A distinct class of homeodomain proteins is encoded by
two sequentially expressed Drosophila genes from the
93D/E cluster

Krzysztof Jagla, Irina Stanceva+, Guy Dretzen, François Bellard and Maria Bellard*
Laboratoire de Génétique Moléculaire des Eukaryotes du CNRS, Unité 184 de Biologie Moléculaire
et de Génie Génétique de l'INSERM, Institut de Chimie Biologique, Faculté de Médecine,
11 rue Humann, 67085 Strasbourg Cedex, France

Received December 22, 1993; Revised and Accepted February 23, 1994
EMBL accession no. Z29082

ABSTRACT

Homeodomains appear to be one of the most frequently
employed DNA-binding domains in a superfamilly of transacting factors. It is likely that during evolution
several sub-types of homeodomain have evolved from
a common ancestral domain, resulting in distinct but
closely related DNA-binding preferences. Here we
describe the conservation of a distinct type of homeo-
domain encoded by the Drosophila lady-bird-late (lbl)
gene, previously named nkch4 (1). Using degenerate
PCR primers corresponding to the most divergent
regions of the first and third helix of the Lbl homeo-
domain we have amplified, from genomic DNA of the
fly, a lady-bird-like homeobox fragment. The Drosophila
PCR products contained both the lbl (1) and a highly
related homeobox sequence, which we named lady-
bird-early (lbe). This new Drosophila gene resides
directly upstream to lbl and together with tinman/NK4
(2, 3, 4, 5), bagpipe/NK3 (2, 4) S59/NK1 (4, 6) and 93Bal
(7) compose the 93D/E homeobox gene cluster. lbe and
lbl are transcribed from the same strand and in a
temporal order corresponding to their 5'–3' chromo-
somal location. Transcripts of both genes are found in
the epiderm of Drosophila embryos, in cells known to
express a segment polarity gene wingless (8), and their
spatial and temporal collnearity of expression strongly
suggests that they cooperate during segmentation. The
amino-acid composition of both Lady-bird homeo-
domains differ from that of Antp-type at several
positions involved in DNA recognition. These substi-
tutions appear to modify DNA-binding preferences
since Lbl homeodomain is unable to recognize the most
common homeodomain binding TAAT motif in gel
retardation experiments.

INTRODUCTION

Many developmentaly active transcription factors are encoded
by genes containing a highly conserved 180 bp motif, called the
homebox (9 reviewed in 7, 10, 11). Among several hundred
homeobox genes discovered up to now in the animal kingdom,
the majority is implicated in either axial pattern formation (12),
establishement of segment polarity (13) or determination of tissue
and cell identity (14, 15). Drosophila axial patterning homeobox
genes: lab, ph, Dfd, Scr, Antp, Ubx, abd-A and abd-B, are located
in the HOM-C cluster (12, 16) and display overlapping expression
domains that determine the segment identity of the developing
embryo. Genes belonging to HOM-C complex have structural
homologues in a wide range of vertebrate species, such as the
genes from Hox complexes in mouse and HOX complexes in
human (11). They have striking similarities in both chromosomal
organisation and in their antero—posterior expression pattern.
They regulate downstream genes via the DNA-binding domain
(homeodomain) located in the C-terminal region of the protein
(17, 18). Crystallography (19) and NMR (20) show that the highly
conserved amino acid residues at position 3, 5, 8, 47, 50, 51
and 54 are the most important for DNA recognition. Two of
them, Ile/Val47 and Asn51 make sequence-specific contacts with
a TAAT core motif (19, 20 and reviewed in 3). In vitro and in vivo experiments show that single amino acid
substitutions at positions 50 (22, 23) or 47 (24) strongly modulate
the DNA-binding specificities of homeodomains. On the other
hand, evolutionary conservation of the homeodomain between
cognate HOM and Hox proteins determines their closely related
DNA-binding preferences. Indeed, the human Dfd-like HOX4B
protein can specifically activate the autoregulated Dfd
transcription unit in Drosophila embryos (25), and mouse Hox1.3
is able to substitute for its Drosophila counterpart Scr in regulation
of Scr target genes (26).

