Development 130, 1243-1254

doi:10.1242/dev.00348

Genome-wide identification of in vivo *Drosophila* Engrailed-binding DNA fragments and related target genes

Pascal Jean Solano[1,*], Bruno Mugat[1,*], David Martin[2], Franck Girard[1], Jean-Marc Huibant[1], Conchita Ferraz[1], Bernard Jacq[2], Jacques Demaille[1] and Florence Maschat[1,†]

[1]Institut de Génétique Humaine (UPR 1142). 141 rue de la Cardonille, 34396 Montpellier, France
[2]Laboratoire de Génétique et Physiologie du Développement (UMR 6545), IBDM, Parc Scientifique de Luminy, 13288 Marseille, Cedex 9, France

*These authors contributed equally to this work
†Author for correspondence (e-mail: fmaschat@igh.cnrs.fr)

Accepted 18 December 2002

SUMMARY

Chromatin immunoprecipitation after UV crosslinking of DNA/protein interactions was used to construct a library enriched in genomic sequences that bind to the Engrailed transcription factor in *Drosophila* embryos. Sequencing of the clones led to the identification of 203 Engrailed-binding fragments localized in intergenic or intronic regions. Genes lying near these fragments, which are considered as potential Engrailed target genes, are involved in different developmental pathways, such as anteroposterior patterning, muscle development, tracheal pathfinding or axon guidance. We validated this approach by in vitro and in vivo tests performed on a subset of Engrailed potential targets involved in these various pathways. Finally, we present strong evidence showing that an immunoprecipitated genomic DNA fragment corresponds to a promoter region involved in the direct regulation of *frizzled2* expression by *engrailed* in vivo.

Key words: Engrailed, Chromatin immunoprecipitation, In vivo targets, *Drosophila*

INTRODUCTION

Identification of target genes that are directly regulated by transcription factors is a key issue in developmental biology, and has been the purpose of several recent studies. Indeed, the genome-wide location of DNA-binding proteins using genomic microarrays has been performed in yeast (Iyer et al., 2001; Lieb et al., 2001; Ren et al., 2000). In mammalian cells, CpG island microarrays have allowed the identification of promoter regions capable of binding to the E2F transcription factor (Weinmann et al., 2002). Recently, whole-genome microarray assays associated with bioinformatic methods have also been successfully performed to identify direct target genes of the Dorsal transcription factor in *Drosophila* (Markstein et al., 2002; Stathopoulos et al., 2002). Identifying the genes that are directly regulated by transcription factors, rather than merely in the downstream pathways, remains essential for understanding gene function (Liang and Biggin, 1998; Mannervik, 1999; Furlong et al., 2001; Egger et al., 2002). Homeodomain transcription factors play key roles during development by coordinating the behavior of most cells within their domains of expression (Garcia-Bellido, 1975; Lawrence and Morata, 1992), and identifying their target genes is challenging (Biggin and McGinnis, 1997). Interestingly, whereas homeodomain proteins recognize closely related binding sites, they are involved in specific genetic pathways and their absence produces very specific phenotypic effects (Wolberger, 1996). Therefore, to understand how homeodomain proteins function, it is essential to learn how they are acting in vivo and the identity of the downstream genes that they directly regulate.

To address this issue, we decided to identify genomic fragments that are bound in vivo by the homeoprotein Engrailed in *Drosophila* embryos. *engrailed* is involved in posterior cell identity, as well as in the maintenance of the anteroposterior (AP) boundary throughout development (Lawrence and Morata, 1976; Kornberg, 1981; Maschat et al., 1998). Furthermore, *engrailed* is expressed in a subset of neuroblasts and neurons in the developing central nervous system and in the brain. This suggests a role for *engrailed* during neurogenesis that is believed to be highly conserved during evolution (Bhat and Schedl, 1997; Hanks et al., 1998; Siegler and Jia, 1999; Simon et al., 2001).

We have previously used UV crosslinking and in vivo chromatin immunoprecipitation to analyze Engrailed-binding sequences associated with particular genomic regions such as *polyhomeotic* or *ß3-tubulin* genes (Serrano et al., 1995; Serrano et al., 1997). We present a similar approach, performed on a larger scale, leading to a genome-wide view of direct Engrailed-binding loci in embryos. UV light or formaldehyde crosslinking are currently used to purify protein-DNA complexes, and to isolate specific binding fragments (Graba et al., 1992; Serrano et al., 1995; Serrano et al., 1997; Liang and Biggin, 1998; Cavalli et al., 1999; Toth and Biggin, 2000;

Weinmann et al., 2001; Weinmann et al., 2002). However, UV light is believed to be more efficient in fixing proteins that are directly bound to DNA (Toth and Biggin, 2000).

In the present report, we constructed a library enriched in genomic sequences that bind Engrailed protein in *Drosophila* embryos, by using UV crosslinking and chromatin immunoprecipitation (UV-X-ChIP). Systematic sequencing of the recovered clones led to the identification of 203 potential direct targets of *engrailed* and evidence is presented to show that some of them represent bona fide *engrailed* targets.

MATERIALS AND METHODS

Fly strains

Reference wild-type flies are Oregon R. *engrailed* null mutant strain is *Df(2R)en*^{E}, deleted in both *engrailed* and *invected* (Gustavson et al., 1996). Heterozygous *tramtrack* mutation (*ttk804*) has been used to express *engrailed* ectopically in salivary glands (Fauvarque and Dura, 1993). The following transgenic stocks were used: UAS-Engrailed (referred as to UAS-En) (Tabata et al., 1995); UAS-VP16-HA tagged Engrailed (referred as to UAS-VP16-En) (Lecourtois et al., 2001; Alexandre and Vincent, 2002); (*hs*-Gal4); (*MS1096*-Gal4) (Milan and Cohen, 2000); and (*en*-Gal4) (Tabata et al., 1995).

The 1A4-GFP strain was made for the purpose of this study, by inserting 1A4 DNA fragment (either as a monomer or a trimer) into the WH.GFP (obtained from B. Bello, MRC, Mill Hill, UK). This vector corresponds to a mini white-based P-element vector with multiple cloning sites upstream of the *hsp70* minimal promoter from RHT vector (Bello et al., 1998) and the GFP F64L/65T-drosomycin polyA signal (Levashina et al., 1998). Different transgenic lines were obtained with both 1A4 monomer and trimer and showed the same GFP expression pattern as described in Fig. 5.

Proteins

A Schneider two-cell line, transformed with a gene fusion where the *engrailed* cDNA was placed under the control of the *hsp70* promoter, was grown at 25°C (Gay et al., 1988). To induce *engrailed* expression, cells were placed at 37°C for 45 minutes, followed by 2 hours at 24°C. Soluble nuclear extract (referred as to HS-EN) was prepared as described (Gay et al., 1988). A typical extract contained 2.5 mg/ml of proteins, of which 2% is Engrailed (estimated from western analysis in comparison with FPLC Engrailed purified proteins). HS-EN protein has been used in in vitro immunoprecipitation and in gel shift assays.

Antibodies

Anti-Frizzled 2 antibody is a mouse monoclonal antibody (Cadigan et al., 1998). The following anti-Engrailed antibodies were used: monoclonal antibody 4F11 for immunoprecipitation and super-shift assays, and rabbit polyclonal antibody raised against the entire Engrailed protein for immunostaining (Serrano et al., 1995). Anti-HA 12CA5 is a mouse monoclonal (Roche) and is monitored using a biotinylated secondary antibody, prior to HRP detection (Vector, Vectastain). Cy3-conjugated anti rabbit is from Immunotech.

cDNA

cDNAs used in this study were obtained from different laboratories: *frizzled 2* (Cadigan et al., 1998); *branchless* (Ribeiro et al., 2002); *frazzled* (Kolodziej et al., 1996) and *hibris* (Artero et al., 2001).

Construction of the library

A detailed protocol of X-ChIP is available at http://www.igh.cnrs.fr/equip/WebFM/. Briefly, nuclei were prepared from a 0-16 hour Oregon R embryo collection. Purified nuclei were irradiated with UV light Stratalinker (254 nm at 10 cm for 10 minutes) in order to freeze DNA-protein interactions. Chromatin was recovered from these nuclei by using detergents. Free proteins were separated from the rest of the chromatin by CsCl ultra centrifugation. The supernatant was dialyzed against TE buffer overnight. Chromatin was sonicated in order to shear DNA to 0.1-3.0 kb average length. We usually obtained 100 μg of chromatin from 1 g of embryos. Around 40 μg of chromatin were used for each immunoprecipitation experiment. Chromatin was first incubated for 1 hour at 4°C in the presence of 100 μl of protein A sepharose CL-4B resin (Amersham Pharmacia). The chromatin was then incubated overnight at 4°C, with either 50 μl of resin coupled to 4F11 anti-Engrailed antibody (which corresponds to 'EN probe') or 50 μl of resin with no antibody ('background probe'). After several washes, chromatin bound on the resin was eluted in the presence of 4% Sarkosyl. Samples were dialyzed, and DNA ends were repaired by Klenow before further purification. Samples were treated with RNase, proteinase K and phenol/chloroform extracted before precipitation in the presence of 20 μg glycogen. Linkers containing cloning sites were added and DNA was amplified by PCR, using a primer that covers the linker. Typically, 10-20 μg of amplified DNA were recovered. At that stage, DNA can be either labeled by random priming to be used as a probe on a Southern blot or processed for further purification. Indeed, to construct the library, we performed an additional in vitro immunopurification using a quarter of the PCR amplification after the 'in vivo' step, under the conditions described by Serrano et al. (Serrano et al., 1995). DNA was incubated in the presence of 50 μl resin where 4F11 anti-Engrailed antibody was bound, as well as Engrailed protein isolated from 1 mg of HS-EN cell culture nuclear extracts (containing ~20 μg of Engrailed protein). After overnight incubation at 4°C, resin was washed and DNA was eluted in the presence of 1 M KCl. After phenol/chloroform extraction and precipitation in the presence of glycogen, DNA was PCR amplified using a primer that covers the linker. DNA was then phenol/chloroform extracted and precipitated, digested using *Hin*dIII and cloned into p*Bluescript* KS^{+}. Each clone was sequenced.

