and of the termination codon of the *CHox-cad* translation product reveals that the C33 cDNA contains 250 and 1492 bp of 5′ and 3′ untranslated sequences, respectively. In the 3′ untranslated sequence, 19 bp from the end of the cDNA there is an AATAAA polyadenylation signal (Birnstiel *et al.* 1985) and the cDNA clone ends with a short poly A tail.

Comparison of the genomic and cDNA sequences corroborated our observations regarding the exon–intron organization of the *CHox-cad* gene described earlier. As expected, the cDNA clone did not contain the intron that interrupts the homeobox sequence, giving rise to a contiguous 61 amino acid homeodomain. The *CHox-cad* putative protein product also contains the hexapeptide Pro-Tyr-Glu-Trp-Met-Arg (Fig. 3, underlined) which has been found in a number of homeodomain proteins (Schughart *et al.* 1988). In all cases described, the hexapeptide is present upstream to the homeodomain. Furthermore, the genomic sequence suggests that upstream from the homeodomain there is an intron, an observation supported by the cDNA sequence. Thus, the hexapeptide and the homeodomain in CHox-cad are also in different exons.

Detailed comparison between the *CHox-cad, Cdx1* and *cad* putative protein products showed that they are most homologous in the region of the homeodomain (Fig. 3). This region of homology can be extended up to the hexapeptide sequence. The five amino acids immediately upstream to the homeodomain, Gly-Lys-Thr-Arg-Thr, are identical among the three proteins. Upstream of these five amino acids CHox-cad and Cdx1 have 7 and 9 amino acid regions, respectively, which, though not identical, are composed mostly of conservative changes. Cad shows no similarity to the other two proteins in this region. In the region of the hexapeptide, Cdx1 and CHox-cad share 12 identical amino acids including hexapeptide (Fig. 3). The *cad* protein shows some similarity to CHox-cad in this region with mostly conserved rather than identical amino acids. Upstream of the hexapeptide region, the homology between CHox-cad and the other two proteins decreases, exhibiting only small conserved regions of 4–5 similar amino acids (Fig. 3). A further region of similarity between CHox-cad and Cdx1 was observed when the *Cdx1* cDNA sequence was translated from its 5′ end. This region, Pro-Ala-Lys-Glu-Asp-Trp in CHox-cad, is 56 amino acids downstream of the putative initiator methionine. In Cdx1 the sequence is Ala-Pro-Lys-Asp-Asp-Trp and it begins 9 amino acids upstream of the previously suggested putative initiation site (Duprey *et*

al. 1988). This region of similarity between the two proteins suggests that the *Cdx1* protein initiates further upstream.

Downstream from the cad homeodomain, the protein continues for 169 amino acids until the first in-frame stop codon. In the case of Cdx1 and CHox-cad, the stop codon is localized 54 and 51 amino acids, respectively, downstream from the homeodomain. Searching the *cad* downstream protein sequence revealed short regions of similarity that usually are shared with one but not the other vertebrate protein products. Comparison between CHox-cad and Cdx1 in this region showed that the homeobox homology can be extended six amino acids downstream. The homology between the proteins then decreases and it increases towards the end of the proteins, where they share seven identical and three similar amino acids out of 11.

Expression of CHox-cad *during the first days of embryogenesis*

The temporal pattern of transcription of the *CHox-cad* gene during the first days of chicken embryonic development was studied by northern analysis. RNA was prepared from chicken embryos from the time the egg is laid, to 5 days of incubation. During this time period, the chicken embryo develops from the blastoderm stage (stage 1; Hamburger and Hamilton, 1951; H and H) to the middle stages of organogenesis (stage 26). For stages 1–2 and 4–5, total RNA was utilized; for all the later developmental stages, poly $(A)^+$ RNA was prepared. The northern blot was probed under high stringency with a 673 bp long *Sac*I genomic fragment from the *CHox-cad* gene (probe A, Fig. 1A). The results of this hybridization showed that probe A recognizes two transcripts, 1.6 and 2.6 kb in size (Fig. 4A). Accumulation of the two transcripts begins very early during embryonic development and reaches maximal levels at stages 4–5 (16 h of incubation). At stage 5, the primitive streak in the embryo reaches its maximal length and the embryo begins neurulation (H and H). Following stages 4–5, the transcript levels begin to decrease until they become altogether undetectable, the rate of decrease of the two transcripts, however, is different. The 1.6 kb transcript is no longer present by day 4 of embryonic development (stage 21–22) and the 2.6 kb transcript is undetectable by 5 days of incubation (stage 26).

The northern analysis shown in Fig. 4A suggests that early *CHox-cad* transcription correlates with gastrulation and the formation of the primitive streak. In order to map temporally in detail the first embryonic stages during which *CHox-cad* transcription is evident, RNA was prepared from blastoderm-stage embryos and older. The difference in this instance was that the embryos were staged according to Eyal-Giladi and Kochav (1976) for stages X to XIV which are prestreak stages (a subdivision of stage 1, H and H). Older embryos (stage 2 and older) were staged according to H and H. The blot was probed with a 1119 bp fragment from the *CHox-cad* cDNA (see below) specific for the 2.6 kb transcript (probe B, Fig. 1B). Hybridization with

