The *Drosophila copia* retrotransposon contains binding sites for transcriptional regulation by homeoproteins

Laurent Cavarec and Thierry Heidmann*

Unités de Physicochimie et Pharmacologie des Macromolécules Biologiques, CNRS U147 et INSERM U140, Institut Gustave Roussy, 39 rue Camille Desmoulins, 94805 Villejuif Cedex, France

Received August 16, Revised and Accepted October 7, 1993

ABSTRACT

We have identified in the 5' untranslated region of the *Drosophila copia* retrotransposon, 3' to the left LTR, a sequence for transcriptional regulation by homeoproteins. Co-transfection assays using expression vectors for homeoproteins and reporter vectors containing the lacZ gene under the control of either the entire copia LTR with 5' untranslated sequence, or a minimal heterologous promoter flanked with a 130bp fragment containing the copia untranslated region, disclosed both positive and negative modulations of promoter activity in *Drosophila* cells in culture: a 5–10 fold decrease with engrailed, even-skipped and zerknüllt in DH33 cells, and a 10–30 fold increase with fushi tarazu and zerknüllt in Schneider II cells. In all cases, the regulatory effects were abolished with reporter plasmids deleted for a 58 bp fragment encompassing the putative homeoprotein binding sites. Mobility shift assays with a purified homeodomain-containing peptide demonstrated direct interaction with the 58bp fragment, with an affinity in the 1–10nM range as reported with the same peptide for other well characterized homeodomain binding regulatory sites. Footprinting experiments with the extended LTR demonstrated protection of 'consensus' sequences, located within the 58bp fragment. These homeodomain binding sites could be involved in the developmental regulation of the copia retrotransposon.

INTRODUCTION

Mobile genetic elements—or transposons—are 'endogenous mutagens' that have been found in all plant and animal species studied so far. In *Drosophila*, they represent several percent of the genome in mass and account for up to 90% of the spontaneous mutations detected by classical genetic analysis [compiled in (1)]. Transposition events are rare, but their frequency can be dramatically increased, depending on both genetic (for instance hybrid dysgenesis, presence of suppressor genes) and epigenetic ('stress', heat-shock, UV irradiation) factors [reviewed in (2)].

There are several classes of transposons, but a large fraction of *Drosophila* mobile elements are 'retrotransposons' [reviewed

in (2, 3)], either of the retrovirus-like family—as the prototype 'copia' element, or of the LINE family—as the prototype I element involved in the I-R hybrid dysgenesis syndrome [reviewed in (4, 5)]. Their transposition is replicative and involves reverse transcription of an intermediate RNA transcript of the element, as now demonstrated for the yeast Ty1 (6) and the mammalian IAP (7) retrovirus-like elements, and for the *Drosophila* I LINE element (8, 9). Clearly, regulation of the levels of transcription of these mobile elements should control their transposition frequency, as demonstrated for the yeast Ty1 and the mammalian IAP retrotransposons for which induction of transcription at a high level results in high transposition frequency (6, 10).

The structure and molecular basis of transposition of several of these elements are now well characterized, but little is known about the factors and host genes involved in their transcriptional control. In *Drosophila*, analysis of the transcripts from the whole animal at various stages of development disclosed defined patterns of transcription with, as a rule, only limited expression in the adult (11). The extensive analysis in (11) of the temporal pattern of expression of nine retrovirus-like retrotransposons demonstrated diverse patterns among transposons, with for instance a peak of expression for second-third instar larvae for copia, first instar larvae and young pupae for the 412 and mdg1 elements, and embryos and larval stages for the 297 element. Interestingly, developmentally regulated expression of retrotransposons have also been observed for some mammalian transposable elements, such as the IAP elements—which are expressed almost exclusively at the 2-cell stage of embryo development [and also in tumors; reviewed in (12)], or the ET (Early Transposon) element—which is also expressed only very early in development (13).

It is well established that several developmentally regulated genes in *Drosophila* (as well as in plants and vertebrates) are controlled by homeoproteins, which bind to DNA sequences and act as transcriptional regulators. Homeoproteins are known to be both negative and positive regulators, and to act in association with each other or with other transcription factors to elicit complex patterns of expression, both temporally and spatially, in the developing *Drosophila* [reviewed in (14–18)]. Homeoproteins contain highly conserved 60-aminoacid sequences—

* To whom correspondence should be addressed

referred to as the homeodomains (HD)—which have been isolated from several homeoproteins and were shown to be responsible for DNA binding activity (19–23). The structure of the HD polypeptides has been elucidated by NMR spectroscopy and crystallographic analysis, and discloses a helix-turn-helix structure analogous to that of many prokaryotic transcriptional repressors and activators (24–26). *In vitro* analysis of the binding to DNA of a number of homeoproteins—and of HD peptides—allowed derivation of a minimal consensus DNA binding sequence—as a ‘core’ ATTA sequence—the importance of which was confirmed by the recently determined three-dimensional structure of the Antp and engrailed homeodomain–DNA complexes (26–28). The ATTA core sequence plays a primary role in determining the affinity of binding, with significant secondary contributions deriving from the flanking bases (22, 23, 29, 30).

In this study, we have investigated the transcriptional properties of the *Drosophila* copia element. This prototype retrovirus-like retrotransposon, which has been previously entirely sequenced, contains two 276 bp LTRs, a single ORF with sequence homologies to the integrase, protease and reverse transcriptase of vertebrate retroviruses, and a 145 bp 5′ untranslated region (31). An enhancer sequence—with homologies to the SV40 enhancer—had previously been identified (31, 32), which lies outside of the LTR, in the 5′ untranslated region. We report the identification of a sequence adjacent to this enhancer, which displays (i) direct interaction with a purified Antp homeodomain peptide, as demonstrated by both footprinting experiments and mobility shift assays, and (ii) positive or negative modulation of the level of activity of the copia LTR promoter—as well as of an heterologous minimal promoter—in transient co-transfection assays in cultured cells from either *D.melanogaster* or *D.hydei*, using expression vectors for several homeoproteins.