*To whom correspondence should be addressed

+ Present address: Institut fur Zellbiologie, ETH-Honggerberg, CH-8093 Zurich, Switzerland
Among about 50 Drosophila homeobox genes mapping outside of the HOM-C cluster, five has been found inside the 93D/E region of the third chromosome. Three of them, S59 (6), bagpipe and tinman (msh2) (2, 3, 5), initially called NK1, NK3 and NK4 (4), participate in setting up the differentiation pathways of embryonic mesoderm. S59 is involved in establishment of myoblast and muscle identities (6), whereas tinman and bagpipe determine cell fates in the dorsal mesoderm and cooperate in visceral muscle differentiation (2, 4). Two remaining genes belonging to 93D/E cluster, 93Bal (7) and nkch4/lbl (1) are expressed in the developing CNS. The homeodomains of the 93D/E clustered genes have low homology (50–60%). However, several of these genes have structural homologs in the vertebrate genome (27, 28). For example, Drosophila 93Bal (27) encodes a homeodomain that is strongly homologous (92%) to human HOX11, which is involved in the pathogenesis of T cell leukemia. The tinman gene, required for heart formation in the fly embryo, has a homolog Csx which is specifically expressed in myocardial cells of the mouse embryo (28). We report here the cloning of a new Drosophila homeobox gene, lady-bird-early (lbe), a homolog of the previously described lady-bird-late (lbl) (1). lbe and the neighboring lbl gene (1) are located in the 93D/E cluster and have a similar spotted pattern of expression which demarcates parasegment boundaries in the dorsal and ventral epidermis of the Drosophila embryo. Several criteria, amino acid conservation within and around of the homeodomains, and in vitro affinity to the TAAT motif, suggest that the two lady-bird genes are the members of a new subfamily.

MATERIALS AND METHODS

Genomic PCR

Primers. Degenerate primers A, B and C were used to amplify lbl-related homeobox sequences. Forward primer A and reverse primer C (Fig.1A) were designed to detect both, intron-containing, lbl-like homeoboxes (intron between codons 44 and 45) (1) and those encoded by one exon. The reverse primer B was chosen to amplify, together with primer A, only homeoboxes without the intron.

PCR amplification. 200 ng of each primer with 50 ng of Drosophila genomic DNA in 20 μl of reaction mixture (200 μM dATP, dTTP, dGTP, dCTP, 10 mM Tris–HCl pH 8.3, 5 mM KCl, 1.5 mM MgCl₂, and 1U of Taq polymerase Cetus) were used for 30 cycles in a thermal cycler (Perkin Elmer-Cetus). After an initial denaturation of 5 min at 95°C the cycles were 1 min at 94°C, 2 min at 64°C and 3 min at 72°C. The amplified AC and AB primed DNA fragments (Fig. 1B) were separated in 2% low melting point agarose gel, cut out and purified by centrifugation through the siliconized glass-wool. After digestion with BamHI and HindIII, DNA fragments were cloned in pBluescript and sequenced with an Applied Biosystem sequencer.

Isolation and analysis of genomic clones

Screening of libraries. 5 × 10⁴ plaques of Drosophila genomic library in λ EMBL-3 (kindly provided by V.Pirotta) was screened using a PCR labelled (1) 183 bp probe corresponding to the AC primed DNA fragment of lbe homeobox (see above and legend to Fig.1). Hybridizations were performed in 50% formamide under stringency conditions chosen so as to select the most related homeobox genes.

Analysis of positive clones. DNA from positive λ clones were prepared by liquid culture method (29), restricted and after identification on control Southern blots, homeobox-containing EcoRI fragment was subcloned into pBluescript and sequenced as mentioned above.

Chromosomal localisation of the Drosophila lady-bird genes

In situ analysis. λ lbe and lbl genomic clones were digoxigenine random labelled using the 'DNA labeling and detection kit' (Boehringer Mannheim), hybridized to Drosophila (Oregon R) salivary gland polytene chromosomes and labelled chromosomal bands were visualized by alkaline phosphatase immunostaining.

Alignment of lbl cDNA with lbe genomic clones. Genomic proximity of lbe and lbl was confirmed by finding that the 5’ portion of lbl cDNA clone (our unpublished data) is present in lbe λ1 and λ4 genomic clones (Fig. 2A).

