

A chicken homeo box gene with developmentally regulated expression

Abraham Fainsod and Yosef Gruenbaum

*Department of Cellular Biochemistry, Hebrew University-Hadassah Medical School, Jerusalem 91010 and
Department of Genetics, Hebrew University, Jerusalem 91904, Israel*

Received 5 April 1989

Several homeo box sequences were isolated from a chicken (*Gallus domesticus*) genomic library. One phage, λ GG7, was chosen for further study and the single homeo box sequence it contains was characterized. The sequence of the homeo box revealed it to be a distant member of the *Antp* class of homeo boxes. Expression of the *CHox7* gene results in four transcripts, 1.4, 1.9, 2.4 and 3.5 kb in size. Transcript accumulation of the *CHox7* gene peaks twice during embryogenesis, once with the 1.4 kb transcript during early somitogenesis and the second time with the 1.9 kb transcript during organogenesis.

Homeo box; Embryogenesis; Gene evolution; (*Antennapedia*, *Gallus domesticus*)

1. INTRODUCTION

The homeo box is a protein coding sequence, 183 bp in length, usually present in the last exon of the gene [1–3]. The *Drosophila* homeo box containing genes have been shown via several approaches to be part of the network of genes that directs the early development of the fly embryo [3,4]. Cloning and sequencing of homeo boxes from different organisms have shown them to be highly conserved at both the nucleotide and amino acid levels [3,5]. The high degree of conservation seen between homeo boxes of different organisms has led to the suggestion that, as in the case of *Drosophila*, in vertebrates, homeo box sequences are also part of genes involved in the regulation of embryogenesis.

Homeo box genes have been cloned from several vertebrate classes such as fish [6], amphibia [7], mammals [8,9] and birds [10]. Analysis of the expression patterns of the different homeo box genes

has shown that they can code for one or several transcripts. Detailed analysis of the transcripts has shown that in few cases the same gene can code for mRNAs that do not contain the homeo box [11,11a]. In others instances it was shown that one transcription unit can include several homeo box sequences that can be alternatively spliced [9]. In some cases, once the gene has become transcriptionally active it remains active until adulthood [12], while in others the gene is only transcribed during part of the embryonic development [7]. Analysis of the spatial patterns of expression of the homeo box genes has shown them to be expressed in the embryo in a spatially restricted manner, reminiscent of the patterns of expression of the *Drosophila* developmental genes [13]. Also, as in the case of *Drosophila*, the vertebrate homeo boxes have been shown to encode protein domains that bind DNA in a sequence-specific manner [14–16].

The chicken embryo is one of the most thoroughly studied vertebrate embryos. Its development is similar to that of mammalian embryos [17]. When the egg is laid the chicken embryo is at the blastoderm stage allowing the study

Correspondence address: A. Fainsod, Department of Cellular Biochemistry, Hebrew University-Hadassah Medical School, Jerusalem 91010, Israel

of subsequent developmental stages [18]. Chicken embryos can be manipulated *in vitro* and have the additional advantage that they can be incubated *in vitro* from fertilization to hatching [19]. Recently, we described the cloning of two chicken homeo box genes belonging to a distant subfamily of *Antp*-type homeo boxes [10]. This work describes the sequence of the homeo box region and the developmentally regulated transcription of a third chicken homeo box gene, *CHox7*.

2. EXPERIMENTAL

A chicken oviduct genomic library prepared in λ EMBL4 (kindly provided by Dr B.W. O'Malley) was screened with the *Drosophila Antp* and *Scr* homeo boxes together with the murine *Hox 1.5* and *3.1* homeo boxes under conditions of reduced stringency [1,10]. Three genome equivalents, about 3×10^5 phage, were plated for screening. DNA fragments were subcloned and sequenced as described [10].

Embryos for RNA isolation were staged [18]. RNA was extracted from embryos at different developmental stages, electrophoresed, blotted and hybridized as in [10].

3. RESULTS

3.1. Isolation and characterization of the chicken *CHox7* homeo box sequence

In order to clone as many chicken homeo box sequences as possible, a chicken oviduct genomic library was probed with four homeo box probes together, under reduced stringency conditions [1]. The probes used all belong to the *Antp* type of homeo boxes and include the *Drosophila Antp* and *Scr* and the murine *Hox 1.5* and *3.1* homeo boxes [10]. About 15 independent homeo boxes were