MATERIALS AND METHODS

DNA constructions

The copia element used is the ‘white apricot insertion’ [λ w^a5.9 clone in (33), sequenced in (31)], that we subcloned into the Bluescript vector (Stratagene) as a *Ban*I fragment, leaving only 77bp and 87bp of white DNA respectively 5′ and 3′ to the copia element (pBScopia). Four oligonucleotides were synthesized on an Applied Biosystems DNA synthesizer: Hom1=(5′)CAGTCCATGCCTAATAAAC; Hom2=(5′)CACAATTTATTTTCA-CAATC; Enh1=(5′)AAATAGCATTTTTTTCACATTC; Enh2=(5′)CTGAGAAGGAAATAATTTC; they were used to PCR-amplify from the copia-containing plasmid a 130bp fragment containing both the consensus homeoprotein binding sites and the enhancer sequences (using Hom1 and Enh2), a 58bp fragment containing the homeoprotein binding sites (using Hom1 and Hom2), and a 72bp fragment containing the two dyad symmetry of the putative enhancer region (using Enh1 and Enh2).

Construction of the copia LTR-lacZ responder plasmids. LTR lacZ: a 490bp *Ban*I–*Dde*I fragment from the pBScopia plasmid (containing the entire copia LTR and 5′ untranslated sequence) was inserted at the unique *Sal*I site of pGEMnslacZ [which contains a lacZ gene with nuclear location signal (34)], after Klenow treatment of both vector and insert; the polyadenylation sequence from the hsp70 gene [as a *Bam*HI–*Eco*RI fragment from HZ50 (35)] was then inserted at a unique *Bam*HI site 3′ to the nslacZ gene, after Klenow treatment of both vector and insert. LTR Δ HpRE lacZ: a 360bp *Ban*I–*Apa*I fragment (LTR

alone) from pBScopia was inserted at the unique *Sal*I site of a pGEMnslacZ plasmid containing the polyadenylation sequence above, after Klenow treatment of both vector and insert (which reconstitutes the *Sal*I site at the *Sal*I–*Apa*I junction); the 72bp fragment obtained by PCR amplification using the Enh1 and Enh2 primers (enhancer region) was then inserted 3′ to the copia LTR at the reconstituted *Sal*I site, after Klenow treatment of the vector and T4 kinase treatment of the insert.

Construction of the hsp70-lacZ reporter plasmids. The hsp70-lacZ responder plasmids were constructed by inserting the copia regulatory sequences at the unique *Not*I site of a plasmid containing a lacZ gene under control of an hsp70 enhancerless promoter and polyadenylation sequence [HZ50 plasmid (35)]; the inserted sequences were PCR-amplified fragments, either 130bp long (Hom1 and Enh2 primers) or 72bp long (Enh1 and Enh2 primers), which were treated with T4 kinase before ligation with the Klenow-treated vector.

Expression vectors for homeoproteins. These were gifts from Drs K.Han and M.S.Levine, and are described in (36); they contain the coding sequence for homeoproteins inserted between the Actin-5C promoter and polyadenylation signal; the control pAct vector corresponds to the same plasmids, but with no insert [pAct5C-PPA in (36)].

Cells, transfections, β -galactosidase assays and Northern blot analysis

Schneider II cells from *D.melanogaster* (37) and DH33 cells from *D.hydei* (38) were grown in Schneider medium (GIBCO) with 10% fetal calf serum (GIBCO) at 25°C. Transfections were performed by the calcium phosphate method (39), with 500ng and 5 μ g of reporter plasmid for DH33 and Schneider II cells respectively (for 2.5×10^6 cells), and varying amounts of expression vector (between 100ng and 5 μ g; the total amount of transfected DNA being maintained constant with Bluescript plasmid, Stratagene). β -galactosidase activity was measured two days post transfection after protein extraction as described in (40), using a spectrophotometric assay with CPRG (Boehringer) as a β -galactosidase substrate. In standard assays, 125 μ l of protein extract (0.1 mg/ml) and 4.5 mM CPRG were used for Schneider II cells, and 5 to 50 μ l (0.01 mg/ml) of protein extract with 1mM CPRG for DH33 cells; β -galactosidase is expressed as O.D. units per μ g of protein extract per minute. For analysis of RNA levels, transfections were performed as above, but with 10^7 cells, and 15 μ g and 30 μ g of expression vector and responder plasmid respectively. Total cellular RNAs were extracted two days post-transfection and analysed as in (8), upon electrophoresis of 10 μ g aliquots on 1% agarose-formaldehyde gels and transfer to nylon filters (Hybond N; Amersham) in 0.15M ammonium acetate. Filters were hybridized with a 32 P-labelled lacZ probe in Church solution (41) with final wash in 0.1 $\times$ SSC, 0.1% SDS at 65°C.

Mobility shift assays

The 58bp fragment amplified by PCR with the Hom1 and Hom2 primers was radioactively labelled with γ - 32 P dATP (3000Ci/mmol; Amersham) using T4 polynucleotide kinase (Biolabs). The radiolabelled fragment was then purified on a native polyacrylamide gel. When indicated, isotopic dilutions (at least 100-fold) were performed with unlabelled 58bp fragment whose concentration was determined by direct absorption measurement at 260nm. The Antp homeodomain is a 6154 g/mol purified

peptide described in (42). Binding reactions were carried out at $20 \pm 1^\circ\text{C}$ for 4hr in $20\mu\text{l}$ of a solution containing 20mM Tris-HCl, pH7.6, 75mM KCl, $50\mu\text{g}$ bovine serum albumin per ml, 1mM dithiothreitol, and 10% glycerol, i.e. as in (21). The samples were then run on native 9% polyacrylamide gels at room temperature for 2h30 at 240V. When indicated, irrelevant 22-mer double-stranded oligonucleotides were added in competition assays (SP1 site: 5'-GATCGATCGGGGCGGGGCGATC-3'; AP1 site: 5' CTAGTGATGAGTCAGCCGGATC-3').