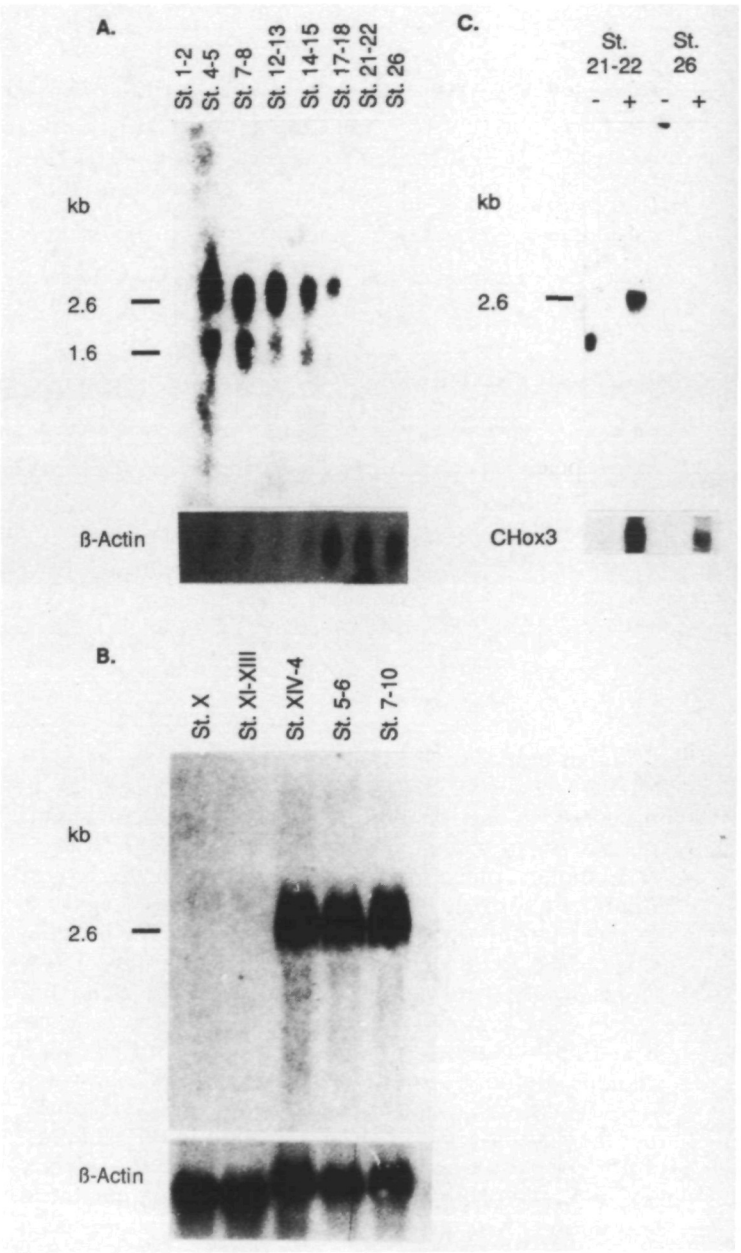

Fig. 4. Expression of *CHox-cad* during early embryogenesis. (A) Expression of *CHox-cad* during the first five days of incubation. For stages 1–2 (unincubated eggs) and 4–5 (16 h of incubation), 20 μg of total RNA were loaded. For stages 7–8 (26 h of incubation), 12–13 (2 days), 14–15 (2.5 days), 17–18 (3 days), 21–22 (4 days) and 26 (5 days), 1 μg of poly $(A)^+$ RNA was loaded The blot was hybridized with probe A. (B) Onset of *CHox-cad* transcription. RNA was prepared from whole embryos. In all lanes 30 μg of total RNA were loaded. The blot was probed with probe B. (C) Disappearance of the *CHox-cad* transcripts between day 4 (stages 21–22) and day 5 (stage 26) of incubation. In all lanes 8 μg of RNA were loaded. For each developmental stage, poly $(A)^+$ (+) and poly $(A)^-$ (−) RNA was loaded The blot was hybridized with probe A.

this probe could not detect any *CHox-cad*-specific transcripts in embryos from stages X to XIII. The first appearance of the 2.6 kb transcript is in embryonic

stages that correspond to the formation of the primitive streak (stages XIV-4, Fig. 4B). This result further strengthens the correlation between the onset of gastrulation and the onset of *CHox-cad* transcription.

To increase the sensitivity of detection of the *CHox-cad* transcript, the northern analysis of *CHox-cad* expression at stages 21–22 and 26 was repeated using 8 μg of poly $(A)^+$ RNA per lane instead of 1 μg as in Fig. 4A and probed with probe A (Fig. 4C). This result supports the previous observation that by day 4 (stage 21–22) the 1.6 kb transcript is absent and by day 5 (stage 26) no *CHox-cad* transcripts remain. Fig. 4C shows the control hybridization of the same blot to the *CHox 3* probe which is expressed at constant levels during these developmental stages (Rangini *et al.* 1989). Northern analysis performed on mRNA prepared from embryos after 6 to 10 days of incubation and adult tissues, such as ovaries, brain, heart, liver and kidney showed no expression of *CHox-cad* (data not shown).

CHox-cad *expression in the primitive streak*

In order to establish the site of transcription of *CHox-cad* in the early embryo, stage 5 and 6 (H and H) chicken embryos were analyzed by *in situ* hybridization. Chicken embryos incubated for about 16 h (stage 5, H&H) were dissected out and processed for *in situ* hybridization as described (Materials and methods). Serial cross sections of chicken embryos were probed with strand-specific RNA probes prepared from either probe A (Fig. 1A) or probe B, which lacks the homeobox sequence (Fig. 1B), both yielding the same results. The *in situ* hybridization results show that the *CHox-cad* transcripts are localized to epiblast cells and to cells in the primitive streak (Fig. 5C and 5D). Once cells migrate out of the primitive streak and become mesoderm cells, they become negative for *CHox-cad* transcripts by *in situ* hybridization (Fig. 5D). Hybridization to serial sections of the same embryos revealed that maximal levels of *CHox-cad* transcripts are present in the caudal part of the embryo (Fig. 5C), and more rostral sections show decreasing transcript levels (Fig. 5A and 5B). In all cases, parallel sections were hybridized to sense RNA probes as a negative control. These results suggest that *CHox-cad* is expressed in the primitive streak and in cells from the epiblast that are being recruited into the primitive streak. In addition, the *CHox-cad* transcripts are restricted along the anteroposterior embryonic axis and their maximal level is in the caudal region of the embryo. Anterior to Hensen's node where the neural plate has formed, no *CHox-cad* transcripts could be detected by this technique. Detailed analysis of the *in situ* hybridization results of sections of 10 embryos in 5 independent experiments revealed a thinly populated layer of cells under the mesoderm which also contains low levels of *CHox-cad* transcripts as judged by the grain density in each cell (Figs 5C and 5D; arrows). This *CHox-cad*-positive layer is one-cell thick and the cells are loosely packed suggesting that these cells are part of the definitive ('gut') endoderm as they migrate out of the primitive streak (Stern and Canning, 1988). This observation suggests that *CHox-cad* might be transcriptionally active in the early endodermal lineage from the onset of gastrulation.