cloned, one of which was described in [10]. In order to characterize more chicken homeo box genes, we picked at random another phage, λ GG7, which contains one homeo box sequence that cross-hybridized with the *Scr* and *Hox 1.5* probes during screening of the library (fig.1). The homeo box region was sequenced according to the strategy shown in fig.1. The 183 bp of the homeo box sequence, 70 bp of upstream sequence, and 121 bp of downstream sequence are shown in fig.2. The sequence of the homeo box begins at nucleotide no.1, and the putative translation of the homeo box is shown underneath. Using the sequence of the homeo box as a reference for the reading frame, we can see that all of the 374 bp code for one contiguous open reading frame. Comparisons of the *CHox7* homeo box sequence with those of homeo boxes from chicken and other organisms revealed a homology of about 67% or lower; also no corresponding homeo box has been cloned from other organisms. When the putative translations of the homeo boxes were compared, the levels of homology observed were 54% and lower. The highest levels of similarity were seen in comparisons with *CHox3* [10], *Hox 1.5* [20], *Hox 7.1* [21] and *msh* [21]. As in the case of the two previously reported chicken homeo boxes, the *CHox7* homeo box belongs to a distantly related family of *Antp*-type homeo boxes [3,5]. It is also interesting to point out that in the region upstream of the homeo box, out of 23 amino acids eight are glycines, and downstream of the homeo box there are two putative *N*-glycosylation sites (fig.2, overlined).

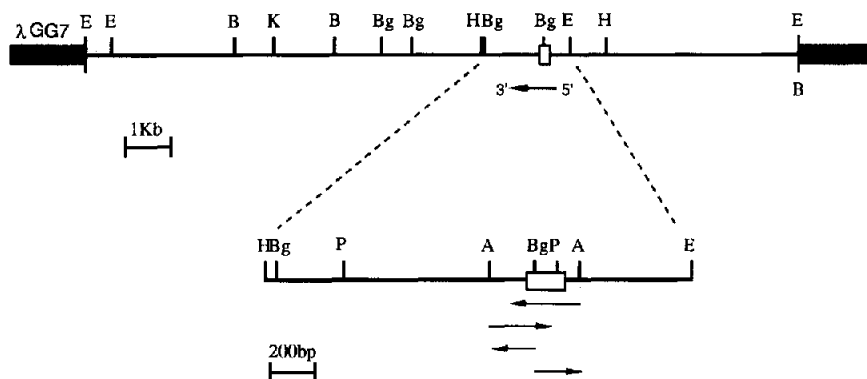


Fig.1. Restriction map and sequencing strategy of *CHox7*. The homeo box is shown as open boxes. The 5' to 3' direction of transcription is shown. A, *Apal*; B, *BamHI*; Bg, *BglII*; E, *EcoRI*; H, *HindIII*; K, *KpnI*; P, *PvuII*.

```

-70
      G GGC CCC AAG
      Gly Pro Lys

-60
GGG AAG GGG AAG GGC GGC CCC GCG GCC GAG CAG CCT CCG CCG GGC TCC GGG GCC GGG AAG
Gly Lys Gly Lys Gly Gly Pro Ala Ala Glu Gln Pro Pro Pro Gly Ser Gly Ala Gly Lys

1
AGC CGC CGG CGG CGC ACG GCC TTT ACG AGC GAG CAG CTG CTG GAG CTG GAG AAG GAG TTC
Ser Arg Arg Arg Arg Thr Ala Phe Thr Ser Glu Gln Leu Leu Glu Leu Glu Lys Glu Phe

61
CAC TGC AAG AAG TAC CTC TCG CTG ACG GAG CGC TCG CAG ATC GCG CAC GCC CTG AAG CTG
His Cys Lys Lys Tyr Leu Ser Leu Thr Glu Arg Ser Gln Ile Ala His Ala Leu Lys Leu

121
AGC GAG GTG CAG GTG AAG ATC TGG TTC CAG AAC CGC CGC GCT AAG TGG AAA CGC ATC AAG
Ser Glu Val Gln Val Lys Ile Trp Phe Gln Asn Arg Arg Ala Lys Trp Lys Arg Ile Lys

181
GCG GGC AAC GTC AGC AAC CGC TCG GGA GAG CCC GTC CGC AAC CCC AAG ATC GTG CGT GCC
Ala Gly Asn Val Ser Asn Arg Ser Gly Glu Pro Val Arg Asn Pro Lys Ile Val Arg Ala

241
CAT CCC GGT GCA CGT CAA TCG CTT CGC CGT GCG CAG CCA GCA CCA GCA GAT CGA GCA GGG
His Pro Gly Ala Arg Gln Ser Leu Arg Arg Ala Gln Pro Ala Pro Ala Asp Arg Ala Gly

301
GCC C
Ala

```

Fig.2. Sequence of the *CHox7* homeo box and flanking region. The homeo box starts at nucleotide 1 and is underlined. The putative protein translation of the open reading frame that includes the homeo box is shown beneath the sequence. The lines above the sequence represent putative *N*-glycosylation sites.