Developmental expression pattern of lbe and lbl genes

RT-PCR analysis. Total RNA from staged embryos, larvae, pupae and adult flies were prepared by the guanidine thiocyanate–LiCl method (31) and reverse transcribed according to Promega protocols (30) using following primers: 5’-TTTGGGCACGCCCTCCC-3’ for lbe and 5’-CGCTTTAAGTTTGGCCCTTCGTTTGG-3’ for lbl. Subsequently, the same primers were used, together with forward primers 5’-GATCGCCGCCTCACTGGGACTG-3’ for lbl and 5’-CCCGGCCAGTTGGAACATATTTGCCAC-3’ for lbl to amplify the RT products, through 30 cycles of PCR (94°C for 1 min, 68°C for 2 min and 72°C for 3 min). Control RT-PCR using specific primers for the rp49 gene, was performed as described previously (1). Amplified fragments were separated on 2% low melting point agarose gel, blotted and hybridized with an 32P end-labelled internal oligonucleotide probe.

In situ hybridisation. The digoxigenine labelled probes 111 bp and 99 bp in length, corresponding respectively to specific 3’ coding regions of lbe and lbl genes, were synthetized by PCR using 'Dig DNA Labeling Mix' (Boehringer Mannheim). The lbe and lbl transcripts were localized on Drosophila (Oregon R) embryos, after dechorionation and removal of vitelline membranes, by in situ hybridization (32).

Interaction of Lbl homeodomain peptide with TAAT motif

Overexpression and purification of his-tagged peptides. Lbl homeodomain encoding sequences and those of Dfd and HOX11, taken as a control, were amplified by RT or genomic PCR and cloned into Nhel site of the His-pET vector (33). E.coli BL21 strains (pLyS S or pLyS E) were electropropared in presence of the homeodomain expression vectors, grown, and IPTG induced as described previously (34). Histidine-tagged peptides were purified in non-denaturing conditions by chromatography through Ni2+–agarose affinity column (33).

Gel-shift assay. Homeodomain binding assays were performed by standard techniques (35, 36), by mixing different amount of purified peptides (Fig. 4B) with 2 × 10⁴ cpn (about 2 fmol) of 32P end-labelled TAAT-containing double-stranded oligonucleotide 5’-TAATGTTAAATGTTAATGTGTTAAGAG-3’ or 63 bp control nucleotide (37) in 20 mM Tris–HCl pH 7.6, 75 mM KCl, 1 mM dithiothreitol, 50 μg/ml BSA in final volume of 20 μl. After 20 min incubation at 4°C, complex and free probe were separated on a 5% (w/v) polyacrylamide gel containing TBE.
buffer (45 mM Tris-borate pH 7.5, 1 mM EDTA). Electrophoresis were performed at 25 mA for 2 h at room temperature.

RESULTS

Detection of lb-like homeoboxes and isolation of lbe genomic clones

Comparison of Lbl, originally called Nkch4 homeodomain (1) with those of known Drosophila and vertebrate homeoproteins revealed a striking divergency of the Lbl homeodomain in first and third helix. Using degenerate PCR primers (Fig. 1A) corresponding to the most divergent regions of Lbl homeodomain we amplified Drosophila, genomic DNA fragments 757, 183 or 150 bp in length, depending on the primers used (Fig. 1A and B). Primers A+C were designed to detect both the intron-containing and intron-less homeoboxes, when primers A+B to amplify homeoboxes without an intron between position 44/45. Sequence analysis revealed that the 183 bp AC amplification product correspond to novel homeobox fragment, highly related to that of lbi gene (Fig.1B). The 757 bp PCR product, flanked by primers AC (Fig.1B), corresponds, as expected, to the lbi homeobox, interrupted by a 574 bp intron (1). Southern blot analysis (Fig.1C) confirmed the PCR results (Fig.1B) and revealed the presence of two Drosophila genes that hybridise to 183 bp AC homeobox fragment. Alignment of both Drosophila PCR clones (Fig. 1D) showed that lbe and lbi carry at position 47 an ACC/T codon for Thr, which, according to (18), should

![Figure 1](image1.png)

The probe was Ibe digested Drosophila DNA. The filled box indicates the position of the Ibe homeobox and the open box with arrowhead.

![Figure 2](image2.png)

The probes correspond to first amino acid of the homeodomain.

The in-frame TGA stop codon is indicated at position +382 and putative AATAAA polyadenylation signal is italicized and underlined. The homeobox/homeodomain (filled box) and regions conserved between lbe and lbi genes are boxed (compare Fig.2C). (C) Comparison of deduced amino acid sequence of Lbl and Lbe homeodomains. Percentage of homology is calculated taking homeodomain of Lbl as reference. Other conserved regions are indicated by open boxes and amino acids identical to that of Lbi are represented by dashes. Shaded boxes denote chain of poly-Pro or poly-Ala, respectively, in carboxy-terminal region of Lbl and Lbe. Number (+1) corresponds to first amino acid of the homeodomain.
Clones are at silent positions (Fig. 1D), suggesting that these homeoboxes belong to the same subfamily. Moreover, most of the sequence differences between these two clones are at silent positions (Fig. 1D), suggesting that these homeoboxes belong to the same subfamily.