Further evidence that *CHox-cad* transcript levels are different anterior or posterior to Hensen's node was obtained by northern analysis. Chicken embryos were incubated to stages 5–6 (16–20 h of incubation) at which time the embryos are beginning neurulation. The embryos were dissected out and sectioned into anterior and posterior parts by cutting the embryo into two parts at right angles to the embryonic axis at the level of Hensen's node. This dissection at Hensen's node results in the developing neural plate and notochord being contained in the rostral section and the primitive streak in the caudal section irrespective of the precise position of Hensen's node along the axis. RNA was prepared from the two regions of the embryos and probed for the 2.6 kb *CHox-cad*-specific transcript with probe B (Fig. 6). Posterior to Hensen's node, the 2.6 kb transcript is present in high levels, while in the anterior regions of the embryos there is a marked decrease in the level of the 2.6 kb transcript. These data further support the observations obtained by *in situ* hybridization of early embryos, which indicate the accumulation of *CHox-cad* transcripts in the caudal region of the embryo.

Expression of CHox-cad *in the developing gut*

The spatial pattern of expression of *CHox-cad* was also studied in embryos at stage 19 (3.5–4 days of incubation). By this stage of development, the 2.6 kb transcript is still present but it is in very low abundance (Fig. 4A). Stage 18 (Fig. 7A, 7C and 7E) and stage19 (Fig. 7B, 7D and 7F) embryos were sectioned and analyzed by *in situ* hybridization with either probe A or probe B as described (Materials and methods). Both probes gave the same results. At these developmental stages, *CHox-cad* transcripts are limited to the developing gut. *CHox-cad*-specific transcripts can be seen in the foregut (Fig. 7A and 7B) and in the hindgut (Fig. 7E and 7F), where expression is limited to the epithelial lining of the gut, which is of endodermal origin. In embryos at this stage, the gut is forming into a tube with the ventral side of the midgut opening into the yolk sac. The yolk sac is lined by an epithelia of endodermal origin which is also positive for *CHox-cad* expression (Fig. 7C and 7D). These results show that *CHox-cad* is expressed in the endodermal lining of the embryonic gut throughout its length.

Discussion

Expression of CHox-cad *during gastrulation and organogenesis*

In this paper, we describe *CHox-cad,* a novel homeobox gene isolated from the chicken genome. The expression of *CHox-cad* in primitive-streak-stage embryos is limited to cells of the epiblast, primitive streak and the definitive endoderm. In addition, analysis of serial sections of embryos in several experiments