3.2. Expression of *CHox7* during embryogenesis

In order to study the involvement of the *CHox7* gene during development of the chicken embryo we followed by means of Northern blotting expression of the gene from the time at which the primitive streak is completely formed up to 5 days of incubation. Poly(A)⁺ and poly(A)⁻ RNA samples from the different developmental stages were probed with an *ApaI* fragment, which contains the homeo box, and whose sequence is shown in fig.1. As shown in fig.3, the *CHox7* gene codes for four transcripts, 1.4, 1.9, 2.4 and 3.5 kb in size. It can be seen that expression of *CHox7* is evident at the time the primitive streak is formed. *CHox7* specific transcripts can be observed as early as blastoderm stage embryos (not shown). Expres-

sion of the *CHox7* gene peaks twice during the first 5 days of incubation, but the most abundant transcript is different at each peak. The first time the *CHox7* gene peaks in expression is at about stage 8–9, the embryo having then been incubated for about 1 day, and it has 4–7 somites [18]. As shown in fig.3, at this time the most abundant transcript is that of 1.4 kb. After the first peak expression declines, subsequently reaching the second peak around stage 17–18, at which the embryo has been incubated for 3 days and is well into organogenesis [18]. During this peak in expression the 1.9 kb transcript is the most abundant (fig.3). During embryogenesis the 2.4 and the 3.7 kb transcripts are of low abundance and also show some variation in their level of expression.

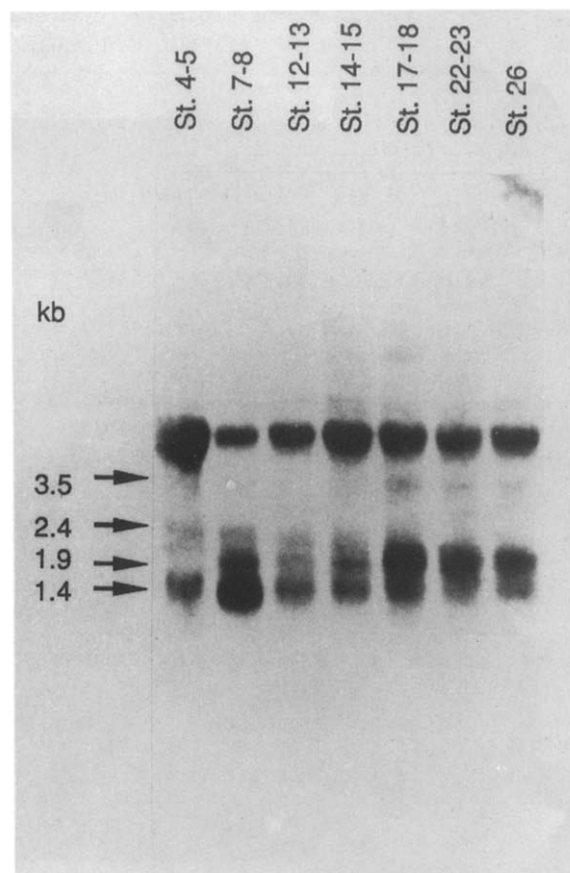


Fig.3. Developmental expression of the *CHox7* transcripts. RNA was prepared from fertilized eggs incubated for 16 h (for embryos at stages 4–5), 26 h (stages 7–8), 2 days (stages 12–13), 2.5 days (stages 14–15), 3 days (stages 17–18), 4 days (stages 22–23), and 5 days (stage 26). For stages 4–5, 25 μ g total RNA were loaded and for all other stages 5 μ g poly(A)⁺ RNA.

The *CHox7* probe additionally displays a band at the size of the 28 S rRNA seen also in poly(A)⁺ RNA (not shown). Cross-hybridization to the 28 S rRNA has been described for other homeo box probes [10].

4. DISCUSSION

Here, the sequence and developmentally regulated expression of a new chicken homeo box containing gene, *CHox7*, have been described. The *CHox7* homeo box sequence shows the same 9 invariant amino acids as most homeo boxes from

other organisms [3–5], making it a bona fide homeo box sequence. Comparison of the *CHox7* homeo box to other known homeo boxes revealed homologies at the DNA level of 67% or lower, and at the protein level of 54% or lower, making it a diverged member of the *Antp* class [3,5]. This is interesting, since the situation is usually reversed for most known homeo boxes [3–5]. However, previously we described two chicken homeo box genes which, like *CHox7*, were selected at random using *Antp*-type homeo boxes as probes [10]. Also, for the two previously described chicken homeo box sequences, the homologies are in the range of 60%, and are even lower when the amino acid sequences are compared. It is of interest to note that the best conserved regions of *CHox7* are amino acids 10–21 and 42–52, which correspond to helices 1 and 3 of the DNA-binding domain [22]. For all three chicken genes, no cognates have been found in other organisms. Together, all of these results suggest that the homeo box gene family in chickens may have diverged from the *Antp* class, however, more chicken homeo box sequences require cloning and sequencing in order to gain a better understanding of the evolution of this gene family in birds.