Southern blots

For each of the 315 clones of the library, 1 μg of plasmid DNA was digested by *Hin*dIII, separated on 1% agarose gel in 0.5× TBE and transferred onto GeneScreen Plus membranes (NEN Life Science). Membranes were hybridized and washed following manufacturer instructions. Probes were prepared by ^{32}P labeling of 75 ng PCR amplified DNA (isolated after 'in vivo' immunoprecipitation), using the Rediprime kit labeling system (Amersham Pharmacia). Signals have been quantified using a phosphoimager.

Gel shift assays

DNA probes were synthesized by PCR amplification, using specific primers that were ^{32}P end labeled using T4 kinase. After gel purification, binding assays were performed in the presence of 1-5 ng labeled DNA, corresponding to 2000 cpm. Different quantities of HS-EN protein, isolated from cell culture nuclear extracts, were incubated with DNA probe for 30 minutes at 4°C in 10 μl of 25 mM HEPES (pH 7.6), 10% glycerol, 100 mM KCl, 1 mM DTT, 1% PVA, 1% NP40, 0.1% BSA and 200 ng of poly(dI:dC). DNA-protein complexes were resolved on 6% native polyacrylamide minigels in 0.5× TBE buffer (pH 8.3). For competition experiments, 750 ng unlabelled competitor DNA were added to the mix, and incubated with the protein and labeled DNA for 30 minutes at 4°C, before loading on gel. For supershift experiments, 4F11 antibody was incubated together with the labeled DNA for 30 minutes at 4°C, in the absence (–) or in the presence (+) of HS-EN protein, before loading on gel.

RNA in situ hybridization

Embryo fixation and in situ hybridization using DIG labeled antisense RNA probes were performed as described previously (Alexandre et al., 1996). Dissected larvae were fixed for 20 minutes in fixation buffer [30 mM PIPES (pH 7.4), 160 mM KCl, 40 mM NaCl, 4 mM EGTA,

1 mM spermidine, 0.4 mM spermine, 0.2% β-mercaptoethanol, 0.1% Triton X-100 and 4% paraformaldehyde], then washed four times in PBS + 0.1% Tween. The conditions for in situ hybridization were the same as for embryos. For double staining in embryos, in situ hybridization was performed first with an alkaline phosphatase-based detection system (Roche), followed by incubation with polyclonal anti-Engrailed antibody (dilution 1:300) overnight at 4°C. Detection of the immune signal was carried out with biotinylated secondary antibody, prior to HRP detection (Vectastain). Embryos were dehydrated and mounted in Canada Balsam for observation.

Immunostaining and in situ hybridization of polytene chromosomes

Squashes and hybridization were performed in *tramtrack* mutant background (*ttk804*), allowing *engrailed* expression in the salivary glands, as described by Serrano et al. (Serrano et al., 1995), with the following modifications: the 1A4-GFP DNA, labeled using Bionick labeling system (Invitrogen), was used as a probe and detected using fluorescein anti-biotin antibody (Vector) (1:200). Polyclonal anti-Engrailed antibody (1:200), secondary detected by Cy3 anti-rabbit antibody (1:200) has been used to identify Engrailed-binding sites. Chromosomal banding was detected with DAPI.

Computational analysis

A series of Perl programs were specifically written to analyze UV-X-ChIP datasets (D. Martin, F. M. and B. J., unpublished). The sequences of all immunoprecipitated fragments were automatically compared with the *Drosophila* genome sequence (BDGP), using the blastn program (Altschul et al., 1990). The output was automatically treated using scripts in order to remove contaminating vector or non-*Drosophila* sequences and to detect repeated elements or chimeric fragments. Using this approach, 357 distinct Engrailed-binding loci have been identified, from 542 clones sequenced. Functional assignments for potential target genes were automatically performed through a script querying Gene Ontology (GO) terms associated to them in FlyBase (Ashburner et al., 2000).

In order to discover over-represented motifs in these sequences, only clones without internal *Hin*dIII restriction site were considered. These genomic sequences might contain all the information necessary for Engrailed recognition. Out of the 203 positive clones, 107 sequences agreed with this criteria and were analyzed using the MEME algorithm, according to Bailey and Elkan (Bailey and Elkan, 1995). This led to the identification of 49 related motifs present in 204 hits and compiled in a position weight matrix presented on Fig. 2A.

RESULTS

In vivo isolation of Engrailed-binding fragments

Chromatin immunoprecipitation on UV-treated *Drosophila* embryonic nuclei was performed in parallel with either anti-Engrailed monoclonal antibody or with no antibody (Fig. 1A). This step is referred as to 'in vivo' because the source of Engrailed protein corresponds to the endogenous embryonic protein (see Materials and Methods). We then performed an additional immunoprecipitation cycle with an exogenous source of Engrailed protein (referred as to 'in vitro' IP, in Fig. 1A). We assumed that this step reduces background and purifies immunoprecipitated DNAs that bind Engrailed directly, as previous studies have show that Engrailed is able to bind DNA with high affinity in vitro, even though co-factors may be required in vivo for full activity (Saenz-Robles et al., 1995; Serrano and Maschat, 1998). We have also previously shown that Engrailed-binding fragments that had been specifically immunoprecipitated in vivo after UV crosslinking were highly purified during further in vitro immunoprecipitation (Serrano et al., 1995; Serrano et al., 1997). After the 'in vitro' step and a final PCR amplification, the fragments were cloned, leading to a library enriched in genomic Engrailed-bound sequences. From this library, 542 individual clones were isolated and sequenced (Fig. 1A). Sequence analysis revealed a total of 357 distinct genomic loci (see Materials and Methods).

To determine if these fragments were specifically enriched through immunoprecipitation, and do not simply represent nonspecific binding to the protein A sepharose resin, we performed Southern blots on 323 clones. We compared hybridization signal intensity with EN probe and a Background probe (see Materials and Methods) (Fig. 1A). With this assay, only the fragments for which the ratio EN probe/Background probe was higher than 2 were considered for further analysis. This was the case for 65% of the 315 intronic and intergenic fragments, whereas the eight DNA fragments lying in exonic sequences showed signals comparable with both probes (data not shown). As an example, Fig. 1B shows the results for 14 positive clones. Thus, 203 clones isolated in the library correspond to DNA that was specifically enriched by UV-X ChIP.

Localization of the Engrailed-binding fragments within the *Drosophila* genome and assignment for potential target genes

In order to localize precisely the 203 Engrailed-binding fragments, the corresponding sequences were compared with the published *Drosophila* genome sequence (Adams et al., 2000). This allowed the identification of genes in the same location, which were categorized according to Gene Ontology (GO) annotation (see Materials and Methods, Table 1 and http://www.igh.cnrs.fr/equip/WebFM/).

Forty-seven percent of the fragments were localized within gene introns. In this case, we assumed that the corresponding intron is a part of the *engrailed* regulated target gene. Fifty-three percent were present in intergenic regions. In this case, we restricted our analysis to the nearest transcription unit, whatever its orientation and its distance with respect to the Engrailed-binding fragment. Half of the intergenic fragments are localized at less than 5 kb upstream of the genes, suggesting that the Engrailed-binding fragment may be a part of their promoter region. In rare cases (5%), when the Engrailed-binding fragment lies between two transcription units among which only one encodes a known function, we considered the latter as the putative target.

In 55% of the cases, the Engrailed-binding fragments could be associated with a gene whose function is known or that contains a recognizable protein domain. In all the other cases (45%), the binding fragments were associated to genes with an unknown function, which is approximately the ratio of this category in the *Drosophila* genome (Adams et al., 2000). For an overview, 81 genes of known or predicted function are presented here and have been grouped into functional classes (Table 1). The other 30 genes encode proteins with recognizable domains, the function of which is unknown, and have not been listed here.