In order to determine the complete lbe homeobox sequence, a Drosophila genomic library was screened with the lbe AC fragment (Fig. 1A) as a probe and five positive λ clones were isolated (Fig. 2A). Further restriction mapping allowed us to orient these clones and identify homeobox-containing genomic fragments using a PCR labelled lbe probe. Sequencing of the 5kb lbe EcoRI fragment (Fig. 2A) showed that it contains the homeobox sequence identical to that obtained by PCR. The coding sequences are flanked by a splice acceptor site upstream and by stop codon downstream of the homeobox (Fig. 2B). The deduced Lbe and Lbl amino acid sequences (Fig. 2C and Fig. 2D) are highly homologous in the homeodomains (93%) and conserved outside the homeodomain (Fig. 2B and Fig. 2C). Thus, the amino-terminus of Lbl homeodomain is flanked by a PKK motif that is also present in Lbe protein (Fig. 2C). Similarly, the ELKDDV prolongation of Lbl homeodomain is retained in its homolog Lbe (Fig. 2C). Both Drosophila Lady-bird gene products additionally conserve a NNVNLILK sequence and a chain of poly-Ala (Lbe), or poly-Pro (Lbl) in carboxy-terminal region (Fig. 2C).

Chromosomal location of the Drosophila Lady-bird genes

We have previously shown that the lbe (nkh4) gene maps to bands 93D9 - E2 on the right arm of the Drosophila third chromosome, close to the NK (93D/E) homeobox gene cluster (1). To localize lbe, we have hybridized digoxigenin labeled λ1 lbe and λ1 lbe specific probe (1) to polytene chromosomes (Fig. 3A upper panels) and shown that lbe and lbe map to the same chromosomal...
region. This result was confirmed by a double probe in situ hybridization experiment which revealed that lbe and lbl label two closely positioned bands (Fig.3A lower panels). In addition to the cytological mapping we attempted to estimate, at the molecular level, the distance between the genes by alignment of the 5' portion of lbl cDNA (our unpublished data) to the λ lbe genomic clones (Fig.2A and Fig.3B). Two of them, lbe λ1 and λ4 were shown to contain the 5' region of lbl gene (Fig.2A). This result allowed us to localize lbe directly upstream to lbl and to find that both genes are transcribed from the same strand (Fig.3B).

Developmental expression pattern of lady-bird genes

RT-PCR analysis (Fig.3C,D) revealed that neighbouring lady-bird genes have an overlapping temporal profile of expression. Indeed, transcripts from the most 5' located lady-bird-early gene (Fig.3B) appear about 2 h earlier in the developing Drosophila embryo than those of lady-bird-late (Fig.3C,D). In situ hybridization data (Fig.3E,F) suggest that the same group of segmentally repeated epidermal cells express both genes. In addition double staining experiments (data not shown) reveal that lbe and lbl transcripts, appear in the dorsal and ventral epidermis and Ibl addition double staining experiments (data not shown) reveal that lbe and lbl hybridization experiment which revealed that Ibe label in situ region. This result was confirmed by a double probe in situ hybridization experiment which revealed that lbe and lbl label two closely positioned bands (Fig.3A lower panels). In addition to the cytological mapping we attempted to estimate, at the molecular level, the distance between the genes by alignment of the 5' portion of lbl cDNA (our unpublished data) to the λ lbe genomic clones (Fig.2A and Fig.3B). Two of them, lbe λ1 and λ4 were shown to contain the 5' region of lbl gene (Fig.2A). This result allowed us to localize lbe directly upstream to lbl and to find that both genes are transcribed from the same strand (Fig.3B).

DISCUSSION

Transcription factors containing the highly conserved homeodomain are involved in numerous regulatory pathways. The accomplishment of those functions requires precise and subtle interactions between the recognition region of the homeodomain and its target sequence. The substitution of amino acids involved in these interactions is certainly one of the ways to modulate the regulatory role of these proteins.