DNase I footprinting

About $3\mu\text{g}$ of a 578bp HindIII fragment from the LTR-lacZ responder was end-labelled by filling in the HindIII sites with $\alpha^{32}\text{P}$ dCTP with the Klenow fragment of DNA polymerase I. This fragment was then restricted with PstI to remove the 5' labelled end, and purified on a Sephadex G75 column. The probe was then run at room temperature on a native preparative 5% polyacrylamide gel and recovered by the 'crush and soak' method (39), followed by phenol extraction and ethanol precipitation. 3 fmol (1ng) of this fragment was incubated with varying amounts

of Antp homeodomain in $50\mu\text{l}$ of incubation buffer [20mM Tris-HCl, pH7.6, 75mM KCl, $50\mu\text{g}$ bovine serum albumin per ml, 1mM dithiothreitol, 10% glycerol, $2\mu\text{g}/\text{ml}$ poly (dI-dC).(dI-dC)], for 30 min at 25°C . Two microliters of 375U/ml DNaseI was added and incubated for 1 minute at 25°C in the presence of 1mM MgCl_2 and 0.5mM CaCl_2 . The reaction was stopped upon addition of $140\mu\text{l}$ of 'stop buffer' (190mM sodium acetate, 30mM EDTA, 0.15% SDS and $65\mu\text{g}/\text{ml}$ yeast RNA). DNA was isolated by phenol-chloroform and chloroform extractions, followed by ethanol precipitation, and analyzed on a 8% sequencing gel. A 'G+A' reaction was performed as in (43).

RESULTS

Identification of a functional homeoprotein-responsive element in the copia 5' untranslated region

To test for the ability of homeoproteins to modulate transcription from the copia promoter, we used a transient co-transfection assay in *Drosophila* cells in culture, with constructs containing the copia promoter linked to a readily assayable reporter gene (the lacZ

Figure 1. Regulatory elements in the copia LTR and 5' untranslated region, and structure of the responder plasmids for cotransfection assays with homeoprotein producing vectors. (A) Schematic representation of the copia LTR with U3-R-U5 organization and RNA start site (+1), and of the 5' untranslated sequence with the 22bp repeats (grey boxes) displaying dyad symmetry and homology to SV40 core enhancer and the ATTA putative 'core' sequences for homeoprotein binding (open boxes); sequences with homologies to the NP consensus are boxed with dotted lines, with perfect matches in capital letters. The 5' ends of the oligonucleotides used for PCR amplification of the 130 bp untranslated regulatory sequence, of the 58 bp sequence containing the homeoprotein responsive element (HpRE), and of the SV40 enhancer-like sequence (ENH) are indicated with horizontal arrows. (B and C) Structure of the two 'responder' plasmids used in the cotransfection assays with the homeoprotein-producing vectors. The copia LTR-based responder plasmid with the lacZ gene for β -galactosidase contains the entire 130 bp untranslated sequence deleted or not for the HpRE sequence (deletion indicated by the brackets). The hsp70-based responder plasmid contains a 'minimal' hsp70 promoter (enhancerless promoter, with TATA box and RNA start site indicated) with either the complete or the HpRE-deleted copia untranslated sequence 5' to the heterologous promoter.

gene coding for β -galactosidase), and expression vectors for homeoproteins (36). A first reporter plasmid was constructed, which contained the entire copia LTR plus the 5' untranslated region, inserted as a *Ban*I–*Dde*I 490bp fragment 5' to a *lacZ* gene with the polyadenylation signal from the *hsp70* gene (LTR-*lacZ* responder; see Figure 1); the copia 5' untranslated sequence was retained as it contains two previously identified 22-nucleotide sequences at positions 354 to 375 and 382 to 403, which display dyad symmetry and present homologies with a sequence occurring in the simian virus 40 enhancer (31, 32). In addition, it contains 3 ATTA 'core' sequences on either strand, two of them being part of a near perfect (9/10 and 8/10 matches) NP sequence (20, 44) for binding of the engrailed homeoprotein (TCAATTAAAT; boxed in Figure 1). Three other reporter plasmids were constructed (see Figure 1): the same as above but with a 58bp deletion in the untranslated sequence encompassing the three ATTA cores (from position 287 to position 345), and two plasmids (*hsp70-lacZ* responders) in which the untranslated sequence alone (a 130 bp fragment deleted or not for the 58bp fragment above), was inserted 5' to a minimal heterologous promoter, the *hsp70* promoter deleted for its own enhancer elements [HZ50; (35)]. In standard assays, DH33 or Schneider II cells were co-transfected with a fixed amount of reporter plasmid and increasing amounts of expression vectors coding for either the homeoprotein fushi tarazu (FTZ), zerknullt (ZEN), paired (PRD), engrailed (EN), or even-skipped (EVE), according to the procedures described in Materials and Methods. β -galactosidase activity of the responder element was measured by a quantitative assay using chlorophenol b-red D galactopyranoside (CPRG; see Materials and Methods) as a substrate. An expression vector without homeoprotein coding sequence (pAct) was used in parallel as a control, and each transfection experiment was repeated from two to five times.