According to GO annotation, the majority of these genes are involved in cell communication and developmental processes (Table 1). As expected from previous work, potential Engrailed

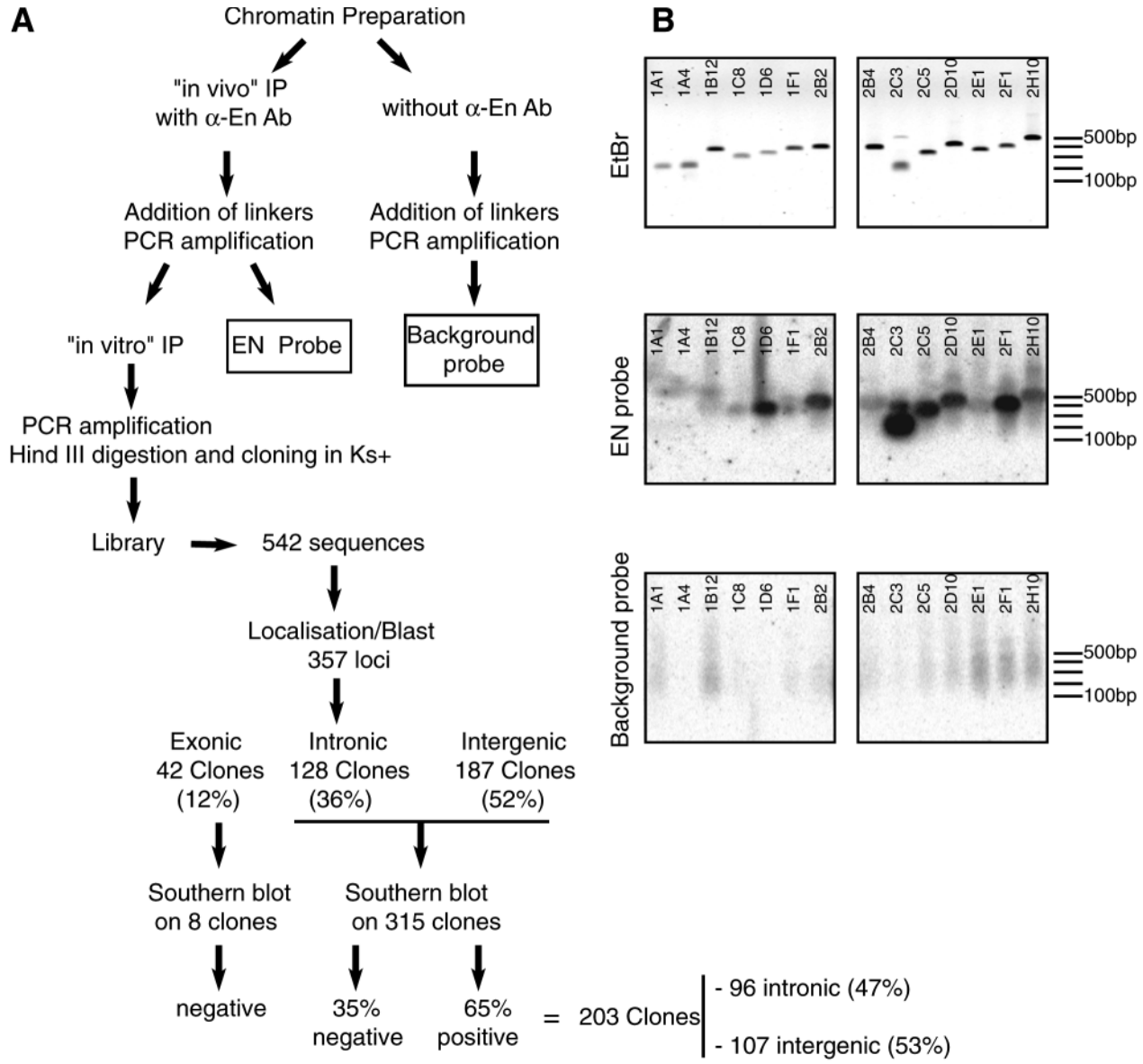

Fig. 1. (A) Strategy of Engrailed chromatin immunoprecipitation, cloning procedure and design of the probes used on Southern blot to test the specificity of the procedure. The En probe corresponds to the chromatin preparation used to construct the library. (B) Southern blots performed on 14 inserts isolated from the library. Clone inserts, visualized with ethidium bromide (EtBr), have been transferred onto nitrocellulose. Blots have been hybridized with EN probe and with Background probe, as indicated and described in A.

targets identified using this approach include genes that are involved in the establishment and the maintenance of the AP axis body (Kornberg, 1981; Vincent and O'Farrell, 1992; Serrano et al., 1995; Alexandre and Vincent, 2002). We also identified several genes involved in wing development (Hidalgo, 1994; Maschat et al., 1998), tracheal development, muscle development (Serrano et al., 1997) and axon guidance (Siegler and Jia, 1999). Furthermore, different categories of genes encoding proteins involved in signal transduction were found (signal proteins, receptors, protein kinases, protein phosphatases, transcription factors and cell adhesion protein). Interestingly, cell adhesion proteins and receptors were particularly well represented (Table 1; http://www.igh.cnrs.fr/equip/WebFM/). This suggests that *engrailed* could act at different molecular levels in several developmental processes.

In vitro analysis of specific Engrailed recognition motifs

A motif research analysis was performed on a subset of 107 sequences from the 203 clones selected in the UV-X-ChIP library (see Materials and Methods). It revealed that the most frequent motifs were a group of 49 related octanucleotides, compiled in a position weight matrix presented in Fig. 2A and resolved as a 'YAATYANB' consensus. This consensus sequence largely overlaps those already described for Engrailed (Desplan et al., 1988; Kissinger et al., 1990).

In order to verify that Engrailed binds to this consensus in vitro, we performed a gel shift assay on the most represented motif 'CAATTAGC', used as a pentamer. Several retarded complexes are formed in the presence of HS-EN protein with

Table 1. Listing of potential Engrailed target genes

AP patterning
slp2, fz2, nkd, arm, Wnt2, scm, corto, ash2

Wing and tracheal development
vvl, knrl, sty, vn, bnl, px

Muscle development
Gsc, twi, mam, hbs, nrm, aret, Ca-alpha 1D, Pka-C2, CG1890

Cytokinesis
tensin, CG15158, insc, pebble, Grip84, Klp54D, CG12908, p120ctn

Neurogenesis
eg, acj6, onecut, Mio, fra, comm, beat-Va, Or22c, Or42b, Or83c, Shal, Ptp4E, HD-14, Cad89D, CG4509, fat2, stan, 18W, con, Cha, CG5559, unc-13-4A, ine, SNF4A PK, twins, camKII, CG9811, G-ia65A, robl62A, huntingtin, enb, Fur1, ben, tomosyn, Leucokinin, trio

Eye development
ed, Calx, CdsA, lama, Pkg21D, inaD, CG12731, cno, drk

Others
p53, Ets98B, dd4, Acp33A, Ubp64E

Eighty-one putative target genes are listed according to the signaling pathway where they are known to act. 'Others' indicates genes involved in other developmental pathways or more general factors. The complete listing with the chromosomal localization of Engrailed-binding fragments isolated from the library, the associated target gene, and 'molecular function' and 'biological process' where the potential targets are involved are defined according to Gene Ontology (GO), and are available at http://www.igh.cnrs.fr/equip/WebFM/.

an affinity close to 10^{-9} M (Fig. 2B). The formation of these complexes is competed in the presence of either a single double strand unlabelled motif 'CAATTAGC' (referred as to C) or with a known specific Engrailed binding fragment D2 (Serrano et al., 1995). Furthermore, super-shifts of the complexes are observed in the presence of a specific anti-Engrailed antibody 4F11 (Fig. 2B). By contrast, addition of either a mutated form of the motif 'CAGCCGGC' (referred as to Cm) or *polyhomeotic* N fragment, which does not bind Engrailed (Serrano et al., 1995), had no effect on the formation of the complexes. Together, these data show that Engrailed binds specifically to this motif.

Gel shift assays were then performed on 14 Engrailed-binding fragments isolated from the library, and the results are shown here in four cases where the associated target genes are involved in different signaling pathways (Fig. 3). 1A4 clone corresponds to a genomic fragment lying 5 kb downstream of *frizzled 2* (*fz2*), which encodes one of the *wingless* (*wg*) receptors (Cadigan et al., 1998). 2H10 clone corresponds to a genomic fragment lying within *hibris* (*hbs*), which encodes a member of the immunoglobulin superfamily involved in muscle guidance (Artero et al., 2001; Dworak et al., 2001). 1B12 clone corresponds to a genomic DNA fragment lying within the first intron *of branchless* (*bnl*), encoding the *Drosophila* homologue of the Fibroblast Growth Factor (FGF) involved in tracheal morphogenesis (Sutherland et al., 1996). 2C5 clone corresponds to a genomic fragment lying in the first intron of *frazzled* (*fra*), which encodes a netrin receptor involved in motor axon guidance (Kolodziej et al., 1996).

In each case, two sets of experiments were performed, either with the entire immunoprecipitated fragment (150 bp to 350 bp), or with a shorter 100 bp fragment, surrounding the YAATYANB motifs, previously defined in Fig. 2A. Similar results (number of complexes, affinity) were obtained for both types of probes, and the data presented in Fig. 3 correspond to the shorter 100 bp fragments.