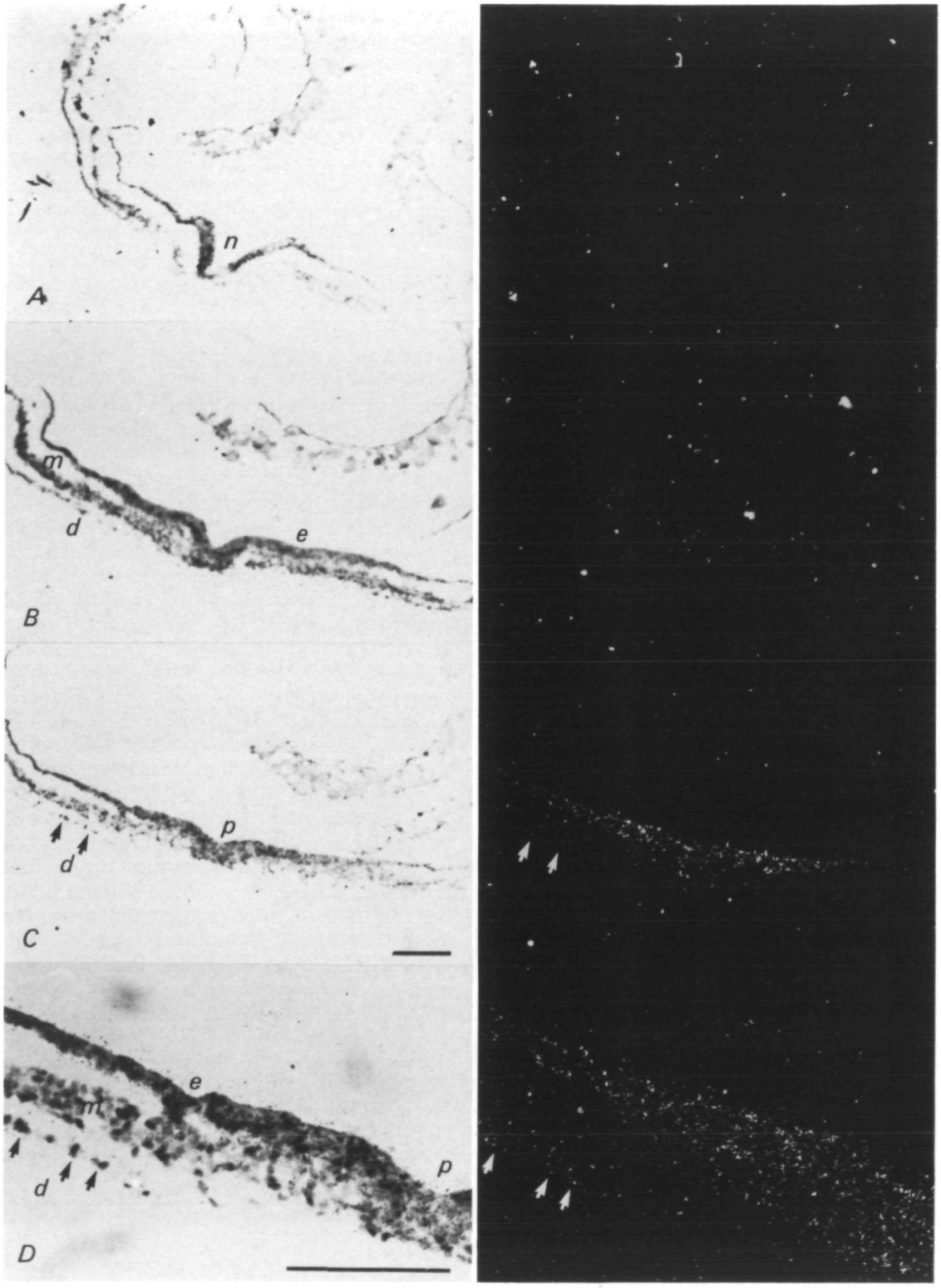

Fig. 5. Spatial localization of the *CHox-cad* transcripts in head-fold-stage embryos Stage 5 embryos (16 h of incubation) were serially cross sectioned. The sections were hybridized *in situ* with single-stranded RNA probes. All sections are from the same embryo. The bright-field and dark-field pictures of the sections are shown The sections are arranged from anterior to posterior and top is dorsal and bottom is ventral. (A) Rostral cross section anterior to Hensen's node where neurulation is evident. (B) Cross section posterior to Hensen's node midway along the primitive streak (C) Caudal cross section where the gastrulation process is evident. (D) High magnification of the section in C, the three germ layers are evident. Definitive endoderm cells are shown with arrows Abbreviations: e, epiblast; d, definitive endoderm; m, mesoderm, n, neural folds, p, primitive streak. Panels A, B and C are at the same magnification, bar equals 100 μm.

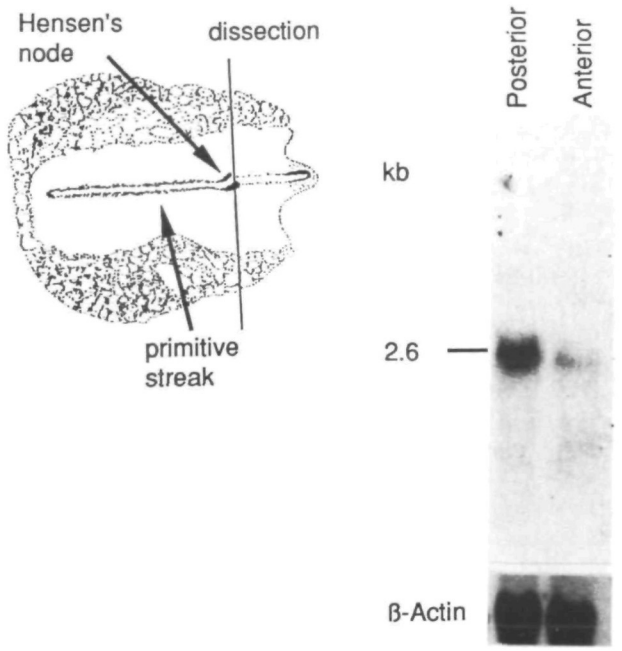

Fig. 6. Posterior localization of the *CHox-cad* transcription Stage 6 embryos were obtained and divided into anterior and posterior regions relative to Hensen's node. The regions were obtained by cutting the embryo at right angles to the primitive streak at the level of Hensen's node as described in the diagram. In the diagram anterior is to the right. 30 μg of total anterior or posterior RNA were loaded The blot was hybridized with probe B.

suggested that *CHox-cad* expression in the epiblast and primitive streak is also restricted with respect to the anterior–posterior axis of the embryo, with maximal transcription levels at the caudal end. Several vertebrate homeobox genes expressed during gastrulation have been reported, such as the murine *Hox 1.5* (Gaunt, 1987), *Hox 1.6* (Sundin *et al.* 1990) *Hox 2.9* (Frohman *et al.* 1990) and *Evx1* (Bastian and Gruss, 1990) genes, the *Xenopus* Xhox3 (Ruiz i Altaba and Melton, 1989) and *Xhox1A* (Harvey *et al.* 1986) and the chicken *CHox 3* (Rangini *et al.* 1989), *CHox 7* (Fainsod and Gruenbaum, 1989) and *Ghox-lab* (Sundin *et al.* 1990). For some of these genes, in addition to their temporal pattern of expression, the spatial localization of their transcripts in the gastrulating embryo is known. *Hox 1.5* is expressed in the ectoderm and mesoderm of primitive streak embryos and exhibits a predominantly posterior localization. A very similar pattern was found for the *Evx1* gene. *Xhox3* is also expressed in a posteroanterior gradient, but its expression is restricted to the mesoderm. *Ghox-lab* is expressed in the epiblast, primitive streak and mesoderm of primitive-streak-stage embryos with a predominant posterior localization. *Hox 2.9* is expressed along the length of the primitive streak and in the mesoderm in the posterior half of the embryo. Comparison of the patterns of expression reveals that *CHox-cad* shares with some of