Even-skipped, engrailed and zerknullt act as repressors of copia in DH33 cells. As illustrated in Figure 2, cotransfection of the reporter plasmid containing the extended copia LTR with expression vectors for even-skipped (EVE), engrailed (EN) or zerknullt (ZEN) had a severe effect on the level of expression of the copia reporter, which was reduced to approximately 20% of its initial value. The effect was dose-dependent, was not observed with the control expression vector even at the highest concentration of vector tested (5mg), and was homeoprotein-specific since it was not observed with fushi tarazu (or paired, not shown). To determine if the ATTA core sequences included between position 305 and position 341 are involved in these regulations, cotransfections were performed with the LTR-*lacZ* responder plasmid deleted for the 58bp fragment described above. Under these conditions, the regulatory effects of zerknullt (not shown), even-skipped and engrailed (see Figure 2) were not observed any more, but removal of the ATTA core sequences also resulted in a decrease of the basal activity of the reporter plasmid (about 10-fold). One possible interpretation for these effects could be that homeoproteins with an 'activator' effect on copia expression are already present in DH33 cells and their positive effect is abolished upon removal of the ATTA sites; accordingly, the regulatory effects of even-skipped, engrailed and zerknullt would result from the displacement of these putative endogenous homeoproteins from the identified homeoprotein responsive sequences. An alternative and more plausible explanation (see the experiment reported below) could be that removal of the 58bp fragment—which is located downstream to

Figure 2. Even-skipped (EVE), engrailed (EN) and zerknullt (ZEN)—but not fushi tarazu (FTZ)—are specific repressors of the copia LTR-*lacZ* responder in DH33 cells. Cotransfection assays were performed with homeoprotein expression vectors (1 to 5 μ g, see abscissa) and 5 μ g of the copia LTR-*lacZ* responder plasmids in Figure 1B either complete (filled symbols) or deleted (open symbols) from the 58bp homeoprotein responder element (HpRE). Control assays are with an expression vector (pAct) containing the same promoter but no coding sequence; total plasmid DNA concentrations in the cotransfections were maintained constant by complementation with Bluescript plasmid. β -galactosidase activity is expressed relative to that of the responder plasmid alone taken as 1. Each value is the mean of at least two independent experiments, or more when error bars (standard deviations) are indicated. The background level of the assay (mock-transfected cells) is approximately 0.01 (i.e. 10 times lower than the smallest value in the Figure).

the promoter start site—disrupts specific interactions between elements of the transcriptional machinery, or modifies mRNA stability and/or translation.

A definite evidence for the role of the ATTA-containing copia untranslated region was obtained by using the *hsp70-lacZ* responder plasmids in Figure 1C, in which the regulatory elements of the copia untranslated region have been isolated (as a 130bp PCR-amplified fragment) and linked to an heterologous 'minimal' promoter [HZ50 (35)]. In this construct, the *lacZ* reporter gene is under the control of the *hsp70* promoter deleted for its own heat shock consensus and enhancer sequences, and the copia 130bp untranslated sequence is inserted upstream to the promoter start site. With this reporter, cotransfection assays as in Figure 2 disclosed a pattern of regulation identical to that observed with the extended copia promoter (Figure 3): again, engrailed, even-skipped and zerknullt severely reduced the level of expression of the reporter plasmid; as found previously, the effect was dose-dependent, as well as homeoprotein-dependent since it was not observed with fushi tarazu (or paired, not shown). Finally, removal from the reporter plasmid of the ATTA-containing 58bp fragment (Figure 3B) totally abolished the regulation by homeoproteins, without any significant effect, in that case, on the basal level of the reporter plasmid. An HZ50-derived construct with the 58bp fragment alone as a regulatory element (i.e. without the SV40-like enhancer elements of the untranslated region) was also tested, but its level of expression was close to background, thus preventing the detection of any possible negative effect of homeoproteins under these conditions.

These results therefore demonstrate the existence of an homeoprotein-responsive element (HpRE) in a 58bp fragment

Figure 3. The copia untranslated region contains all the elements for regulation of an heterologous promoter by homeoproteins. Cotransfection assays were carried out in DH33 cells as in Figure 2, with the hsp70-lacZ responder plasmids shown in Figure 1C; each value is the mean of at least two independent experiments, or more when error bars (standard deviations) are indicated. **(A)** Even-skipped, engrailed, zerknüllt, but not fushi tarazu, are repressors as observed for the copia LTR-based responder plasmid; same conditions as in Figure 2. **(B)** Regulation by homeoproteins of the responder plasmid depends on the presence of the copia 58 bp homeoprotein responsive element (HpRE); activity of the responder plasmids (5 μ g) with (+HpRE) or without (-HpRE) the copia 58bp HpRE was assayed with either 1 or 5 μ g of even-skipped producing vector (dark filled) or the pAct control vector (control, open); the activity of the HpRE-containing responder plasmid with no expression vector was taken as the common reference equal to unity.

from the copia 5' untranslated region, which can act in association with either the native or an heterologous promoter—de facto either upstream or downstream to the RNA start sites. Incidentally, they also show that the copia LTR and the flanking 77bp of the white gene remaining in the LTR-lacZ responder are not involved in homeoprotein regulation, since their removal—in the hsp70-lacZ responder—had no effect.

Fushi tarazu and zerknüllt are transcriptional activators of the copia homeoprotein responsive element in Schneider II cells. It has been observed [(32), and see below] that the expression of copia is much higher (at least 10 times) in DH33 cells (derived from *D. hydei*, which do not contain copia elements), than in Schneider II cells (derived from *D. melanogaster*, which possess approx. 100 copia elements), possibly because different sets of transcription factors are expressed in each cell line [for instance, homeoprotein expression has never been detected in Schneider II cells, but other transcription factors as well can be differentially expressed, (45)]. Schneider II cells might therefore constitute a simple cellular model to detect direct effects of exogenous homeoproteins, and cotransfection experiments as above were performed with the simplified reporter gene (hsp70-lacZ responders, see Figure 1C). As expected and at variance with transfections of DH33 cells, 10 times more responder plasmid had to be used to detect signals significantly above the background level of the assay. Under these conditions, we found that fushi tarazu and zerknüllt (Figure 4A, and paired, not shown) had a strong positive effect on the level of expression of the reporter gene: the effect was dose-dependent, with an increase of at least 10-fold (see Figure 4A, B); activation was not observed with the control plasmid pAct, and is homeoprotein-dependent since engrailed and even-skipped had no detectable effects (or a small negative effect, if any, for engrailed at low concentrations). Finally, cotransfection experiments with a responder from which the 58bp homeoprotein responsive element had been deleted, did

not show any effect of the same homeoproteins (Figure 4B). These results unambiguously confirm the role of the previously identified homeoprotein-responsive element in the observed regulations. Interestingly, we observed that fushi tarazu and zerknüllt were able to mediate their enhancer activity (although to a lesser extent) on the 58bp HpRE alone, when placed upstream to the minimal hsp70 promoter (data not shown).