All these DNA fragments formed retarded complexes in the presence of HS-EN protein (Fig. 3). Addition of D2 DNA was able to compete the formation of the complexes, while addition of N DNA, as expected, had no effect. Moreover, addition of 4F11 antibody super-shifted the complexes. These data show the specificity of Engrailed binding, which was also confirmed using purified Engrailed protein (data not shown). The

A

position / base	1	2	3	4	5	6	7	8
A	0,10	1	1	0	0	0,82	0,29	0
C	0,65	0	0	0,01	0,26	0	0,16	0,47
G	0	0	0	0	0,01	0,07	0,39	0,27
T	0,25	0	0	0,99	0,73	0,11	0,16	0,26
Consensus	Y	A	A	T	Y	A	N	B

B

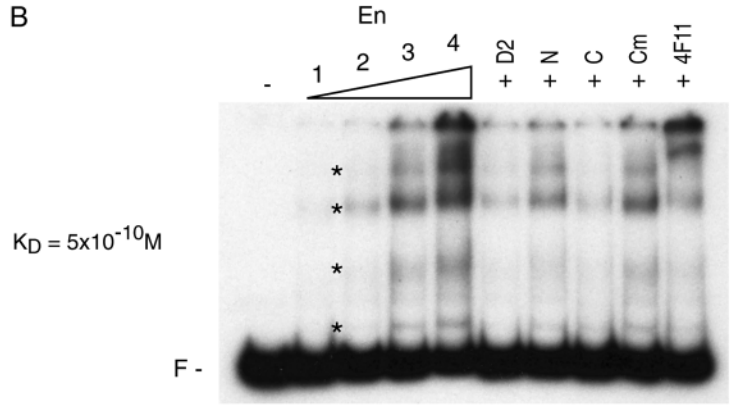

Fig. 2. (A) Identification of Engrailed-binding consensus sequence. The YAATYANB consensus was deduced from sequence analysis of 107 selected clones, as defined in the Material and Methods. For each position, the ratio of A, C G or T is indicated. (B) Gel shift assay was performed on a pentamer of the motif CAATTAGC, the sequence of which is shown below the gel. Labeled DNA fragment was analyzed in the absence (–) or in the presence of different amounts of Engrailed protein: 1=2×10^{-10} M; 2=3×10^{-10} M; 3=5×10^{-10} M; 4=10^{-9} M. Competition experiments were performed in the presence of 5×10^{-10} M En protein (+) and in the presence of 300-fold excess of different DNAs: D2, *polyhomeotic* D2 fragment, corresponding to a specific Engrailed-binding fragment (Serrano et al., 1995); C, double strand monomer 'CAATTAGC'; N, *polyhomeotic* N fragment that does not bind Engrailed specifically in vitro (Serrano et al., 1995); Cm, double strand mutated monomer 'CAGCCGGC'. Supershift assays were performed in the presence of 5×10^{-10} M En protein (+) and of 4F11 monoclonal anti-Engrailed antibody. F indicates free DNA. The asterisks indicate the Engrailed protein-DNA complexes.

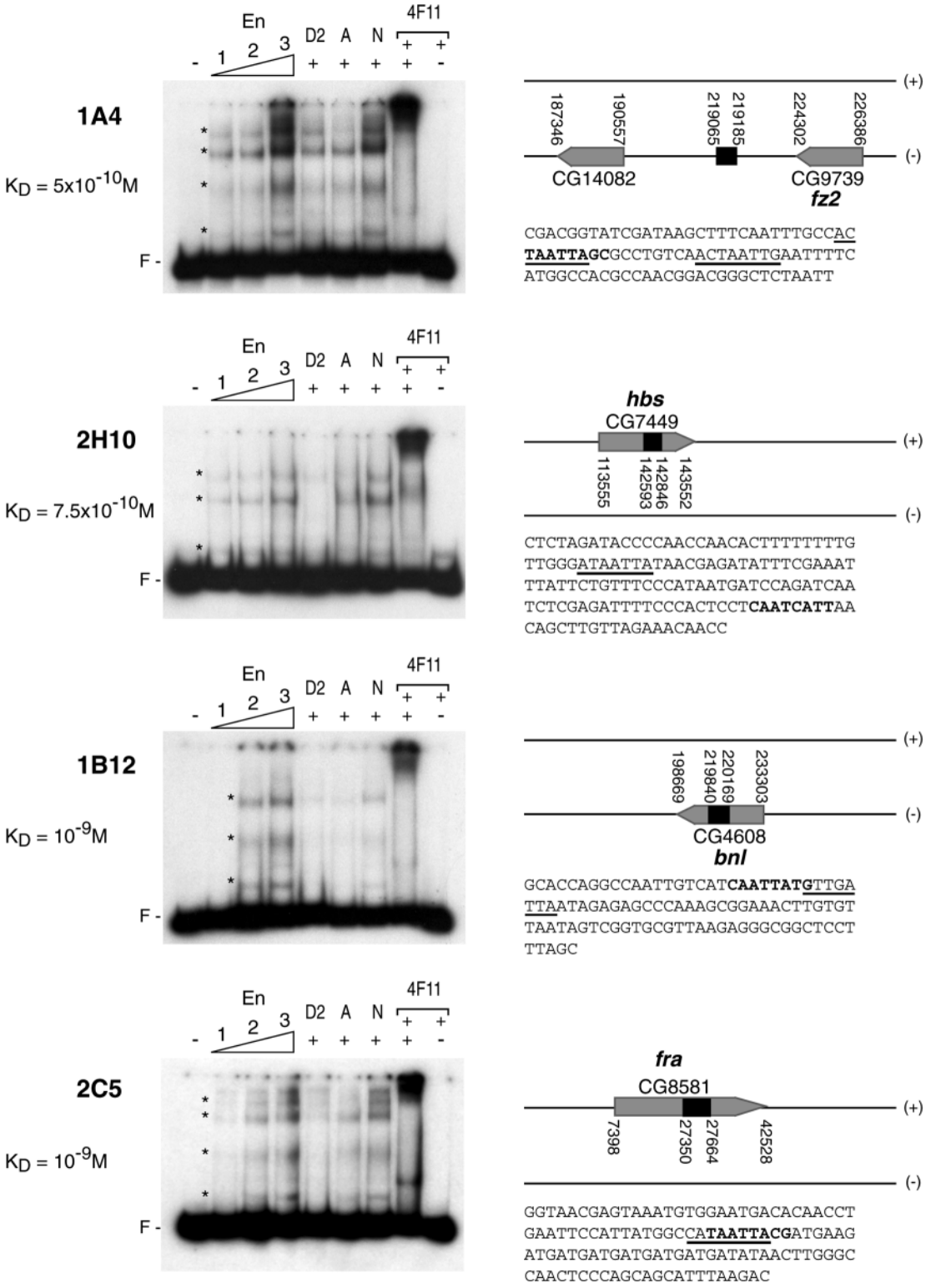

Fig. 3. Specific Engrailed binding to the isolated genomic DNA fragments. The ability of Engrailed to bind the immunoprecipitated fragments was tested in vitro by gel shift assays, as shown here for the 1A4, 2H10, 1B12 and 2C5 fragments. In each case, the sequence of the fragment is shown, as well as a map of the surrounding transcription units. The fragments are around 100 bp in size, surrounding the consensus Engrailed-binding motifs (highlighted in bold in the + strand, and underlined in the opposite strand). Labeled DNA fragments were analyzed in the absence (–) or in the presence of different amounts of Engrailed protein: 1=3×10^{-10} M; 2=5×10^{-10} M; 3=10^{-9} M. Competition experiments were performed in the presence of 10^{-9} M En protein (+), and in the presence of a 300-fold excess of different DNAs: D2, *polyhomeotic* D2 fragment; A, the cold fragment itself; N, *polyhomeotic* N fragment. Supershift assays were performed in the presence of 4F11 monoclonal anti-Engrailed antibody, and in the presence (+) or in the absence (–) of 10^{-9} M En protein. F indicates free DNA. The asterisks mark the Engrailed protein-DNA complexes.

addition of the cold DNA fragment itself allowed us to compare the affinity of Engrailed on this fragment to the affinity of the strong Engrailed-binding fragment D2 (compare lanes D2 and A in Fig. 3). The affinities are at least 10^{-9} M and are closely related to the affinity of the motif 'CAATTAGC' (compare Fig. 2B with Fig. 3).

In conclusion, we have shown that Engrailed is able to bind specifically to these four in vivo immunoprecipitated DNA fragments, which lie close to genes involved in different developmental processes, most probably via the 'YAATYANB' consensus sequence that we have identified.