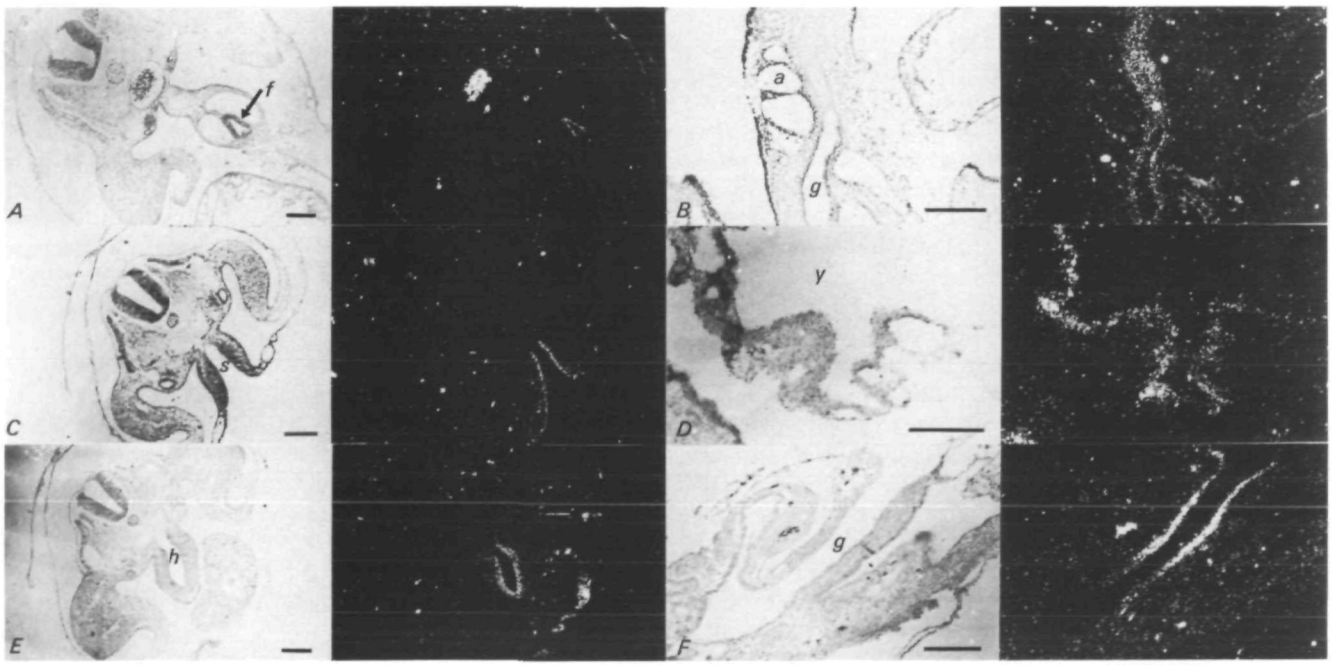

Fig. 7. *CHox-cad* expression in the embryonic gut Sections of stage 18 (A, C and E) and stage 19 (B, D and F) embryos (3 days of incubation) were *in situ* hybridized with single stranded RNA probes. The bright-field and dark-field pictures of the sections are shown. (A) Cross section in the region of the foregut (B) Parasagittal section at the level of the aortic arches. (C) Cross section in the region of the midgut and its opening to the yolk sac. (D) Section through the yolk sac. Top represents where the yolk was originally localized. (E) Cross section in the region of the hindgut. (F) Section through the caudal region of the embryo. Abbreviations a, aortic arch, f, foregut; g, embryonic gut cavity, h, hindgut; s, yolk sac, y, yolk In panels A, B, C, E and F the bar equals 100 μm, in panel D the bar equals 50 μm.

these genes, the rostrocaudal restriction of transcript accumulation. However, *CHox-cad* exhibits a novel spatial pattern of germ layer specificity, its expression being restricted to the epiblast, primitive streak and definitive endoderm. Northern analysis of the *CHox-cad* rostrocaudal gradient revealed that the relative abundance of the transcripts is substantially different in analysis of mRNA extracted from embryonic regions anterior or posterior to Hensen's node, increasing noticeably posterior to Hensen's node.

In addition to the expression of *CHox-cad* in the epiblast, primitive streak and definitive endoderm of early embryos, at later stages, the gene is expressed in the endodermal lining of the embryonic gut including the yolk sac. Several homeobox genes expressed in part or all of the gut of the embryo have been described, such as *Hox 1.3* (Dony and Gruss, 1987), *Hox 1.4* (Galliot *et al.* 1989), *Hox 1.6* (Duboule and Dollé, 1989), *Hox 2.1* (Holland and Hogan, 1988), *Hox 5.1* (Featherstone *et al.* 1988), *Hox 6.1* (Sharpe *et al.* 1988), *Cdx1* (Duprey *et al.* 1988) and *XlHbox 8* (Wright *et al.* 1988). Apart from *Cdx1* and *XlHbox 8,* all the other homeobox genes mentioned are expressed in mesoderm-derived tissues in the gut. In contrast, *Cdx1* and *XlHbox 8,* like *CHox-cad,* are found in endoderm-derived tissues of the gut, but are expressed at different times during development. The cell type specificity and, in some instances, the spatial restriction of expression of these homeobox genes suggest that several homeobox genes are involved in the differentiation of the vertebrate gut.

Analysis of the fate map of the chicken embryo at the time of gastrulation reveals that endodermal cells are of epiblast origin (Nicolet, 1970). At specific stages during gastrulation, the cells that migrate through the anterior regions of the primitive streak are in their majority destined to become endodermal cells, while in other regions of the primitive streak the contribution to the endoderm is smaller (Nicolet, 1970). These endodermal cells will initially form the lining of the gut, which eventually will give rise to the endodermal epithelia of other organs. The expression of *CHox-cad* in the epiblast, primitive streak and definitive endoderm during gastrulation and later in the epithelia of the gut correlates with the pathway that the precursor endodermal cells follow from gastrulation to the gut. These observations raise the possibility that *CHox-cad* becomes transcriptionally active at the onset of gastrulation in endoderm precursor cells in the epiblast. *CHox-cad* then remains active in the same cells as they migrate and begin differentiating up to day 5 of embryogenesis. This possibility is further supported by the observations of Stern and Canning (1990), which showed that precursor mesoderm and endoderm cells can be labelled with the HNK-1 antibody before the onset of gastrulation. At present, we cannot rule out the possibility that *CHox-cad* is activated also in mesodermal precursors, but if this is the case the gene is turned off as they migrate out of the primitive streak. This possibility arises from the fact that maximal *CHox-cad* expression is found in the caudal regions of the primitive streak, which gives rise mainly to mesodermal cells (Nicolet, 1970).

In *Drosophila*, *cad* expression begins as a maternal transcript gradient that is replaced by a zygotic transcripts localized to the posterior end of the embryo. At later stages of development, the *cad* transcripts are localized to the posterior midgut and Malpighian tubules, the posterior midgut being of endodermal origin. Several aspects of the *CHox-cad* pattern of expression are reminiscent of the *cad* pattern of expression. Early in embryogenesis, both *CHox-cad* and *cad* exhibit a pattern of expression in the form of transcript accumulation in the caudal region of the embryo. Somewhat later in embryonic development both genes are expressed in cells of endodermal origin in the gut and *CHox-cad* expression is turned off by day 5. *Cdx1,* on the other hand, is expressed in the differentiating epithelial lining of the intestine in older embryos when *CHox-cad* expression is undetectable. In contrast to *cad* maternal expression, neither *Cdx1* nor *CHox-cad* are expressed in the ovary. *Cdx1* and *CHox-cad*, therefore, appear to implement different aspects of the *cad* expression pattern.