The Lady-bird-late (Lbl) homeodomain, which has unusual amino acids in both helix 1 and the DNA-recognition helix 3 (1), was found to be conserved in a second Drosophila protein, encoded by lady-bird-early (lbe) gene. Beyond the strong likeness (93%) of homeodomains, the lady-bird genes also conserve coding sequences adjacent to homeobox and located in a more 3' region. Sequence comparison with other homeodomains, revealed notable divergency and suggests that the Lady-bird homeodomains belong to a distinct subfamily. The lady-bird genes are located together with tin (2, 3), bap (2), S59 (6) and 93Bal (7) inside of the Drosophila 93D/E gene cluster. This gene clustering at 93D/E region may have a functional significance since their homeodomains are evolutionarily conserved. For example the tinman and bagpipe counterparts, Imx and Krx, reside tandemly in murine chromosome 14 (28) and the second mouse homolog of tinman, called Csx, is also involved in heart development (28).

The exact role of the 93D/E complex remains to be determined, however the tin mutation abolishes bap expression (2) indicating, that the 93D/E genes are clustered to facilitate interactions. tin and bap, and the neighbouring lady-bird genes, are expressed temporally according to their 5'—3' chromosomal position. The same phenomenon of temporal colinearity of expression has also been observed for the tandemly located segment polarity genes gooseberry and gooseberry-neuro (39) in Drosophila, for the Bx-C genes in short germ band insects (Schistocerca gregaria) (40) and the Hox genes in mouse (11, 41). The sequential activity of lbe and lbl appears in the common epidermal regions of the Drosophila embryo. In addition the lady-bird genes have a similar transcript distribution as segment polarity genes, wingless (8) and gooseberry (39, 42), further suggesting that they may also be involved in segmentation regulation pathways. Homeodomains encoded by 93Bal and both Lady bird genes have Thr instead of ile at position 47. This substitution is rare and has been observed, up to now, only in a few homeodomains: BarH1, BarH2 (14), Bsh (15), Hox11 (24,27). It has been postulated that ile47 contacts the TAAT motif directly (19, 20) and substitution of ile47 by Thr modify the DNA binding preferences of the Hox11 homeodomain (24). The Drosophila gene lbe codes for a Thr at position 47 as well as for a number of other substitutions which are either common with HOX11, or specific for Lady-bird. According to the En—DNA cocrystal structure (19) and NMR analysis of Anrp—DNA complexes (20), Arg5, Arg9, and Tyr8, from the N-terminal region of homeodomain, together with ile47 and Asn51 from the third helix, contribute to TAAT subtype recognition. Like HOX11, both Lady-bird homeodomains conserve Arg5 and Asn51, but differ from Antp-type homeodomain at position 3 (Lys), 8 (Phe) and 47 (Thr). These common features would influence the affinity of the
HOX11 and Lady-bird homeodomains for the TAAT motif. Our mobility shift experiments show that the HOX11 homeodomain has a 40 times lower affinity for the TAATGG multimer compared to Dfd (an Antp-type homeodomain). This finding is also supported by the observation that the Gtx homeodomain, compared to Dfd (an Antp-type homeodomain). This finding is has a 40 times lower affinity for the TAATGG multimer mobility shift experiments show that the H0X11 homeodomain HOX11 and Lady-bird homeodomains for the TAAT motif. Our positions seem to only partially explain the binding preferences carrying only substitutions at position 3 binds extremly weakly possesses ability to bind to the control oligonucleotide. From alignment of Lbl and Lbe homeodomains, one can conclude that there are several, amino acid substitutions specific for Lady-bird including those at the highly conserved positions 42 (Asp/Glu→Asn), 46 (Lys→Ile) and 54 (Met→Ala). Although only Met54 is known to contact the target sequence directly (20), the other modifications present in the recognition helix may modulate the overall binding preferences of the Lbl homeodomain.

Determination of the DNA-binding repertoire of Lady-bird homeodomains, followed by search for the target genes, should yield insights into mechanisms of their action and provide some answer to the question whether or not members of this new homebox gene family are involved in analogous developmental decisions.

ACKNOWLEDGEMENTS

We thank P. Chambon for support, B. Wasylyk and P. Dollé for critical reading of the manuscript. We are grateful to F. Ruffenach for oligonucleotides synthesis, S. Vicaire for sequencing and M. Boulay for excellent photo service. This work was supported by grants no. 92N60/0694 from Ministère de la Recherche, from CNRS, the INSERM, The Fondation pour la Recherche Médicale and the Association pour la Recherche sur le Cancer.

REFERENCES