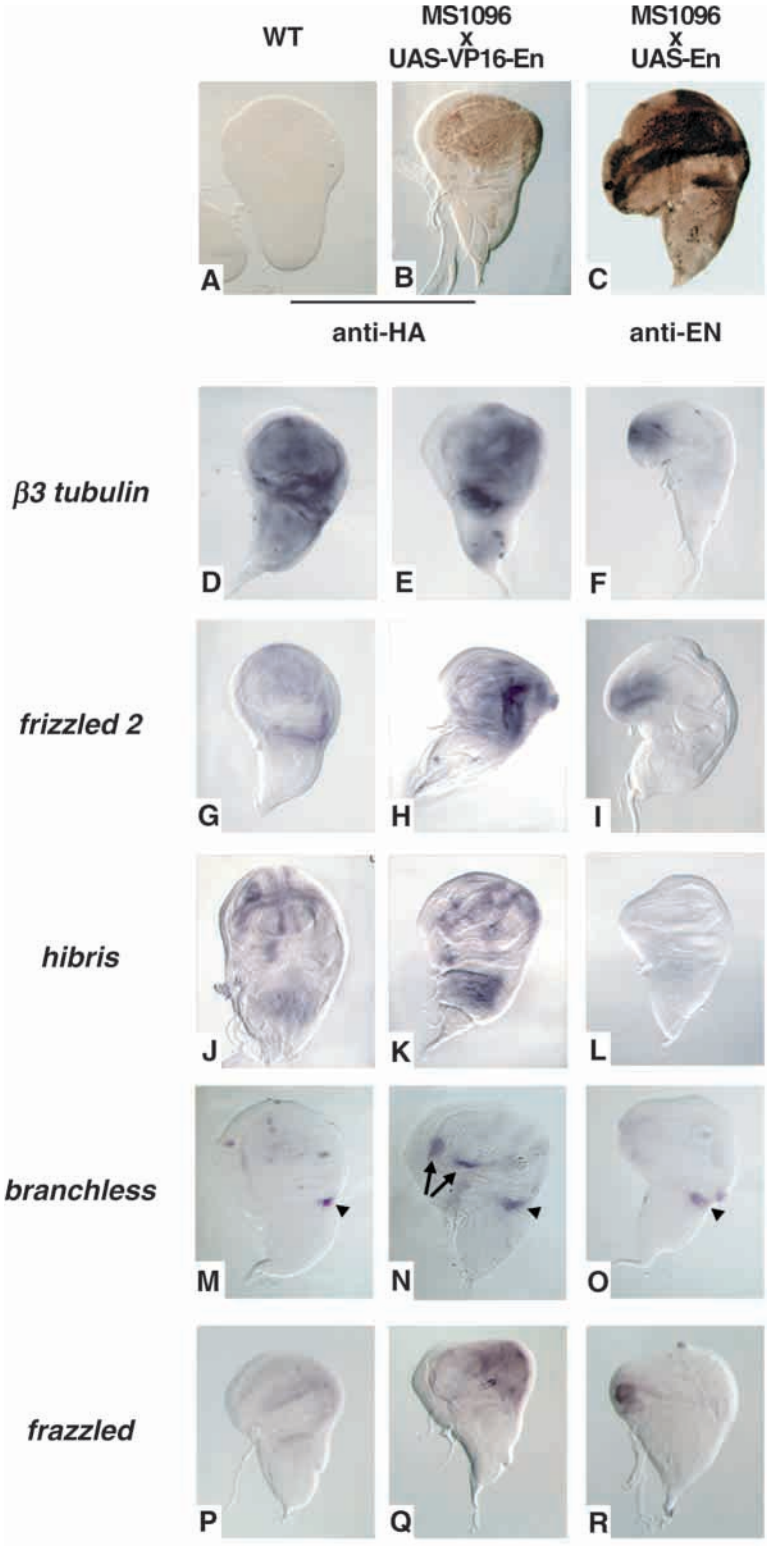

Fig. 4. Engrailed misexpression induces changes in the expression of potential Engrailed target genes. In situ hybridization using anti-sense RNA probes were carried out on late third instar wing imaginal discs in different genetic backgrounds: wild-type (WT) (D,G,J,M,P); *MS1096 X UAS-VP16-En* (E,H,K,N,Q); *MS1096X UAS-En* (F,I,L,O,R). Overexpression of (VP16-En), driven by *MS1096-Gal4*, was detected in the wing pouch, using anti-HA antibody (B), when compared with wild type (A). Overexpression of the normal form of Engrailed was detected with polyclonal anti-En antibody (C). Note that (VP16-En) overexpression leads to a posterior distortion of the disc, whereas Engrailed (En) overexpression leads to an anterior distortion. (D-F) In situ hybridization with *β3-tubulin* probe. (G-I) In situ hybridization with *frizzled 2* probe. (J-L) In situ hybridization with *hibris* probe. (M-O) In situ hybridization with *branchless* probe. Normal expression of *bnl* is indicated by an arrowhead. Arrows indicate the ectopic expression of *bnl* in the wing pouch. (P-R) In situ hybridization with frazzled probe. Discs are oriented anterior leftwards, ventral upwards.

The expression of potential target genes depends on *engrailed* regulation in vivo

In order to discriminate among the list of putative targets, the ones responding to *engrailed* regulation in vivo, we used a simple screen. We monitored by in situ hybridization, the expression of several potential target genes, after ectopic expression of Engrailed using the UAS-GAL4 system. Because Engrailed can act as a repressor or an activator (Serrano et al., 1997; Serrano and Maschat, 1998), we overexpressed either the wild-type Engrailed protein (UAS-En) or a chimeric activator form (UAS-VP16-En) (Lecourtois et al., 2001; Alexandre and Vincent, 2002), under the control of *MS1096*-Gal4, in third instar wing imaginal discs (Fig. 4). We first tested this approach on *β3-tubulin*, which we have previously shown to be directly repressed by *engrailed* (Serrano et al., 1997). As expected, overexpression of wild-type Engrailed protein led to a repression of endogenous *β3-tubulin* in the wing disc (Fig. 4F), whereas overexpression of the activator form of Engrailed had no detectable effect, probably because of the strong expression of endogenous *β3 tubulin* in the discs (Fig. 4D,E).

Using this assay, we studied the expression of 14 genes that are localized close to the genomic DNA fragments isolated in the library and tested previously for their Engrailed-specific binding ability (Fig. 1B). The results are shown for four genes (*frizzled2*, *hibris*, *branchless*, *frazzled*) that are representative of the different pathways where *engrailed* seems to be involved (Table 1; Fig. 3). *frizzled 2* expression is activated in the presence of (VP16-En) (Fig. 4H) and repressed in the presence of En (Fig. 4I) (see wild-type expression in the wing pouch for comparison, Fig. 4G). This suggests that *engrailed* might act as a repressor on *fz2* expression. *hibris* is expressed along the wing margin and in the presumptive region of wing vein L3 and L4 in wild type (Fig. 4J). This expression is slightly activated in the presence of (VP16-En) (Fig. 4K), but strongly repressed when En is overexpressed (Fig. 4L), suggesting that *hbs* expression is regulated by *engrailed* in vivo. *branchless* is essentially expressed in a dorsal/posterior territory surrounding the wing pouch in wild type (Fig. 4M). In the presence of (VP16-En), several additional patches of *bnl* expression are detected within the wing pouch (Fig. 4N), whereas no activation of *bnl* is observed after wild type En overexpression (Fig. 4O). As expected, because *MS1096* drives

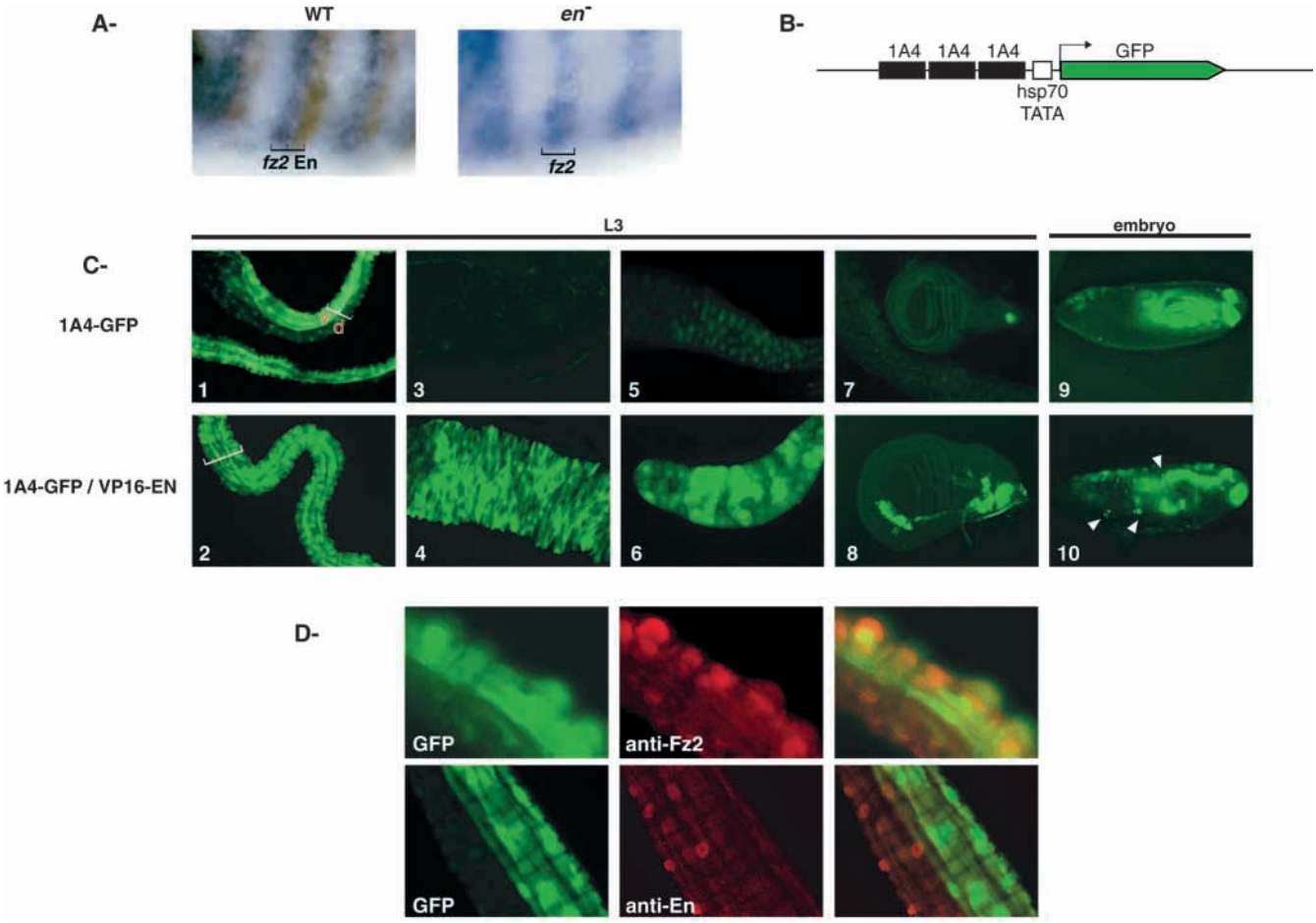

Fig. 5. Engrailed is a direct repressor of *frizzled 2* expression. (A) *fz2* expression was analyzed by in situ hybridization (in blue) in wild-type (WT) and *engrailed* mutant (*en*$^{-}$) genetic backgrounds. Embryos were double stained with anti-Engrailed antibody (in brown). (B) Schematic representation of the (1A4-GFP) transgene, where a trimer of 170 bp 1A4 fragment has been cloned upstream to *hsp70* minimal promoter and GFP reporter gene. (C) GFP expression is shown in different genetic backgrounds: (1A4-GFP) corresponds to the normal expression of the transgene (1, 3, 5, 7, 9); (1A4-GFP/VP16-En) corresponds to GFP expression in the presence of the activator form of Engrailed, driven either by (*hs*-Gal4) (2, 4, 6, 8) or by (*en*-Gal4) (10). GFP expression was analyzed in either late L3 larval tissues (1-8) or in embryos (9, 10): (1, 2) hindgut (d, dorsal; v, ventral); (3, 4) midgut; (5, 6) salivary gland; (7, 8) wing imaginal disc; (9, 10) embryo. White arrowheads show the position of stripes. Bracket in 2 indicates GFP expression in both dorsal and ventral compartments. (D) GFP expression of late third instar larvae is shown in the hindgut (GFP) and can be compared with endogenous *fz2* expression, detected by an anti-Frizzled 2 antibody in red (anti-Fz2) or with *engrailed* expression, detected by polyclonal anti-Engrailed antibody in red (anti-En). Merged images are shown.