The CHox-cad *protein product*

Comparison of the putative protein product of *CHox-cad* with other homeodomain proteins reveals that it belongs to the *cad* family of homeobox genes (Scott *et al.* 1989). This family includes the murine *Cdx1* (Duprey *et al.* 1988) and the *C. elegans ceh-3* (Burglin *et al.* 1989) homeobox genes as well as *CHox-cad* and *cad*. The highest degree of homology was localized in the region of the homeodomain but, in all cases, was also extended by a number of amino acids upstream to the homeobox. The presence of an intron that interrupts the CHox-cad homeodomain between amino acids 44 and 45 is a relatively rare observation particularly in vertebrate homeobox genes. Only three vertebrate homeobox genes, out of at least 70 members cloned, have been reported to contain a homeobox whose sequence is interrupted by an intron: *Xhox3* (Ruiz i Altaba and Melton, 1989) *Evx1* and *Evx2* (Bastian and Gruss, 1990). These three genes belong to the *eve* subfamily of homeobox genes and the intron splits the homeodomain in all three genes between amino acids 46 and 47. In *Drosophila*, a number of homeoboxes have been identified whose sequence is interrupted by an intron. The homeobox introns in the fly have been localized to two locations. In the *engrailed* and *invected* genes, the intron is localized between amino acids 17 and 18 (Poole *et al.* 1985) A second location for homeodomain introns in *Drosophila* homeobox genes is between amino acids 44 and 45 as in *Labial* (Mlodzik *et al.* 1988), *Abdominal-B* (DeLorenzi *et al.* 1988), *Distal-less* (Cohen *et al.* 1989) and *NK-1* (Kim and Nirenberg, 1989). Therefore, the intron in the *CHox-cad* homeobox is in a position that is not uncommon for fly homeobox genes.

Comparison of CHox-cad *to* Cdx1 *and* caudal

The comparison between *CHox-cad* and *Cdx1* is of

particular interest as they are both vertebrate homeobox genes. It is important to establish whether they represent the same gene in two evolutionary distant organisms, or whether an ancestral homeobox gene underwent duplications and they represent different members of the vertebrate *cad* gene family. The *CHox-cad* and *Cdx1* proteins were found to be the most similar in the region that extends from the hexapeptide to several amino acids downstream to the homeodomain. In this region, which in CHox-cad is 93 amino acids long and in Cdx1 is 94 amino acids long, both proteins share 82 identical amino acids and 5 conservative changes. From the putative initiator methionine to the hexapeptide region, the *Cdx1* protein has 52 residues, of which 13 residues are identical and 12 are conservative changes when compared to CHox-cad. Downstream from the extended homeodomain, a similar level of homology is observed. Further information as to the relation between *CHox-cad* and *Cdx1* comes from the analysis of their temporal patterns of expression. *CHox-cad* expression was found to be maximal during gastrulation and the beginning of neurulation by northern analysis and *in situ* hybridization, whereas, *in situ* hybridization of 7 and 8 day *post-coitum* (p.c.) mouse embryos, which represent gastrulation and neurulation stages, were found to be negative for *Cdx1* expression (Duprey *et al.* 1988). Northern analysis of *Cdx1* expression showed low levels at 10 days p.c. which disappeared until 14 days p.c. when it began to increase, reaching maximal levels at 17 days p.c. Between days 5 and 10 of chicken embryo development (stages 26 through 36), no *CHox-cad* transcripts could be detected by northern analysis. Stage 36 chicken embryos (10 days of incubation) roughly parallel in development 16 day p.c. mouse embryos (Sundin *et al.* 1990). Therefore, the comparison of the temporal patterns of expression suggests that at developmental stages during which *CHox-cad* is maximally expressed, *Cdx1* expression is undetectable. Also, the reverse situation holds true even though the establishment of parallel developmental stages is more complicated later in embryogenesis.

In summary, the comparison of the *CHox-cad* and *Cdx1* protein sequences reveal two related proteins whose evolutionary relation is not clear. Comparison of their temporal patterns of expression showed non-overlapping patterns. These observations suggest that the genes are involved in different processes during embryonic development. In addition, it is interesting to note the possible source of the two transcripts recognized by probe A. While probe A hybridized to two transcripts, 1.6 and 2.6 kb in size, probe B recognizes only the larger of these. The *CHox-cad* cDNA clone that was isolated and the fact that probe B lacks the homeobox identify the 2.6 kb transcript as a real *CHox-cad* mRNA. Other regions of the *CHox-cad* cDNA were also utilized as probes for northern analysis; in all instances probes lacking the homeobox region only hybridized to the 2.6 kb transcript (data not shown). Regarding the 1.6 kb transcript, even though it is recognized under high stringency, and its temporal pattern of expression is similar to that of the larger transcript, its source remains unclear. The only probe that detects this transcript contains the *CHox-cad* homeobox sequence, raising the possibility of homeobox cross-hybridization. If this is the case then, the 1.6 kb transcript originates from a different homeobox gene whose homeobox sequence is very similar to *CHox-cad* judging from the hybridization stringency. Support for the possibility that the vertebrate *cad* family may contain several members comes from the analysis of other homeobox genes in vertebrates. Analysis of the murine *Hox* complexes suggested that they arose as a result of duplications during evolution (Schughart *et al.* 1989). Furthermore, two murine *eve* type genes have been cloned, *Evx1* and *Evx2*, suggesting that duplication of homeobox genes during evolution was not limited to the *Hox* clusters (Bastian and Gruss, 1990). In a similar manner an ancestral *cad* may have undergone duplications and divergence and the different genes may have undertaken different functions, but only cloning and analysis of the different members will provide the ultimate answer.

The evidence presented here shows that early *CHox-cad* expression begins with the onset of gastrulation and reaches maximal levels when the primitive streak reaches its full length. This temporal correlation between the gastrulation process and *CHox-cad* expression raises the possibility that this gene may be involved in gastrulation. This suggestion is further supported by the expression of *CHox-cad* in the early endodermal lineage and evidence for a posterior localization of *CHox-cad* transcription. Several homeobox genes are expressed during gastrulation and they can be divided into those expressed in ectoderm and mesoderm, and those expressed only in the mesoderm. In addition, some of the homeobox genes exhibit a posterior pattern of expression. The expression of multiple homeobox genes during gastrulation exhibiting different germ layer restriction, raises the possibility that a network of homeobox genes is being employed at this early stage of embryonic development as is the case in the *Drosophila* embryo.