Northern blot analyses were finally performed to determine whether the positive regulations observed in these cells actually reflect increases in RNA levels and to eliminate the possibility of translational effects. Co-transfections with the responder plasmid above (30 μ g) and the different expression vectors (15 μ g) were performed in 10^7 Schneider II cells, RNAs were extracted 48hrs post-transfection and Northern blots hybridized with a lacZ probe. As can be seen in Figure 4C, the level of lacZ transcripts was enhanced when the reporter gene was co-transfected with plasmids coding for fushi tarazu and zerknüllt, whereas engrailed and evenskipped vectors had no effect—or a small negative effect—in agreement with the data on β -galactosidase activities.

Identification of homeodomain binding sites in the copia homeoprotein-responsive element

Co-transfection experiments with expression vectors for homeoproteins do not indicate whether homeoproteins act directly or indirectly (ie via other intermediates) at the level of the identified homeoprotein-responsive elements in copia. To get further insight into the role of these elements, we therefore looked for the presence of homeoprotein binding sites specifically located in the functionally defined regulatory region. *In vitro* assays were performed (DNase I protection experiments and mobility shift assays) using a purified Antennapedia homeodomain (Antp HD) peptide as a specific ligand. This Antp HD is a 60-aminoacid protein which has been previously demonstrated to specifically interact, *in vitro*, with already known homeoprotein binding sites, with an affinity in the 1–10 nanomolar range (19, 21).

Figure 4. Fushi tarazu and zerknullt are specific activators of the copia homeoprotein responsive element in Schneider II cells. (A) Cotransfection assays were as in Figure 3, but in Schneider II cells, with 0.5 µg of the hsp70-lacZ responder plasmid shown in Figure 1C; differential regulatory effects are observed in Schneider II versus DH33 cells for zerknullt (strong positive/strong negative effects), fushi tarazu (strong positive/null effects) and engrailed (nul/strong negative effects); control refers to pAct plasmid and activities are expressed relative to that of the responder plasmid alone taken as unity; each value is the mean of two independent experiments. (B) Homeoprotein positive regulations are dependent on the presence of the copia Hpre in the responder plasmid. Same conditions as in 4A; activities of the responder plasmids with (+Hpre) or without (-Hpre) the copia 58bp homeoprotein responsive element were assayed with 1 or 5 µg of either pAct control plasmid (control, open), fushi tarazu (grey) or zerknullt (dark) expression vectors; the activity of the Hpre-containing responder cotransfected with 1 µg of control plasmid was taken as unity; each value is the mean of two independent experiments. (C) Northern blot analysis of responder plasmid expression. 15mg of responder plasmid and 30 µg of the indicated homeoprotein producers were co-transfected in 10⁷ Schneider II cells: responder alone (lane 1), even-skipped (lane 2), engrailed (lane 3), fushi tarazu (lane 4), zerknullt (lane 5). Each lane corresponds to 10mg of total RNA extracted 48hrs post-transfection; the 3.9kb 'responder' transcript (upon hybridization of the blot with a lacZ radiolabelled probe) and position of the calf 28S ribosomal RNA (r) are indicated.

Figure 5. DNase I footprinting analysis of the copia LTR and untranslated region using a 60 aminoacid homeodomain-containing polypeptide (HD). Footprinting reactions were performed on a 565bp PstI-HindIII fragment containing the entire copia LTR and untranslated 5' region (labelled at the 3' end of the transcribed strand), and a 60 aminoacid homeodomain-containing polypeptide (Antennapedia HD, cf Materials and Methods). A G+A sequencing reaction was run in the adjacent lane. Amounts of Antp peptide used are as follows: lanes 1 and 7 (none), lane 2 (25nM), lane 3 (50nM), lane 4 (75nM), lane 5 (90nM), lane 6 (115nM). The sequence of the strongly protected region (I) is indicated with the ATTA core boxed in solid lines, as well as that of the adjacent region (II) with weaker protection; the U3, R and U5 domains of the copia LTR are indicated.

DNase I protection experiments. DNase I protection experiments were carried out with a 565bp PstI-HindIII fragment from the copia lacZ construct, which contains the complete copia LTR and untranslated sequences. End-labelled DNA (1nM) was incubated with increasing concentrations of Antp HD in the presence of polydI-dC, submitted to DNase I digestion, and run on a sequencing gel as described in Materials and Methods. As illustrated in Figure 5, a major region is protected from DNaseI digestion, namely the AATAAACAATTAAATTG sequence located in the 58bp homeoprotein responsive element. This protected region, approximately 15bp long, is centered around an ATTA core embedded in the near-perfect NP sequence indicated in Figure 1 (9/10 match, box I). Increasing HD concentration (up to 100nM) resulted in a weaker and incomplete protection of the second previously noted NP-like sequence located just 3' to the previous one (8/10 match, box II). In the LTR domain, no protection can be observed even in the 3' end of the LTR where three ATTA cores are present, consistent with the results from the co-transfection experiments on the absence of regulatory sequences in this region. The absence of protection of these ATTA sequences, as well as the reduced protection of

Figure 6. Mobility shift assay of Antp-homeodomain/copia-HpRE interactions. Binding of the Antp HD peptide, at the indicated concentrations, to the 58bp end-labelled copia HpRE was assayed as indicated in Materials and Methods. (upper) Increasing concentrations of Antp HD were incubated with a fixed amount of end-labelled copia HpRE and run on a 9% polyacrylamide gel; F: free DNA, B1 and B2: bound DNA-HD complexes. (lower) Specificity of binding; end-labelled copia HpRE (0.1nM) and Antp HD at the indicated concentrations were incubated in the presence of competitor DNA—either the unlabelled copia HpRE or an unrelated 22-mer double-stranded oligonucleotide (SP1 site, same results with AP1 site, cf Materials and Methods)—at the indicated concentrations.

the second more divergent NP-like sequence above, constitute internal controls for the specificity of the observed HD protection.