Gal4 expression only in the wing pouch (Fig. 4B), endogenous *bnl* expression outside the wing pouch is not affected (Fig. 4M-O), showing the specificity of the experiment. Finally, *frazzled* is slightly expressed in wild-type wing disc (Fig. 4P). This expression is activated when (VP16-En) is overexpressed (Fig. 4Q), and repressed upon En overexpression (Fig. 4R).

In conclusion, these data demonstrate that the expression of several potential target genes identified via UV-X-ChIP is modulated when *engrailed* is misexpressed. This test has been successfully performed on 12 genes over 14 tested (Fig. 4 and data not shown).

frizzled2 is a direct target of *engrailed* regulation

Interactions between *engrailed* and the *wingless* signaling pathway have been extensively described (DiNardo et al., 1988; Martinez-Arias and White, 1988; Bejsovec and Martinez Arias, 1991; Heemskerk et al., 1991; Hatini and DiNardo, 2001). A direct regulation of *frizzled* receptor expression by *engrailed* has been shown (Lecourtois et al., 2001). In this study, we found that the other *wingless* receptor gene, *frizzled2* (*fz2*), might also be directly regulated by *engrailed*. A high-affinity Engrailed-binding fragment (1A4) was detected in the close vicinity of the *fz2* transcription unit (Fig. 3). In wild-type embryos, *fz2* expression becomes segmentally repeated around stage 9, in two or three rows of cells just anterior to *engrailed*. In stage 9 *engrailed* mutant embryos, *fz2* expression is extended posteriorly, being detected in 4 rows of cells (Fig. 5A). This shows that Engrailed acts as a repressor of *fz2* expression in embryos, as has been suggested with the previous test in the wing disc (Fig. 4). We verified whether the 1A4 Engrailed-binding fragment was able to drive the expression of a reporter gene in vivo and whether it was responding to *engrailed* regulation. For this purpose, this 170 bp fragment, either as a monomer or a trimer, was cloned upstream of a GFP

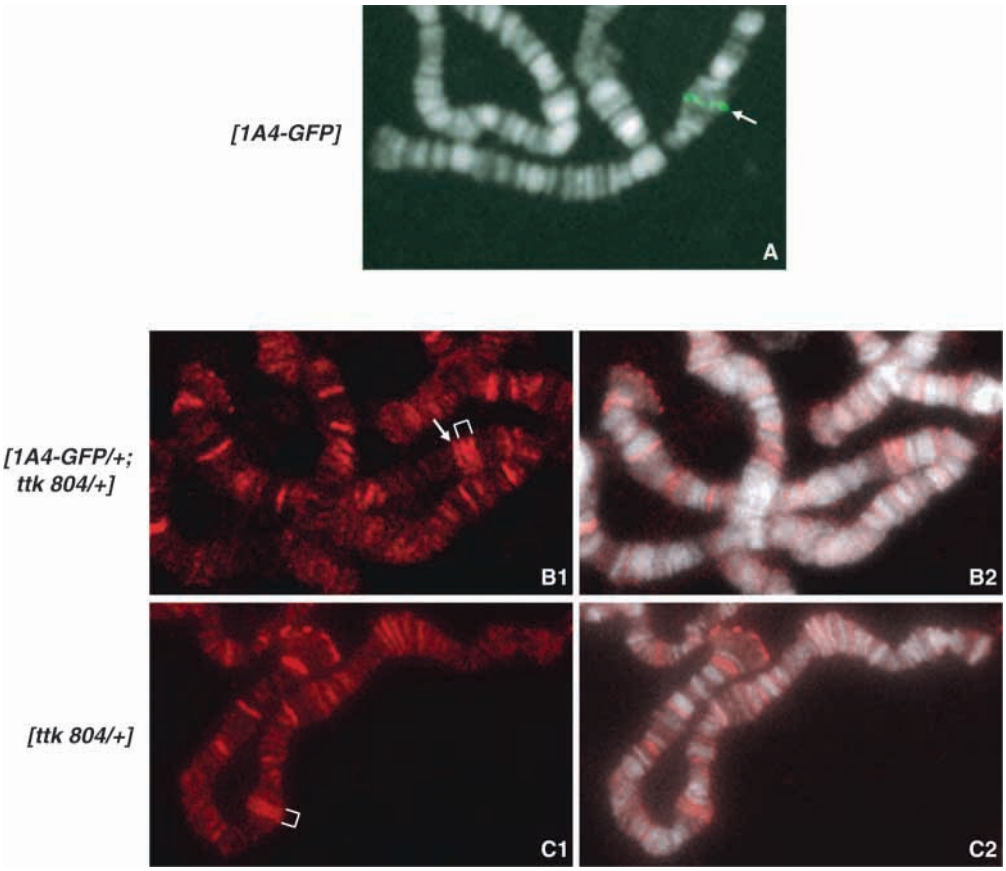

Fig. 6. Engrailed directly binds 1A4 fragment on polytene chromosomes. (A) In situ hybridization on polytene chromosomes. Localization of the transgene on 2L chromosome, in position 25C (shown by the arrow), using 1A4-GFP biotinylated DNA probe, detected by FITC anti-biotin antibody. Banding of the chromosome is shown by DAPI staining. (B1,B2,C1,C2) Immunodetection of Engrailed-binding sites on 2L polytene chromosome, in *tramtrack* heterozygous background [*ttk804*/+], identified by anti-Engrailed antibody (in red). (B1) Engrailed-binding sites of [1A4-GFP/+; *ttk804*/+] salivary glands. (B2) Same as B1 with DAPI staining. (C1) Engrailed-binding sites of [*ttk804*/+] salivary glands. (C2) Same as C1 with DAPI staining. The white arrow shows an additional Engrailed binding site at the insertion point of the transgene and the bracket indicates two sites close to it.

reporter gene and *hsp70* minimal promoter and introduced into the *Drosophila* genome by P element-mediated transposition (Fig. 5B). In these transgenic lines, GFP expression was essentially detected in the embryonic hindgut (Fig. 5C9) and in half of the larval hindgut (Fig. 5C1). GFP is expressed in the ventral cells of the larval hindgut that do not express *engrailed*, which mimics endogenous *fz2* expression (Fig. 5D). This demonstrates that the 1A4 DNA fragment might be a part of endogenous *fz2* regulatory sequences. Overexpression of (VP16-En) fusion protein driven by *h*s-Gal4 leads to ectopic GFP expression in the entire hindgut (Fig. 5C2), but also in tissues that do not express the transgene in wild type, such as the midgut (Fig. 5C3,C4), the salivary glands (Fig. 5C5,C6), and the wing disc (Fig. 5C7,C8). Overexpression of (VP16-En) fusion protein driven by *en*-Gal4 in embryos leads to ectopic GFP expression in a striped pattern (Fig. 5C9,C10). Such activation does not occur with overexpression of wild-type Engrailed, confirming a repressor role of Engrailed on *fz2* expression through this 1A4 fragment (data not shown). These results show that 1A4 is able to respond to *engrailed* regulation in vivo.

Finally, in order to verify that the Engrailed-binding on 1A4 was direct, we analyzed the pattern of Engrailed immunostaining on polytene chromosomes, in the transgenic 1A4-GFP line. An additional Engrailed-binding site was detected on polytene chromosomes in the locus of the transgene (25C), clearly showing a direct fixation of Engrailed on the 1A4 fragment, in vivo (Fig. 6).

Altogether, these data show that the 1A4 fragment that was isolated by UV-X-ChIP is a part of the *fz2* regulatory regions and is able to directly respond to *engrailed* regulation in vivo.

DISCUSSION

In this report, we used UV-X-ChIP to prepare a genomic library enriched in DNA fragments that bind the Engrailed transcription factor in vivo. A systematic sequencing of the clones led to the isolation of 203 Engrailed-binding fragments that can be assigned to potential targets, because they lie either in an intron (in 47% of the cases) or in the close vicinity of a gene. The gel shift assay analysis, presented here in four examples but verified in 12 cases (over 14 tested), confirmed that these DNA fragments bind Engrailed with high affinity in vitro. Furthermore, in vivo tests on a subset of potential target genes showed that in most cases (12 of 14 tested), the expression of the selected genes was sensitive to *engrailed* misexpression. The results obtained on this sample indicate that X-ChIP may indeed be an efficient method with which to isolate direct targets. The identification of in vivo Engrailed binding fragments, and of the related target genes, constitutes a first step for the further analysis of *engrailed* direct targets.