We wish to thank Hefzibah Eyal-Giladi for her help in the interpretation of the *in situ* hybridizations Rebecca Haffner for the invaluable discussions. Joel Yisraeli and Kenneth Robzyk for critically reading the manuscript. This work was supported by grants from the United States-Israel Binational Science Foundation (86–00014) and grants from the Fund for Basic Research, administered by The Israel Academy of Sciences and Humanities No. 241/87 to A F. and No 193/87 to Y.G. Z.R. was supported by the Israel Ministry of Research and Development and the National Council for Research and Development.

References

Acampora, D , D'Esposito, M., Faiella, A., Pannese, M , Migliaccio, E , Morelli, F , Stornaiuolo, A , Nigro, V , Simeone, A and Bonceinelli, E. (1989) The human HOX family *Nucl Acids Res* **17**, 10385–10402.

Bastian, H and Gruss, P (1990). A murine *even-skipped*

homologue, *Evx 1*, is expressed during early embryogenesis and neurogenesis in a biphasic manner *EMBO J* **9**, 1839–1852

Birnstiel, M L., Busslinger, M and Strub, K (1985) Transcription termination and 3′ processing the end is in site! *Cell* **41**, 349–359

Burglin, T R , Finney, M , Coulson, A and Ruvkun, G (1989) *Caenorhabditis elegans* has scores of homoeobox-containing genes *Nature* **341**, 239–243

Chirgwin, J. M , Przybyla, A. E., MacDonald, R. J and Rutter, W. J (1979) Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease *Biochemistry* **18**, 5294–5299

Cohen, S M , Bronner, G , Kuttner, F Jurgens, Gerd and Jackle, H (1989) *Distal-less* encodes a homoeodomain protein required for limb development in *Drosphila Nature* **338**, 432–434

DeLorenzi, M , Ali, N , Saari, G , Henry, C , Wilcox, M and Bienz, M (1988) Evidence that the *Abdomal-B* r element function is conferred by a *trans*-regulatory homeoprotein *EMBO J* **7**, 3223–3231

Dony, C and Gruss, P (1987) Specific expression of the Hox 1 3 homeobox gene in murine embryonic structures originating from or induced by the mesoderm *EMBO J* **6**, 2965–2975

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

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

Eyal-Giladi, H. and Kochav, S (1976) From cleavage to primitive streak formation a complementary normal table and a new look at the first stages of the development of the chick I General morphology. *Devl Biol* **49**, 321–337

Fainsod, A and Gruenbaum, Y (1989). A chicken homeobox gene with developmentally regulated expression *FEBS Lett* **250**, 381–385

Featherstone, M S , Baron, A , Gaunt, S J , Mattei, M G and Duboule, D (1988) *Hox-5 1* defines a homeobox-containing gene locus on mouse chromosome 2. *Proc natn Acad Sci U S A* **85**, 4760–4764

Fritz, A F , Cho, K W Y , Wright, C V E , Jegalian, B G and De Robertis, E M (1989) Duplicated Homeobox Genes in *Xenopus Devl Biol* **131**, 584–588

Frohman, M A , Boyle, M and Martin, G. R (1990) Isolation of the mouse *Hox 2 9* gene, analysis of embryonic expression suggests that positional information along the anterior–posterior axis is specified by mesoderm *Development* **110**, 589–607

Galliot, B , Dolle, P , Vigneron, M , Featherstone, M S , Baron, A. and Duboule, D (1989) The mouse Hox-1 4 gene primary structure, evidence for promoter activity and expression during development *Development* **107**, 343–359

Garber, R L , Kuroiwa, A and Gehring, W J (1983) Genomic and cDNA clones of the homeotic locus *Antennapedia* in *Drosophila*. *EMBO J* **2**, 2027–2036

Gaunt, S J. (1987) Homoeobox gene *Hox-1 5* expression in mouse embryos earliest detection by *in situ* hybridization is during gastrulation *Development* **101**, 51–60

Gaunt, S. J., Sharpe, P T and Duboule, D (1988) Spatially restricted domains of homeo-gene transcripts in mouse embryos. relation to a segmented body plan *Development* **104**, 169–179.

Gehring, W J (1987*a*). The homeobox Structural and evolutionary aspects In *Molecular Approaches to Developmental Biology* pp 115–129

Gehring, W J. (1987*b*) Homeo boxes in the study of development *Science* **236**, 1245–1252

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

Graham, A , Papalopulu, N , Lorimer, J , McVey, J H , Tuddenham, E G D and Krumlauf, R (1988) Characterization of a murine homeobox gene, *Hox-2 6*, related to the *Drosophila Deformed* gene *Genes and Dev* **2**, 1424–1438

Gübler and Hoffman, B J (1983) A simple and very efficient method for generating cDNA libraries *Gene* **25**, 263–269

Hamburger, V and Hamilton, H L (1951) A series of normal stages in the development of the chick embryo *J Morph* **88**, 49–92

Hart, C P , Fainsod, A and Ruddle, F H (1987) Sequence analysis of the murine *Hox-2 2, -2 3*, and *-2 4* homeoboxes evolutionary and structural comparisons *Genomics* **1**, 182–195

Harvey, R P , Tabin, C J and Melton, D A. (1986) Embryonic expression and nuclear localization of *Xenopus* homeobox (Xhox) gene products *EMBO J* **5**, 1237–1244

Henikoff, S (1984) Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing *Gene* **28**, 351–359

Hill, R E , Jones, P F , Rees, A R , Sime, C. M , Justice, M J , Copeland, N. G., Jenkins, N A , Graham, E and Davidson, D R (1989) A new family of mouse homeobox-containing genes molecular structure, chromosomal location, and developmental expression of Hox-7 1 *Genes and Dev* **3**, 26–37

Holland, P W. H. and Hogan, B L M (1988) Spatially restricted patterns of expression of the homeobox-containing gene *Hox 2 1* during mouse embryogenesis *Development* **102**, 159–174.