Mobility shift assay. The results above were confirmed, in a quantitative manner, by mobility shift assays. Binding of the Antp HD to the previously characterized 58bp fragment could be readily detected by the retarded electrophoretic mobility of the end-labelled HpRE DNA fragment. Indeed, in the nanomolar range of Antp HD concentrations, a band of retarded mobility could be clearly observed (B1, see Figure 6), whose intensity increased in a dose-dependent manner with HD concentration. This retarded band corresponds to a specific interaction of the Antp HD to saturable sites since it was competed out by an excess of unlabelled binding sequence added prior to the labelled DNA, and not by an identical amount of irrelevant unlabelled oligonucleotides (22-mer, see Figure 6 bottom left and Materials and Methods). In the high concentration range, a second retarded complex could be observed (B2, Figure 6 upper part), which is also associated with specific interaction to saturable sites since it could be competed, as well as the B1 complex, by an excess of unlabelled HpRE DNA (see Figure 6, bottom right). The equilibrium dissociation constant (K_d) for binding to the major B1 site was accurately determined from Antp homeodomain saturation experiments by varying the concentration of the 58bp HpRE fragment, as illustrated in Figure 7; labelled DNA was isotopically diluted (see Materials and Methods), so that DNA concentration could be precisely determined. Accordingly, the K_d was found equal to 4 nM at $20 \pm 1^\circ\text{C}$ (Figure 7). This value

Figure 7. Determination of equilibrium binding coefficients by electrophoretic mobility shift assay. Antp HD peptide at a fixed concentration of 10nM was incubated for 4 hr in binding buffer containing increasing concentrations of labelled HpRE (0.39 to 100nM with regular two-fold increase, from lane 1 to 9). Bound (B) and free (F) DNAs were visualized by autoradiography (upper part), and data obtained from the titration plotted (lower part); an equilibrium dissociation constant of 4.1nM can be derived upon non linear least-square fit of the data (theoretical binding isotherm in solid line).

is closely related to that in (21) for the same peptide and an oligodeoxynucleotide containing an Antp consensus binding site (1–2 nM).

DISCUSSION

Functional binding sites for homeoproteins in the copia retrotransposon

We have demonstrated that the copia retrotransposon possesses binding sites for transcriptional regulation by homeoproteins, and this is the first report for the occurrence of such sites on a mobile element. These binding sites are located 3' to the left LTR, in the copia untranslated region, close to a previously identified enhancer element first described in (31). DNaseI footprinting experiments and gel mobility shift assays with a purified minimal Antennapedia homeodomain protein demonstrate high affinity binding to a nucleotide sequence with a near perfect NP consensus, with a K_d of 4 nM in good agreement with values previously measured for the same homeodomain to other homeoprotein binding sites [1–10 nM (19, 21)]; binding with a lower affinity was also observed to a nucleotide sequence, adjacent to the primary sequence, closely related to the consensus NP sequence. The involvement of these binding sites in copia transcriptional regulation was demonstrated by a test using *Drosophila* cells in culture, from either *D.hydei* or *D.melanogaster*. Co-transfection experiments with expression vectors for different homeoproteins and responder plasmids containing the copia untranslated region under the control of either the copia

LTR or an heterologous minimal promoter, disclose either positive or negative regulations, whose occurrence is dependent of the presence of the HD binding site-containing fragment. Regulation is homeoprotein-dependent, concentration-dependent, and cell type-dependent, and takes place at the RNA level as shown by Northern blot analysis.

Rather paradoxically, the copia LTR has been extensively used as a control for transfection efficacy in *Drosophila* cells, and particularly in transfection assays using expression vectors for homeoproteins and reporter plasmids [see for instance (36, 44)]. In those cases, no regulatory effects of either the engrailed, even-skipped, zerknullt, fushi tarazu or paired homeoproteins on the level of expression of the 'control' copia reporter had ever been reported. This apparent discrepancies can be easily accounted for by the difference between the constructs which had been used and ours, as the former copia-containing constructs were deleted from the untranslated region and only possessed the 5'LTR *stricto sensu* [as an EcoRI–HindIII fragment from the pCOPneo construction (44)].

The structural organization of the copia LTR and untranslated sequence is reminiscent of that of the yeast Ty retroelement, where enhancer elements have also been characterized in the untranslated region, 3' to the left LTR (46, 47). Phylogenetic analysis have previously emphasized the close relationship between these two retrotransposons (48), and recent analysis of the *Drosophila* retrovirus-like retrotransposon mdg1 (49) and LINE element promoters [reviewed in (50)] have further evidenced that transcriptional control sequences are often located in these elements 3' to the RNA start site. Finally, it has been extensively documented that homeoprotein DNA binding sites from several homeotic genes can act as regulatory sequences in either orientation, as for classical enhancers, and also when linked to heterologous promoters [references in (14, 51)], as observed in the present study.