Using this approach, we found that potential *engrailed* target genes are involved in different developmental processes, such as AP patterning, neurogenesis, wing or tracheal development, and muscle development. In these different pathways, we identified a number of target genes encoding cell adhesion molecules and all types of receptors, which is in agreement

with *engrailed* involvement in cell-cell contact events (Dahmann and Basler, 2000). The identification of targets that also encode signal proteins, enzymes such as protein kinases or phosphatases and, to a lesser extent, transcription factors, suggest that *engrailed* can act at different levels on a regulatory cascade.

Using correct markers, we notice that *engrailed* mutants exhibit severe pleiotropic phenotypes. This includes defects in axon migration (Siegler and Jia, 1999), but also in the attachment of the muscles along the AP axis of the body (Serrano et al., 1997) and in the pathfinding of the tracheal network (data not shown). These different phenotypes could result from the abnormal setting of the compartments in *engrailed* mutant embryos. The identification of potential targets involved in these processes (Table 1) suggests rather that *engrailed* may be more directly implicated in these phenotypes. This confirms the contribution of *engrailed* in the orderly assembly and migration of cells during morphogenesis and pattern formation, leading to the normal positioning of the tissues along the AP axis.

The identification of Engrailed-binding fragments close to *connectin*, *18 Wheeler* or *eagle*, which have already been defined as genetic *engrailed* targets involved in neurogenesis, confirms a direct implication of *engrailed* in axon guidance (Dittrich et al., 1997; Eldon et al., 1994; Siegler and Jia, 1999). Furthermore, we identified that the netrin *frazzled* receptor gene, which is involved in motor axon guidance, might also be a direct target of *engrailed*.

We isolated several Engrailed-binding fragments closely related to genes involved in muscle development, and, in particular, genes that ensure connections between the epidermis or PNS and the muscles. Indeed, our data suggest strongly that the involvement of *engrailed* in myotube guidance (Serrano et al., 1997) might result, in part, from the direct regulation of *hibris* (Dworak et al., 2001).

We also report a direct link between *engrailed* and tracheal development. We indeed isolated several target genes that are involved in directing tracheal cell migration, such as the FGF-like secreted molecule, *bnl*. The tracheal system originates from placodes that consist in part of ectodermal cells. Interestingly, *bnl* is not expressed in the trachea, but in the ectodermal cells that overlie the migrating and branching trachea, thus acting as a guidance molecule that controls tracheal cell migration (Sutherland et al., 1996). This result suggests a direct effect of *engrailed* on tracheal guidance, which correlates with the *engrailed* mutant phenotype (data not shown), although *engrailed* involvement in tracheal migration has never been previously suggested.

Finally, one important and well-conserved function of *engrailed* concerns its relationship with the *wingless* signaling pathway. The spreading of the Wingless (Wg) signal is crucial for establishing the pattern of differentiated cell types within tissues and organs (DiNardo et al., 1988; Martinez-Arias and White, 1988; Bejsovec and Martinez Arias, 1991; Heemskerk et al., 1991; Lecourtois et al., 2001; Hatini and DiNardo, 2001). This might depend in part on the level of expression of Wg receptors. Our screen strongly suggests that the Wg receptor *frizzled 2*, is likely to be directly regulated by *engrailed*. During hindgut morphogenesis, the morphogen *dpp* is repressed in the dorsal *engrailed* expressing cells, but is activated in the opposite ventral cells and it has been suggested that *wg* might be responsible for this activation (Takashima and Murakami, 2001). We show that *fz2* is only expressed in the ventral cells of the larval hindgut. Furthermore, we show that the 1A4 genomic fragment, isolated in the library, was directly bound by Engrailed in vivo and was effective in driving the expression of a GFP reporter gene in these *fz2*-expressing cells. Our results strongly suggest that 1A4 is responsible for the repression of *fz2* by *engrailed* in the dorsal cells of the hindgut, which validates the criteria of proximity to assign a gene to one immunoprecipitated DNA fragment. The results obtained for 1A4 and in two other cases (data not shown) demonstrate that this approach allowed us to identify genomic fragments that are functional Engrailed-binding sites in vivo.

As shown in this report, the UV-X-ChIP technique, when associated with sequencing, provides a means to enable the rapid collection of a large data set of high-affinity binding sites used by a transcription factor during development. This method is general enough to be used to identify binding sites and targets for other transcription factors (F. G. and F. M., unpublished). Among the 203 sequences that localize within unique genomic regions, only 40 were found two or three times, indicating a low level of redundancy. This suggests that Engrailed is able to bind to a large number of genomic sites. Therefore, a more exhaustive genome-wide localization analysis should rather combine the same chromatin immunoprecipitation procedure with genomic DNA microarrays, which are not yet available in *Drosophila*.

We thank Karène Mahtouk and Sophie Vidal for their technical support, and Andrew Goldsborough for his comments concerning the manuscript. Thanks to G. Cavalli, G. Chanas and N. Nègre for helpful discussions during this work; C. Alexandre and J. P. Vincent for (UAS-VP16-En) fly strain; B. Bello for WH.GFP vector; R. Nusse for providing anti-Frizzled 2 antibody and *frizzled 2* cDNA; M. Affolter for *branchless* cDNA; C. Goodman for *frazzled* cDNA; and M. K. Baylies for *hibris* cDNA. Funding for the research was supported by CNRS, FRM and by ARC grant 5341 to F. M.; a Singer-Polignac fellowship to P. S.; and a MENRT fellowship to D. M.

REFERENCES

Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., Scherer, S. E., Li, P. W., Hoskins, R. A., Galle, R. F., George, R. A. et al. (2000). The genome sequence of Drosophila melanogaster. *Science* **287**, 2185-2195.

Alexandre, C., Jacinto, A. and Ingham, P. W. (1996). Transcriptional activation of *hedgehog* target genes in *Drosophila* is mediated directly by the Cubitus interruptus protein, a member of the Gli family of zinc finger DNA binding proteins. *Genes Dev.* **10**, 2003-2013.

Alexandre, C. and Vincent, J. P. (2002). Requirements for transcriptional repression and activation by Engrailed in Drosophila embryos. *Development* **130**, 729-739.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol*. **215**, 403-410.

Artero, R. D., Castanon, I. and Baylies, M. K. (2001). The immunoglobulin-like protein Hibris functions as a dose-dependent regulator of myoblast fusion and is differentially controlled by Ras and Notch signaling. *Development* **128**, 4251-4264.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T. et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet*. **25**, 25-29.

Bailey, T. L. and Elkan, C. (1995). The value of prior knowledge in

discovering motifs with MEME. *Proc. Int. Conf. Intell. Syst. Mol. Biol*. **3**, 21-29.

Bejsovec, A. and Martinez Arias, A. (1991). Roles of *wingless* in patterning the larval epidermis of Drosophila. *Development* **113**, 471-485.

Bello, B., Resendez-Perez, D. and Gehring, W. J. (1998). Spatial and temporal targeting of gene expression in Drosophila by means of a tetracycline-dependent transactivator system. *Development*, **125**, 2193-2202.

Bhat, K. M. and Schedl, P. (1997). Requirement for *engrailed* and *invected* genes reveal novel regulatory interactions between *engrailed/invected*, *patched*, *gooseberry* and *wingless* during *Drosophila* neurogeneis. *Development* **124**, 1675-1688.

Biggin, M. D. and McGinnis, W. (1997). Regulation of segmentation and segmental identity by Drosophila homeoproteins: the role of DNA binding in functional activity and specificity. *Development* **124**, 4425-4433.

Cadigan, K. M., Fish, M. P., Rulifson, E. J. and Nusse, R. (1998). Wingless repression of Drosophila *frizzled* 2 expression shapes the Wingless morphogen gradient in the wing. *Cell* **93**, 767-777.

Cavalli, G., Orlando, V. and Paro, R. (1999). Mapping DNA target sites of chromatin-associated proteins by formaldehyde cross-linking in Drosophila embryos. In *Chromosome Structural Analysis: A Practical Approach* (ed. W. A. Bickmore), pp. 20-37. Oxford: Oxford University Press.

Dahmann, C. and Basler, K. (2000). Opposing transcriptional outputs of Hedgehog signaling and Engrailed control compartmental cell sorting at the Drosophila A/P boundary. *Cell* **100**, 411-422.

Desplan, C., Theis, J. and O'Farrell, P. H. (1988). The sequence specificity of homeodomain-DNA interaction. *Cell*, **54**, 1081-1090.

DiNardo, S., Sher, E., Heemskerk-Jongens, J., Kassis, J. A. and O'Farrell, P. H. (1988). Two-tiered regulation of spatially patterned *engrailed* gene expression during Drosophila embryogenesis. *Nature* **332**, 604-609.

Dittrich, R., Bossing, T., Gould, A. P., Technau, G. M. and Urban, J. (1997). The differentiation of the serotonergic neurons in the Drosophila ventral nerve cord depends on the combined function of the zinc finger proteins Eagle and Huckebein. *Development* **124**, 2515-2525.

Dworak, H. A., Charles, M. A., Pellerano, L. B. and Sink, H. (2001). Characterization of Drosophila *hibris*, a gene related to human nephrin. *Development* **128**, 4265-4276.