Huynh, T V., Young, R A and Davis, R W (1985) Constructing and Screening cDNA Libraries in λgt10 and λgt11 In *DNA Cloning Techniques A Practical Approach*, Vol I, IRL Press pp 49–78

Joyner, A L and Martin, G R (1987) *En-1* and *En-2*, two mouse genes with sequence homology to the *Drosophila* engrailed gene expression during embryogenesis. *Genes and Dev* **1**, 29–38

Kappen, C , Schughart, K. and Ruddle, F H (1989) Two steps in the evolution of Antennapedia-clas vertebrate homeobox genes *Proc natn Acad Sci U S A* **86**, 5459–5463

Kim, Y and Nirenberg, M (1989) *Drosophila* NK-homeobox genes *Proc natn Acad Sci U S A* **86**, 7716–7720

Kozak, M (1986) Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes *Cell* **44**, 283–292

Lehrach, H , Diamond, D , Wozney, J M and Boedtker, H (1977) RNA molecular weight determination by gel electrophoresis under denaturing conditions, a critical reexamination *Biochemistry* **16**, 4743–4750

Macdonald, P. M , Ingham, P and Struhl, G. (1986) Isolation, structure, and expression of *even-skipped* a second pair-rule gene of *Drosophila* containing a homeobox *Cell* **47**, 721–734

McGinnis, W , Garber, R L , Wirz, J , Kuroiwa, A and Gehring, W J (1984) A homologous protein-coding sequence in *Drosophila* homeotic genes and its conservation in other metazoans *Cell* **37**, 403–408

Mlodzik, M., Fjose, A and Gehring, W. J. (1985) Isolation of *caudal*, a *Drosophila* homeo box-containing gene with maternal expression, whose transcripts form a concentration gradient at the pre-blastoderm stage *EMBO J* **4**, 2961–2969

Mlodzik, M., Fjose, A and Gehring, W J (1988) Molecular structure and spatial expression of a homeobox gene from the *labial* region of the Antennapedia-complex *EMBO J* **7**, 2569–2578

Mlodzik, M and Gehring, W J (1987) Expression of the *caudal* gene in the germ line of *Drosophila* formation of an RNA and protein gradient during early embryogenesis *Cell* **48**, 465–478

Nicolet, G (1970) Analyse autoradiographique de la localisation des differentes ebauches presomtives dans la ligne primitive de l'embryon de Poulet *J. Embryol exp Morph.* **23**, 79–108

Njølstad, P R., Molven, A , Apold, J and Fjose, A (1990) The zebrafish homeobox gene *hox-2 2*: transcription unit, potential regulatory regions and *in situ* localization of transcripts *EMBO J* **9**, 515–524

Njølstad, P R , Molven, A , Hordvik, I , Apold, J and Fjose, A. (1988) Primary structure, developmentally regulated

expression and potential duplication of the zebrafish homeobox gene ZF-21. *Nucl. Acids. Res.* **16**, 9097–9111.

OKAYAMA, H. AND BERG, P. (1982). High efficiency cloning of full-length cDNA. *Molec. cell. Biol.* **2**, 161–170.

POOLE, S. J., KAUVAR, L. M., DREES, B. AND KORNBERG, T. (1985). The *engrailed* locus of *Drosophila*: structural analysis of an embryonic transcript. *Cell* **40**, 37–43.

RANGINI, Z., FRUMKIN, A., SHANI, G., GUTTMANN, M., EYAL-GILADI, H., GRUENBAUM, Y. AND FAINSOD, A. (1989). The chicken homeobox genes *CHox1* and *CHox3*: cloning, sequencing and expression during embryogenesis. *Gene* **76**, 61–74.

ROBERT, B., SASSOON, D., JACQ, B., GEHRING, W. AND BUCKINGHAM, M. (1989). *Hox-7*, a mouse hoemeobox gene with a novel pattern of expression during embryogenesis. *EMBO J.* **8**, 91–100.

RUIZ I ALTABA, A. AND MELTON, D. A. (1989). Bimodal and graded expression of the *Xenopus* homeobox gene *Xhox3* during embryonic development. *Development* **106**, 173–183.

SANGER, F., NICKLEN, S. AND COULSON, A. R. (1977). DNA sequencing with chain terminating inhibitors. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467.

SCHUGHART, K., KAPPEN, C. AND RUDDLE, F. H. (1989). Duplication of large genomic regions during the evolution of vertebrate homeobox genes. *Proc. natn. Acad. Sci. U.S.A.* **86**, 7067–7071.

SCHUGHART, K., UTSET, M. F., AWGULEWITSCH, A. AND RUDDLE, F. H. (1988). Structure and expression of *Hox-2.2*, a murine homeobox-containing gene. *Proc. natn. Acad. Sci. U.S.A.* **85**, 5582–5586.

SCOTT, M. P., TAMKUN, J. W. AND HARTZELL, G. W. II (1989). The structure and function of the homeodomain. *Biochim. biophys. Acta* **989**, 25–48.

SHAPIRO, M. B. AND SENAPATHY, P. (1987). RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucl. Acids Res.* **15**, 7155–7175.

SHARPE, P. T., MILLER, J. R., EVANS, E. P., BURTENSHAW, M. D. AND GAUNT, S. J. (1988). Isolation and expression of a new mouse homeobox gene. *Development* **102**, 397–407.

STERN, C. D. AND CANNING, D. R. (1988). Gastrulation in birds: a model system for the study of animal morphogenesis. *Experientia* **44**, 651–657.

STERN, C. D. AND CANNING, D. R. (1990). Origin of cells giving rise to mesoderm and endoderm in chick embryo. *Nature* **343**, 273–275.

SUNDIN, O. H., BUSSE, H. G., ROGERS, M. B., GUDAS, L. J. AND EICHELE, G. (1990). Region-specific expression in early chick and mouse embryos of Ghox-lab and Hox 1.6, vertebrate homeobox-containing genes related to *Drosophila* labial. *Development* **108**, 47–58.

WEDDEN, S. E., PANG, K. AND EICHELE, G. (1989). Expression pattern of homeobox-containing genes during chick embryogenesis. *Development* **105**, 639–650.

WILKINSON, D. G., BAILES, J. A. AND MCMAHON, A. P. (1987). Expression of the proto-oncogene *int*-1 is restricted to specific neural cells in the developing mouse embryo. *Cell* **50**, 79–88.

WRIGHT, C. V. E., SCHNEGELSBERG, P. AND DE ROBERTIS, E. M. (1988). XIHbox 8: a novel *Xenopus* homeo protein restricted to a narrow band of endoderm. *Development* **104**, 787–794.

(Accepted 4 February 1991)

Note added in proof

The sequences shown in Figs 2 and 3 have been submitted to the EMBL database and have the accession numbers X57760 and X57761.