Diversity of the homeoprotein-mediated regulations of copia

The complexity of the regulations mediated by homeoproteins is illustrated by the differential effects of several homeoproteins depending on the cell line used. For instance, zerknullt decreases the activity of the lacZ reporter gene in DH33 cells whereas it enhances its activity in Schneider II cells; fushi tarazu enhances the activity in Schneider II cells but has no effect in DH33 cells. Several interpretations for these differential effects can be hypothesized, among which that of the presence in DH33 cells of endogenously expressed homeoproteins that would be simply competed out by the products of the exogenously introduced homeoprotein expression vectors. This interpretation, however, cannot simply account for the absence of effect of the removal of the HpRE 58bp fragment on the basal level of expression in the hsp-lacZ responder (see Results, figure 3B). Another possible source of regulation diversity could be the presence of more than one binding site for homeoprotein in the copia regulatory domain. Actually, *in vitro* studies with the Antp homeodomain have revealed a major binding site, but both mobility shift assay and foot-printing experiments have suggested possible interactions—with a significantly lower affinity for Antp HD—to a second binding site adjacent to the main site, and which further discloses similarities with the consensus NP sequence. Accordingly, the combinatory binding of both endogenously and exogenously expressed homeoproteins could result in differential effects [although we did not observe, in co-transfection assays using

simultaneously two different homeoprotein expression vectors, any cooperative or synergistic effects (data not shown)]. Finally, other sites on the copia regulatory element could also be involved. For instance, we have demonstrated by mobility shift assays (45) that DH33 cell nuclear extracts contain a protein which binds specifically to the copia SV40-like 'enhancer' sequence, whereas Schneider II cell nuclear extracts do not. This enhancer sequence is adjacent to the homeodomain binding sites (cf Figure 1), and the presence or not of a factor bound to this enhancer sequence—and further possibly directly interacting with homeoproteins, as observed for the herpes virus VP16 protein (52) or the yeast GRM protein (53)—could severely modulate the homeoprotein response. It has also been demonstrated that copia gene products are able to negatively regulate their own promoter at the level of sites present within the LTR *sensu stricto* (54). It is therefore possible that this self-regulation interferes with the homeoprotein response in the copia-containing Schneider II cells. Actually, we have detected for these cells significantly weaker responses with the copia LTR reporter than with the hsp-lacZ reporter—where the homeodomain binding sites are placed upstream of the heterologous hsp70 enhancerless promoter—a result consistent with the observation in (54) that copia elements do not negatively regulate the hsp70 promoter.

It should be finally emphasized that the diversity of the regulatory effects of homeoproteins as demonstrated in this study, is a common major feature of this class of transcription factors (see Introduction); differential effects have also been reported for other genes, with for instance the *Drosophila dorsal* morphogen which can mediate both activation and repression in a context-dependent manner (55). As discussed above, this complexity possibly involves multiple binding sites and factors, as well as direct protein–protein interactions with factors involved in the transcription machinery [reviewed in (14); see also (56) for evidence of multi-protein complexes between nuclear factors and engrailed].

Perspectives

The presence of active homeoprotein-responsive elements in the copia retrotransposon raises several questions. A first one concerns their possible role in the developmental regulation of copia expression. Clearly now, identification of functional binding sites should allow to devise experiments in which the effect of their removal or mutation can be analyzed *in vivo*, using transgenic *Drosophila* containing copia-driven reporter genes, in relation with alterations in their spatial and temporal patterns of expression in the course of *Drosophila* development. Actually, preliminary experiments using transgenic *Drosophila* with altered expression for given homeoproteins—for instance *Drosophila* containing homeoprotein genes under the control of the heat-shock inducible hsp promoter—have revealed a 2–3 fold increase of endogenous copia transcripts upon overexpression of the fushi tarazu gene, when monitoring the overall copia expression by dot blot analysis, and a 'patterned' expression of a copia-driven lacZ reporter gene under conditions of forced engrailed expression, using an *in situ* assay for β -galactosidase activity (57). Another related question concerns the generality of the observed regulations: other developmentally regulated *Drosophila* retrotransposons (412, 297, gypsy) have in their untranslated region, close to the 5' LTR, AT-rich sequences where putative homeoprotein binding sites can be identified. Experiments similar to those reported in this paper are now in progress with these

elements, which should help in answering the intriguing question of the *raison d'Être* of homeoprotein binding sites in retrotransposons.

ACKNOWLEDGEMENTS

We wish to specially acknowledge Drs K.Han and M.S.Levine for the generous gift of the homeoprotein expression vectors, and Drs A.Joliot and A.Prochiantz for that of purified Antennapedia homeodomain peptide. We are very grateful to S.Jensen for initiating one of us to *Drosophila* and for helpfull discussions and suggestions, and to M.-P.Gassama for her help in the maintenance of the fly stocks and technical assistance. We acknowledge Drs C.Lavialle and J.F.Nicolas for critical reading of the manuscript. This work was financed by the Centre National de la Recherche Scientifique (URA 147), the Institut National de la Santé et de la Recherche Médicale (U 140), the Association pour la Recherche sur le Cancer (contract 6552 to T.H.), and the Ligue contre le Cancer (to L.C.).