Egger, B., Leemans, R., Loop, T., Kammermeier, L., Fan, Y., Radimerski, T., Strahm, M. C., Certa, U. and Reichert, H. (2002). Gliogenesis in Drosophila: genome-wide analysis of downstream genes of glial cells missing in the embryonic nervous system. *Development* **129**, 3295-3309.

Eldon, E., Kooyer, S., D'Evelyn, D., Duman, M., Lawinger, P., Botas, J. and Bellen, H. (1994). The Drosophila 18 wheeler is required for morphogenesis and has striking similarities to Toll. *Development* **120**, 885-899.

Fauvarque, M. O. and Dura, J. M. (1993). Polyhomeotic regulatory sequences induce developmental regulator-dependnt variegation and targeted P-element insertions in Drosophila. *Genes Dev.* **7**, 1508-1520.

Furlong, E. E., Andersen, E. C., Null, B., White, K. P. and Scott, M. P. (2001). Patterns of gene expression during Drosophila mesoderm development. *Science* **293**, 1629-1633.

Garcia-Bellido, A. (1975). Genetic control of wing disc development in Drosophila. *Ciba Found. Symp.* 161-182.

Gay, N. J., Poole, S. J. and Kornberg, T. B. (1988). The *Drosophila engrailed* protein is phosphorylated by a serine-specific protein kinase. *Nucleic Acids Res.* **16**, 6637-6647.

Graba, Y., Aragnol, D., Laurenti, P., Garzino, V., Charmot, D., Berenger, H. and Pradel, J. (1992). Homeotic control in *Drosophila*: the *scabrous* gene is an in vivo target of *Ultrabithorax* proteins. *EMBO. J.* **11**, 3375-3384.

Gustavson, E., Goldsborough, A. S., Ali, Z. and Kornberg, T. (1996). The *Drosophila engrailed* and *invected* genes: partners in regulation, expression and functions. *Genetics* **142**, 893-906.

Hanks, M. C., Loomis, C. A., Harris, E., Tong, C. X., Anson-Cartwright, L., Auerbach, A. and Joyner, A. (1998). *Drosophila engrailed* can substitute for *engrailed 1* function in mid-hindbrain, but not limb development. *Development* **125**, 4521-4530.

Hatini, V. and DiNardo, S. (2001). Divide and conquer: pattern formation in Drosophila embryonic epidermis. *Trends Genet.* **17**, 574-579.

Heemskerk, J., DiNardo, S., Kostriken, R. and O'Farrell, P. H. (1991). Multiple modes of *engrailed* regulation in the progression towards cell fate determination. *Nature* **352**, 404-410.

Hidalgo, A. (1994). Three distinct roles for the *engrailed* gene in *Drosophila* wing development. *Curr. Biol.* **4,** 1087-1098.

Iyer, V. R., Horak, C. E., Scafe, C. S., Botstein, D., Snyder, M. and Brown, P. O. (2001). Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. *Nature* **409**, 533-538.

Kissinger, C. R., Liu, B., Martin-Blanco, E., Kornberg, T. B. and Pabo, C. O. (1990). Crystal structure of an *engrailed* homeodomain-DNA complex at 2.8 A resolution: a framework for understanding homeodomain-DNA interactions. *Cell* **63**, 579-590.

Kolodziej, P. A., Timpe, L. C., Mitchell, K. J., Fried, S. R., Goodman, C. S., Jan, L. Y. and Jan, Y. N. (1996). *frazzled* encodes a Drosophila member of the DCC immunoglobulin subfamily and is required for CNS and motor axon guidance. *Cell* **87**, 197-204.

Kornberg, T. (1981). *engrailed*: a gene controlling compartment and segment formation in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **78**, 1095-1099.

Lawrence, P. and Morata, G. (1976). Compartments in the wing of *Drosophila*: A study of the *engrailed* gene. *Dev. Biol.* **50**, 321-337.

Lawrence, P. A. and Morata, G. (1992). Developmental biology. Lighting up Drosophila. *Nature* **356**, 107-108.

Lecourtois, M., Alexandre, C., Dubois, L. and Vincent, J. P. (2001). Wingless capture by Frizzled and Frizzled2 in Drosophila embryos. *Dev. Biol*. **235**, 467-475.

Levashina, E. A., Ohresser, S., Lemaitre, B. and Imler, J. L. (1998). Two distinct pathways can control expression of the gene encoding the Drosophila antimicrobial peptide metchnikowin. *J. Mol. Biol*. **278**, 515-527.

Liang, Z. and Biggin, M. D. (1998). *eve* and *ftz* regulate a wide array of genes in blastoderm embryos: the selector homeoproteins directly or indirectly regulate most genes in *Drosophila*. *Development* **125**, 4471-4482.

Lieb, J. D., Liu, X., Botstein, D. and Brown, P. O. (2001). Promoter-specific binding of Rap1 revealed by genome-wide maps of protein-DNA association. *Nat. Genet.* **28**, 327-334.

Mannervik, M. (1999). Target genes of homeodomain proteins. *BioEssays* **21**, 267-270.

Markstein, M., Markstein, P., Markstein, V. and Levine, M. S. (2002). Genome-wide analysis of clustered Dorsal binding sites identifies putative target genes in the Drosophila embryo. *Proc. Natl. Acad. Sci. USA* **99**, 763-768.

Martinez-Arias, A. and White, R. A. H. (1988). *Ultrabithorax* and *engrailed* expression in *Drosophila* embryo mutant segmentation genes of the pair rule class. *Development* **102**, 325-338.

Maschat, F., Serrano, N., Randsholt, N. B. and Geraud, G. (1998). *engrailed* and *polyhomeotic* interactions are required to maintain the A/P boundary of the Drosophila developing wing. *Development* **125**, 2771-2780.

Milan, M. and Cohen, S. M. (2000). Temporal regulation of *apterous* activity during development of the Drosophila wing. *Development* **127**, 3069-3078.

Ren, B., Robert, F., Wyrick, J. J., Aparicio, O., Jennings, E. G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E. et al. (2000). Genome-wide location and function of DNA binding proteins. *Science* **290**, 2306-2309.

Ribeiro, C., Ebner, A. and Affolter, M. (2002). In vivo imaging reveals different cellular functions for FGF and Dpp signaling in tracheal branching morphogenesis. *Dev. Cell* **2**, 677-683.

Saenz-Robles, M. T., Maschat, F., Tabata, T., Scott, M. P. and Kornberg, T. B. (1995). Selection and characterization of sequences with high affinity for the Engrailed protein of *Drosophila*. *Mech. Dev*. **53**, 185-195.

Serrano, N., Brock, H. W., Demeret, C., Dura, J. M., Randsholt, N. B., Kornberg, T. B. and Maschat, F. (1995). *polyhomeotic* appears to be a target of Engrailed regulation in *Drosophila*. *Development* **121**, 1691-1703.

Serrano, N., Brock, H. W. and Maschat, F. (1997). *β3-tubulin* is directly repressed by the Engrailed protein in *Drosophila*. *Development* **124**, 2527-2536.

Serrano, N. and Maschat, F. (1998). Molecular mechanism of *polyhomeotic* activation by Engrailed. *EMBO J*. **17**, 3704-3713.

Siegler, M. V. and Jia, X. X. (1999). Engrailed negatively regulates the expression of cell adhesion molecules connectin and neuroglian in embryonic Drosophila nervous system. *Neuron* **22**, 265-276.

Simon, H. H., Saueressig, H., Wurst, W., Goulding, M. D. and O'Leary, D. D. (2001). Fate of midbrain dopaminergic neurons controlled by the *engrailed* genes. *J. Neurosci*. **21**, 3126-3134.

Stathopoulos, A., van Drenth, M., Erives, A., Markstein, M. and Levine, M. (2002). Whole-genome analysis of dorsal-ventral patterning in the Drososphila embryo. *Cell* **111**, 687-701.

Sutherland, D., Samakovlis, C. and Krasnow, M. A. (1996). *branchless*

encodes a Drosophila FGF homolog that controls tracheal cell migration and the pattern of branching. *Cell* **87**, 1091-1101.

Tabata, T., Schwartz, C., Gustavson, E., Ali, Z. and Kornberg, T. B. (1995). Creating a *Drosophila* wing de novo, the role of *engrailed*, and the compartment border hypothesis. *Development* **121**, 3359-3369.

Takashima, S. and Murakami, R. (2001). Regulation of pattern formation in the Drosophila hindgut by *wg*, *hh*, *dpp*, and *en*. *Mech. Dev.* **101**, 79-90.

Toth, J. and Biggin, M. D. (2000). The specificity of protein-DNA crosslinking by formaldehyde: in vitro and in drosophila embryos. *Nucleic Acids Res*. **28**, E4.

Vincent, J. P. and O'Farrell, P. H. (1992). The state of *engrailed* expression is not clonally transmitted during early Drosophila development. *Cell* **68**, 923-931.

Weinmann, A. S., Bartley, S. M., Zhang, T., Zhang, M. Q. and Farnham, P. J. (2001). Use of chromatin immunoprecipitation to clone novel E2F target promoters. *Mol. Cell Biol*. **21**, 6820-6832.

Weinmann, A. S., Yan, P. S., Oberley, M. J., Huang, T. H. and Farnham, P. J. (2002). Isolating human transcription factor targets by coupling chromatin immunoprecipitation and CpG island microarray analysis. *Genes Dev*. **16**, 235-244.

Wolberger, C. (1996). Homeodomain interactions. *Curr. Opin. Struct. Biol*. **6**, 62-68.