The *CHox7* gene codes for four transcripts, each exhibiting a different temporal pattern of expression. We have previously described the developmentally regulated expression of two other chicken homeo box genes, *CHox1* and *CHox3* [10]. Like *CHox7*, *CHox3* is already expressed at the time the egg is laid, and the level of expression of its abundant transcripts decreases after stages 7–8 [10]. The time of decline in *CHox7* expression between the peaks and that of decrease in *CHox3* expression overlap, suggesting possible common regulatory signal(s) on both genes. The striking developmental regulation of the chicken homeo box transcripts during embryogenesis supports the idea that they perform regulatory functions during the development of the chick embryo.

Acknowledgements: We wish to thank Dr B.W. O'Malley for the chicken oviduct genomic library, Dr W. McGinnis for the *Drosophila* probes and Dr F.H. Ruddle for the mouse probes. This work was supported by grants from the United States-Israel Binational Science Foundation (BSF), Jerusalem, Israel (no.86-00014), and the Fund for Basic Research, administered by The Israel Academy of Sciences and Humanities.

REFERENCES

- [1] McGinnis, W., Levine, M.S., Hafen, E., Kuroiwa, A. and Gehring, W.J. (1984) *Nature* 308, 428–432.
- [2] Scott, M.P. and Weiner, A.J. (1984) *Proc. Natl. Acad. Sci. USA* 81, 4115–4119.
- [3] Gehring, W.J. (1987) in: *Molecular Approaches to Developmental Biology* (Firtel, R.A. and Davidson, E.H. eds) pp.115–129, A.R. Liss, New York.
- [4] Gehring, W.J. (1987) *Science* 236, 1245–1252.
- [5] Hart, C.P., Fainsod, A. and Ruddle, F.H. (1987) *Genomics* 1, 182–195.
- [6] Njølstad, P.R., Molven, A., Eiken, H.G. and Fjose, A. (1988) *Gene* 73, 33–46.
- [7] Fritz, A. and De Robertis, E.M. (1988) *Nucleic Acids Res.* 16, 1453–1469.
- [8] Duboule, D., Galliot, B., Baron, A. and Featherstone, M.S. (1988) in: *Cell to Cell Signals in Mammalian Development* (De Laat, S.W. et al. eds) NATO ASI series, Springer, Berlin.
- [9] Simeone, A., Pannese, M., Acampora, D., D'Esposito, M. and Boncinelli, E. (1988) *Nucleic Acids Res.* 16, 5379–5390.
- [10] Rangini, Z., Frumkin, A., Shani, G., Guttmann, M., Eyal-Giladi, H., Gruenbaum, Y. and Fainsod, A. (1989) *Gene* 76, 61–74.
- [11] Colberg-Poley, A.M., Voss, S.D., Chowdhury, K. and Gruss, P. (1985) *Nature* 314, 713–718.
- [11a] Wright, V.E., Cho, K.W.Y., Fritz, A., Bürglin, T.R. and De Robertis, E.M. (1987) *EMBO J.* 6, 4083–4094.
- [12] Awgulewitsch, A., Utset, M.F., Hart, C.P., McGinnis, W. and Ruddle, F.H. (1986) *Nature* 320, 328–335.
- [13] Holland, P.W.H. and Hogan, B.L.M. (1988) *Genes Dev.* 2, 773–782.
- [14] Fainsod, A., Bogarad, L.D., Ruusala, T., Lubin, M., Crothers, D.M. and Ruddle, F.H. (1986) *Proc. Natl. Acad. Sci. USA* 83, 9532–9536.
- [15] Cho, K.W.Y., Goetz, J., Wright, C.V.E., Fritz, A., Hardwicke, J. and De Robertis, E.M. (1988) *EMBO J.* 7, 2139–2149.
- [16] Odenwald, W.F., Garben, J., Arnheiter, H., Tournier-Lasserre, E. and Lazzarini, R.A. (1989) *Genes Dev.* 3, 158–172.
- [17] Hogan, B., Holland, P. and Schofield, P.S. (1985) *Trends Genet.* 1, 67–73.
- [18] Hamburger, V. and Hamilton, H.L. (1951) *J. Morphol.* 88, 49–92.
- [19] Perry, M.M. (1988) *Nature* 331, 70–72.
- [20] McGinnis, W., Hart, C.P., Gehring, W.J. and Ruddle, F.H. (1984) *Cell* 38, 675–680.
- [21] Robert, B., Sasson, D., Jacq, B., Gehring, W. and Buckingham, M. (1989) *EMBO J.* 8, 91–100.
- [22] Otting, G., Qian, Y., Müller, M., Affolter, M., Gehring, W. and Wüthrich, K. (1988) *EMBO J.* 7, 4305–4309.