REFERENCES

- Sankaranarayanan, K. (1988) *Banbury Rep.* **30**, 319–336.
- Echalier, G. (1989) *Advances in Virus Research* **36**, 33–105.
- Finnegan, D.J. (1990) *Current Opinion in Cell Biology* **2**, 471–477.
- Finnegan, D.J. (1989) In *Mobile DNA* (Berg and Howe, eds), 503–517. American Society for Microbiology Washington, D.C.
- Bucheton, A. (1990) *Trends In Genetics* **6**, 16–21.
- Boeke, J.D., Garfinkel, D.J., Styles, C.A. and Fink, G.R. (1985) *Cell* **40**, 491–500.
- Heidmann, O. and Heidmann, T. (1991) *Cell* **64**, 159–170.
- Jensen, S. and Heidmann, T. (1991) *EMBO J.* **10**, 1927–1937.
- Pélisson, A., Finnegan, D.J. and Bucheton, A. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 4907–4910.
- Agnel, M. and Heidmann, T. unpublished data.
- Parkhurst, S.M. and Corces, V.G. (1987) *EMBO J.* **6**, 419–424.
- Kuff, E.L. and Lueders, K.K. (1988) *Adv. in Cancer Research* **51**, 183–276.
- Brulet, P., Condamine, H. and Jacob, F. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 2054–2058.
- Hayashi, S. and Scott, M.P. (1990) *Cell* **63**, 883–894.
- Affolter, M., Schier, A. and Gehring, W.J. (1990) *Current Opinion in Cell Biology* **2**, 485–495.
- Gehring, W.J., Muller, M., Affolter, M., Percival-Smith, A., Billeter, M., Qian, Y.Q., Otting, G. and Wüthrich, K. (1990) *Trends in Genetics* **6**, 323–329.
- McGinnis, W. and Krumlauf, R. (1992) *Cell* **68**, 283–302.
- StJohnston, D. and Nüsslein-Volhard, C. (1992) *Cell* **68**, 201–219.
- Muller, M., Affolter, M., Leupin, W., Otting, G., Wüthrich, K. and Gehring, W.J. (1988) *EMBO J.* **7**, 4299–4304.
- Desplan, C., Theis, J. and O'Farrell, P.H. (1988) *Cell* **54**, 1081–1090.
- Affolter, M., Percival-Smith, A., Muller, M., Leupin, W. and Gehring, W.J. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 4093–4097.
- Ekker, S.C., Young, K.E., von Kessler, D.P. and Beachy, P.A. (1991) *EMBO J.* **10**, 1179–1186.
- Florence, B., Handrow, R. and Laughon, A. (1991) *Mol. Cell. Biol.* **11**, 3613–3623.
- Otting, G., Qian, Y.Q., Muller, M., Affolter, M., Gehring, W.J. and Wüthrich, K. (1988) *EMBO J.* **7**, 4305–4309.
- Qian, Y.Q., Billeter, M., Otting, G., Muller, M., Gehring, W.J. and Wüthrich, K. (1989) *Cell* **59**, 573–580.
- Kissinger, C.R., Liu, B., Martin-Blanco, E., Kornberg, T.B. and Pabo, C.O. (1990) *Cell* **63**, 579–590.
- Otting, G., Qian, Y.Q., Billeter, M., Muller, M., Affolter, M., Gehring, W.J. and Wüthrich, K. (1990) *EMBO J.* **9**, 3085–3092.
- Billeter, M., Qian, Y.Q., Otting, G., Muller, M., Gehring, W.J. and Wüthrich, K. (1990) *J. Mol. Biol.* **214**, 183–197.
- Hanes, S.D. and Brent, R. (1991) *Science* **251**, 426–430.
- Ekker, S.C., von Kessler, D.P. and Beachy, P.A. (1992) *EMBO J.* **11**, 4059–4072.
- Mount, S.M. and Rubin, G.M. (1985) *Mol. Cell. Biol.* **5**, 1630–1638.
- Sneddon, A. and Flavell, A.J. (1989) *Nucl. Acids Res.* **17**, 4025–4035.
- Levis, R., Bingham, P.M. and Rubin, G.M. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 564–568.
- Bonnerot, C., Rocancourt, D., Briand, P., Grimber, G. and Nicolas, J.-F. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 6795–6799.
- Hiromi, Y. and Gehring, W.J. (1987) *Cell* **50**, 963–974.
- Han, K., Levine, M.S. and Manley, J.L. (1989) *Cell* **56**, 573–583.
- Schneider, I. (1972) *J. Embryol. Exp. Morphol.* **27**, 353–356.
- Sondermeijer, P.J.A., Derksen, J.W.M. and Lubsen, N.H. (1980) *In Vitro* **16**, 913–914.
- Hanahan, D. (1983) *J. Mol. Biol.* **166**, 557–580.
- Simon, J.A., Lis, J.T. and (1987) *Nucl. Acids Res.* **15**, 2971–2988.
- Church, G.M. and Gilbert, W. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 1991–1995.
- Joliot, A., Pernelle, C., Deagostini-Bazin, H. and Prochiantz, A. (1990) *Proc. Natl. Acad. Sci. USA* **88**, 1864–1868.
- Maxam, A.M. and Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 560.
- Jaynes, J.B. and O'Farrell, P.H. (1988) *Nature* **336**, 744–749.
- Cavarec, L., Jensen, S. and Heidmann, T. in preparation
- Fulton, A.M., Rathjen, P.D., Kingsman, S.M. and Kingsman, A.J. (1988) *Nucl. Acids Res.* **16**, 5439–5458.
- Farabaugh, P., Liao, X.B., Belcourt, M., Zhao, H., Kapakos, J. and Clare, J. (1989) *Mol. Cell. Biol.* **9**, 4824–4834.
- Xiong, Y. and Eickbush, T. (1990) *EMBO J.* **9**, 3353–3362.
- Arkhipova, I.R. and Ilyin, Y.V. (1991) *EMBO J.* **10**, 1169–1177.
- Martin, S.L. (1991) *Current Opinion in Genetics and Development* **1**, 505–508.
- TenHarmsel, A., Austin, R.J., Savenelli, N. and Biggin, M.D. (1993) *Mol. Cell. Biol.* **13**, 2742–2752.
- Stern, S., Tanaka, M. and Herr, W. (1989) *Nature* **341**, 624–630.
- Passmore, S., Elble, R. and Tye, B.-K. (1989) *Genes & Development* **3**, 921–935.
- Bryant, L.A., Brierley, C., Flavell, A.J. and Sinclair, J.H. (1991) *Nucl. Acids Res.* **19**, 5533–5536.
- Pan, D. and Courey, A.J. (1992) *EMBO J.* **11**, 1837–1842.
- Gay, N.J., Poole, S. and Kornberg, T. (1988) *EMBO J.* **7**, 4291–4297.
- Cavarec, L., Jensen, S. and Heidmann, T. unpublished data.